

**Elements of Caption Quality: Determining the priority of qualitative closed
caption elements**

Alison Whiting, BSc, Queen's University, 2010

A thesis
presented to Ryerson University
in partial fulfillment of the
requirements for the degree of
Master of Science in Management
in the program of
Master of Science in Management, Information Technology Management

Toronto, Ontario, Canada, 2018

© Alison Whiting, 2018

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

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.

I understand that my thesis may be made electronically available to the public.

Abstract

Elements of Caption Quality: Determining the priority of qualitative closed caption elements, Master of Science in Management (MScM) 2018, Alison Whiting, Ryerson University.

In Canada, the Canadian Radio Television and Communications Commission requires that all English- and French-language broadcasters caption 100% of their programs, and that live-produced programming – such as news broadcasts, sports events, and award shows – are captioned with a 95% accuracy rate for English-language. However, measuring caption quality as a purely objective count of the number of errors in the text means that many qualitative factors of quality are not considered. This research explored what priority Deaf and hard of hearing viewers place on non-quantitative elements of caption quality, namely caption display speed, missing words, spelling and grammar errors, and speaker identification. Using a survey tool based off the principals of the NASA-TLX workload assessment tool, participants were asked to watch two television clips with their original live-produced captions and provide feedback on how the captions impacted their viewing pleasure. The main findings suggested that viewers place equal value on verbatim accuracy and caption display speed, and that a trade off between the two cannot easily be made. This research provides a starting point for measuring caption quality using subjective quality factors.

Table of Contents

Author’s Declaration	ii
Abstract	iii
List of Tables.....	vi
List of Figures	vii
List of Appendices	viii
1. Introduction	1
2. Literature Review	5
2.1. Closed captioning.....	5
2.2. Limitations of closed caption.....	8
2.3. Emotive captioning	9
2.4. Policies and guidelines for closed captioning quality.....	12
2.4.1. Canadian policies and guidelines.....	12
2.4.2. International policies and guidelines.....	13
2.5. Caption quality.....	15
3. Theoretical Framework.....	17
3.1. Inclusive design theory	17
3.2. Inclusive design and closed captions.....	18
4. Research Methodology	21
4.1. Research questions and hypothesis	21
4.2. NASA-TLX and questionnaire development	22
4.2.1. NASA-TLX.....	22
4.2.2. Pre-study questionnaire.....	23
4.2.3. Post-study questionnaire	24
4.3. Content used in this study.....	24
4.4. Data Collection.....	25
4.4.1. Experimental setup.....	25
4.4.2. Data collection	25
4.5. Data analysis	26
4.6. Data observations and caption preference scores.....	27
5. Results.....	32
5.1. Caption preference scores.....	32
5.2. Results of D/HOH group	33
5.2.1. Video comparison	32
5.2.2. Trade-offs and impact on viewing pleasure	34
5.3. Results of hearing group.....	37
5.3.1. Video comparison	37

5.3.2. Trade-offs and impact on viewing pleasure	38
5.4. Thematic analysis.....	39
6. Discussion.....	43
6.1. Limitations	52
6.1.1. Recruitment	52
6.1.2. Experimental design.....	53
7. Recommendations and Conclusion	55
7.1. Conclusion.....	55
7.2. Contributions	57
7.3. Future research.....	58
Appendices.....	60
References and Work Sited.....	80

Tables

4.1: Caption information for the two videos used in this study.....	25
4.2: Themes.....	26
5.1: Caption preference scores for D/HOH group	32
5.2: Caption preference scores for hearing group.....	33
5.3: Paired t-test for City and CTV rating scores by D/HOH group	34
5.4: Mean and standard deviation for City and CTV rating scores by D/HOH group	34
5.5: Chi-square results for weighted pair analysis.....	35
5.6: Impact of caption quality factors for D/HOH group	36
5.7: Mean and standard deviation for City and CTV rating scores by hearing group	37
5.8: Chi-square results for weighted pair analysis.....	38
5.9: Impact of caption quality factors for hearing group	39
5.10: Frequency of each category of speed comments.....	41
5.11: Positive and negative comments by video for D/HOH group.....	41
5.12: Positive and negative comments by video for hearing group	42
8.1: Shapiro-Wilkes test for normality on data used to determine learning effect	71
8.2: Paired t-test for first video (R1) and second video (R2) rating scores.....	71
8.3: Shapiro-Wilkes test for normality on data used to determine significance of rating scores .	71
8.4: Shapiro-Wilkes test for normality on data used to determine if any differences between City (V1) and CTV (2) rating scores	72

Figures

Figure 2.1: Examples of emotive captioning	12
Figure 4.1: Relationship of participants to D/HOH community	29
Figure 4.2: Level of education for D/HOH and hearing groups	29
Figure 4.3: Ages of participants for D/HOH and hearing groups	30
Figure 4.4: Hours of TV watched per week for D/HOH and hearing groups	30
Figure 5.1: Demographic breakdown of participants selecting Accuracy compared to Speed	35
Figure 5.2: Frequency of theme occurrences for each video for D/HOH group	40
Figure 5.3: Frequency of theme occurrences for each video for hearing group	40
Figure 6.1a: Screen capture of City	49
Figure 6.1b: Screen capture of CTV	49

Appendices

Appendix A: Video caption transcripts	65
Appendix B: Caption preference score calculation methodology.....	70
Appendix C: Additional statistical results.....	71
Appendix D: Survey questionnaire	73
Appendix E: Ethics letter of approval	83

1. Introduction

In many parts of the world television programs are made available with closed captions, defined as the verbatim text equivalent of the audio speech component of programming, thus making that content available to audiences who are Deaf or hard of hearing (D/HOH). Closed captions appear on-screen in conjunction with the equivalent verbal expressions and can be turned off or on by the viewer through the television settings. While difficult to determine exact numbers due to poor and limited available data, it is estimated that there are approximately 357,00 culturally Deaf Canadians and 3.21 million hard of hearing Canadians (Canadian Association of the Deaf, 2015). These individuals make up the main target audience for closed captions and are the end user for which captions are designed. However, captions can also be used as aids for hearing viewers, such as when watching a sports game in a noisy bar or watching a video in a quiet library, as well as in education or informational settings for learning a second language (Burnham et al., 2008; Zdenek, 2015).

As of 2007, the Canadian Radio, Television and Telecommunications Commission (CRTC) requires that all English- and French-language broadcasters caption 100% of their programs over the broadcast day, with the exception of advertising and promos (CAB, 2008). In order to assist broadcasters in fulfilling the regulatory requirements, standards such as EIA 608 (analogue), EIA 708 (digital), and The Canadian Association of Broadcasters (CAB) provide standards and guidelines for captioning around accuracy, consistency, and clarity, as well as placement, characters and line limitations, and text attributes and graphics. Programs that are pre-recorded – such as dramas, sitcoms, etc. – must be captioned with 100% accuracy (captioning for these types of programs is called off-line or post-production captions). Programs that are aired live and captioned in real-time (called live captions) – such as news broadcasts, sports events, and award shows – must be captioned with 95% accuracy for English-language, and 85% accuracy for French-language (CRTC, 2015). According to CRTC regulation 2012-362, the verbatim accuracy score is calculated using an equation that takes into consideration the number of suppressed, substituted, and inserted words in the audio. The limitation of this method of calculating quality is that it only takes into consideration differences between the

words that are spoken and the text version produced in the captions. Other qualitative qualities outlined in the guidelines and standards, such as placement of captions, text attributes, graphics, legibility, and readability, are not considered in this calculation. It also does not allow for any user input.

This research project explores what priority viewers place on non-quantitative elements of closed captions. Thus, the research question asked is, what priority do Deaf or hard of hearing (D/HOH) viewers place on verbatim accurate captions as compared to other caption elements, specifically caption speed, missing words, spelling/grammar errors, and speaker identification? The purpose of this research is two-fold, (1) to measure the value that D/HOH viewers place on verbatim accurate captions when other caption elements are affected to achieve that level of accuracy, and (2) to explore how caption quality is measured and valued by viewers beyond the CRTC's verbatim accuracy score.

Exploring a more inclusive measure of caption quality, one that takes into consideration the opinions and experiences of the caption user, is in line with the social model of disability. In recent years societal perception of disability has shifted from a medical model, where people are seen as disabled or incapacitated by their physiological impairments and mental limitations, to a social model, in which people have their disability thrust upon them by poor design, exclusionary services, and oppressive environments (Clarkson & Coleman, 2015; Llewellyn & Hogan, 2000; Marks, 1997). The social model of disability puts the problem back into the collective responsibility of society as a whole (Llewellyn & Hogan, 2000), and there is a shift towards designing and engineering services and environments to meet the needs of all members of the community, without the need for excessive adaptive tools or medical interventions. This social model of disability is further driven by the need for governments to meet the needs of a rapidly gaining global population, as well as increasingly vocal organisations of disabled people (Clarkson & Coleman, 2015; Vanderheiden, 2000). Finally, the need for a more inclusive and subjective measure of quality for closed captions is guided by the criticism from the D/HOH communications that are excluded from the existing quality assessment process. Deaf people have criticized the existing system in large part due to lack of contextual features, such as emotion in dialogue, background music, music tempo, mode or

depth, non-speech items, speaker identification, and timing of jokes (Hersh, 2013), all of which are not and cannot be measured using a verbatim accuracy score. As such, the CRTC's existing verbatim accuracy score, while being one measure for assessing caption content accuracy and clarity, cannot by itself provide a comprehensive measure of caption quality as experienced by the viewer. Understanding the value that viewers place on the various non-accuracy caption elements will help drive towards a more inclusive and comprehensive measure of caption quality. Furthermore, providing improved closed captioning can elevate the consumer experience for all people that use captioning services.

The basic premise behind the data collection and analysis techniques used in this study is the NASA-TLX workload assessment tool (Hart & Staveland, 1988), a subjective tool used to measure the mental workload of participants performing various tasks. The NASA-TLX has been used before in studies involving the evaluation of visual and/or auditory displays, and has grown in popularity among researchers for its ease of implementation (Cao, Chintamani, Abhilash, & Ellis, 2009). The NASA-TLX assessment tool is a survey that consisting of two parts, and focuses on six subjective dimensions of mental workload. In the first part of the survey, participants select between pairs of the dimensions which one they think is more relevant to their definition of mental workload. The second part of the survey is administered after participants have performed a task and it requires them to rate how mentally taxing the task was in relation to the six dimensions. This basic structure forms the basis of the methodology in this study; asking participants to select which caption quality factor between two factors they feel is most important to their viewing pleasure. Next they are asked to rate on a numerical scale how positively or negatively each caption quality factor impacted their viewing pleasure for two videos. This information is then combined to generate a caption preference score.

This research study consists of two parts, firstly, participants are asked to complete a pre-study questionnaire which asks for information regarding demographic information, television and caption usage habits, and asks participants to weight caption elements by selecting between pairs. Participants will then watch two television clips recorded with original live captions and complete a post-study questionnaire after each video. The post-study questionnaire asks participants to rate how positively or negatively each of the studied caption

elements affected their viewing pleasure. The data is then used to calculate the overall caption preference score for each caption element, as per the NASA-TLX calculations. The study uses a within-subject design method, in which all participants watch both videos, the two videos are shown in randomised order to mitigate any learning effects.

The quantitative data collected in the pre- and post-study questionnaires was analysed using SPSS software and parametric statistics. Statistical analyses were conducted to determine if there was significant difference in how participants rated the positive or negative impact of each caption element, as well as determine if participants enjoyed one style of captioning over the other. The post-study questionnaires also included open-ended questions, and the responses to them were analysed using thematic analysis and NVivo software. The quantitative data from the thematic analysis underwent non-parametric statistical tests and the results were used to corroborate earlier findings.

This research will be of interest to social scientists in the field of design science research and accessible media technology, adding to a growing body of knowledge pertaining to inclusive design and closed captioning. From a practical perspective, the outcomes of this exploratory research will be of interest to media corporations and broadcasters, legislators of disability rights, as well as D/HOH television viewers. The remainder of this thesis has the following organisation; chapter 2 consists of a literature review of current closed caption quality research and inclusive design theory, chapter 3 discusses the theoretical framework, chapter 4 describes the methodology used in this research study, chapter 5 reports on the results, chapter 6 includes a discussion of the results and their relevance to the field of study, and chapter 7 contains concluding remarks.

2 Literature Review

The following section provides a brief history of closed captioning, explains limitations of post-production captions and live captions, describes emotive captioning and how it attempts to address the limitations of traditional captions, and reviews current policies, guidelines and research in the field of quality assurance of closed captions.

2.1 Closed captioning

Closed captions have been in existence since the 1960s, when D/HOH educators created captions using a new closed-circuit cable system for use in their own classrooms at Deaf residential and day schools (Downey, 2007). It was not until the 1980s that captioning became more publicly available. It was at this time that a national captioning system was set up and the media industry was invited, but not forced, to fund and transmit captioned programs (Downey, 2007), that decoder boxes became available for purchase so that D/HOH viewers could view captions on their own television sets (Fels, Lee, & Branje, 2005), and that a wide variety of closed captioning programming became available on a regular basis (Sillman, 1984). Captioning availability continued to progress in the late 20th century, and in 1993, the Television Decoder Circuitry Act was introduced that required all new television sets 13 inches or larger manufactured for sale in the U.S. to contain captioning technology (Lee, Fels, & Udo, 2007).

Currently, equal access to media programs and services is mandated by different disability rights legislations, such as the Americans with Disabilities Act (1990) in the United States, the Disability Discrimination Act (1995) in the United Kingdom, and the Accessibility for Ontarians with Disabilities Act (2005) in Ontario, Canada. Most recently, the 21st Century Communications and Video Accessibility Act of 2010 was passed, that requires all captioned television programs to also be captioned when delivered over the internet (Zdenek, 2011). Additionally, the act expands the scope of devices that must display captions from just television sets 13 inches or larger to all devices that receive or display video programming transmitted simultaneously with sound. The depth and breadth of caption accessibility has developed substantially over the past five decades, providing D/HOH viewers with better access to media and television content.

Closed captions were originally developed to provide access to sound and speech information in film and television programs to D/HOH individuals. They are a form of assistive text-based technology intended to make the auditory components of television accessible to D/HOH viewers (Burnham et al., 2008), without the need for an aid or additional assistive devices. Closed captions were originally designed as white upper-case letters, a single mono-spaced font, a single size, on black background (VITAC 2003), however these days many television sets and other media devices allow the user to customise font, colour, and size of the captions. Although these options exist, caption aesthetics have remained relatively constant, primarily using white lettering on a black background, despite many other aspects of television and media technology advancing in the 21st century.

Captions are used not only to communicate speech and spoken words, but also to communicate the non-speech information in program audio. Often referred to as descriptive captions (reference), these types of closed captions are used to describe music, background noises and sound effects, and can be a useful tool for creating the mood, setting the scene, and establishing the context for dialogue (CAB, 2008). In most cases, it is the captioner that will determine how and when to use descriptive captions. Pre-recorded scripted television is captioned as part of the post-production process, thus allowing for the potential of including the program creative team's input and direction with regards to music, mood, background noise and context. The production team may also be part of the decision process for deciding how best to communicate this information to D/HOH audiences through descriptive captions. In contrast, live-televised programs (e.g. news, sports, live-events) must be captioned in real-time, which adds a level of difficulty and complexity to producing quality captions, as there is a limited amount of time in which to create the captions, and there is little opportunity to edit or correct mistyped words. Additionally, the captioner must decide in real-time what non-speech information can and should be communicated to D/HOH audiences through descriptive captions.

Although originally designed for use by D/HOH viewers, closed captions are also used by hearing individuals in a number of situations. Captions may be useful as an aid in noisy sports bars, when learning a new language, or when studying in a quiet space, such as the library (Udo

& Fels, 2010; Zdenek, 2011). Although shown to be useful and beneficial to many different types of people and many different types of situations, closed captions are not without criticism. Snell, (2012) has argued that for captions to be successful they must enable emotional experience while being viewed, in addition to communicating language-based understanding, while many researchers have identified the limitations of captions in communicating emotive information. Background noise and music, which help to set mood and scene, are often omitted due to limitations in television display technology and reading speed (Lee et al., 2007), and many researchers have found that D/HOH viewers are dissatisfied with the amount of information currently available through captions, including information about emotions, background music, and non-speech information (e.g. Hersh, 2013; Ohene-Djan, Wright, & Combie-Smith, 2007; Rashid, Vy, Hunt, & Fels, 2008). An additional challenge to the limitations of closed captions, is that while D/HOH viewers agree that captions should be improved, there is often disagreement on where the priority should be placed. While often grouped together for research, Deaf or hard of hearing people make up distinct communities; those people who identify as culturally Deaf, and those that identify as having impaired hearing (Neves, 2008; Romero-Fresco, 2009). People who identify as hard of hearing, deaf or Deaf may require different captioning solutions related to display speed, accuracy and sound synchronization, in part due to their different types and degrees of hearing loss (Neves, 2008). In the early 2000s, the Deaf community successfully lobbied the CRTC to implement policies that require broadcasters to meet caption quantity and quality standards, including a 95% verbatim accuracy rate (Ladurantaye, 2012), indicating that they place a high value on requiring the captions to provide a full and true representation of the content of the program. In contrast, hard of hearing viewers often have some level of hearing ability and may place a higher value on caption synchronicity with the audio.

Further to this is the notion of captions as an example of sensory substitution. Sensory substitution is the concept of transforming stimuli characteristics of one sensory modality into stimuli of another sensory modality (Lenay, Gapenne, Hanneton, Margue Catherine, & Genouëlle, 2003). Captions are an example of sensory substitution in that they turn speech, music, and sounds into text so that the information can be read instead of heard. In doing so

however, this is likely to change the overall viewing experience for the user (Snell, 2012). Research shows that different parts of the brain are activated for listening versus reading comprehension (Buchweitz, Mason, Leda, & Just, 2009), thus, changing sounds and speech into text changes the cognitive processes required to process that information from an auditory task to a visual task, subsequently impacting the overall experience for the viewer. Finally, captions may not present an ideal method of communicating auditory information due to the generally poor literacy comprehension levels of deaf people compared to their hearing peers (e.g. Albertini & Mayer, 2011; Wilbur, Goodhart & Fuller, 1989; Parault & Williams, 2010; Strong et al., 1997). Deaf viewers, for example, may find it challenging to read the captions at the rate at which they are presented, and the addition of descriptive captions may potentially further increase the amount of text the viewer must read and comprehend while watching a television program or film. While captions are praised for their ability to make accessible to viewers auditory information in circumstances where sound cannot be heard, there is still a strong need to evaluate their quality critically and explore opportunities for development and improvement.

2.2 Limitations of closed captioning

Captions provide an important and necessary service to D/HOH television viewers, but they are not without their limitations. As mentioned, captions are primarily used to capture the speech dialogue, and non-speech information, such as background noise and music, is often omitted or restricted due to limitations in television display technology and reading speed, and does not sufficiently meet all the needs of the community (Fels et al., 2005; Lee et al., 2007). Many studies find that D/HOH viewers are dissatisfied with the amount of information currently available through captions, including information about emotions, background music, and non-speech information (e.g., Hersh, 2013; Ohene-Djan, Wright, & Combie-Smith, 2007; Rashid, Vy, Hunt, & Fels, 2008). Another limiting factor is the way that captions are processed in the brain compared to auditory information. Captions are an example of sensory substitution, the concept of transforming stimuli characteristics of one sensory modality into stimuli of another sensory modality (Lenay et al., 2003). In the case of captions, it is the process of turning speech into text, such that the information can be read instead of heard, and in the case of descriptive captions, non-speech sounds are turned into descriptive text to be read instead of

heard. These changes are likely to change the overall viewing experience of the user (Snell, 2012) due to different parts of the brain being activated for listening versus reading comprehension (Buchweitz et al., 2009). Further to this is the generally poorer literacy comprehension levels of deaf people compared to their hearing peers (e.g. Albertini & Mayer, 2011; Wilbur, Goodhart & Fuller, 1989; Parault & Williams, 2010; Strong et al., 1997). The volume of information that must be quickly and effectively made available through captions is such that it can be challenging for the viewers to read and comprehend all the text before it disappears from the screen.

