

1-1-2012

# Improving and Evaluating a Software Tool for Providing Animated Text Enhancements to Close Captions

Jorge Mori  
*Ryerson University*

Follow this and additional works at: <http://digitalcommons.ryerson.ca/dissertations>



Part of the [Software Engineering Commons](#)

---

## Recommended Citation

Mori, Jorge, "Improving and Evaluating a Software Tool for Providing Animated Text Enhancements to Close Captions" (2012). *Theses and dissertations*. Paper 1415.

This Thesis is brought to you for free and open access by Digital Commons @ Ryerson. It has been accepted for inclusion in Theses and dissertations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact [bcameron@ryerson.ca](mailto:bcameron@ryerson.ca).

IMPROVING AND EVALUATING A SOFTWARE TOOL FOR PROVIDING ANIMATED  
TEXT ENHANCEMENTS TO CLOSE CAPTIONS

by

Jorge Mori

BSc, Ryerson University, Toronto, Ontario, 2008

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Science

in the Program of

Computer Science

Toronto, Ontario, Canada, 2012

©Jorge Mori 2012

## 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.

# IMPROVING AND EVALUATING A SOFTWARE TOOL FOR PROVIDING ANIMATED TEXT ENHANCEMENTS TO CLOSE CAPTIONS

Jorge Mori

MSc, Computer Science, Ryerson University, 2012

## **ABSTRACT**

While television and film technologies have changed according to user preferences, Closed Captions (CC) have suffered from a lack of innovation since their inception in the 1970's. For the Deaf and Hard of Hearing communities CC provides only limited access to non-speech audio information. This thesis explores the usability of a new captioning application, EnACT that provides animated text for non-speech audio information such as the emotions portrayed and their corresponding intensities. Reactions from software users were collected and evaluated. Participants found the software easy to use and a suitable alternative to conventional CC options for non-speech audio however, they disliked the amount of time it took for them to adjust timing for the animations of the captions. Overall, participants rated EnACT easy to use and the task of assigning emotions and their corresponding intensities to the video script as relatively simple, however, additional emotional labels were requested by participants overall.

## **ACKNOWLEDGEMENTS**

I would like to thank many people, without whom I would not have been able to complete the research and implementation of my thesis over the course of my time at Ryerson.

First and foremost I would like to express my sincere gratitude and appreciation to my thesis supervisor Deborah Fels who supported and guided me through my research and studies at Ryerson University.

I would also like to thank the members of my thesis committee, Sophie Quigley, Abdolreza Abhari and Eric Harley for their time and effort in reviewing my thesis and providing valuable feedback.

I would like to thank my wife and best friend Leshanne Pretty because her love, patience and motivation kept me going on finishing this project, my parents Jorge Mori and Roxana Saavedra, my sister Lizbeth Mori, my brother Aldo Mori. Without their constant positive, emotional, moral, and loving support, I would not be where I am today.

Finally, thank you to all my friends and colleagues within the Center for Learning Technologies (CLT) at Ryerson University, who provided such a pleasant environment to work in, showed interest as my thesis work progressed, and were constantly by my side to offer help.

## Table of Contents

Chapter I – Introduction.....	1
1.1 Contributions of the Thesis .....	3
1.2 Thesis outline .....	5
Chapter II - Literature Review .....	6
2.1 Universal Design Theory .....	6
2.1 Closed Captioning .....	7
2.1.2 Closed Caption Standards and Regulations.....	9
2.1.3 EIA-608 formats .....	11
2.1.4 CEA-708 format (formerly EIA-708).....	12
2.1.5 Other Captioning standards .....	14
2.1.6 Captioning Types.....	15
2.2 CC and Literature .....	16
2.3 Use of Graphics and Animations .....	17
2.3.1 Animated Text / Kinetic text .....	21
2.4 Emotions.....	24
2.4.1 Emotions in sound and music .....	24
2.4.2 Use of EC to provide emotions through music.....	25
Chapter III – Methodology and Implementation .....	28
3.1 Target Users .....	28