Real-time captions have other unique limitations due to the fact that they are created in real-time and the caption stenographer has little knowledge of what will happen next in the program. Due to speed or density of the material it is not always possible to provide verbatim captions; for instance, sports events can be the most challenging type of program to caption because the commentary is frequently too fast for verbatim transcription (CAB, 2008). In these instances, the stenographer must decide what information to relay in the captions while ensuring that the viewer does not miss important information. Further to this, is that verbatim captions, in instances of fast-paced live programming, are often too fast for many deaf viewers to read and comprehend (Romero-Fresco, 2009). Another challenge for the stenographer is that they are required to not only caption the speech information, but the non-speech information as well in real-time. Due to the fast pace of real-time captioning it can be extremely difficult to incorporate descriptive captions, as the stenographer must not only recognise the need for a descriptive caption but also decide what it should say (CAB, 2008). To help mitigate these problems, the United States has introduced best practices that include video programmers providing captioning vendors with access to show notes, proper names, and song lyrics to make it easier to caption live programs (FCC, 2014b). The criticisms and limiting factors of captions mentioned are not ones that can be measured or assessed through a verbatim accuracy score and are much more subjective in nature for the viewer, thus further demonstrating a need for a tool that can assess the quality of captions based on non-accuracy related factors.

2.3 Emotive captioning

Emotive captions have been developed as an alternate way to present sounds, music, and emotive information to viewers, through the use of graphics and other features. Research indicates that traditional captions are often unable to provide important emotional information, and that an “emotional gap” is subsequently generated (Ohene-Djan et al., 2007). These limitations are further exaggerated when characters cannot be seen on the screen or when multiple characters are speaking at once (Ohene-Djan et al., 2007). As a result, research has been conducted into the possibility of using graphical, animated, and tactile displays as a way of conveying the additional non-verbal information in television and film (Fels et al., 2005). By using graphics, colour, and animation to illustrate sound information, it will be possible to reduce the volume of descriptive captions required, thus reducing the amount of text being displayed on the screen at any given time. Graphics are also more visual than text, thus more akin to the visual sign language used to communicate by D/HOH individuals.

A number of research studies have been conducted in recent years into the use of graphics, colours, typography, and dynamic text to convey music, emotion, and other paralinguistic information not conveyed in traditional captions. Early research by Silverman and Fels (2002) explored the use of speech bubbles and icons with descriptions about background sounds and music in order to provide the viewer with more information than is conveyed by captions alone. They also had contact with the video’s creators, which allowed them to gain knowledge regarding the director’s choice of sound and overall mood. Subsequent research focused on the communication of emotion through enhanced captions. Research studies by Fels et al. (2005) and by Lee et al. (2007) utilised graphics, colour, and animation in different ways to illustrate sound and emotive information. Both studies also employed an inclusive design approach, and engaged the film’s creative teams in the identification of emotions in the script (Lee et al., 2007), and the design and development of the emotive captions (Fels et al., 2005). Later research explored the use of kinetic typography and dynamic text to relate emotive information present in the speech (e.g. Brown, Jones, & Crabb, 2015; Hong et al., 2011; Hu, Kautz, Yu, & Wang, 2013; Rashid, Aitken, & Fels, 2006). Kinetic typography provided information about tone of voice and emotion by adding animation and movement to the type (Rashid et al., 2006), and dynamic text moved the subtitles around the screen in accordance with video content and/or

speaker identification (Brown et al., 2015; Hong et al., 2011). The goal of these types of enhanced captions was to provide more meaning to the captions, thus enhancing the viewer's experience. Overall, the feedback from test participants was mixed. Some D/HOH individuals found that the emotive captions did add to their viewing experience in a positive way, while others have noted that these types of captions would be better suited for children's television (e.g. Fels et al., 2005; Lee et al., 2007).

Emotive captions have the potential to provide viewers with more auditory information than traditional captions, however they too have their limitations. Studies into the use of emotive captions for film and television have received mixed responses from test participants (e.g. Fels et al., 2005; Lee et al., 2007), demonstrating that there is still improvement to be made regarding how to communicate emotive information to viewers.

One of the challenges of descriptive captioning is that someone must decide what information to convey; this challenge is not eliminated with the use of emotive captions, where someone must still decide what information to convey, how best to convey it, and what graphics to use (Fels et al., 2005). Thus far, emotive captioning has only been used in test environments and for post-production captioning. There is yet to be research done into the impact of emotive captioning on real-time captioning. Providing emotive captioning in real-time poses a number of unique challenges; the captioner must not only record the speech and dialogue, but must also interpret the emotion, music, and background noise, and apply the appropriate emotive feature to the caption, while ensuring that quality standards are maintained. Research is needed in order to determine what impact adding emotive captioning has on the quality of the captions, as well as the viewing enjoyment and video comprehension for the user.

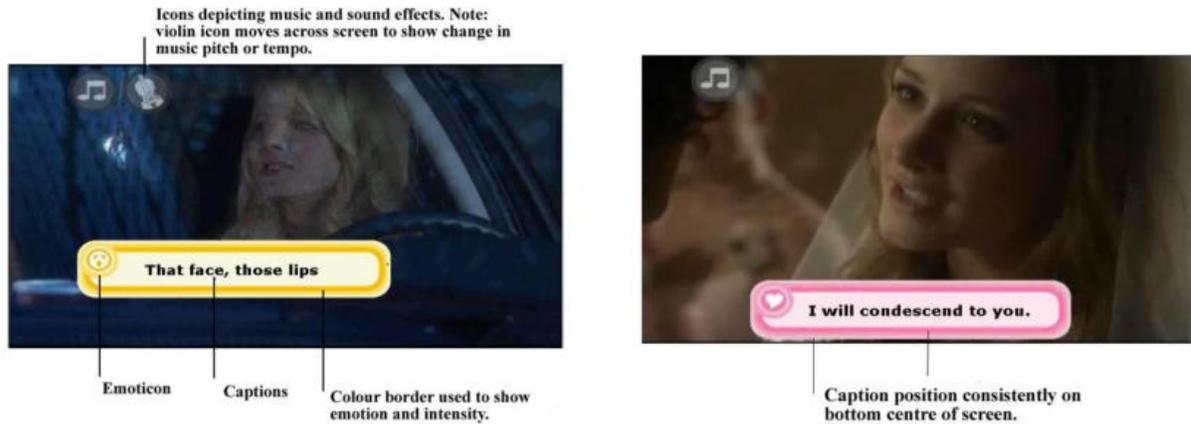


Figure 2.1: Examples of emotive captioning (Fels et al., 2005, p. 2333 with permission)

2.4 Policies and guidelines for closed captioning quality

2.4.1 Canadian policies and guidelines

Countries are responsible for their own closed caption regulations and quality assurances, as there is no universal or international governing body regulating rights for people with disabilities. Policies and guidelines for closed captions specifically often fall under a disability rights act or a media governing body. In Canada, the Canadian Radio-television and Telecommunications Commission (CRTC) requires that pre-recorded programs are captioned with 100% accuracy, and that English-language live programming be captioned with 95% accuracy. The CRTC measures accuracy as the “level of exactness between captions and the audio content of a program” (CRTC, 2015, p. 2), and includes correct spelling for pre-recorded programs. A program’s accuracy score is calculated by the broadcaster using the following CRTC published formula, *errors* includes word substitutions, word omission, and word insertion (see equation 1) (CRTC, 2016).

$$Accuracy\ Rate = \frac{(words - errors)}{words} \times 100 \quad (1)$$

In addition to regulating caption accuracy, the CRTC publishes other policies and guidelines to address caption lag time, correcting errors before a program is re-broadcast, providing adequate information on screen, and formatting aspects (e.g., the use of hyphens). The CRTC outlines four basic principles of captioning, including accuracy, responsibility (preservation of

meaning and intent), consistency (style, format, placement, description, speaker designation, rate of display), and clarity (complete and true rendering of the program audio – who is speaking, non-speech information) (CAB, 2008). These four basic principles outline a number of important characteristics of closed captions that are not related to caption accuracy but have to potential to impact the viewing pleasure of a D/HOH audience member.

Canadian broadcasters are required to monitor and report on their closed caption mechanisms and procedures to the CRTC. The primary mechanism the CRTC uses to validate caption quality is the verbatim accuracy score. To monitor and regulate caption quality in Canada, the CRTC requires broadcasters to calculate and report monthly the accuracy rate for two programs containing live content, and that every two years broadcasters report their efforts for improving accuracy rates. This system of regulation and reporting holds Canadian broadcasters to a high standard of closed captioning as it relates to verbatim accuracy, but does not provide any quality measures for non-accuracy caption elements, such as format, placement, and speaker designation. The CRTC acknowledges the value that these types of caption elements have on the viewing pleasure of D/HOH viewers by listing them as fundamental basic principles, but does not provide any mechanism for assessing quality assurance as it relates to these elements.

2.4.2 International policies and guidelines

In the UK, the policies and guidelines for captioning fall under anti-discrimination laws and the Equality Act of 2010 (EQA) and are regulated by Ofcom's Code of Television Access Services. These regulations require public broadcasters to meet certain accessibility milestones by set years, including closed captions, sign language, and audio descriptions availability, as well as yearly benchmarks ("Previous Code on Television Access Services - Ofcom," 2012). Some broadcasters have been quick to adopt these guidelines and have dramatically increased their provision of closed captions. For instance, since April 2008, the BBC aims to caption 100% of its programmes, including live events (Romero-Fresco, 2009). In addition to setting guidelines and benchmarks for caption availability, Ofcom provides recommendations and best practices regarding caption placement, layout, communication of non-speech information,

synchronisation, speed, and accuracy (“Ofcom’s Code on Television Access Services,” 2015). Ofcom’s Code on Television Access Services also outlines monitoring and compliance standards, requiring broadcasters to submit reports every six months, and to retain a recording in sound and vision of every program for a period of 60 days. Broadcasters are also encouraged to regularly monitor playout of programming to ensure accessibility standards are being provided correctly.

In the United States, Congress and the Federal Communications Commission (FCC) regulate closed caption requirements. These governing bodies provide rules and policies for closed caption availability and, since 2014, quality control aspects pertaining to accuracy, synchronicity, completeness, and placement. Additionally, the FCC distinguishes between pre-recorded, live, and near-live programming and explains how the standards apply to each type of programming, while recognising the greater challenges involved with captioning live and near-live programming (Federal Communications Commission, 2015). Disability advocacy groups, including the National Association for the Deaf (NAD), have played a significant role in advancing the rights of people with disabilities in the United States. The NAD has advocated for improved closed captioning legislation and policy, and as a result of their efforts, 100 percent of all new, non-exempt English-language television programs must be closed captioned in the US. Additional efforts have seen the passing of the Twenty-First Century Communications and Video Accessibility Act (2010), which requires programs first shown on television with captions to also be captioned when made available online. The NAD will continue to advocate for D/HOH rights relating to closed captions, including availability of real-time captions for all live-televised programming.

In Australia, the Australian Communications and Media Authority (ACMA) is the governing board that oversees the rules and regulations pertaining to the availability and quality control of closed captions. These rules outline how much television must be captioned, the types of programs that must be captioned, exemptions, quality standards, as well as annual reporting requirements for broadcasters (“Web Accessibility and Closed Captioning in Australia and New Zealand,” 2015). The ACMA’s Television Captioning Quality Standard was developed in 2012 in collaboration with caption users, advocacy groups, broadcasters, and captioning providers. It

outlines best practices and guidelines relating to readability of captions (e.g., use of colour, font styles, line breaks, and positions), accuracy of captions (i.e., verbatim capture of spoken content), and the comprehensibility of captions (e.g., speaker identification, speed of display, and spelling/grammar errors) (Australian Communications and Media Authority, 2013).

Media accessibility and closed caption access are not regulated by any international board or act. As a result, each country mandates and regulates closed caption access and quality within their own country. Some countries provide strict and clear rules, others provide guidelines and best practices, while others do very little to set standards or requirements, relying on international human rights laws to effectively protect the rights of people with disabilities. Where rules, guidelines and best practices are set, there are differences in what each country considers as priority, whether it be access, verbatim accuracy, or speed, as well as inconsistent requirements for monitoring and reporting. Additionally, it is not always clear what penalties, if any, a broadcaster will face for failing to meet the set standards or guidelines. Consequently, D/HOH viewers have a different viewing experience depending on their geographical location. Without consistent and clear accessibility requirements, the subjective nature of caption quality is further evidence that a verbatim accuracy calculation cannot fully assess the qualitative viewing experience of D/HOH audiences.

2.5 Caption quality

Little academic research has been conducted on the topic of caption quality. While there is considerable evidence that D/HOH viewers are displeased with caption quality and access (e.g., Hersh, 2013; Ohene-Djan, Wright, & Combie-Smith, 2007; Rashid, Vy, Hunt, & Fels, 2008), there is limited academic research assessing the validity and effectiveness of caption quality control measures currently in place. The most notable research studies in this field focus on caption speed and accuracy.

As closed captions became more widely available in the 1980s, D/HOH viewers began to examine the quality of the captions more closely and demand access to all spoken audio content; this was interpreted by captioning companies and broadcasters as a demand for verbatim accurate captions (Jensema, Mccann, & Ramsey, 1996). In a 1996 study, Jensema et

al. looked at caption speed by program type and the verbatim accuracy of the caption text as a measure of quality. They examined a total of 205 programs, ranging in variety from children's animation programs to prime-time dramas, documentaries, and sports, among others. The study looked at caption speed as a function of television program type, and the accuracy levels of the caption text to determine level of editing done by the captioners. The study found that when editing was performed on captions it was largely attributable to program circumstances and technological limitations, rather than careless captioning or deliberate editing policies (Jensema et al., 1996).

More recently is a study by Romero-Fresco (2009), who examines a number of real-time captions produced in the UK for their speed compared to the speed of speech in the program, the amount of editing carried out, and the information lost due to this process. The research concludes that creating verbatim real-time captions at a speed that is legible to the viewer is challenging, and as a result many real-time captions in the UK are edited. The edits that do take place are argued to create minimal loss of information compared to the information viewers would miss attempting to read real-time captions at speeds required for verbatim accuracy (Romero-Fresco, 2009). The primary limitation of this study is that it examines speed and accuracy of captions only and does not take into account any of the other caption elements that impact overall quality and viewer experience, such as font size and colour, caption placement, and graphic attributes. Additionally, the study is largely based on quantitative data regarding caption speeds and reading speeds, and does not incorporate a qualitative user perspective. Thus, a study is needed that will examine caption quality from the perspective of the viewer, and measures quality based on elements other than speed and accuracy.

3. Theoretical Framework

3.1. Inclusive design theory

Designing for disability has a long history, with significant advances often being made as a result of large scale war conflicts (Clarkson & Coleman, 2015). The concept of inclusive design theory first appeared in the 1980s as an attempt to rethink attitudes towards aging and to encourage designers to consider the needs of older people in their designs (Clarkson & Coleman, 2015). It shares many of the same principles and ideologies as human-centered design and universal design theories, such as placing the user at the center of the design and designing for a broad user base. Inclusive design however, not only considers the typical user or even a broad range of users, but specifically considers the needs of people with reduced functional capability during the design process (Persad, Langdon, & Clarkson, 2007). The goal of inclusive design is to design mainstream products and/or services that are accessible to as many people as reasonably possible, without being stigmatising or requiring adaptations (Cremers, Neerincx, & Jong, 2013; Persad et al., 2007). Inclusive design research has gained more attention in recent years, the interest being driven primarily by the aging population and a growing movement to integrate disabled people into society (Clarkson & Coleman, 2015).

The main principles of inclusive design include designing for as many people as possible and doing so from the beginning of the design process. Inclusive design also preferences the concept of customisability over the idea that a single design can suit the needs of people with a variety of capabilities; it reflects a view that inclusivity is more achievable than universality (Newell, Gregor, Morgan, Pullin, & Macaulay, 2010). Two important concepts in inclusive design are the type of user considered and the way that they are considered in the design process. Inclusive design rejects the notion of an 'average user' and instead considers the diverse nature of users' sensory, cognitive, and motor characteristics (Persad et al., 2007). It is also essential that this diverse range of user groups is present and involved in the design process (Cremers et al., 2013). Inclusive design not only means producing a product or service that is inclusive and accessible, but also that the process in and of itself is accessible and inclusive (Bontoft & Pullin, 2003).

Inclusive design presents a strong perspective for designing for people with disabilities, but it is not without criticism. To incorporate a truly inclusive process, users must be recruited to participate in the design process and not just as test participants at the end. Finding appropriate users can be difficult; it can be challenging to recruit them and their participation in the design process can be expensive (Dong, Clarkson, Cassim, & Keates, 2005). Even when end users are recruited, they may be unaware of their needs or unable to articulate them for the designers (Steen, 2011; van Kleef, van Trijp, & Luning, 2005). It is also a design process that challenges the norms, and that at first glance may appear peripheral, which can be uncomfortable for users and designers (Bontoft & Pullin, 2003). Another major challenge to the adoption of inclusive design is encouraging designers to adopt inclusive practices. Research indicates that there is little industry awareness of the benefits of inclusive design, despite the apparent need for more inclusive products (Keates, Clarkson, Harrison, & Robinson, 2000). Furthermore, in industries that are aware of inclusive design practices, they report significant barriers to the implementation of these types of processes (Dong, Keates, & Clarkson, 2004). Reported barriers include fears and concerns, inadequate training and resources, lack of interest and lack of a business case. These challenges are further compounded by the fact that different industries reported different barriers to adopting inclusive design processes (Dong et al., 2004), thus making it difficult to develop any one coping mechanism or solution.

3.2. Inclusive design and closed captions

The adoption of an inclusive design perspective while studying assistive technology helps to ensure that the needs and desires of the end users, normally marginalised and excluded from mainstream design processes, are kept at the forefront of the research objectives. It also ensures that the end product is accessible and useful to a large number of people, not just those in the D/HOH communities, given that good design for people with impairments is often good design for everyone (Gill & Perera, 2003). Several studies that explore ways to address the limitations of close captions have utilised an inclusive approach (e.g. Fels et al., 2005; Lee et al., 2007). In doing so, these research studies have attempted to address the limitations of closed captioning identified by D/HOH viewers by considering them throughout the design process. For captions to be a truly inclusive product, they must be considered and designed as part of

the film or television production process with the involvement of the production team, not in post-production by a captioner alone (Udo & Fels, 2010). The research studies that have taken an inclusive approach have collaborated with the production and creative teams of the video clips being captioned in order to gain insight into the producer's intentions and the non-speech elements that need to be captured. The basic tenets of inclusive design and the prior research that has successfully implemented an inclusive approach to closed captioning indicate the relevance of this theoretical perspective to the research questions at hand.