3.2 Research Questions .....	29
3.3 Study Design .....	29
3.3.1 Usability study with Amateur Captionists.....	32
3.3.2 Case study with Professional Captionists.....	34
3.3 Equipment .....	34
3.4 Data Collection and Analysis.....	35
3.4.2 Usability questionnaire .....	36
3.4.3 Use Cases.....	39
3.5 System Description/Design .....	39
3.5.1 EnACT: Version 0.5 .....	40
3.6 Development Platform .....	42
3.6.1 C#.NET Framework 3.5 .....	42
3.6.2 Adobe Flash and ActionScript 2.0.....	43
3.6.3 Extensible Markup Language (XML) .....	44
3.7 History of EnACT development .....	45
3.7.1 EnACT Engine - initial Versions (1.0 and 2.0) .....	45
3.7.2 EnACT Editor Prototypes.....	47
3.8 EnACT Editor Version 3 (September 2008 – Present) .....	53
3.8.1 Resolutions implemented in EnACT 3.0.....	53
3.8.2 EnACT Script Properties .....	60

Chapter IV - Evaluation .....	63
4.1 Usability .....	63
4.2 Case study .....	68
4.2.1 Participant 1 .....	68
4.2.2 Participant 2 .....	71
4.2.3 Participant 3 .....	72
4.3 Discussion .....	73
4.3.1 Working with Emotions.....	74
4.3.2 EnACT User Interface .....	77
4.3.3 Confidence and Comfort Level using EnACT .....	78
4.3.4 Participant suggestions and opinions on EnACT .....	79
4.3.5 Limitations of the research .....	85
Chapter V - Conclusion, summary and future work .....	87
5.1 Summary .....	87
5.2 Future Research.....	89
Appendices.....	91
References.....	126



## **List of Tables**

Table 1. Problems of EnACT 2.0.....	52
Table 2. Problems and Solutions that were solved between Version 2 to Version 3 .....	58
Table 3. Default colours for emotions in EnACT.....	62
Table 4. Chi-square table for ratings in the difficulty of task attempted where 1 = Difficult and 3 = Easy .....	64
Table 5. Chi-square table for the rate of opinions of the location of elements where 1=Poor and 3=Good .....	65
Table 6. Chi-square results of the confidence rating from participants from using EnACT and participant's comfort rating when using the software where 1 = low and 3 = high. ....	67

## List of Figures

Figure 1. Closed Caption example.....	9
Figure 2. Closed Caption channels [4].....	11
Figure 3. CEA-708 capabilities [16].....	13
Figure 4. A screen shot of a Teletext system called Ceefax .....	15
Figure 5. A comic book art approach to represents emotions and intensities.....	19
Figure 6. Use of color, graphics, icons and animations to represent sound information.....	20
Figure 7. KIM displays incoming messages and replays messages in the main conversation [6] .....	22
Figure 8. Examples of animations used in [25] .....	23
Figure 9 Kinetic text used in [7] .....	24
Figure 10. EC showing lead singer upper left and background singer on the bottom right .....	26
Figure 11. System Design for EnACT [40] .....	41
Figure 12. Relationship of the different EnACT system components. ....	46
Figure 13. The EnACT captioning tool is divided into two major components that are needed for the EnACT engine to render the EC .....	46
Figure 14. The First EnACT prototype developed by Zhang, Hunt and Mori, 2006 .....	48
Figure 15. Interface elements of the EnACT system.....	49
Figure 16. The script contains four dialogues but only 2 appear on the SEA .....	50
Figure 17. Example code of error in the ``emotion type`variable .....	51
Figure 18. Screenshot of EnACT Editor, Version 3 .....	53
Figure 19. A code sample from the parseDialogue method.....	54
Figure 20. The bug fix in the parseDialog method .....	55
Figure 21. WriteDialogue method in pseudo code .....	56