Adopting an inclusive design approach to close captioning is not without challenges. The limitations and challenges of inclusive design are such that it can be challenging and expensive to engage producers of film and television in the design process of captions, just as it is challenging and expensive to recruit end users. Furthermore, just as it can be challenging for users to articulate their needs, the production team may be unsure of how to articulate the emotions, mood, and setting through the use of emotive captioning features, such as graphics and animation. The challenge of incorporating an inclusive design is further complicated when considering a real-time television environment. During real-time televised programming, the production and teams are occupied with producing the television content and likely will not be able to participate in the captioning process. Due to the spontaneity of live television, it is impossible for the production team to communicate the emotions of the dialogue and video content in advance of it occurring. The nature of live-televised programming presents some unique challenges when considering the use of emotive captions and inclusive design approaches that may be addressed through frequent and open communication between the production team and captioner before and during the televised event.

Adopting an inclusive design approach to caption quality control measures ensures that the views, opinions, and concerns of the end users are considered and incorporated into the control tools used to assess the quality and effectiveness of the captions. This is of particular importance when considering the quality measures used to assess non-quantitative elements of closed captions. Subjective elements of captions, such as placement of captions, font size and colour, use of emotive or descriptive captions, are impossible to access using a verbatim accuracy score. The use of these elements and their impact on the viewer's television viewing

experience are subjective in nature and much more challenging to assess from a quality control perspective. Thus, incorporating feedback and input from the viewers themselves who's experience is affected into the design of the quality control tool is imperative in order to ensure that the tool is capturing and measuring quality appropriately.

4. Research Methodology

4.1. Research question and hypothesis

The purpose of this research study was two-fold, (1) to measure the value that HOH viewers place on having verbatim captions in comparison to other caption factors that affect verbatim accuracy, and (2) to explore how viewers subjectively assess caption quality when these factors vary. The overarching research question guiding this research was what priority do hard of hearing viewers place on verbatim accurate captions as compared to other caption elements? Given that trade-offs exist for caption quality, this thesis investigated what compromises to caption quality viewers were willing to accept in order to obtain verbatim captions for live television.

Specifically, the factors of captioning used in this study were speed, missing words, spelling/grammar errors, and speaker identification as these were identified by the Canadian Radio-television Telecommunications Commission as being the prominent factors defining caption quality. In order for captioners to strive for 100% verbatim accuracy while creating captions for a live televised broadcast, it is possible that compromises must be made in other factors. For instance, the rate or speed at which the captions appear on the screen may be too fast for viewers to read and captions must be shortened or paraphrased to be readable, thus they are no longer verbatim. As such, I wanted to attempt to measure what subjective value viewers place on having verbatim captions in relation to the subjective value of other caption elements, and in doing so begin to explore how caption quality can be measured subjectively.

Currently, caption quality in Canada is measured solely by a verbatim accuracy calculation which compares the differences between the caption file and a show's transcript (CRTC, 2016). However, research shows that there are a number of subjective parameters of caption quality that are not currently assessed using only the verbatim accuracy score used, and that D/HOH viewers are dissatisfied with the caption quality currently available. Thus, the following hypothesis is put forth:

H: When presented with a trade-off decision between verbatim accuracy and other qualitative caption elements, Deaf or hard of hearing viewers will place the highest value

on verbatim accuracy, followed by spelling/grammar errors, captions speed, missing words, and finally speaker identification, in that order.

To test this hypothesis, a three-part questionnaire was designed based on the trade-off principles of the NASA-TLX workload assessment tool (Hart & Staveland, 1988). The research study undertaken herein was deductive exploratory research that explored the beginning stages of assessing the impact of different caption elements on viewer enjoyment beyond that of just verbatim accurate captions. This study used a mixed-methods approach, combining quantitative and qualitative data collection and analysis techniques. Data were collected through an online questionnaire that included closed- and open-ended questions to determine viewer opinions and preferences towards the trade-off of caption element quality in live broadcasting. Data analysis consisted of statistical analysis of close-ended questionnaire responses, as well as thematic analysis of the open-ended questionnaire responses. The thematic analysis from the open-ended questions was used only to enhance the interpretation of the results of the statistical analysis.

As the population of interest is D/HOH and it is difficult to recruit sufficient participants from one geographic area, Internet recruitment using list serves and Mechanical Turk allowed for an increased pool from which to recruit participants. This study was approved by the Ryerson Ethics Board (see Appendix E for the ethics approval letter).

4.2. NASA-TLX and questionnaire development

4.2.1. NASA-TLX

The basis for this study's data collection and data analysis structure is the NASA-TLX workload assessment tool (Hart & Staveland, 1988). Known widely as a tool for assessing subjective mental workload, the NASA-TLX has been used extensively and frequently in human performance studies (Noyes & Bruneau, 2007). Since its inception in 1988, The NASA-TLX has been used in a variety of fields, including the evaluation of visual and/or auditory displays (Cao et al., 2009). The NASA-TLX is designed to reduce the variability found in traditional subjective mental workload rating scales, such as the Modified Cooper-Harper scale and the Overall Workload scale, by encouraging participants to express their judgement in a standardised

format and adopt evaluation criteria imposed by the researcher (Hart & Staveland, 1988). If we consider that there is a mental workload associated with reading comprehension, and that it is different than auditory comprehension processes (Buchweitz et al., 2009), it is possible to use the NASA-TLX as a foundation from which to develop the subjective assessment rating tool used in this study.

The NASA-TLX assessment tool is a survey that consists of two parts: rating and weights. First, weights are determined by asking participants to choose between pairs of six different mental workload factors which of the two factors in any specific pair they find to be the most important contributor to their mental workload. Next, participants complete a task and then assign numerical ratings from 1 to 100 (least to most taxing) for each of the mental workload elements they experience during that specific task. The weights and ratings are then combined to calculate the overall mental workload score (Cao et al., 2009). The higher the workload score, the more mentally taxing the task.

For this study, participants were asked to choose the more important factor affecting their viewing pleasure between pairs of caption factors made up from the five factors verbatim accuracy, speed, spelling/grammar errors, missing words, and speaker identification (see Appendix D for the paired comparison questions). Next, they were asked to view two different videos and rate, on a scale of 0 to 10, how negatively or positively they thought each caption element affected their viewing pleasure (0 was the most negative and 10 was the most positive).

4.2.2. Pre-study questionnaire

A three-part questionnaire was developed for this study to collect demographic, television viewing preferences, and forced choice questions for pairs of caption quality factors data using the online software Qualtrics, which allowed for the online collection of survey data. The pre-study questionnaire was 15 questions in length. The first section (questions 1 to 13) asked participants for demographic information, including relationship to the D/HOH community, age, education level, and information regarding television and closed captioning viewing habits. This information was primarily used to determine the demographic distribution of the sample and

its representativeness of the population. The second section (questions 14a-j and 15) introduced the participants to a list of closed caption factors and definitions. These elements included (1) verbatim accuracy, as defined by the CRTC (CRTC, 2016), (2) speed of captions, defined for the purpose of this research as the rate at which captions are displayed on the screen (CRTC, 2012), (3) missing words, defined as words spoken during the program that are missing from the caption text and not part of a paraphrase, (4) spelling/grammar errors, including spelling errors and incorrect or misplaced punctuation in the caption text, and (5) speaker identification, defined as text or graphical imagery used to identify the speaking person. In the style of the NASA-TLX weighting system, these elements were presented to the participants in pairs and the participants were asked to select which of the two factors they thought was most important to their viewing pleasure. A total of 10 combinations of pairs were presented to participants. The pre-study question has been provided in Appendix D.

4.2.3. Post-study questionnaire

The post-study questionnaire consisted of 11 questions and was administered twice, each time after the participants viewed a short television news clip that was recorded with the original live-produced captions. The post-study questionnaire asked participants to rate on a 11-point Likert scale how much each of the five factors affected their viewing pleasure from very negatively to very positively. This represented the rating portion of the NASA-TLX. Finally, participants were asked to list their least and favourite parts of the video and provide an explanation as to why in four open ended questions. Appendix D contains the post-study questionnaire.

4.3. Content used in this study

The two video clips used in this study were selected from two different Canadian news broadcasters for their neutral content, short duration, and use of live produced captions. Two separate broadcasters were used in order to obtain a wider breadth of caption style and quality, and to ensure that each of the five caption elements being studied was represented in at least one of the two videos' captions (see Table 4.1). The clips were of weather forecasts, and this genre was selected for its neutral content, and to mitigate potential genre-biases from

participants. Additionally, the overall structure of weather forecasts is relatively constant between broadcasters, thus the two clips were of comparable content. Live weather forecast reports from two different broadcast news shows were recorded with their original live produced captions, these videos were then edited into two 2 minutes clips and hosted on a private YouTube channel in order to limit access to the clips over the Internet for the purpose of the study.

Table 4.1: Caption information for the two videos used in the study

Video Broadcaster	Verbatim Accuracy Score	Word/Errors	Speed (wpm)	Lag/Delay (seconds)
City TV	87%	411/53	132.20	6.90
CTV	75%	358/90	232.03	4.89

4.4. Data collection

4.4.1. Experimental setup

Participant recruitment and participation in the study occurred entirely online. The pre-study questionnaire required approximately 10-minutes to complete and the post-study questionnaire required approximately 15-minutes to complete.

Requirements for participation included access to a computer and Internet web browser. Participants recruited through Mechanical Turk required an Amazon account in order to log in and access the study link.

4.4.2. Data collection

A total of 159 participants were recruited using D/HOH list serves (18 participants) and Mechanical Turk (141 participants). Participants were 18 years of age or older, self-identified as being a member of, or having a relationship to, the D/HOH community, and use or have used closed captions for watching live television. Consent was obtained prior to starting the pre-study questionnaire using a click-through methodology on the Qualtrics questionnaire platform. After consenting to the study, participants were asked to complete the pre-study questionnaire, and once completed they were shown one of two video clips in random order.

The video clips were recordings of live Canadian weather news segments from two different Canadian broadcasters, CityTV and CTV. After viewing the first video clip, participants were asked to complete the first post-study questionnaire. The process was then repeated with the second video clip. A within-subjects design was used so that all participants saw both videos. The three questionnaires were linked with an assigned ID unique to each participant.

4.5. Data analysis

The results of the quantitative data were analysed using SPSS version 23 software. The pre-study data were analysed using a chi square test to determine if the results of the forced choice questions differed from chance, whereas the post-study quantitative data were analysed using a combination of parametric and non-parametric statistics because the data was normally and not normally distributed depending on what factors were being analysed. The qualitative data collected through open-ended questions of the survey were analysed using NVivo software and thematic analysis techniques (Miles, Huberman, & Saldaña, 2013). Themes included the five caption quality factors that were the focus of the study, as well as three new themes that emerged from the data; placement on screen, display style, and lag (see Table 4.2). An inter-reliability test was conducted with 20 percent of the data. Cohen κ was run to determine if there was an agreement between coders on the themes, the kappa coefficient was above .65 for all themes. Non-parametric statistics were used on the quantitative results of the thematic analysis.

Table 4.2: Themes

Theme	Definition	Sub-Categories	Example
Speed	The pace or rate at which the captions appear and disappear from the screen (CRTC, 2012)	Fast	"I do think it went way to fast to read"
		Slow	"captions were a bit to slow to keep up with what was happening on the screen"
Verbatim Accuracy	The word for word representation of the audio content in the captions. The CRTC requires English-language broadcasters to reach a captioning accuracy rate of at least 95% averaged over a given live program. (CRTC, 2016)	Positive	"I liked how accurate the captions were"
		Negative	"Caption and person they said not match"
Missing Words		Positive	"I liked how most of the words were their"

	Words spoken during the program that are missing from the captions.	Negative	"I'm not sure if it was the speed or missing words that made comprehension difficult"
Grammar/Spelling Errors	Errors in the captions including spelling errors and incorrect or misplaced punctuation. (CRTC, 2012)	Positive	"Most of the words were spelled correctly"
		Negative	"I didn't like the over usage of commas"
Speaker Identification	Text used to provide information about who is speaking. (CRTC, 2012)	Positive	"Character identification, and spelling accuracy are best because they help me to understand which character is speaking what and spelling accuracy avoids confusion"
		Negative	"individual identification becomes difficult when multiple people are talking so fast"
Font Style	The size, colour, and font type used for the captions.	Positive	"Font and the black background made it easy to read"
		Negative	"Colors were not that good"
Placement on Screen	The location of the captions on the screen in relation to the video content (CRTC, 2012)	Positive	"The placement it did not cover up the bottom where names and information was listed"
		Negative	"Better not to have the captioning because it only covered things I could have seen"
Display Style	The stylistic display of the captions appearing and disappearing on screen, this can include scrolling, roll up, etc.	Positive	"I liked the scrolling effect of the captions, far superior to just flashing them on-screen"
		Negative	"The way the captions appeared made it hard to follow the train of text"
Lag	The delay between the audio content and the captions appearing on the screen. (CRTC, 2012)	Positive	"I liked that the captions kept up"
		Negative	"the captioning was way off so it was hard to follow"

4.6. Data observations and caption preference scores

In total, 327 survey responses were collected between December 2017 and March 2018, however, upon examination of the responses, 168 were discarded based on the following criteria. Responses that indicated the participant did not agree to the terms of the survey or exited the survey before completion were discarded, as per the consent agreement.

Participants who completed the survey in less than five minutes were assumed to have not watched the videos in full and their results were discarded. Responses were also discarded from participants who indicated that they never use closed captions while watching live television. A resulting 159 survey responses were kept and underwent further analysis.

Responses were separated into two categories, those who indicated they have a relationship to

the D/HOH community, and those that indicated that they do not. Both sets of data were sorted based on the same criteria; 78 survey responses in the category indicated they belong to or have a relationship to the D/HOH community (D/HOH group), and 81 survey responses in the category indicated no relationship (hearing group).

The survey responses from participants in the D/HOH group were of primary importance to this study. Of this group, 49 participants identified as hard of hearing, 7 participants identified as Deaf, 2 participants identified as Deafened, and 20 participants indicated Other (see Figure 4.1). The Other response was provided for instances where participants did not identify as Hard of Hearing, Deaf, or Deafened, but still felt a strong relationship to the D/HOH communities. Participants who selected Other noted a variety of ties to the D/HOH community, including being deaf in one ear, or having family and/or friends who are members of the community, and having employment in or with the D/HOH community. Of this group, 31 participants identified as male, 92 participants identified as female, and one participant preferred not to answer. All participants in the D/HOH group had attended high school; 12 participants indicated their highest level of education was high school, 48 had completed college or a university undergraduate degree, and 18 had completed graduate school (see Figure 4.2). Regarding caption use for live televised programming, participants were evenly divided between always and sometimes using closed captions. Of this group, 59% of participants were under the age of 40 (see Figure 4.3 for age distribution), and 56% indicated they watched between 0 and 10 hrs of television per week (see Figure 4.4 for distribution of TV watching data).

Of the participants in the hearing group, 30 participants identified as male and 51 identified as female. Regarding caption use for live televised programming, 19 participants always use closed captions, and 62 sometimes use closed captions. Of the hearing participants, 12 indicated their highest level of education was high school, 57 had completed college or a university undergraduate degree, 10 had completed graduate school, and 2 preferred not to answer (see Figure 4.2). Of this group, 73% of participants were under the age of 40 (see Figure 4.3), and 65% indicated they watched between 0 and 10 hours of television per week (see Figure 4.4 for distribution).