Figure 22. WriteDialogue bug fix .....	57
Figure 23. Creating a new project in EnACT .....	60
Figure 24. Editing and selecting emotions and intensities in EnACT .....	61
Figure 25 EnACT - Script Properties.....	61
Figure 26. The study showed positive feedback during specific tasks assigned to the participants .....	65
Figure 27. Experience regarding the GUI.....	66
Figure 28. Participants rating of the comfort level when using EnACT to caption a movie file .	68
Figure 29. Screenshot of the dashboard of the Professional Captioning tool ProCap.....	69
Figure 30. An alternative way to mark-up the script with emotions and intensities uses the right- click.....	78
Figure 31. Windows Media Player does not display the time in the same format that is required for input in the EnACT interface to set the timing for EC.....	81
Figure 32. EnACT Version 3.0 redesigned by an Amateur Captionist participant based on his suggestions for improvement to the interface.....	83

## **List of Appendices**

Appendix A. Definitions .....	91
Appendix B. Problems and Solutions that were solved between Version 2 to Version 3 .....	92
Appendix C. Ethics approval .....	97
Appendix D. Questionnaire .....	98
Appendix E. Training documents .....	105
Appendix F. Study Tasks.....	107
Appendix G. Recruitments emails and posters .....	109
Appendix H. Payment Receipts .....	111
Appendix I. Problems with EnACT 2.0 and solutions implemented in EnACT 3.0 .....	112
Appendix J. Computer Specifications.....	119
Appendix K. Consent Form .....	120
Appendix L. Participants Script.....	123

## **Chapter I – Introduction**

Access to arts and culture in western society is seen as an important aspect of social justice and inclusion. There have been a number of innovations in technology, social and regulatory systems and public attitude that have advanced this notion of access to arts and culture for people with disabilities. This not only includes better access to education, facilities, and production and performing opportunities but also improved access to content by audiences with disabilities. One of the first access technologies to be formalized and regulated was Closed Captioning for people who are deaf or hard of hearing. In this thesis, I will use the term deaf (D) to refer to all individuals who have little or no hearing and hard of hearing (HOH) to refer to individuals who have mild to profound hearing loss. I will use D/HOH when I refer to both groups. For full definitions of terms used to refer to people who have hearing loss, see Appendix A.

Currently, it is estimated that there are approximately 310,000 deaf Canadians and 2.8 million hard of hearing Canadians [1]. It is also estimated that about 1 million Americans are functionally deaf and close to 10 million are HOH. Within this group of D/HOH Americans, about half are reported to be 65 years or older and less than 4 per cent are less than eighteen years of age [2]. Even though [1] acknowledges that no fully credible census has been done to determine the actual number of D/HOH people in Canada and the United States (US), it is believed that between Canada and the US there is approximately 1,310,000 deaf people and 12.8 million HOH.

Closed Captions (CC) are the verbatim translation of the spoken dialog and are overlaid on the video image on screen (often in the lower center of the image) as described in [3]. CC

uses a simple text-based format with a character set built into a television decoder; white characters displayed on a black background with a single font size.

CC have been in existence since the early 1970's however, while television and film technology have evolved dramatically, CCs remained similar to that available in those early days. Recent changes to captions have included the adoption of a very limited symbol set (music note, punctuation and descriptions contained in brackets) in an attempt to convey non-speech information such as music. For example, when music is playing a music note is used and where there is a speaker on the screen communicating by yelling the caption can display "[angry]".

Electric Industries Alliance (EIA) developed EIA-608, a standard for displaying CC that specifies line 21 of the Vertical Blanking Interval (VBI) and fixed bandwidth of 960 bits per second [4] as the transmission specification for analog CC. As a result of the move to digital television (DTV), a CC standard for digital television, CEA-708, has been adopted in North America. This standard advanced the possible configurations for captions considerable. The data bandwidth has been increased to 9600 bits per second [4], and this allow for variable-space fonts, a variety of font sizes and multiple colours and animations. This new standard offers the possibility of innovation for and improvements to CC.

Improvements to CC are warranted not only to keep up with the progress of digital television technology but also to address the numerous issues that have been identified by D/HOH users. Studies have reported that the people in the D and HOH communities believe that they are missing important information with EIA-608 CC particularly the non-speech audio information such as music, speech prosody and sound effects. [5]. It is incumbent on the research and development and the television technology community to begin to address these identified

issues so that inclusion in arts and culture for people who are D/HOH can be maintained and advanced.