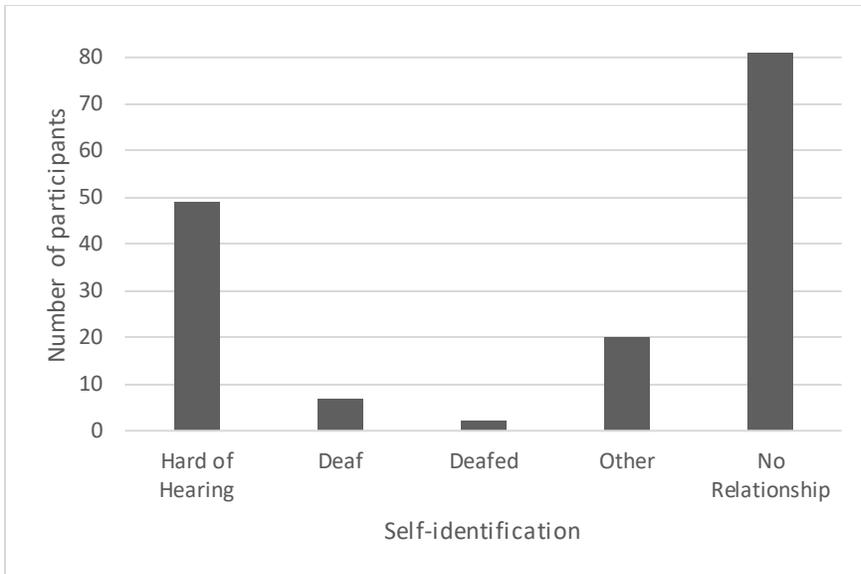


Figure 4.1: Relationship of participants to the D/HOH community

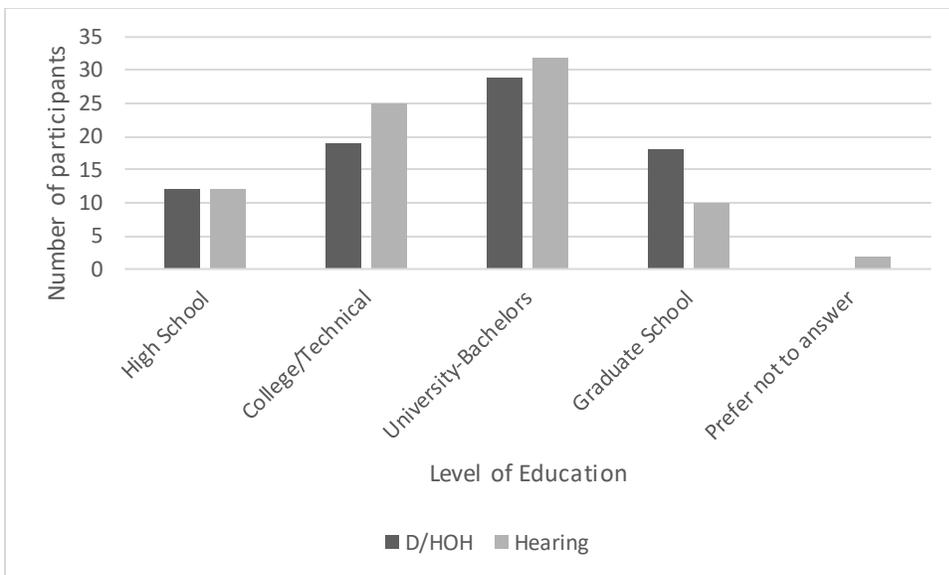


Figure 4.2: Level of education for D/HOH and hearing groups

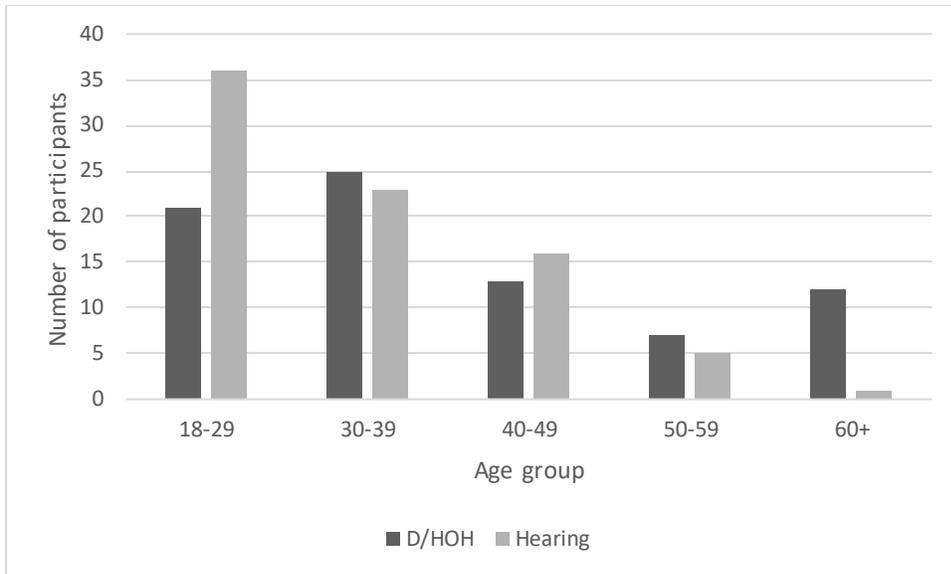


Figure 4.3: Ages of participants for D/HOH and Hearing groups

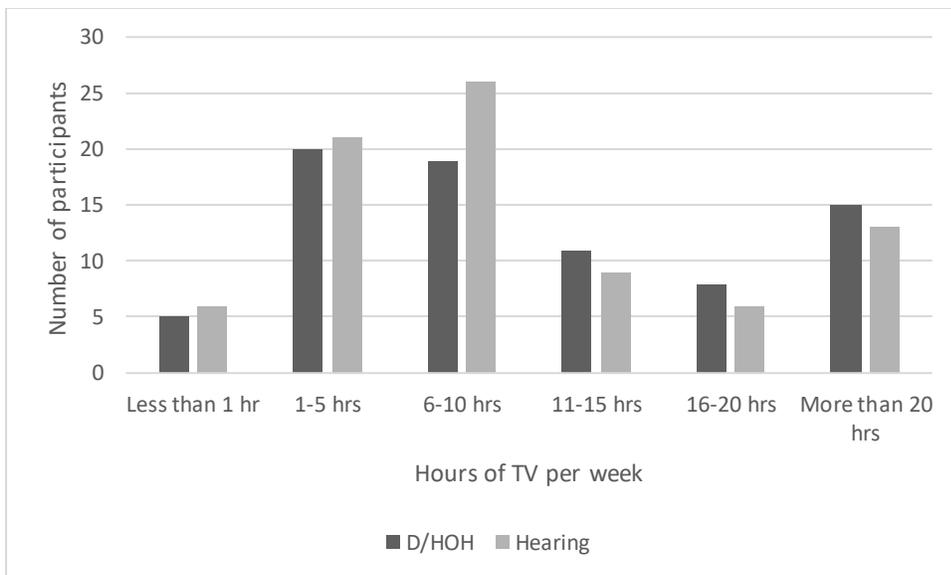


Figure 4.4: Hours of TV watched per week for D/HOH and Hearing groups

Using the NASA-TLX calculations as a guide, the caption preference scores were calculated for each of the caption quality elements. First, the weighted scores were calculated by summing the total number of times each element was selected from the 10 pairs presented in the pre-study questionnaire (see Equation 2). Second, the rating scores were calculated. Since participants were asked to rate each element on a scale of 0 to 10 twice, once for each video, the scores were calculated separately for each element and then an average taken (see

Equation 3). The original Likert scale used in the questionnaire was set as 0 to 10 with 0 indicating the element very negatively effected the participant's viewing pleasure, and 10 indicating the element very positively effective the viewing pleasure. For the purpose of statistical analysis, this scale was reversed and shifted up by one point. Thus, the final scale used in the analysis of the results was set as 1 to 11, where 1 indicates very positively, and 11 indicated very negatively. In doing so, the scale was now similar to that of the NASA-TLX where the higher the number, the more mentally taxing the task. Finally, the caption preference scores were calculated for each element by multiplying the weighted score by the rating score and dividing by the number of pairs (see Equation 4).

$$\textit{Weighted Pairs Score} = \sum \textit{Forced choice selection} \quad (2)$$

$$\textit{Rating Score} = \frac{\sum \textit{Ratings}_{V1} + \sum \textit{Ratings}_{V2}}{2} \quad (3)$$

$$\textit{Caption Preference Score} = \frac{\textit{Weighted Pairs Score} \times \textit{Rating Score}}{\textit{Number of Pairs}} \quad (4)$$

5. Results

The survey results collected from participants were separated into three independent sets of data. The first set of data originated from the pre-study questionnaire and was comprised of the weighted pair scores. The second set of data was from the post-study questionnaire and was comprised of the rating scores for the two real-time videos. The third set of data also originated from the post-study questionnaire and was comprised of the qualitative responses to the open-ended questions.

5.1. Caption preference scores

The caption preference scores were calculated using the data collected in the pre-study and post-study questionnaires. As per the methodology outlined in section 4.6 the weighting scores obtained from the pre-study questionnaire were multiplied by the rating scores obtained from the post-study questionnaire and then divided by the total number of pairs (see Equation 4). Due to the within-subject design all participants watched and rated the caption elements for both videos. As there was no learning effect for any of the caption quality elements (see Appendix C, Table 8.1), an average of the scores was used in the final calculation.

$$\text{Caption Preference Score} = \frac{\text{Weighted Pairs Score} \times \text{Rating Score}}{\text{Number of Pairs}} \quad (4)$$

The results showed that participants within the D/HOH community rated fast moving captions (high speed) as having the most negative impact on their viewing pleasure (10,544), followed closely by verbatim accuracy (10,114), slow moving captions (9,133), missing words (8,771), spelling and grammar mistakes (4,951), and finally speaker identification (3,523).

Table 5.1: Caption preference scores for D/HOH group

Caption Element	Rating Score (V1)	Rating Score (V2)	Rating Score (Average)	Weighted Pairs Score	Caption Preference Score
Fast Captions	484	525	504.5	209	10,544
Verbatim Accuracy	528	504	516	196	10,114
Slow Captions	431	443	437	209	9,133
No Missing Words	521	493	507	173	8,771
No Spelling/Grammar Errors	458	485	471.5	105	4,951
Speaker Identification	363	371	367	96	3,523

Participants with no relationship to the D/HOH community rated slow moving captions as having the greatest impact on their viewing pleasure (10,394), followed by fast moving captions (10,107), verbatim accuracy (8,936), missing words (8,602), spelling and grammar mistakes (5,985), and finally speaker identification (4,680). These results suggest that speed and verbatim accuracy are most importance to participants and their viewing pleasure.

Table 5.2: Caption preference scores for hearing group

Caption Element	Rating Score (V1)	Rating Score (V2)	Rating score (Average)	Weighted Pairs Score	Caption Preference Score
Slow Captions	525	489	507	205	10,394
Fast Captions	503	483	493	205	10,107
Verbatim Accuracy	492	490	491	182	8,936
No Missing Words	494	524	509	169	8,602
No Spelling/Grammar Errors	481	469	475	126	5,985
Speaker Identification	383	378	380.5	123	4,680

5.2. Results of the D/HOH group

5.2.1. Video comparisons

The CTV and City videos (from herein to be referred to as CTV and City respectfully) were selected to be similar in content, both were weather reports, and both had quality issues with their captions, however there were different caption elements represented in different ways in each of the two videos (see Appendix A for caption transcripts and quality issues). Both videos were shown to participants (within-subjects design) and shown in a random order. After conducting a Shapiro-Wilkes test and determining that the data were normally distributed (see Appendix C, Table 8.1) a paired t-Test was conducted to test for learning order effect between the two videos. There was no significant difference between the post-study responses for the first video viewed and the second video viewed (see Appendix C, Table 8.2). The results suggest that there was no order effect between participants watching the first video and the second video presented to them.

A paired t-test was conducted to compare the City rating score results and the CTV rating score results for the rating scores participants assigned each video’s caption elements. There

was a significant difference in the rating scores for fast moving caption, slow moving captions, missing words, spelling/grammar, and verbatim (see Table 5.3). There was no significant difference for speaker ID. Descriptive statistics for mean and standard deviation for each video per element are displayed in Table 5.4. The mean for each caption element is lower for CTV than for City, indicating the caption elements were rated more positively for the CTV video.

Table 5.3: Paired t-test for City and CTV rating scores by the D/HOH group

	t	df	p
Verbatim accuracy	2.78	69	.007
Fast moving captions	3.62	67	.001
Slow moving captions	3.41	62	.001
Missing words	2.31	68	.024
Spelling and grammar errors	2.02	68	.047
Speaker identification	1.78	60	.080

Table 5.4: Mean and standard deviation for City and CTV rating scores by the D/HOH group

	Video	M	SD
Verbatim accuracy	City	7.46	2.57
	CTV	6.63	2.37
Fast moving captions	City	7.60	2.72
	CTV	6.26	2.17
Slow moving captions	City	6.89	2.64
	CTV	5.79	2.40
Missing words	City	7.26	2.50
	CTV	6.55	2.03
Spelling and grammar errors	City	6.78	2.23
	CTV	6.22	1.88
Speaker identification	City	5.93	1.88
	CTV	5.38	1.99

5.2.2. Trade-offs and impacts on viewing pleasure

The pre-study questionnaire data consisted of the weighted pairs scores, where participants were asked to select from each pair the element that they thought had the greatest impact on their viewing pleasure. A chi square test was conducted to determine whether the results of the weighted pair scores differed from chance. The degree of freedom for each test was one. As displayed in Table 5.5, the following pairs had significant chi-square results: verbatim accuracy

vs missing words, verbatim accuracy vs spelling and grammar errors, verbatim accuracy vs speaker identification, spelling and grammar errors vs speed, spelling and grammar errors vs missing words, missing words vs speaker identification, and speaker identification vs speed. Verbatim accuracy vs speed, spelling and grammar errors vs speaker identification, and the missing words vs speed pairs were not significant.

Table 5.5: Chi-square results for weighted pair analysis

	Verbatim v Missing	Verbatim v Spelling/ Grammar	Verbatim v Speaker ID	Spelling/ Grammar v Speed	Spelling/ Grammar v Missing	Missing v Speaker ID	Speaker ID v Speed
Chi-Square	6.21	5.13	13.13	32.05	18.51	14.82	27.13
Sig.	.013	.024	.000	.000	.000	.000	.000

The weighted pair score results were examined in relation to different demographic factors. Figure 5.1 shows the various participant numbers between speed and accuracy for the different demographics. The data suggest that there may be some relationship between education level and preference for accuracy; participants who were college and university educated selected accuracy more frequently than speed, whereas participants with high school level education selected speed more frequently.

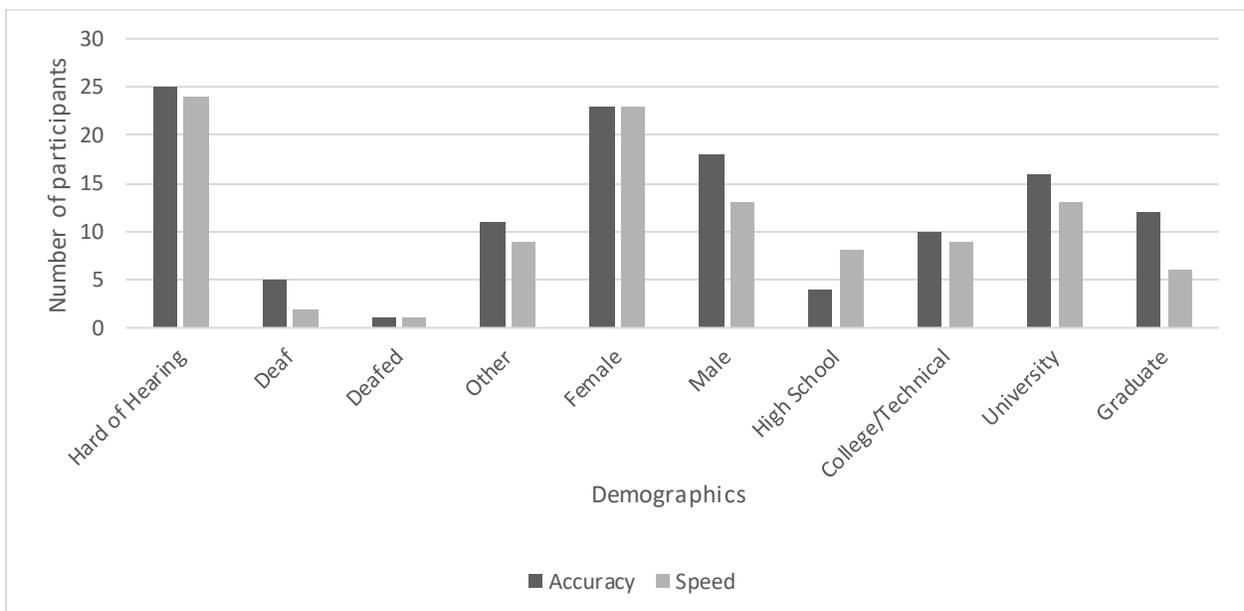


Figure 5.1: Demographic breakdown of participants selecting Accuracy compared to Speed

The post-study questionnaire asked participants to rate on scale of 0 to 10 the impact of each caption element on their viewing pleasure. Because there was no order effect, I combined the data collected on the two videos and analysed how each caption quality element impacted viewer pleasure. A Shapiro-Wilkes test was conducted and the data was found to be of non-normal distribution (see Appendix C, Table 8.3), thus a Friedman test was conducted to compare the rating scores between each caption factor for the combined data set. There was significant difference between the impact of each element on the viewing pleasure of the participants $\chi^2(5)=29.858, p<.001$.

Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significant level set at $p < 0.03$. The results of the Wilcoxon signed-rank indicated that there was significant difference between speaker ID and all other caption elements, as well as between spelling/grammar errors and all elements other than slow moving captions (see Table 5.6).

Table 5.6: Impact of caption quality factors for D/HOH group

	Speaker ID vs Verbatim	Speaker ID vs Fast	Speaker ID vs Slow	Speaker ID vs Missing	Speaker ID vs Spell/Gram
Z	-5.69	-2.45	-3.51	-5.65	-4.95
p	.000	.000	.000	.000	.000

	Spell/Gram vs Fast	Spell/Gram vs Missing	Spell/Gram vs Verbatim	Slow vs Verbatim
Z	-2.45	-2.73	-3.04	-2.74
p	.014	.003	.002	.006

Mean perceived affect of each element on viewing pleasure was verbatim accuracy (M=6.82, SD=2.50), missing words (M=6.72, SD=2.28), fast moving captions (M=6.63, SD=2.51), spelling and grammar errors (M=6.37, SD=2.12), slow moving captions (M=6.19, SD=2.49), and speaker identification (M=5.59, SD=2.05). These results suggested that participants of the D/HOH community found that verbatim accuracy and fast moving captions had the greatest negative impact on their viewing pleasure, which was supported by the findings of the caption preference scores. The results of the Wilcoxon signed-rank suggested that speaker ID had the least impact on viewer enjoyment, compared to the other caption quality factors.

5.3. Results of the hearing group

5.3.1. Video comparisons

The data collected from the hearing group was analysed using the same methodology as the D/HOH group data. As with the D/HOH groups, City and CTV were shown to hearing participants in random order. After determining that the data were normally distributed by conducting a Shapiro-Wilkes test (see Appendix C, Table 8.1), a paired t-test was conducted to test for learning order effect between the two videos. There was no significant difference between the post-study responses for the first video viewed and the second video viewed (see Appendix C, Table 8.2). The results suggested that there was no order effect between participants watching the first video and the second video presented to them.

The rating scores for City and CTV were further analysed to compare the rating score results participants assigned for each video's caption elements. Descriptively, the mean rating scores for CTV were slightly higher than those for the City for all six caption quality factors (see Table 5.7). After determining that the data were of normal distribution using a Shapiro-Wilkes test (see Appendix C, Table 8.1) a paired t-test was conducted to compare the rating scores of City and CTV. There was no significant difference in the rating scores for each of the six caption quality factors. These results suggested that unlike the D/HOH participants, the hearing participants did not significantly prefer one style of captions over the other.

Table 5.7: Mean and standard deviation for City and CTV rating scores by Hearing group

	Video	M	SD
Verbatim accuracy	City	6.40	2.37
	CTV	6.61	2.53
Fast moving captions	City	6.24	2.76
	CTV	6.62	3.07
Slow moving captions	City	6.88	2.47
	CTV	7.01	2.79
Missing words	City	6.59	2.19
	CTV	6.93	2.24
Spelling and grammar errors	City	6.16	1.95
	CTV	6.37	2.25
Speaker identification	City	5.32	2.25
	CTV	5.78	1.96

5.3.2. Trade-offs and impacts on viewing pleasure

The pre-study questionnaire data consisted of the weighted pairs scores, where participants were asked to select from each pair the element that they thought had the greatest impact on their viewing pleasure. A chi square test was conducted on the weighted pairs data to determine whether the results differed from chance. As displayed in Table 5.8 below, the following pairs were shown to be significantly different than chance: verbatim accuracy vs spelling and grammar errors, spelling and grammar errors vs speed, spelling and grammar errors vs missing words, missing words vs speed, and speaker identification vs speed. The test also revealed that verbatim accuracy vs speed was found to not differ than chance, as well as the following three pairs; verbatim accuracy vs speaker identification, spelling and grammar errors vs speaker identification, and missing words vs speaker identification. The results suggested that hearing viewers placed nearly equal value of importance on verbatim accuracy and speed, and on spelling/grammar errors and speaker ID, like their HOH counterparts. However, they also placed nearly equal value of importance on verbatim accuracy and missing words, on verbatim accuracy and speaker ID, and on missing words and speaker ID, unlike the D/HOH participants.

Table 5.8: Chi-square results for weighted pair analysis

	Verbatim v Spelling/ Grammar	Spelling/ Grammar v Speed	Spelling/ Grammar v Missing	Missing v Speed	Speaker ID v Speed
Chi-Square	4.46	10.38	9.00	7.72	15.12
p	.035	.001	.003	.005	.000

A Shapiro-Wilkies test was conducted and determined that the data was not normally distributed (see Appendix C, Table 8.3), thus a Friedmans test was conducted to compare the rating scores between each caption factor for the combined data set (as there was no order effect). There was significant difference between the impact of each element on the viewing pleasure of the hearing participants $\chi^2(5)=34.397, p<.001$.

Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significant level set at $p < 0.03$. There was significant difference between speaker ID and all other caption elements, as well as between slow moving captions and almost all other caption elements. (see Table 5.9). These results suggested that speaker ID, and spelling and grammar errors had the least amount of impact on viewer enjoyment.

Table 5.9: Impact of caption quality factors for Hearing group

	Speaker ID vs Slow	Speaker ID vs Verbatim	Speaker ID vs Missing	Speaker ID vs Spell/Gram	Speaker ID vs Fast
Z	-4.97	-3.97	-5.05	-3.74	-2.24
p	.000	.000	.000	.000	.025

	Slow vs Verbatim	Slow vs Fast	Slow vs Spell/Gram	Spell/Gram vs Missing
Z	-3.11	-3.63	-3.28	-3.16
p	.002	.000	.001	.002

Mean perceived affect of each element on viewing pleasure were missing words (M=6.65, SD=2.34), slow moving captions (M=6.59, SD=2.56), verbatim accuracy (M=6.27, SD=2.45), spelling and grammar errors (M=6.18, SD=2.21), fast moving captions (M=6.03, SD=2.75), and speaker identification (M=5.53, SD=2.20). The results suggested that hearing participants found missing words and slow moving captions had the greatest impact on their viewing pleasure.