## **1.1 Contributions of the Thesis**

According to the literature (explained in more detail in chapter 2), there has been a lack of research investigating ways to improve how captions are produced and displayed for audiences in order to meet the challenges identified by users. Although the new CEA-708 standard allows for improved captions that use colours, animations and graphics, little research has been carried out to determine how best to use these new features as well as to understand the receptivity of captionists and audiences to produce and consume them respectively. However, there is evidence from other areas such as instant messaging and chat applications [6] where this type of text and graphical content is used and accepted; although limited user evaluation results are available. Research performed by [7] was one of the first studies to examine enhancements to captions specifically designed to address user concerns and implements some of the CC attributes of CEA-708 [7]. Based on the success of the study by Rashid et al. (2008) it was decided to extend the research by improving and evaluating the Emotive and Affective Captioning Tool (EnACT) that was designed to make creating and producing enhanced captions (EC) efficient and effective.

The functional specifications of EnACT were then to allow individuals to create EC by selecting words within a script and assigning a desired emotion and intensity. The tagged words would then be rendered into EC that would be displayed on the associated video. A second important specification for EnACT is to be a plug-in or add-on to an existing captioning or video editing tool although basic caption functionality such as screen placement and timing was

necessary in order to allow independent use and evaluation of EnACT by users who did not have any previous captioning experience (Amateur Captionist).

This thesis presents a description of how EnACT was extended and modified from its Version 2.0 to current Version 3.0 to improve the functionality and the usability. My contributions can be grouped in two categories: research contributions and software contributions. They are as follows and have been described in more detail later within this thesis:

#### Research Contributions:

1. Developed EnACT to a usable state.
2. Published EnACT in an open source database.
3. Evaluated EnACT with target users (Amateur and Professional Captionists)
4. Study revealed that it seems feasible not only to use EnACT but also that the process of adding animated captions is possible and even enjoyable.

#### Software Contributions:

1. Fixed major bugs in the software.
2. Modified the user interface:
  - a. Provided a video conversion solution to convert any video file to a flash video format.
  - b. Created and added a preview window to the user interface to display the Enhanced Captions.
3. Created additional functionality to allow users to create new projects.
4. Added keyboard shortcuts to control the timing of the Enhanced Captions.



5. Developed a user study to explore the usability of EnACT to create Enhanced Caption video files.

## 1.2 Thesis outline

The thesis is structured in the following order:

- **Chapter 1:** Serves as an introduction to the thesis. This chapter gives an overview of the goal of the study and background information about the topic.
- **Chapter 2:** Presents the literature review of the thesis. This chapter explains the history of closed captioning, the standards for its development, quality and current state.
- **Chapter 3:** Presents and explains the system architecture, design and implementation of EnACT. This chapter provides a detailed description regarding how the EnACT interfaces are organized and the software capabilities. It also presents the methodology used in this thesis to complete a usability study with the two groups of participants: Professional and Amateur Captionists.
- **Chapter 4:** Presents the results, findings and discussion from the study described in Chapter 3. This chapter provides detailed descriptions of the responses, comments and suggestions from the participants during and after the usability study.
- **Chapter 5:** Presents the conclusions, limitations of the thesis and suggestions for future work.

## **Chapter II - Literature Review**

This chapter discusses the literature explaining the background information about the EnACT system. Research that motivated the creation of animated text to become Enhanced Caption and that has being part of my research work at Ryerson University since 2007.

This section will explain and introduce the theory of Universal Design and how it is applied to Closed Caption. Following this, a brief history of Closed Captions in North America is introduced, research about the problems with current captions and how the use of graphics or animations as potential solution to the problems. Finally this chapter shows how animations inspire the creation of Enhanced Captions and the need to create a software tool that will help the creation of them.

### **2.1 Universal Design Theory**

Universal Design Theory (UDT) relates to the design of buildings, products and environments to be usable by people with disabilities and people without disabilities without the need for adaptation or a specialized design [8].