5.4. Thematic analysis

The open-question responses of the post-study questionnaires underwent thematic analysis using NVivo software version 11. In addition to the five subjective caption elements identified in the hypothesis of this research study (verbatim accuracy, speed, spelling and grammar errors, missing words, and speaker identification), four more caption elements were revealed as themes in the qualitative data: (1) lag time, (2) font style, (3) display style, and (4) placement on screen. The D/HOH group commented most frequently on display style (total of 42 comments), speed (38), lag (35), and verbatim accuracy (15) (see Figure 5.2). The hearing group mostly commented on lag (27) and speed (17) (see Figure 5.3).

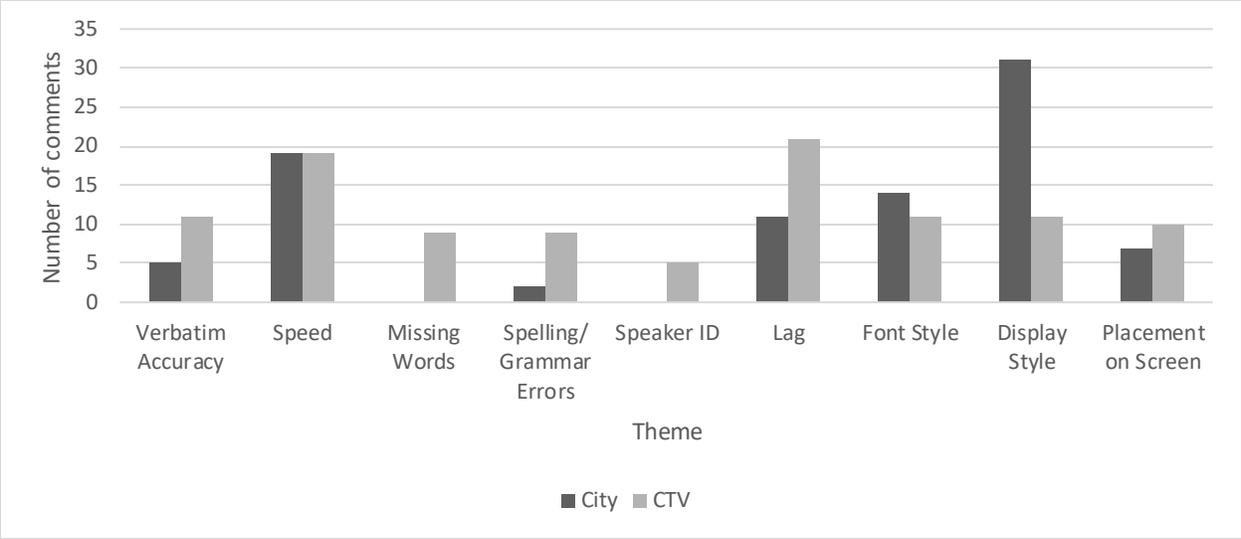


Figure 5.2: Frequency of theme occurrences for each video for D/HOH group

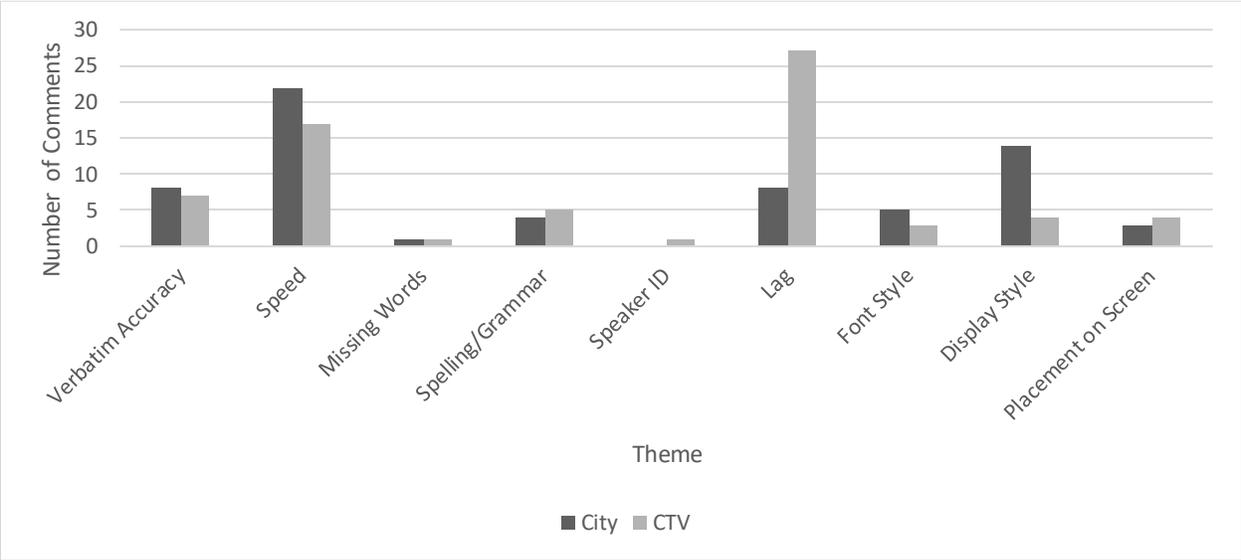


Figure 5.3: Frequency of theme occurrences for each video for hearing group

Speed, being one of the five elements highlighted in this research study, as well as being identified in my survey as one of most impactful caption elements related to enjoyment was examined in more detail. Speed related comments were further categorised into three subthemes; “too fast” (captions display too quickly), “too slow” (captions display too slowly), or “unknown/neutral” (comment on speed was of a general nature and did not fit a too fast or too

slow categorisation). Tables 5.11 and 5.12 shows the frequency of occurrences of comments related to these sub-themes. D/HOH viewers were 1.5 times more likely to comment that the captions were too fast, whereas the hearing viewers were twice as likely to comment that the captions were too slow. This supported the Friedman results that suggested fast moving captions negatively impact D/HOH viewers' pleasure, where as the hearing participants found slow moving captions to affect them most negatively.

Due to the limited data points collected in the thematic analysis non-parametric statistics were used in their analysis. A Mann-Whitney U test was conducted to determine the difference between the groups' comments for each element. There was a no significant difference between the number of comment in the D/HOH group's and the hearing group's comments for all three categories of speed comments (too fast, too slow, and neutral/unknown).

Table 5.10: Frequency of each category of speed comments

Speed	D/HOH Group	Hearing Group
Too Fast	16	9
Too Slow	11	20
Neutral/Unknown	11	10

Examining the frequency of comments between the two videos, the City video received 63 negative comments from the D/HOH groups, and 26 positive comments (see Table 5.11) whereas the CTV video received 57 negative comments and 49 positive comments. For the hearing group, the thematic data would suggest that they too had a slight preference for the CTV video (see Table 5.12). A Wilcoxon sign-rank test was conducted on the frequency of comments and there was no statistical significance. However, the trend in the comments supports the quantitative data (see Sections 4.1 and 4.2).

Table 5.11: Positive and negative comments by video for the D/HOH group

	Positive	Negative	Total
City	26	63	89
CTV	49	57	106

Table 5.12: Positive and negative comments by video for the hearing group

	Positive	Negative	Total
City	11	55	66
CTV	29	39	68

6. Discussion

The primary purpose of this research study was to measure the value that D/HOH viewers place on verbatim accurate captions when other caption elements are affected to achieve that level of accuracy. Using the NASA-TLX as a methodological model for determining user priorities in a forced-choice scenario and the resulting weighted pairs score to modify user rating scores accordingly showed that D/HOH viewers place high importance on multiple caption quality factors, beyond verbatim accuracy. The preferences that were established by the caption preference scores model developed in my research were also seen in the thematic analysis of the open-ended questions in the post-study questionnaire. For example, the post-study questionnaires were able to determine that accuracy and speed were the two factors D/HOH viewers considered to be most importance to their viewing pleasure. Thematic analyses of the open-ended questions also identified two additional caption quality factors, display style and lag time, that impacted viewers' pleasure and understanding of the video content, providing support to the argument that caption quality is impacted by more than just verbatim accuracy, and that a more comprehensive measure of caption quality is needed.

The results of this study indicated that D/HOH viewers care about more than just accuracy for captions, and that they placed about an equal amount of importance on caption speed. The caption preference scores for the D/HOH group placed fast moving captions, verbatim accuracy, and slow moving captions as the three caption quality factors that most negatively impact a viewer's enjoyment. When selecting between the verbatim accuracy and speed pair, D/HOH participants selected verbatim accuracy a total of 42 times and speed a total of 38 times which was not significantly different from chance. While the data collected in this study showed that D/HOH viewers place nearly equal value of important on verbatim accuracy and speed, a conclusive hierarchy of the five caption elements studied was not determinable. However, the results suggested that viewers placed the highest importance on verbatim accuracy and caption speed, and the least amount of importance on speaker identification. Efforts to determine the order of importance that D/HOH viewers placed on missing words and spelling/grammar errors

were inconclusive. Further research with additional data would be required to determine a definitive order of importance.

The findings related to viewer preference for caption display at a readable speed are supported by existing literature on the topics of reading speeds of people who are Deaf, as well as the limited research conducted on caption speed and accuracy. Romero-Fresco (2009) finds that verbatim captions are, in many instances, too fast for Deaf viewers to read and comprehend for real-time captions. He argues that in order to provide captions at readable speed, they must be edited, and that these edits should create minimal loss of information. This presents a trade-off being made on behalf of viewers, one that values speed over verbatim captions. While his results support my findings that caption speed is highly valued by D/HOH viewers, the results of my study suggest that a trade-off between verbatim and speed cannot be easily made. Participants place equal value on speed and verbatim accuracy; half indicate they would compromise accuracy for readable speed, while the other half indicate they would compromise speed in order to have every word spoken captured in the captions. This presents a challenge by not providing clear direction for production companies when they must consider this trade-off. However, this divisiveness supports the argument for a more comprehensive measure of caption quality, the need for user input to be considered when assessing caption quality, and more flexible caption display options. Broadcasters may be able to address this issue with the growing popularity of smartTVs (or interactive televisions), televisions with sophisticated display settings and a range of adjustment options available to viewers so that they can customise their personal viewing experience (Piccolo, Melo, Cecília, & Baranauskas, 2007). Broadcasters may be able to capitalise on these devices by working with television technology to provide captions that can be customised and adjusted by individual viewers to display faster or slower, allowing each viewer to make their preferred individual trade-off. By providing better caption quality, and user customisation options, broadcasters may be providing a better customer and consumer experiences for not only D/HOH viewers, but all viewers that use closed captioning.

Participants who do not identify with the D/HOH community also place a high value of importance on speed and accuracy. The results for the hearing group suggest that they place

the highest amount of importance on slow moving captions, missing words, and verbatim accuracy, and the least amount of importance on spelling/grammar errors and speaker identification. These results are supported, and may be explained, by existing knowledge on hearing and D/HOH reading speeds. A number of studies show that Deaf people, in general, have poor literacy comprehension levels and slower reading speeds than their hearing peers (e.g. Albertini & Mayer, 2011; Wilbur, Goodhart & Fuller, 1989; Parault & Williams, 2010; Strong et al., 1997). This may explain why D/HOH viewers found fast moving captions to negatively impact their viewing pleasure in most cases, whereas hearing viewers found slow moving captions to negatively impact their viewing pleasure in most cases. Hearing viewers may also be more affected by missing words than D/HOH viewers because they are more aware of which words have been missed out from the caption when they can hear the audio. However, it is challenging to draw definitive conclusion due to the disproportionate number of hard of hearing participants compared to Deaf participants.

From the qualitative data collected after viewers had watched the City and CTV videos, speed was the second most commented on caption element by D/HOH participants, after display style (38 out of 195, 19%, were related to the speed of the captions). Viewers seemed to find the captions in the videos moved too quickly to be read and understood comfortably. Of these comments, 45% (17 of 38) stated the captions were too quick, 26% (10 of 38) stated the captions went too slow, and 29% (11 of 38) were of neutral or unknown opinion. In comparison, the hearing group had a similar number of comments on speed as the D/HOH group (39 comments), however, the majority of comments 49% (19 of 39) indicated viewers found the speed too slow, 23% (9 of 39) found them too fast, and 28% (11 of 39) were of neutral or unknown opinion. A Mann-Whitney U test found that the difference between the D/HOH group's comments on speed and those of the hearing group were not significantly different, however this may be due to the small sample size and limited number of data points. Further research that uses a larger sample would be required to determine whether differences exists. Despite not being significantly different, the descriptive results were in keeping with the literature on reading speeds of D/HOH individuals compared to their hearing counterparts. For example:

[Participant ID 44, D/HOH] CTV: The captions moved much faster than I like making it difficult to read and process the information as it was being received.

[Participant ID 76, D/HOH] City: Least favourite was the speed of the captioning was so fast that I can't comment on what effect it any there was with the lack of accuracy, missing words, etc.

[Participants ID 17, D/HOH] CTV: I was hoping the speed would be a little faster.

The difference in how D/HOH participants and hearing participants experienced the speed of the captions is supported and explained by research in caption and reading speeds. Szarkowska, Krejtz, Klyszejko, & Wieczorek, (2011) used eye-tracking methodology to determine how long D/HOH and hearing participants spent reading different types of captions, with length of fixation time being proportionality related to how long it takes the viewer to read the caption. They found that all participants had longer fixation times on verbatim accurate captions than on the edited captions, but that Deaf participants had a significantly longer fixation times than the other groups of participants (Szarkowska et al., 2011). Other studies have found that Deaf persons are often slower readers due to having to read captions in a language other than their primary one (primary language being American Sign Language or other signed language) (Albertini & Mayer, 2011; Parault & Williams, 2010; Strong et al., 1997). Studies in psychology have shown that reading rate was determined by the number of eye fixations per unit of text (Huey, 1968; Javal, 1878), and that faster readers made fewer eye fixations, suggesting they were able to process more text per reading fixation (Gilbert, 1959). This could explain why the D/HOH group made 1.5 times as many comments related to the captions being too fast than being too slow, whereas the hearing group made twice as many comments stating the captions were too slow than too fast. The results corroborate our knowledge that hearing people are, in general, faster readers than people who are Deaf. However, this study did not measure reading speed, and further research that measures individual reading speed as part of a caption study would be required in order to make definitive conclusions.

Szarkowska et al. (2011) also found that while edited captions were the easiest to process for all groups of viewers (based on fixation times), the highest comprehension scores were obtained when the video clips used verbatim accurate captions and lowest for the video clips

that used edited captions. These results may explain why viewers rated verbatim accuracy so highly as a priority, despite needing more time to read the verbatim text.

The thematic analysis of the post-study questionnaire responses to the open-ended questions confirmed existing findings that verbatim accuracy of captions is important to viewers enjoyment and understanding of a television program (Romero-Fresco, 2009; Szarkowska et al., 2011). Of the comments made by the D/HOH group, 16 of 195 (8%) related to verbatim accuracy. Participants made three times more positive comments than negative comments (12 positive and 4 negative) regarding accuracy. Comments suggested that viewers liked when all the audio was captured in the captions, and found it confusing to follow when words and information were missing. The findings also supported the results of the trade-off analysis, which showed that D/HOH viewers place a high level of importance on verbatim accuracy. For example:

[Participant ID 50, D/HOH] City: It seems they tried to caption all the words, makes it easy to understand.

[Participant ID 69, D/HOH] CTV: The missing words made it impossible to know what the full information was... in this case paraphrasing and leaving out the dropped thoughts would definitely have been better for understanding.

The secondary purpose of this research study was to explore how caption quality can be measured and valued by viewers beyond the CRTC's verbatim accuracy calculation (CRTC, 2016). In addition to examining how viewers rate speed, spelling and grammar errors, missing words, and speaker identification in comparison to accuracy, other caption quality factors emerged from the qualitative data. The caption quality factors of display style, lag, font style, and placement on screen were also found as themes in the qualitative comments in the post study question. While the five caption quality factors originally selected for this study were quantifiable (e.g., verbatim accuracy was found using the calculation from the CRTC, speed was calculated as words per minute, spelling and grammar errors can be counted on a per word basis, dropped words can also be counted, and speaker identification was either present or not), of these four new factors three of them, display style, font style, and placement on screen, were subjective. D/HOH viewers were twice as likely to make negative comments about the

display style than a positive one (28 negative comments to 14 positive ones), and three times as likely to make negative comments about the lag/delay (24 negative comments to 8 positive ones). For example:

[Participant ID 62, D/HOH] City: It felt like the sentences were dancing. The lines would switch from top to bottom to top and I didn't know where to look.

[Participant ID 67, D/HOH] CTV: I loved these captions. They were smoothly scrolling more readable and easy to follow.

[Participant ID 9, D/HOH] CTV: Lag in appearing captions and only showing one line at a time. I like when two or three lines will appear in chunks.

The number of comments regarding the subjective quality factors suggest that viewers have strong opinions regarding the overall esthetic of the caption display, and that these opinions are not being captured in current caption quality measurements. These findings support my argument presented in chapter 2 that current measures of caption quality do not take into account all factors of closed captioning that affect viewer pleasure. Broadcasters may want to explore new methods for measuring caption quality, such as the survey presented in this study, that include subjective quality factors and use real-viewer feedback to assess quality.

This study used two video clips from two different Canadian broadcasters in order to include a broader range of caption styles and possible caption quality errors. The ratings D/HOH participants gave each video for caption quality had significant differences (see Table 5.3), suggesting that viewers preferred the style of one set of captions over the other. The mean for each caption element was lower for CTV than for City captions (see 5.4), which suggests that D/HOH viewers preferred the caption style of the CTV video over that of the City video. The results from the hearing group were not significant, suggesting they had no preference of caption style between the two videos. These findings were corroborated by the thematic analysis of the participant comments. Participants in the D/HOH group commented a total of 89 times on the City video, of which 71% were negative and 29% were positive. They commented on the CTV video a total of 106 times, and of these comments 46% were positive while 54% were negative. For example:

[Participant ID 71, D/HOH] City: Captions dropping instead of rising. Very hard to read and follow.

[Participant ID 55, D/HOH] City: I found myself constantly looking up and down at the different lines and felt I was putting in too much work to understand what was going on.

[Participant ID 10, D/HOH] CTV: Font and the black background made it easy to read.

[Participant ID 42, D/HOH] CTV: Liked how clearly the words appeared and how easy they were to read

While not statistically significant, the trend of these results supported the quantitative results that the D/HOH viewers preferred the caption style of CTV, and that the caption style of City more negatively impacted their viewing pleasure. This preference may have been caused by a number of quality factors, including spelling/grammar issues, placement on screen, and/or display style. The display styles of these two videos were very different and likely contributed to the difference in viewer preference. The City captions used pop-on captions, where a full line of caption text would appear, then disappear and reappear below while a new line of text would appear above it (see Figure 6.1a). Many participants indicated they found it difficult to follow this style of captions. Additionally, the captions covered information displayed in a news weather banner at the bottom of the screen. In contrast, the CTV captions used a roll-on style, where a new line of text appeared below the existing line (See Figure 6.1b), the placement of the captions obstructed less information on screen. Since order effect was controlled for, these results demonstrate that while two sets of video captions can meet the Canadian quality standard, D/HOH viewers still preferred one style of caption over another.



Figure 6.1a: Screen capture of City



Figure 6.1b: Screen capture of CTV captions

A complementary study was conducted in 2017 study by Leger (Leger CCAA-User preferences for live captioning in Canada, personal communication, June 22, 2018) to examine

live captioning user preferences. Researchers used an online survey to evaluate participant preferences for various caption quality factors using a card sorting methodology. Participants were presented with a list of 12 caption quality goals and asked to rank them in order of importance. Although the survey findings are not yet published, I have been provided with access to their results. Their survey results and the results from my study showed similar trends in viewer preferences for caption quality elements. The User preferences study found that Deaf participants valued lag, placement on screen, verbatim accuracy, speed and spelling as the top five goals (in that order), whereas hard of hearing participants valued speed, placement on screen, lag, preservation of meaning, and speaker identification as the top five goals (in that order). Both studies found that having captions displayed at a readable speed and verbatim accuracy (having captions that capture every spoken word) were two factors that rated highly with D/HOH viewers. Their study also confirmed that other elements, such as spelling and grammar errors and speaker identification, were rated as less important. Furthermore, their study found that delay (i.e. lag) was ranked as one of the top four most important goals for captions by all participants. This was similar to the analysis of the qualitative comments in my study where lag time was the second most commented on caption quality factor, suggesting that it is of high importance to viewers despite not having been listed as one of the five primary caption elements included in the survey portion of my study.

While there is a dearth of research on caption lag times and their impact on viewer enjoyment or comprehension, the caption standards guideline documents of Canada, Australia, United States and the UK all provide standards and/or recommendations regarding lag or synchronicity of captions with audio (Australian Communications and Media Authority, 2013; CAB, 2008; “FCC moves to upgrade TV closed captioning quality,” 2014; “Ofcom’s Code on Television Access Services,” 2015; Federal Communications Commission, 2015). For instance, Canada requires a lag time of no more than six seconds for live-presented news segments, with best practices suggesting no more than five seconds (CAB, 2008; CRTC, 2012). The Federal Communications Commission in the United States provides suggestions for limiting delay, such as having broadcasters provide captions with background materials in advance of the live television broadcaster, and ensure that captioners are sufficiently trained and skilled to

minimise lag time (FCC, 2014a). These measures taken by television governing bodies around the world to minimise caption display lag times demonstrates the importance of this caption quality element, and the negative impact it may have on viewers when quality is not maintained.

While my study and the User preference study focused on studying viewer preferences for caption quality, there were many differences between them, such as the use of two different methodologies, two different participation groups, and different recruitment strategies (the user preference study was based in Canada, while my study used international online recruitment tools). The User preferences study's sample also included a more equal distribution of Deaf and Hard of Hearing participants, compared to mine, which had a large proportion of hard of hearing participants. Despite these differences, similar results were obtained, providing support for the argument that caption quality should take into consideration more than verbatim accuracy, such as speed and lag. Additionally, the combined results from both studies provides direction for further research. The results of the User preference study found D/HOH viewers would accept preservation of meaning by the captions; the hard of hearing participants ranked it as 4th most important goal, compared to verbatim accuracy which they ranked as 6th. Deaf participants indicated a preference for verbatim accuracy, ranking it 3rd, and preservation of meaning as 6th.

The combination of preference for readable speed of captions, verbatim accuracy, and low lag time presents a difficult, if not impossible, challenge for broadcasters. These three factors are in direct competition of each other, creating a forced choice situation for broadcasters. If verbatim accuracy is chosen as the most important goal, then speed may have to increased in order to capture every spoken word. If speed is not increased, but the television presenter speaks quickly, this will reduce the verbatim accuracy and/or increase the lag time over the course of the program. If speed is prioritised as most important, then some of the captions may need to be edited, thus compromising their verbatim accuracy. Further research is needed to determine what choices viewers would make in a forced-choice situation when presented with these three caption quality factors, and whether edited captions that preserve meaning are an acceptable alternative to verbatim accurate captions. If viewers accept edited captions that

preserve meaning, it may present itself as a possible solution between the speed and verbatim accuracy trade-off.

In summary, this research demonstrates that a caption quality measure that only takes into consideration verbatim accuracy is insufficient to measure caption elements that are significantly meaningful to viewers when they watch television with live captions. Additionally, it demonstrates a need for user input in caption quality control. My findings support what is currently known about viewer dissatisfaction with current captioning styles, and the importance of displaying captions at a speed that is readable for D/HOH viewers. Furthermore, this research supports the premise put forth by other researchers and users that caption quality extends beyond only accuracy, and that viewer's enjoyment and understanding of television programming is affected by other caption quality factors such as speed of caption display and lag time. The results suggest that broadcasters should expand current measures of quality control so that factors such as speed, display style, and lag are represented. Additionally, broadcasters should consider incorporating user feedback in their quality control processes due to the subjective nature of a number of caption elements. One potential solution may be to develop a user feedback system through smart television technology. Finally, some of the challenges presented by individual user preferences for caption factors may be addressed by new television technology if viewers are able to control some caption stylistic settings, such as font size and colour, display style, and a fast or slow speed option. If smartTV settings could be made sophisticated enough to allow users to set their preferences for captions speed and display style, this may help to address the divisiveness that exists in the results pertaining to user preference for speed.

6.1. Limitations

6.1.1. Recruitment

A common challenge faced when studying closed captioning, and shared by this study, was being able to recruit a large enough sample of the target population. This study used online recruitment strategies which allowed for access to a wider spread of the population, but also had limitations. Recruitment was limited to individuals already using Mechanical Turk and who

saw the recruitment post, and to members of the Canadian Hard of Hearing Association (CHHA). In total, the study was able to recruit 78 participants from the D/HOH community and a further 81 hearing participants. While not as small a sample as many studies in the area of closed captioning, it still may not be representative enough of the population at large. As a result, although these results make suggestions about the preferences of the D/HOH community with regards to subjective quality elements of closed captioning, a larger study would need to be carried out in order to draw more generalizable conclusions.

An additional limitation of this study was the disproportionate representation of hard of hearing and Deaf participants. The D/HOH group was comprised of 49 hard of hearing participants, 7 Deaf participants, 2 deafened participants, and 20 participants that indicated “Other”. While all participants that identified themselves as members of or with relationship to the D/HOH community were included in the D/HOH group for statistical analysis, the larger number of participants that identify as hard of hearing, and limited number of Deaf and Deafened participants, may have affected results of the analysis. As such, the results of this study may be more representative of the opinions and viewers of the HOH community. The literature shows that hard of hearing viewers often do not share the same caption preferences that Deaf or deafened viewers have (Fels et al., 2005; Lee et al., 2007), and that due to social and cultural reasons they must be considered as two distinct groups (Romero-Fresco, 2009). These potential differences need to be examined in a study that has more equal representation from the Deaf community and does not combine them into a single group.

6.1.2. Experimental design

In addition to recruitment challenges, the experimental design of the research study faced some limitations. The study itself took place entirely online. This limitation required participants to have access to a computer and internet connection, it also assumed a certain level of computer literacy and acumen. The research design did not include compensation or an incentive to participate in the study. Some potential participants may have been deterred or unwilling to volunteer their time in order to participate. Finally, the design asked people to self-

identify their relationship to the D/HOH community, relying on participant honesty, and had no way to verify if participants were actually Deaf or hard of hearing.

The experimental tasks posed some challenges. The study was designed such that participants responded to the pre-study survey, watch the first video clip, respond to 11 questions, watch a second video clip, and then respond to the same set of questions again. Using an online platform to host such a study placed a lot of trust that participants would watch the video clips in full before moving on to respond to the questions. An attempt was made to mitigate this limitation by discarding survey responses that were completed in under five minutes. These participants were assumed to have not watched the video clips in full, given that the combined total length of video is three minutes and seventeen seconds. Future research may want to consider conducting studies at a specific physical location where participants can be monitored for video viewing compliance.

It is also possible that the video selection used in this study posed some limitations and challenges for the participants. Each video was under two minutes in length and was a recording of a live presentation of a weather forecast, recorded with the original live produced captions. The weather forecast was a segment of a larger newscast program and was thus taken out of context of the longer television program. The normal delay that occurs with live produced captions meant that the captions displayed on the screen at the start of the segment were in reference to audio not present in the video clip. This likely caused some confusion for participants. Additionally, the video content, a long-range weather forecast, may not have been interesting to some viewers, thus viewer television preferences may have had an impact on their survey responses. And finally, due to the shortness of each video clip, less than two minutes in length each, means there was likely no fatigue factor in the study, which may have an impact on viewer's enjoyment of captions in a longer television program. These limitations in the study, should be considered when conducting any future research in the field of caption quality controls.

7. Recommendations and conclusion

7.1. Conclusion

The research question studied was what priority do hard of hearing viewers place on verbatim accurate captions as compared to other caption elements, specifically caption speed, missing words, spelling/grammar errors, and speaker identification? The hypothesis put forth was:

H: When presented with a trade-off decision between verbatim accuracy and other qualitative caption elements, viewers will place the highest value on verbatim accuracy, followed by spelling/grammar errors, captions speed, missing words, and finally speaker identification.

This research study was able to partially confirm this hypothesis and presented findings regarding which caption quality factors are of highest importance to D/HOH viewers. However, the results did not show a conclusive and complete hierarchy of quality factors. Based on the information gathered from participants, it appears that D/HOH people place equal value on verbatim accuracy and caption speed and place the least value on speaker identification. The order of the remaining caption elements studied was inconclusive. Additional analysis of the qualitative data indicated that lag time and caption display style were two more caption quality factors that impacted participants' viewing enjoyment. Finally, the findings from this study were corroborated by the CCAA-User preferences for live captioning in Canada, which used a different methodology and sample group, but found similar results around viewer preferences for speed, accuracy, and lag.

The purpose of this thesis work was to investigate an alternative method for measuring caption quality, in the form of a user survey for caption quality. The NASA-TLX workload assessment tool for measuring mental workload was used as a framework when developing the caption quality survey tool. For the purpose of this study, a survey was designed in which participants had to first make forced choices of priorities between five caption quality elements (the weighted score), then complete a task (watch a video) and then assign a numerical score from 1 to 11 (positive to negative) for how each caption quality element impacted their viewing

pleasure (the rating score). These were then multiplied together and divided by the number of forced choice pairs (10).

Analysis of the thematic data identified lag to be an important caption element to viewers. These findings present a challenge for broadcasters due to the fact that providing verbatim accurate captions will impact the speed at which they are presented and vice versa. If the speaking rate is too fast (above 170 WPM) then the captions may be moving too quickly to read; but conversely, if captions are displayed too slowly, the lag between audio and caption text will increase over the duration of the program unless sufficient or frequent pauses in audio allow for the captions to catch up to the audio. Some of the challenge in the trade-offs that must be made between speed, verbatim accuracy and lag occur because of the causes and controls (or lack thereof) behind each caption quality factor. Verbatim accuracy is a caption factor controlled entirely by captioners, speed is a factor of how many words need to be presented at any one-time interval and is controlled by the captioner/broadcasters and television technology, whereas lag is an artefact of the hardware/broadcast signal as well as how long it takes the captioner to begin captioning. Whereas it may be possible improve television and broadcasting technology to allow for viewer-controlled caption speed and mitigate the lag time problems that relate to technology limitations, there is very little, if anything, that can be done to fully eliminate lag. By the very nature of live-production captioning, the captioner must first hear some of the audio before they can begin to caption it, thus creating an inevitable lag or delay between audio content and caption text appearing on the screen. Although further research must be done to determine what choices viewers would make when faced with a forced-choice scenario of these three caption factors, broadcasters could potentially address some of the issues by improving television technology to allow viewers to adjust the caption speed settings on their television device – for instance, being able to select a pre-determined “slow” or “fast” caption speed. It should be noted however, that selecting a slow caption speed may in turn impact lag time, if captions are displayed at a slower rate, the lag time will likely increase over the course of the program, unless there are sufficient gaps in the audio content. Regardless, providing viewers with this choice would put control in

their hands, as opposed to having a broadcaster making a decision regarding caption speed on their behalf.

The second purpose of this study was to explore a caption quality measurement tool that incorporated more than just an accuracy score. This study, and the survey methodology used, was able to collect meaningful data about viewer preferences on five factors that impact caption quality. Further research must be carried out to determine how best to incorporate user-feedback into caption quality control measures, but possible options include hosting surveys online, or through smart television technology allowing users to provide feedback at the end of a broadcast. The results of this study are encouraging and suggest that it is possible to assess caption quality using subjective measuring tools. Furthermore, the growing popularity of smart televisions could be leveraged to improve caption quality measures and allow for viewer feedback.

7.2. Contributions

This research made a number of contributions to the field of closed captioning research. While exploratory in nature, the results demonstrated a need for a more comprehensive method of measuring caption quality that takes into consideration more than just verbatim accuracy. The methodology used in this research was designed based on the principles of the NASA-TLX workload assessment tool, and provide a unique method for measuring subjective caption quality outside of current quality measurements used by the media industry. This research successfully demonstrated that a subjective measurement tool, like the one designed in this study, that collects viewer feedback can provide valuable information about caption user preferences.

Additionally, the results of this research made a scientific contribution to closed captioning research with the findings on verbatim accuracy and speed. Limited research exists in the field of caption quality and viewer preferences. Currently, broadcasters make assumptions about viewer preferences and measure caption quality using verbatim accuracy scores. The findings of this research suggest that such an assumption cannot easily be made. This research adds to a growing body of literature about how users of closed captions use and interact with closed

captions. These findings may be of significance to broadcasters, and aid in the development of more inclusive policies for closed caption quality.

7.3. Future research

There are a number of future research directions that could stem from my research methodology and results. My study used two short weather forecast clips for the videos shown to participants; a more comprehensive study using the same methodology, but which includes all genres of television with live-produced captions and longer clips, could be used to determine if the same conclusions are drawn. Such a study would account for viewer biases around genre preferences and the lack of sufficient context for the video clips.

Another avenue of study would be to look at other subjective caption quality elements that were not considered in the scope of this project but emerged as important themes in the qualitative data. These includes lag time, display style, font style, placements of captions on screen. These caption elements were included in the User preferences study, and results were obtained regarding viewer priorities regarding these other elements. These caption elements could be studied using the methodology outlined in my study to determine if similar results are obtained. Furthermore, the trade-off relationships and choices around speed, verbatim accuracy and lag time should be studied in greater detail.

One trend that emerged from the thematic data was around viewer preferences on speed. The statistical results comparing the comments made on the speed of the captions by the D/HOH group and the hearing group were not statistically significant. However, the descriptive results support the current knowledge about the reading speeds of D/HOH people and hearing people. A larger study with more participants should be conducted to assist in determining differences in preference between these two groups.

Finally, the results of this study are currently being used in a PhD study that will look at developing a method of assessment for caption quality which reflects D/HOH viewers' opinions. An artificial intelligence network will be trained to assess the quality of closed captions based on the subjective quality assessment provided by D/HOH audiences. The goal is to develop a

system that can automate the procedure of quality assessment and may potentially reduce the human efforts involved.

Appendices

A. Video Caption Transcripts

City New Caption Transcript

Caption Line	Video Time Stamp	Caption Text	Error Type
1	00:00:00,220 -- 00:00:01,270	PARADE, THEY RAISE MONEY FOR	
2	00:00:01,270 -- 00:00:03,539	COMMUNITY CENTRE 55'S SHARE A	
3	00:00:03,539 -- 00:00:05,440	CHRISTMAS PROGRAM.	
4	00:00:05,440 -- 00:00:07,040	LOTS OF SANTA CLAUS PARADES	
5	00:00:07,040 -- 00:00:08,430	GOING ON.	
6	00:00:08,430 -- 00:00:09,000	GREY CUP IS GOING ON AND Z YOU	Spelling/grammar (omit Z)
7	00:00:09,000 -- 00:00:11,960	HEARD IN SPORTS IT'S SNOWING IN	
8	00:00:11,960 -- 00:00:13,990	OTTAWA. It's about two to five centimeters THIS IS A SMALL AREA OF LOW	Not verbatim
9	00:00:13,990 -- 00:00:17,910	PRESSURE SYSTEM THAT IS SINKING	
10	00:00:17,910 -- 00:00:18,820	ACROSS SOUTHEASTERN ONTARIO.	
11	00:00:18,820 -- 00:00:19,559	THE DARKER THE BLUE, THE MORE	
12	00:00:19,559 -- 00:00:21,280	INTENSE THE SNOW AND THEY'RE	
13	00:00:21,280 -- 00:00:22,420	INTO THE MOST INTENSE PART NOW	
14	00:00:22,420 -- 00:00:23,940	IN OTTAWA.	
15	00:00:23,940 -- 00:00:24,899	IT'S ONLY GOING TO TAPER OFF	
16	00:00:24,899 -- 00:00:26,830	FROM THIS POINT FORWARD.	
17	00:00:25,830 -- 00:00:26,070	IT'S two to four, maybe UP TO 5 CENTIMETRES OF	Not verbatim
18	00:00:26,070 -- 00:00:26,670	SNOW.	
19	00:00:26,830 -- 00:00:28,670	THE BACK END OF IT though IS also BRINGING	Missing words

20	00:00:28,670 -- 00:00:30,420	FLURRIES ACROSS LAKE SIMCOE.	
21	00:00:30,420 -- 00:00:32,350	SOME OF THE FLURRIES JUST	Spelling/grammar ("the" should be "these")
22	00:00:32,350 -- 00:00:33,090	STARTING TO EDGE INTO NORTHERN	
23	00:00:33,090 -- 00:00:35,240	YORK REGION OVER THE NEXT HOUR	
24	00:00:35,240 -- 00:00:36,300	OR TWO. So areas such as Whitchurch-Stouffville	Not verbatim
25	00:00:36,770 -- 00:00:42,800	POSSIBLY as far south as UNIONVILLE, MARKHAM AND	Not verbatim
26	00:00:42,800 -- 00:00:45,470	NORTHERN AJAX AND WHITBY MAY BE	Spelling/grammar ("be" should be "see")
27	00:00:45,470 -- 00:00:47,500	a few scattered FLURRIES BEFORE THIS EVENING IS	Not verbatim
28	00:00:47,500 -- 00:00:47,900	OUT.	
29	00:00:47,900 -- 00:00:48,830	OTHER WISE WE ARE NOT EXPECTING	Spelling/grammar ("other wise" should be "otherwise")
30	00:00:48,830 -- 00:00:49,000	ACCUMULATION	
31	00:00:49,000 -- 00:00:51,000	BY THE WAY, FUTURE CAST,	
32	00:00:51,000 -- 00:00:51,500	NOTHING.	
33	00:00:51,500 -- 00:00:51,900	I GOT NOTHING.	
34	00:00:51,900 -- 00:00:53,500	I GOT A LITTLE BIT OF CLOUD TO	
35	00:00:53,500 -- 00:00:54,230	SHOW YOU AND THAT'S IT.	
36	00:00:54,230 -- 00:00:56,200	IN FACT OVER THE NEXT FEW DAYS	Spelling/grammar ("few" through be "three")
37	00:00:56,200 -- 00:00:58,200	WE'RE STARTING A DRY STRETCH,	
38	00:00:58,270 -- 00:01:00,539	a fairly SUNNY STRETCH WITH JUST A FEW	Missing words
39	00:01:00,539 -- 00:01:01,440	DLOUDZ HERE AND THERE, THAT'S	Spelling/grammar ("dlouz" should be "clouds")
40	00:01:01,440 -- 00:01:02,040	MONDAY AND TUESDAY AND WEDNESDAY	
41	00:01:02,040 -- 00:01:03,430	AS WELL.	
42	00:01:03,430 -- 00:01:04,000	CURRENT TEMPERATURES ARE	

43	00:01:04,000 -- 00:01:05,960	just SLIGHTLY ABOVE THAT FREEZING	Missing words
44	00:01:05,960 -- 00:01:07,990	MARK, WE'VE GOT A WIND CHILL,	
45	00:01:07,990 -- 00:01:09,910	THOUGH, FEELS LIKE OF ABOUT MINUS 2	
46	00:01:09,440 -- 00:01:10,040	TO MINUS 5 OUT THERE.	
47	00:01:10,040 -- 00:01:11,430	so yes SLIGHT WIND CHILL AND YOU HAVE	Missing words
48	00:01:11,430 -- 00:01:13,000	TO DRESS FOR THAT IN THE	
49	00:01:13,000 -- 00:01:15,960	WE HAVE WIND GUSTS NEAR TO 40	
50	00:01:15,960 -- 00:01:16,990	KILOMETRES PER HOUR, A LITTLE	
51	00:01:16,990 -- 00:01:18,910	BIT STRONGER DOWN AT THE ISLAND	
52	00:01:18,910 -- 00:01:18,910	BECAUSE THE WIND IS PUSHING OVER	
53	00:01:18,910 -- 00:01:22,240	THE LAKE, LESS FRICTION, TO IT	
54	00:01:22,240 -- 00:01:23,300	so TENDS TO BE STRONGER.	Missing words
55	00:01:23,770 -- 00:01:26,800	and AS WE GO THROUGH THE DAY	Missing words
56	00:01:26,800 -- 00:01:27,470	TOMORROW, I MEANT TO PUT A PAUSE	
57	00:01:27,470 -- 00:01:28,500	IN THERE and slow that down for you, BUT WE ARE LOOKING AT	Not verbatim
58	00:01:28,500 -- 00:01:30,900	GUSTY CONDITIONS TONIGHT AND	Spelling/grammar (omit "and")
59	00:01:30,900 -- 00:01:31,830	Start EASING THROUGH THE DAY TOMORROW.	Missing word
60	00:01:31,830 -- 00:01:34,000	NOT MUCH OF A WIND CHILL TO DEAL	Not verbatim up ("not" should be "So don't have")
61	00:01:34,000 -- 00:01:35,800	WITH tomorrow. JUST IN THE MORNING, FEELS	Missing words Spelling/grammar ("feels" should be "it should feel")
62	00:01:35,800 -- 00:01:36,500	LIKE about MINUS 6.	Missing words
63	00:01:36,500 -- 00:01:37,960	TOMORROW CLOSE TO SEASONAL	
64	00:01:37,960 -- 00:01:38,990	TEMPERATURES, THERE IS YOUR	
65	00:01:38,990 -- 00:01:40,910	SUNNY, DRY STRETCH.	