UDT was created as an initiative to aid designers, architects and builders make built environments more accessible to individuals with disabilities. These groups realized that the change required for the people with disabilities benefited many others. As a result of this, UDT was expanded to include all individuals so that the goal became “design for all” [8]. The implementation of UDT in the creation of products and services has shown that it can decrease the need for costly adaptations and/or retrofits for each group with different usability requirements [9].

The seven underlying principles of UDT are as follows:

1. **“Equitable use:** The design is useful and marketable to people with diverse abilities.
2. **Flexibility use:** The design accommodates a wide range of individual preferences and abilities.
3. **Simple and intuitive use:** Use of design is easy to understand, regardless of user’s experience, knowledge, language skills, or current concentration level.
4. **Perceptible information:** The design communicates necessary information effectively to the user, regardless of ambient conditions or the user’s sensory abilities.
5. **Tolerance for error:** The design minimizes hazards and the adverse consequences of accidental or unintended actions.
6. **Low physical effort:** The design can be used efficiently and comfortably and with a minimum fatigue.
7. **Size and space for approach and use:** Appropriate size and space is provided for approach, reach, manipulation, and use regardless of user's body size, posture, or mobility” ([8] pp. 189).

Since Closed Captioning is considered to be a service to the public all the rules of UDT apply with the exception of rules 6 and 7 since Closed Caption are not a physical entity.

## **2.1 Closed Captioning**

CC is the process of transcribing spoken dialogue and non-speech information into verbatim text equivalents and symbols (see Figure 1 for an example of CC)[10]. The text is electronically encoded into the content files (digital) or Vertical Blanking Interval (analogue) by the captionist. It is then transmitted to the television or cinema where it is encoded by the

hardware at the user end (TV set in the case of the television or specialized captioning decoding equipment in the cinema) [4]. In North America, the National Television System Committee (NTSC) specifies 525 scan lines for each image that is displayed on the TV screen [11]. The Vertical Blanking Interval (VBI) is the time between the last scanning beam scanning a horizontal line and the beginning of the next scanning beam process. Analogue captions are typically allocated on line 21 of the vertical blanking interval (VBI) [4]. In North America, captions are typically displayed as white text on a black background according to the EIA-608 formatting standard (further discussion of caption formatting standards is provided in Section 2.1.2).

CC was created in 1970's to benefit the D/HOH communities as part of social justice movements at that time. CC was to provide equivalent access to publically available culture and to enable viewers who were D/HOH to understand and enjoy TV shows and movies. Captioning standards and regulations were introduced later to ensure that there were common approaches to caption displays and some form of quality and quantity control. In addition, as time progressed, it was discovered that captions also served other purposes and communities such as second language learning and accessing television content in noisy locations such as pubs and gyms [12].



Figure 1. Closed Caption example

## 2.1.2 Closed Caption Standards and Regulations

### 2.1.2.1 Canada

In May 2007, the CRTC released a new policy with respect to CC [13]. Not only did the quantity of captions required by all French and English language broadcasters increase to 100% with the exception of advertising and station promotions, they also stipulated that there be some measure of quality. They wanted to have created minimum quality standards to ensure consistency across the entire broadcasting system for the benefit of caption audiences.

The CRTC requested the Canadian Association of Broadcasters (CAB) to coordinate the establishment of French and English language working groups to design and implement universal standards for CC that will deliver solutions and guidelines to maintain the same quality [14]. Among other recommendations, this preliminary report on CC suggested the preference of CAB is to Roll-Up captions instead of pop-on captions (see Section 2.1.6 for definitions and example of these caption styles) for pre-recorded programs. The preliminary report was opened to the public for scrutiny, and on January 19<sup>th</sup>, 2009, and consumer and caption advocates argued against many aspects of this report including caption style and quality definition and measures.

As a result, a consensus on the definition and measures for quality remained outstanding and controversial.

The CRTC then released a ruling stating that all Canadian pre-recorded programming must use the pop-on captioning format in July 2009. They also stated that the standards submitted by the Working Groups were incomplete and required further attention. The CRTC advised the Working Groups to re-submit revised and complete standards addressing the following areas ([15] Sec 84 – Sec 90):

- “Speed of captions
- Captions that block or are blocked by on-screen information
- Acceptable rate of error in the captions
- Standards for digital broadcasting, including in high definition”.

The CRTC requested that the CAB provide academically sound evidence supporting their proposed standards with respect to all of the requested areas and to provide validation exercises to justify their recommendation. As part of the request for validation exercises, the CRTC also requested complete descriptions of the methodologies used and complete evidence that the results achieved were statistically valid and representative of all user communities.

In 2010, the CAB stated that programming shows including dramas and documentaries would be displaying pop on captions; however other pre-recorded shows will display roll-ups due to time constraints [16].

In 2011, the CAB provided their final report on English language CC standards on February but the CRTC was not satisfied with the clarifications provided. The CRTC called for

comments from the public to appropriateness of the CC quality standard provided by the CAB and any related actions to be taken in the future.

This action taken by the CRTC demonstrated that there was a need to involve the public in captioning decisions and represent their interests in the development of CC in Canada

### 2.1.3 EIA-608 formats

As introduced in section 2.1, to transmit broadcasted CC, text representing audio dialogue is encoded into a broadcast signal, decoded and then displayed in the picture area of a television set. The encoding process relies on an operator (called a captionist) who is responsible for transforming verbal speech within a program into text. In the NTSC and Standard Definition Serial Digital Interface (SD-SDI) television system in North America, the captioning data is transmitted through the VBI line 21, outside the normal viewing area of the picture. The decoder in the television set then strips the captioning information from line 21 and displays it on screen. A new method of encoding has been created for HDTV, and will be described later in the paper.

EIA-608 contains four channels as shown in Figure 2 for transmitting CC.

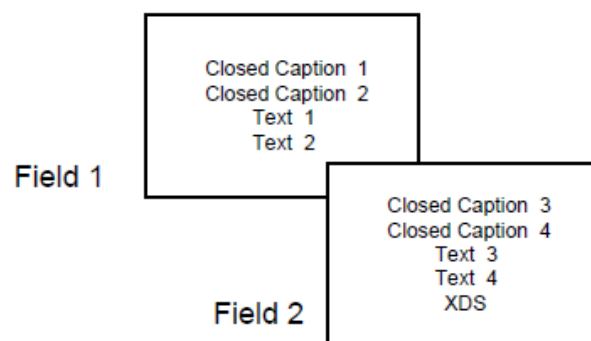


Figure 2. Closed Caption channels [4]

At the TV station, a CC encoder places the text data on line 21. At the place of TV viewing, the decoder built into the TV or set top decoder is used to decode the CC and display it onscreen. For this process to take place, Field 1 and Field 2 are used for this encoding and decoding process. “Field 1 carries the data through the VBI Closed Caption 1, Closed Caption 2, Text channel 1 and Text Channel 2. Field 2 carries Closed Caption 3, Closed Caption 4, Text 3, Text 4 and Extended Data Service (XDS)” [4].

This form of CC uses simple text based format, consisting of a single white colour font size, displayed against a black background and when the system was first created, the CC was only displayed in white uppercase letters. These days CC can now be used with a mix of upper and lower case letters, a small set of text colours with a few special characters (e.g. music notes) [3].

In EIA-608 there are 60 fields per second, so the whole system can transmit a total of 120 characters per second (the captions field changes constantly while the XDS and text occasionally).

The bit rate in EIA-608 is 960 bits per second (bps) since there are 120 characters per second and each text character is 8 bits ( 7 bits plus 1 parity bit) [4]. Because of this limited bandwidth the system was initially limited to only use white uppercase font. However several options have been added allowing the use of mixed case letters and a small set of colours although they are usually absent as the users have grown used to the white font uppercase letters.

#### **2.1.4 CEA-708 format (formerly EIA-708)**

CEA-708 was developed by the Electronic Industries Alliance (EIA) and is the CC standard for digital broadcast content and technology. While the CC from the EIA-608 standard



consists of an analog waveform inserted on line 21 of the NTSC VBI; “DTV is transmitted as a logical data channel in the DTV digital bit stream” [17].