66	00:01:40,910 -- 00:01:41,040	WE HAVE GUSTY WINDS TUESDAY,	
67	00:01:41,040 -- 00:01:42,430	THAT'S COMING FROM THE SOUTHWEST	
68	00:01:42,430 -- 00:01:45,000	AND THAT IS A 14 THAT YOU ARE	
69	00:01:45,000 -- 00:01:45,960	SEEING.	
70	00:01:45,960 -- 00:01:47,990	THE AVERAGE -- SORRY, THE RECORD	
71	00:01:47,990 -- 00:01:48,910	HIGH FOR THAT DAY IS JUST OVER 18.	
72	00:01:49,410 -- 00:01:51,240	IT'S NOT going to be RECORD BREAKING BUT	Not verbatim
73	00:01:51,240 -- 00:01:53,300	WE'LL BE INTO EVENING RAIN AND	
74	00:01:53,300 -- 00:01:54,800	FRIDAY IS THE DAY TO BRING THE	

CTV News Caption Transcript

Caption Line	Video Time Stamp	Caption Text	Error Type
1	00:00:00,000 -- 00:00:01,590	Least stable.	
2	00:00:01,590 -- 00:00:03,090	Live AT QUEENS PARK, NOW BACK TO	
3	00:00:03,090 -- 00:00:04,859	YOU.	
4	00:00:04,859 -- 00:00:05,540	THANK YOU, PAUL.	
5	00:00:05,540 -- 00:00:08,220	DENISE ANDREACCHI IS IN FOR	
6	00:00:08,220 -- 00:00:09,330	ANWER Knight TODAY.	Spelling/grammar and Missing words (should be Anwar Knight)
7	00:00:09,330 -- 00:00:11,000	WE LIKE THE TEMPERATURES TO GO	
8	00:00:11,000 -- 00:00:12,750	UP. But it was a decent start to the day [Denise] Yes it certainly is, and lots of sunshine too Ken,	Not verbatim Speaker identification
9	00:00:12,750 -- 00:00:14,610	AND LIKE YOU SEE behind me, CTV we have some active WEATHER just	Not verbatim
10	00:00:14,610 -- 00:00:15,960	SKIRTING THE BRUCE PENINSULA.	

11	00:00:15,960 -- 00:00:17,279	AND RAINFALL IN THROUGH PARTS OF	
12	00:00:17,279 -- 00:00:18,810	COTTAGE COUNTRY, AND even we have a WEATHER	Not verbatim
13	00:00:18,810 -- 00:00:21,000	STATEMENT FOR TIMMINS, AND PARTS	
14	00:00:21,000 -- 00:00:22,619	OF NORTHEASTERN ONTARIO GETTING	
15	00:00:22,619 -- 00:00:23,810	IN into some FREEZING RAIN, IT'S ALL	Missing words
16	00:00:23,810 -- 00:00:25,560	ABOUT THE HIGH PRESSURE FOR THE	
17	00:00:25,560 -- 00:00:25,939	SOUTHERN TIER OF THE PROVINCE,	
18	00:00:25,939 -- 00:00:27,140	AND THE GTA.	
19	00:00:27,140 -- 00:00:28,210	PLUS we're getting A BEAUTIFUL SOUTHWEATERLY	Missing words
20	00:00:28,210 -- 00:00:30,530	FLOW OF AIR, GUSTY AT TIMES, AS	
21	00:00:30,530 -- 00:00:32,600	WE CAN SEE, AND ANYWHERE FROM about 30	Missing words
22	00:00:32,600 -- 00:00:34,569	TO 50 KILOMETRES AN HOUR, ON AND	
23	00:00:34,569 -- 00:00:35,620	OFF THROUGHOUT THE MORNING AND	
24	00:00:35,620 -- 00:00:38,270	THE AFTERNOON, AND IT IS COMING	Spelling/grammar ("and" should be "but")
25	00:00:38,270 -- 00:00:39,010	FROM THE SOUTHWEST, AND IT'S	
26	00:00:39,010 -- 00:00:40,530	DRAWING TEMPERATURES WAY UP.	
27	00:00:40,530 -- 00:00:42,300	YESTERDAY FOR THE GTA, we AVERAGING	Spelling/grammar ("averaging" should be "averaged")
28	00:00:42,300 -- 00:00:43,950	3 TO 4 degrees, AND CURRENTLY we are sitting IN THE	Not verbatim ("the" should be "those")
29	00:00:43,950 -- 00:00:45,719	HIGH SINGLE DIGITS ACROSS PARTS	
30	00:00:45,719 -- 00:00:46,550	OF THE SOUTHWEST, AND ALREADY AT	Spelling/grammar ("and" should be "we're")
31	00:00:46,550 -- 00:00:48,930	9 degrees IN HAMILTON, NIAGARA, AND	
32	00:00:48,930 -- 00:00:49,219	OAKVILLE.	
33	00:00:49,219 -- 00:00:50,870	WE DO HAVE THE RISK OF SOME	

34	00:00:50,870 -- 00:00:51,609	ACTIVE WEATHER.	
35	00:00:51,609 -- 00:00:52,930	AS WE CAN SEE now , IT'S NORTH OF US,	Missing words
36	00:00:52,930 -- 00:00:54,520	AND IT'S VERY SLOW MOVING, AND most T	Missing words
37	00:00:54,520 -- 00:00:58,769	OF THE DAY TODAY, we will be BASKING IN some	Not verbatim
38	00:00:58,769 -- 00:00:59,590	NOVEMBER SUNSHINE, AND in through the east a little but cooler but still WELL	Not verbatim
39	00:00:59,590 -- 00:01:00,280	ABOVE THE SEASONAL AVERAGE WHERE	
40	00:01:00,280 -- 00:01:02,020	WE SHOULD BE.	
41	00:01:02,020 -- 00:01:04,700	PEARSON YESTEDAY, 3 and a half DEGREES FOR AN	
42	00:01:04,700 -- 00:01:05,689	AVERAGE HIGH, we're already sitting and nine, we should AND BUMPING UP A	Not verbatim ("and bumping" should be "bump")
43	00:01:05,689 -- 00:01:07,220	COUPLE MORE DEGREES IN THE NEXT	
44	00:01:07,220 -- 00:01:07,990	COUPLE OF HOURS, AND A MIX OF	
45	00:01:07,990 -- 00:01:09,310	SUN AND CLOUD.	
46	00:01:09,310 -- 00:01:11,780	WE WILL CONTINUE WITH THAT GUSTY	
47	00:01:11,780 -- 00:01:12,000	AIR.	
48	00:01:12,000 -- 00:01:13,300	But WHY WOULD YOU WANT TO BE INDOORS	Missing words
49	00:01:13,300 -- 00:01:14,800	SHOPPING ON A DAY LIKE TODAY?	
50	00:01:14,800 -- 00:01:16,100	THESE ARE THE type of DAYS YOU NEED TO	Missing words
51	00:01:16,100 -- 00:01:19,200	TAKE ADVANTAGE OF, AND PARTLY	Spelling/grammar (omit "and")
52	00:01:19,200 -- 00:01:21,000	CLOUDY SKIES for most of the day, AND we're starting off with a pretty good weekend, we'll TALKING more ABOUT	Not verbatim Spelling/grammar ("talking" should be "talk")
53	00:01:21,100 -- 00:01:22,200	THE WEEKEND WEATHER AND THE	
54	00:01:22,200 -- 00:01:23,510	LONG RANGE FORECAST COMING UP a little later on , Ken.	Not verbatim

B. Caption Preference Score Calculation Methodology

Calculating the verbatim accuracy caption preference score.

$$\textit{Weighted Pairs Score} = \sum \textit{Forced choice selection} = 196$$

$$\textit{Rating Score} = \frac{\sum \textit{Ratings}_{v_1} + \sum \textit{Ratings}_{v_2}}{2} = \frac{528 + 504}{2} = 516$$

$$\begin{aligned} \textit{Caption Preference Score} &= \frac{\textit{Weighted Pairs Score} \times \textit{Rating Score}}{\textit{Number of Pairs}} = \frac{196 \times 516}{10} \\ &= 10,114 \end{aligned}$$

C. Additional Statistical Results

Table 8.1: Shapiro-Wilkes test for normality on data used to determine learning effect

	D/OH Group			Hearing Group		
	Statistic	df	p	Statistic	df	p
Verbatim R1	.95	51	.023	.94	50	.013
Fast R1	.96	51	.128	.95	50	.029
Slow R1	.97	51	.197	.96	50	.062
Missing Words R1	.95	51	.032	.95	50	.028
Spelling/Grammar R1	.95	51	.047	.88	50	.000
Speaker ID R1	.94	51	.019	.91	50	.001
Verbatim R2	.95	51	.044	.95	50	.024
Fast R2	.94	51	0.017	.94	50	.017
Slow R2	.95	51	.025	.95	50	.047
Missing Words R2	.97	51	.214	.94	50	.011
Spelling/Grammar R2	.95	51	.048	.93	50	.005
Speaker ID R2	.91	51	.001	.90	50	.000

Table 8.2: Paired t-test for first video (R1) and second video (R2) rating scores

	D/OH Group			Hearing Group		
	t	df	p	t	df	p
Verbatim R1 – Verbatim R2	.51	68	.608	-.77	69	.445
Fast R1 – Fast R2	-1.67	67	.099	-1.02	71	.313
Slow R1 – Slow R2	-1.91	62	.060	-.40	64	.690
Missing Words R1 – Missing Words R2	-.14	68	.892	-1.14	70	.258
Spelling/Grammar R1 – Spelling/Grammar R2	-.76	68	.451	-.81	69	.421
Speaker ID R1 – Speaker ID R2	.41	60	.684	-1.48	59	.145

Table 8.3: Shapiro-Wilkes test for normality on data used to determine significance of rating score data

	D/OH Group			Hearing Group		
	Statistic	df	p	Statistic	df	p
Verbatim	.96	111	.003	.96	118	.001
Fast	.97	111	.009	.95	118	.000
Slow	.96	111	.005	.96	118	.002
Missing Words	.96	111	.005	.94	118	.000
Spelling/Grammar	.96	111	.001	.91	118	.000
Speaker ID	.93	111	.000	.91	118	.000

Table 8.4: Shapiro-Wilkes test for normality on data used to determine if any differences between City (V1) and CTV (V2) rating scores

	D/OH Group			Hearing Group		
	Statistic	df	p	Statistic	df	p
Verbatim V1	.95	52	.028	.86	50	.000
Verbatim V2	.96	52	.067	.95	50	.024
Fast V1	.92	52	.002	.95	50	.029
Fast V2	.96	52	.096	.96	50	.062
Slow V1	.95	52	.027	.95	50	.028
Slow V2	.97	52	.318	.88	50	.000
Missing Words V1	.94	52	.014	.91	50	.001
Missing Words V2	.93	52	.005	.94	50	.013
Spelling/Grammar V1	.97	52	.164	.94	50	.017
Spelling Grammar V2	.92	52	.002	.95	50	.047
Speaker ID V1	.92	52	.002	.94	50	.011
Speaker ID V2	.93	52	.004	.93	50	.005

D. Questionnaire

Caption Quality Survey

Elements of Caption Quality

Principal Investigator: Alison Whiting, MScM candidate, IMDC. alison.whiting@ryerson.ca

Working under the supervision of thesis supervisor:

Deborah Fel, P.Eng., Ph.D., Professor at Ryerson University and Director at the Inclusive Media and Design Centre (IMDC), (416) 979-5000 x7619, dfels@ryerson.ca

Consent

You are being asked to participate in a research study. Before you give your consent to be a volunteer, it is important that you read the following information and ask as many questions as necessary to be sure you understand what you will be asked to do.

Purpose of Study

In this study, we would like to hear of hearing viewers' opinions on different closed caption elements that affect a viewer's experience and viewing pleasure when watching television with live captioning. If you agree to participate in the study, you will be asked to view two pre-recorded live videos with captions. You will also be asked to complete two questionnaires; one prior to beginning the study, and one after. The pre-study questionnaire will ask basic demographic information such as age and gender, as well as your television viewing habits. The post-study questionnaire will ask for your impressions and opinions of the impact of different caption elements on your viewing pleasure.

Description of the Study

First, you will be asked to read the consent preamble, this will take about 5 to 10 minutes. If you agree to continue to participate in the study, you will be taken to the pre-study questionnaire. The pre-study questionnaire is used to collect background information and to collect your opinions on and experience with closed captions. This will take approximately 10 minutes.

Following the pre-study questionnaire, you will be asked to view two 2-minute video clips from different live weather news shows with their original captions as aired. After each video, we will ask you to complete a questionnaire to collect information about how different caption characteristics affected your viewing pleasure. This will take about 20-30 minutes.

The total time for this study is about one hour.

Confidentiality

Consent and questionnaire responses will be used exclusively for research and educational purposes. Questionnaire response data will be stored separately from the consent data and questionnaires. Electronic questionnaire answers given by participants will be encrypted and stored on a password protected server at the IMDC. Identifying information (e.g. name, email address, telephone number, IP address) will not be collected from participants.

Risks and Discomforts

The risks associated with the study are minimal. You may experience some fatigue while viewing the videos and completing the questionnaires. You may discontinue the study at any time without penalty.

You may also experience some discomfort in answering some of the personal questions in the study, such as age, gender, and education level. You are not required to answer any questions that make you feel uncomfortable. Any data collected will only be used to analyse the impact of caption elements on viewing pleasure. Only members of the project team will have access to all the data, which will be encrypted and stored on a password protected server in the IMDC at Ryerson University.

Expected Benefits

You may not directly benefit from your participation in this study; however, the data will be used to assess the limitations of the current method of measuring caption quality, and may impact the development and availability of captions, or at least better quality captions, for Deaf and hard of hearing audiences in the future.

Costs and/or Compensation for Participation

There are no costs associated with your participation. The study can be done remotely, at a location convenient to you. Access to a computer and internet connection is required.

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 or any of the agencies or organisations through which you were recruited. If you decide to participate, you are free to withdraw your consent and to stop participation at any time without penalty or loss of benefits to which you are allowed. At any particular point in the study, you may refuse to answer any question or to stop participation altogether. If you stop participating in the study your questionnaire will not be used in the research and will be destroyed immediately. To stop participating in the study at any time, close the browser and no data will be collected.

Questions About the Study

If you have any questions or concerns, please do not hesitate to call Alison Whiting or Deborah Fels at (416) 979-5000 x7619. In addition to the co-researchers and their supervisors, the Research Ethics Board (REB) may also be contacted if you have any questions or concerns about your rights as a research participant. Please contact the Office of the Vice President, Research and Innovation, Ryerson University, 350 Victoria St., Toronto, ON, M5B 2K3, Tel: (416) 979-5042, email: rebchair@ryerson.ca

Clicking on "I agree" below indicates that you have read the information in this agreement, have had a chance to ask any questions you have about the study, and know that your participation is entirely voluntary. Your acknowledgment 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 told that by agreeing to this consent agreement you are not giving up any of your legal rights.

<input type="checkbox"/>	I agree
<input type="checkbox"/>	I do not agree

Purpose of Study

The purpose of this questionnaire is to gather demographical information about the participants, as well as their television viewing habits and caption preferences. The information gathered here will be used to analyse the results of this study.

Q1: What statement best describes your relationship to the Deaf and/or hard of hearing communities?

	I have no relationship to these communities
	I identify as Deaf
	I am Deafened
	I am Hard of Hearing
	Other, please specify below

Other: _____

Q2: What is your gender?

	Male
	Female
	Other
	Prefer not to answer

Q3: What is your age?

	18-29
	30-39
	40-49
	50-59
	60+

Q4: What is your highest level of education completed?

	Elementary school
	High School
	College/Technical (diploma, 2 or 3 years)
	University (bachelor's degree, 4 years or equivalent)
	Graduate school
	Prefer not to answer

Q5: Please describe your current employment status.

	Employed full time outside the home
	Employed part time outside the home
	Self-employed
	Homemaker
	Student
	Retired
	Other

Q6: How many hours of television do you watch a week, on average?

	Less than 1 hour
	1-5 hours
	6-10 hours
	11-15 hours
	16-20 hours
	More than 20 hours

Q7: What type(s) of live-televised programming do you watch?

	Less than 1 hour	1-5 hours	6-10 hours	11-15 hours	16-20 hours	20+ hours	Do not watch
News							
Sports							
Weather							
Talk Shows							
Award Shows							
Special Events							
Other, please specify below							

Other: _____

Q8: How often do you watch television alone?

	Always
	Frequently
	Sometimes
	Seldom
	Never

Q9: How often do you watch television with friends and/or family?

	Always
	Frequently
	Sometimes
	Seldom
	Never

Q10: How often do you use closed captions when watching television?