CEA-708 contains features for using alternative fonts, colours, caption positioning and other options related to text-based enhancements [3] that considerably expands the style options from EIA-608 as shown in Figure 3.

CEA-708 allocates a data rate ten times greater (9600 bps) than the EIA-608 standard’s analog version [4]. The increased capacity afforded by the higher data rate opens up the possibility for simultaneous transmissions of captions in multiple languages or styles [17].



Figure 3. CEA-708 capabilities [18]

EIA-708 is also able to use a variety of increased horizontal and vertical aspect ratios such as: 704x480, 1280x720 and 1920x1080 in comparison to the 525 horizontal scan line used on the NTSC analog format. This increase in flexibility of display properties, transmission rate and aspect ratios means that there is a flexibility for captions not possible with legacy analog technology.

The introduction of digital television technology and the resulting increased technical and creative flexibility CC was a catalyst for the EC project. Images, colour, animation and different screen locations were now possible, and could be developed and evaluated. .

### **2.1.5 Other Captioning standards**

Teletext is a service, mainly available in Europe and Australia [19], that consists of pages of text-based information it was used to retrieve information about sporting news, weather, as well as subtitles for the hard of hearing (the equivalent of CC) unlike the North American CC which is only used to provide captions for the D/HOH (as shown in Figure 4). This method of captioning began in the early 1970's when the British Broadcasting Corporation (BBC) and the Optional Reception of Announcements by Coded Line Electronics (ORACLE) started the first test services. Teletext can display colour, different fonts, mixed case lettering and animations however no study or projects have involved using those animations to provide extra information for the captions. The research done at Ryerson University and presented in this thesis is the only project in the world that provides an alternative solution to what CC cannot do today.

The VBI is a common method used to decode Teletext data. Most of the Teletext systems adopt 625 lines instead of the 525 lines used in NTSC systems [20]. Teletext has a higher transmission rate that is able display more information than CC that use the EIA-608 standard and currently uses different fonts and animations, however their standards have also missed the opportunity to research the way in which these capabilities can be used to effectively

communicate information for audiences in a more meaningful way.



Figure 4. A screen shot of a Teletext system called Ceefax

### 2.1.6 Captioning Types

There are three main types of captioning according to [21]:

- Off-line captioning: This refers to captions that are created for and applied to pre-recorded media such as TV shows or documentaries and often created by third party companies. Currently there are two main types of off-line captioning and they are used widely in pre-recorded media:
  - Pop-on captions: the entire caption appears on the screen at once and remains there until it disappears or is replaced by another caption.
  - Roll-up captions: This caption appears on the screen by displaying the words from left to right and one line at a time in a continuous motion. Once the entire line is complete, it scrolls up to make way to another caption, as this happens the line on the top is erased. Usually two or three lines of text appear at one time.

- On-line or Live Captions: These types of captions refer to captions that are provided simultaneously with a broadcast. These captions normally appear as roll-up captions.
- Real-Time Captions: This type of caption refers to captions that are created and transmitted at the same time of the broadcast. They are done by experienced Real-Time captionists using a stenotype machine and appear as roll-up onscreen.

The EC proposed in the study of this thesis deals with improving the current pop-on captions as these types of captions are done before broadcast.

## **2.2 CC and Literature**

Currently, CC displays the verbatim or paraphrased transcript of the spoken words and the non-speech information such as tone of voice, inflection, rate of speech, volume or emotion of speech are not often included in either 608 or 708 captions. Occasionally, and time and space permitting, emotions may be labeled with a single descriptor such as “[happily]” or punctuation such as “!”. Background sound may be described with one or two words when important. However adding more text can affect the readability of the caption, and there are no standards regarding the formatting of added words (e.g., some captions contain emotion words in square brackets, some in italics, etc.). In addition, words describing the emotions likely do not produce the same effect on the viewer as expressing the emotion through other means.

[22] reported that missing words, spelling errors and captions moving too quickly caused dissatisfaction, confusion and unnecessary cognitive load for the audience. Furthermore, the interpretation required by captionists in translating audio information to the audience cannot be standardized because of its subjectivity. [5] reported that caption viewers wanted captions to be explicit rather than implied. This further defines the role of a captionist as a fine balance between

delivering more meaningful information to the audience without sacrificing the usefulness of captioning most basic function, to display dialogue accurately.