	Always
	Sometimes
	Never

Q11: How often do you use closed captions when watching live-televised programming?

	Always
	Sometimes
	Never

Q12: What do you like about the closed captions on television for live programming? (check all that apply)

	Rate or speed of display
	Verbatim accuracy

	Paraphrasing
	Placement on screen
	Font of text
	Size of text
	Colour
	Speaker identification
	Other, please specify below

Other: _____

Q13: What do you dislike about the closed captions on television for live programming? (check all that apply)

	Rate or speed of display
	Verbatim accuracy
	Paraphrasing
	Placement on screen
	Font of text
	Size of text
	Colour
	Speaker identification
	Other, please specify below

Other: _____

Q14: Think about your experiences with live captioning. Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

A)

Verbatim Accuracy	Speed of Captions
-------------------	-------------------

Verbatim accuracy: The Canadian Radio-television and Telecommunications Commission (CRTC) requires English-language broadcasters to reach a captioning accuracy rate of at least 95% averaged over a given live program.

Speed of captions: The rate at which the captions are displayed on the screen. In order to achieve verbatim accuracy, the caption speed may need to be increased.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

B)

Verbatim Accuracy	No Missing Words
-------------------	------------------

Verbatim accuracy: The CRTC requires English-language broadcasters to reach a captioning accuracy rate of at least 95% averaged over a given live program.

Missing words: Words spoken during the program that are missing from the captions. In an attempt to achieve verbatim accuracy the stenographer may miss words while captioning the program.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

C)

Verbatim Accuracy	No Spelling/Grammar Errors
-------------------	----------------------------

Verbatim accuracy: The CRTC requires English-language broadcasters to reach a captioning accuracy rate of at least 95% averaged over a given live program.

Spelling/grammar errors: May include spelling errors and incorrect or misplaced punctuation. In an attempt to achieve verbatim accuracy the stenographer may make spelling and/or grammatical errors in the transcription of the captions.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

D)

Verbatim Accuracy	Speaker Identification
-------------------	------------------------

Verbatim accuracy: The CRTC requires English-language broadcasters to reach a captioning accuracy rate of at least 95% averaged over a given live program.

Speaker identification: Text used to provide information about who is speaking. In order to achieve verbatim accuracy the stenographer may not include speaker identification.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

E)

No Spelling/Grammar Errors	Speed of Captions
----------------------------	-------------------

Spelling/grammar errors: May include spelling errors and incorrect or misplaced punctuation. In an attempt to achieve verbatim accuracy the stenographer may make spelling and/or grammatical errors in the transcription of the captions.

Speed of captions: The rate at which the captions are displayed on the screen. In order to achieve verbatim accuracy the caption speed may need to be increased.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

F)

No Spelling/Grammar Errors	No Missing Words
----------------------------	------------------

Spelling/grammar errors: May include spelling errors and incorrect or misplaced punctuation. In an attempt to achieve verbatim accuracy the stenographer may make spelling and/or grammatical errors in the transcription of the captions.

Missing words: Words spoken during the program that are missing from the captions. In an attempt to achieve verbatim accuracy the stenographer may miss words while captioning the program.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

G)

No Spelling/Grammar Errors	Speaker Identification
----------------------------	------------------------

Spelling/grammar errors: May include spelling errors and incorrect or misplaced punctuation. In an attempt to achieve verbatim accuracy the stenographer may make spelling and/or grammatical errors in the transcription of the captions.

Speaker identification: Text used to provide information about who is speaking. In order to achieve verbatim accuracy the stenographer may not include speaker identification.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

H)

No Missing Words	Speed of Captions
------------------	-------------------

Missing words: Words spoken during the program that are missing from the captions. In an attempt to achieve verbatim accuracy the stenographer may miss words while captioning the program.

Speed of captions: The rate at which the captions are displayed on the screen. In order to achieve verbatim accuracy the caption speed may need to be increased.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

I)

No Missing Words	Speaker Identification
------------------	------------------------

Missing words: Words spoken during the program that are missing from the captions. In an attempt to achieve verbatim accuracy the stenographer may miss words while captioning the program.

Speaker identification: Text used to provide information about who is speaking. In order to achieve verbatim accuracy the stenographer may not include speaker identification.

Q14: Select between the following pairs of characteristics the one which is of most importance to your viewing pleasure.

J)

Speaker Identification	Speed of Captions
------------------------	-------------------

Speaker identification: Text used to provide information about who is speaking. In order to achieve verbatim accuracy the stenographer may not include speaker identification.

Speed of captions: The rate at which the captions are displayed on the screen. In order to achieve verbatim accuracy the caption speed may need to be increased.

Q15: The following is a list of characteristics of caption quality. Please check off the five characteristics that you believe are most important to caption quality.

	Rate or speed of display
--	--------------------------

	Verbatim translation
	Use of paraphrasing
	Placement on screen
	Use of colour in captions for emphasis, emotion, or tone
	Use of text descriptions for emotional information
	Use of text descriptions for emotional information
	Use of graphics or symbols to denote background elements, such as applause or music
	Use of text descriptions for background noise or music
	Use of different fonts and/or text sizes
	Speaker identification

Thank you for completing the pre-study questionnaire. The purpose of this next stage is to gather information about each participant's preferences on specific characteristics of closed captions for live televised programming. The information obtained through this questionnaire will be used to analyse the trade-off viewers are willing to make in order to have verbatim captions. All videos had instances of problems with various characteristics.

On the next slide, you will be presented with a link to a video. Please watch the video and then return to this survey to complete a set of questions about what you watched. Afterwards, you will be presented with a link to a second video. Please watch the video and then return to this survey to complete a set of questions about what you watched.

Please click on the link below to watch the video, then proceed to the next stage of the survey.

[CityNews Weather](#)
 Or (shown at random)
[CTV News Weather](#)

Q16: Please select the video you just viewed.

	CityNews Weather
	CTV News Weather

Q17: How much do you think the fast appearing and disappearing captions affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q18: How much do you think the slow appearing and disappearing captions affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q19: How much do you think the missing words affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q20: How much do you think the spelling and/or grammar errors affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q21: How much do you think the use of speaker identification affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q22: How much do you think the lack of verbatim accurate captions affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q23: Think back to the video and briefly list your favourite part and why.

Q24: Think back to the video and briefly list your least favourite part and why.

Q25: Please add any additional comments.

Please click on the link below to watch the video, then proceed to the next stage of the survey.

[CityNews Weather](#)

Or (shown at random)

[CTV News Weather](#)

Q26: Please select the video you just viewed.

<input type="checkbox"/>	CityNews Weather
<input type="checkbox"/>	CTV News Weather

Q27: How much do you think the fast appearing and disappearing captions affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q28: How much do you think the slow appearing and disappearing captions affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q29: How much do you think the missing words affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q30: How much do you think the spelling and/or grammar errors affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q31: How much do you think the use of speaker identification affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q32: How much do you think the lack of verbatim accurate captions affected your viewing pleasure?

Very negatively			No effect				Very positively				Not applicable
0	1	2	3	4	5	6	7	8	9	10	X

Q33: Think back to the video and briefly list your favourite part and why.

Q34: Think back to the video and briefly list your least favourite part and why.

Q35: Please add any additional comments.

E. Ethics Letter of Approval



To: Alison Whiting
Ted Rogers School of Management
Re: REB 2017-247: Elements of Caption Quality: Determining the Priority of Closed Caption Elements
Date: August 18, 2017

Dear Alison Whiting,

The review of your protocol REB File REB 2017-247 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 2017-247) on future correspondence.

Congratulations and best of luck in conducting your research.

A handwritten signature in black ink, appearing to read "Patrizia Albanese". The signature is stylized and written over a light blue horizontal line.

Dr. Patrizia Albanese, PhD
Chair, Ryerson University Research Ethics Board

References and Work Cited

- Albertini, J., & Mayer, C. (2011). Using miscue analysis to assess comprehension in Deaf college readers. *Journal of Deaf Studies and Deaf Education*, 16(1), 35–46. <https://doi.org/10.1093/deafed/enq017>
- Australian Communications and Media Authority. (2013). Broadcasting Services (Television Captioning) Standard 2013. Retrieved August 2, 2017, from <https://www.legislation.gov.au/Details/F2016C00225>
- Bontoft, M., & Pullin, G. (2003). What is an inclusive design process? In *Inclusive Design* (pp. 250–231). London: Springer.
- Brown, A., Jones, R., & Crabb, M. (2015). Dynamic subtitles: the user experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (pp. 103–112). ACM.
- Buchweitz, A., Mason, R. A., Leda, T., & Just, M. A. (2009). Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Psychology & Neuroscience*, 2(2), 111–123.
- Burnham, D., Leigh, G., Noble, W., Jones, C., Tyler, M., Grebennikov, L., & Varley, A. (2008). Parameters in television captioning for Deaf and Hard-of-Hearing adults: effects of caption rate versus text reduction on comprehension. *Journal of Deaf Studies and Deaf Education*, 13(3), 391–404. <https://doi.org/10.1093/deafed/enn003>
- CAB. (2008). Closed Captioning Standards and Protocol for Canadian English Language Television Programming Services. Retrieved April 2, 2017, from <http://www.cab-acr.ca/english/social/captioning/captioning.pdf>
- CAD. (2015). Statistics on Deaf Canadians. Retrieved September 7, 2018, from <http://cad.ca/issues-positions/statistics-on-deaf-canadians/>
- Cao, A., Chintamani, K. K., Abhilash, P. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods*, 41(1), 113–117.
- Clarkson, P. J., & Coleman, R. (2015). History of inclusive design in the UK. *Applied Ergonomics*, 46, 235–247. <https://doi.org/10.1016/j.apergo.2013.03.002>
- Cremers, A. H. M., Neerincx, M. A., & Jong, J. G. M. De. (2013). Inclusive design: bridging theory and practice. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 323–332). Springer Berlin Heidelberg.
- CRTC. (2012). Quality standards for English-language closed captioning. Retrieved July 31, 2018, from www.crtc.gc.ca
- CRTC. (2015). Call for comments on the English-language closed captioning quality standard related to the accuracy rate for live programming. Retrieved July 31, 2018, from <https://crtc.gc.ca/eng/archive/2015/2015-325.pdf>

- CRTC. (2016). English-language closed captioning quality standard related to the accuracy rate for live programming. Retrieved July 31, 2018, from <https://crtc.gc.ca/eng/archive/2016/2016-435.pdf>
- Dong, H., Clarkson, P. J., Cassim, J., & Keates, S. (2005). Critical user forums - an effective user research method for inclusive design. *The Design Journal*, 8(2), 49–59.
- Dong, H., Keates, S., & Clarkson, P. J. (2004). Inclusive design in industry: barriers, drivers and the business case. *LNCS*, 3196, 305–319.
- Downey, G. (2007). Constructing closed-captioning in the public interest: from minority media accessibility to mainstream educational technology. *Info*, 9(2–3), 69–82. <https://doi.org/10.1108/14636690710734670>
- FCC. (2014a). FCC Adopts Non-quantitative Quality Standards for Closed Captioning of Video Programming - ProQuest. Retrieved March 31, 2017, from <http://search.proquest.com.ezproxy.lib.ryerson.ca/docview/1511823395/D8A7DA42329A49FCPQ/5?accountid=13631>
- FCC. (2014b, February 20). FCC moves to upgrade TV closed captioning quality. Retrieved April 2, 2017, from https://apps.fcc.gov/edocs_public/attachmatch/DOC-325695A1.pdf
- FCC moves to upgrade TV closed captioning quality. (2014, February 20). Retrieved August 2, 2017, from https://apps.fcc.gov/edocs_public/attachmatch/DOC-325695A1.pdf
- Federal Communications Commission. (2015). Closed Captioning on Television. Retrieved January 25, 2017, from <http://transition.fcc.gov/cgb/consumerfacts/closedcaption.pdf>
- Fels, D. I., Lee, D., & Branje, C. (2005). Emotive captioning and access to television. *Americas Conference On*, 2330–2337.
- Gilbert, L. C. (1959). Speed of processing visual stimuli and its relation to reading. *Journal of Educational Psychology*, 50(1), 8.
- Gill, J., & Perera, S. (2003). Accessible universal design of interactive digital television. In *Proceedings of the 1st European conference on interactive television: from viewers to actors* (pp. 83–89).
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183.
- Hersh, M. (2013). Deaf people's experiences, attitudes and requirements of contextual subtitles: A two-country survey. *Telecommunications Journal of Australia*, 63(2), 23.1-23.14. <https://doi.org/10.7790/tja.v63i2.406>
- Hong, R., Wang, M., Yuan, X.-T., Xu, M., Jiang, J., Yan, S., & CHua, T. (2011). Video accessibility enhancement for hearing-impaired users. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 7(1), 421–430.
- Hu, Y., Kautz, J., Yu, Y., & Wang, W. (2013). Speaker-following video subtitles. *ACM Trans.*

- Multimedia Com-Put. Commun. Appl*, 2(1), 32. <https://doi.org/10.1145/0000000.0000000>
- Huey, E. B. (1968). *The psychology and pedagogy of reading*. Cambridge: MIT Press.
- Javal, E. (1878). Essai sur la physiologie de la lecture. *Annales d'Oculistique*, 80, 61–73.
- Jensema, C., Mccann, R., & Ramsey, S. (1996). Closed-Captioned Television Presentation Speed and Vocabulary. *American Annals of the Deaf*, 141(4), 284–292. <https://doi.org/10.1353/aad.2012.0377>
- Keates, S., Clarkson, P. J., Harrison, L.-A., & Robinson, P. (2000). Towards a practical inclusive design approach. In *Proceedings on the 2000 conference on Universal Usability* (pp. 45–52). ACM. <https://doi.org/10.1145/355460.355471>
- Ladurantaye, S. (2012, July 5). Closed captioning: Get it right, the CRTC tells broadcasters. *The Globe and Mail*.
- Lee, D. G., Fels, D. I., & Udo, J. P. (2007). Emotive captioning. *Computers in Entertainment*, 5(2), 11. <https://doi.org/10.1145/1281329.1281344>
- Lenay, C., Gapenne, O., Hanneton, S., Margue Catherine, & Genouëlle. (2003). Sensory substitutions: limits and perspectives. In A. Streri & E. Gentaz (Eds.), *Touching for Knowing : Cognitive Psychology of Haptic Manual Perception* (pp. 275–292).
- Llewellyn, A., & Hogan, K. (2000). The use and abuse of models of disability. *Disability & Society*, 15(1), 157–165. <https://doi.org/10.1080/09687590025829>
- Marks, D. (1997). Models of disability. *Disability and Rehabilitation*, 19(3), 85–91. <https://doi.org/10.3109/09638289709166831>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative Data Analysis A Methods Sourcebook* (3rd ed.). London: Sage.
- Neves, J. (2008). 10 fallacies about Subtitling for the d/Deaf and the hard of hearing 1. *The Journal of Specialised Translation*, (10), 128–143.
- Newell, A. F., Gregor, P., Morgan, M., Pullin, G., & Macaulay, C. (2010). User-sensitive inclusive design. *Universal Access in the Information Society*, 10(3), 235–243. <https://doi.org/10.1007/s10209-010-0203-y>
- Noyes, J. M., & Bruneau, D. P. J. (2007). A self-analysis of the NASA-TLX workload measure. *Ergonomics*, 50(4), 514–519. <https://doi.org/10.1080/00140130701235232>
- Ofcom's Code on Television Access Services. (2015). Retrieved July 30, 2017, from https://www.ofcom.org.uk/__data/assets/pdf_file/0016/40273/tv-access-services-2015.pdf
- Ohene-Djan, J., Wright, J., & Combie-Smith, K. (2007). Emotional subtitles: a system and potential applications for Deaf and Hearing Impaired people. In *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments*.

- Parault, S. J., & Williams, H. M. (2010). Reading motivation, reading amount, and text comprehension in Deaf and hearing adults. *Journal of Deaf Studies and Deaf Education*, 15(2), 120–135. <https://doi.org/10.1093/deafed/enp031>
- Persad, U., Langdon, P., & Clarkson, P. J. (2007). Characterising user capabilities to support inclusive design evaluation. *Universal Access in the Information Society*, 6(2), 119–135. <https://doi.org/10.1007/s10209-007-0083-y>
- Piccolo, L., Melo, A. M., Cecília, M., & Baranauskas, C. (2007). Accessibility and interactive TV: design recommendations for the Brazilian scenario. *Human-Computer Interaction – INTERACT 2007*, 361–374. https://doi.org/10.1007/978-3-540-74796-3_34
- Previous Code on Television Access Services - Ofcom. (2012). Retrieved May 9, 2017, from <https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/tv-access-services/code-tv-access-services-2013>
- Rashid, R., Aitken, J., & Fels, D. I. (2006). Expressing emotions using animated text captions. *Lncs*, 4061, 24–31. <https://doi.org/10.1007/11788713>
- Rashid, R., Vy, Q., Hunt, R., & Fels, D. I. (2008). Dancing with words: using animated text for captioning. *International Journal of Human-Computer Interaction*, 24(5), 505–519. <https://doi.org/http://dx.doi.org/10.1080/10447310802142342>
- Romero-Fresco, P. (2009). More hast less speed: Edited versus verbatim respoken subtitles. *Vigo International Journal of Applied Linguistics*, 6(1), 109–133.
- Sillman, D. (1984). Television captioning for the Deaf. *IEEE Transactions on Consumer Electronics*, 30(2), 62–65.
- Silverman, C., & Fels, D. I. (2002). Emotive captioning in a digital world. *LNCS*, 2398, 292–294.
- Snell, N. E. (2012). *Toward a theory of media reconciliation: A closed captioning exploratory study*. Clemson University.
- Steen, M. (2011). Tensions in human-centred design. *CoDesign*, 7(1), 45–60. <https://doi.org/10.1080/15710882.2011.563314>
- Strong, M., Prinz, P. M., Helman, E., Keast, M., Kessler, N., Kuntze, L., ... Yang, J. (1997). A study of the relationship between American Sign Language and english literacy. *Journal of Deaf Studies and Deaf Education*, 2(1), 37–46.
- Szarkowska, A., Krejtz, I., Klyszejko, Z., & Wieczorek, A. (2011). Verbatim, Standard, or Edited?: Reading patterns of different captioning styles among Deaf, Hard of Hearing, and hearing viewers. *American Annals of the Deaf*, 156(4), 363–378. <https://doi.org/10.1353/aad.2011.0039>
- Udo, J. P., & Fels, D. I. (2010). The rogue poster-children of universal design: closed captioning and audio description. *Journal of Engineering Design*, 21(2–3), 207–221. <https://doi.org/10.1080/09544820903310691>

- van Kleef, E., van Trijp, H. C. M., & Luning, P. (2005). Consumer research in the early stages of new product development: a critical review of methods and techniques. *Food Quality and Preferences*, 16(3), 181–201.
- Vanderheiden, G. (2000). Universal design of consumer products: Current industry practice and perceptions. In *Proceedings of the IEA/HFES Congress* (pp. 19–22).
- Web Accessibility and Closed Captioning in Australia and New Zealand. (2015). Retrieved January 26, 2017, from www.3playmedia.com
- Zdenek, S. (2011). Which sounds are significant? towards a rhetoric of closed captioning. *Disability Studies Quarterly*, 31(3), n.p.
- Zdenek, S. (2015). *Reading Sounds: Closed-Captioned Media and Popular Culture*. Chicago: University of Chicago Press.