The recommended caption speed is 141 to 150 words per minute with many viewers not experiencing difficulty until captions reach 170 words per minute [23]. This could mean that it is possible to add more text to describe the non-speech audio information but this may then add extra processing cognitive load and it could cause exhaustion. There may be other ways to express this information, such as through the new CEA708 features of colour, animation and graphics that would not add more text and resulting reading load, and still be effective.

Whilst the basic function of captioning is to display speech input, research by [24] found that the much of the semantic information to be gained from language comes primarily from communication cues outside of the words spoken in a dialogue. This study breaks semantic communication down as such: 7% words, 38% paralinguage (the non-verbal part of speech like emotions and intensities) and 55% body language. Paralinguage provides five time times more information about language than words alone. Current captions have provided little to no improvement in representing the critical information that paralinguage conveys. As a result of this lack of paralinguage information expressed through captions, D/HOH viewers must compensate by relying on visual cues such as body language and gestures combined with text captions to understand the show's content. This can be a problem for dialogue where the speaker is not displayed on the screen like a narrative voice or a background actor.

### **2.3 Use of Graphics and Animations**

As it was previously explained, CC describes background noises or emotions from dialogue by describing it with text between two square brackets characters “[ ]”; sometimes the text is also italicized. This technique increases the number of words per minute (wpm) displayed

and may therefore decrease the readability or speed of display, potentially making the captions more difficult to follow or crowded.

Graphics, however, may be able to overcome the limitations of text in describing non verbal information. [5] experimented with the conventional design of captions by displaying graphics instead of text to provide some of the paralinguistic and sound effect information. [3] suggested that graphics could be used to decrease the amount of text-based captioning required which, as a result, could reduce the wpm. Graphics could also assist in capturing sound information that cannot be described easily using text. Speech bubbles used in comic books are a good example on how graphics can help the reader understand the mood and emotion from a dialogue.

One variation of graphic displays were studied in [5] where researchers experimented with a design following comic book conventions for a video of an comedic spoof on opera that contained dialogue and music. The graphic captions consisted of using speech bubble shapes and text styling to represent four basic emotions (happy, sad, anger and fear), and intensities of these emotions (see Figure 5) as well as music and sound effects. The rounded rectangle represents dialogue and the oval speech bubble represents background sound or music.



**Figure 5. A comic book art approach to represents emotions and intensities**

Study results showed that while this approach increased the participant's understanding of the content, several participants disliked the use of comic book conventions because they associated it with children's content.

In a second study by the same authors [3] used graphics, colour, icons and animations to accompany text as shown in Figure 6. The design of the enhancements was carried out by the graphic artist associated with the production and the director of the show. To evaluate the impact of the enhancements on audiences, the authors presented D/HOH participants with a version of the video containing conventional CC and another version using the enhanced version of the captions. In this study six emotions were represented including: fear, anger, sadness, happiness, disgust and surprise. The specific discrete emotion and the intensity of the emotion were identified and rated for four different segments of the show.



**Figure 6. Use of color, graphics, icons and animations to represent sound information**

This study showed that D and HOH groups seemed to diverge considerably on how the information should be expressed. The use of graphics, colours and face icons had more positive reactions from HOH participants than deaf participants. HOH participants liked the use of face icons while deaf participants did not. A similar result occurred for the graphical representation of the emotions; HOH responses really enjoyed them while deaf participants did not. Deaf participants reported that they associated the use of face icons with children's content and were therefore unable to take the content seriously when watching a drama or action show. This study also showed that the use of colour was not an effective tool for conveying additional meaning to the text as it confusing and has different meanings for different people or cultural groups. In an attempt to facilitate speaker identification and due to the size of the captions and screen dimension, some captions were placed close to the speaker's mouth. Participants in this study believed that this forced them to lip read when they did not want to, or that it was slightly covering up the mouth so they were unable to see the speaker's lips moving.

The study concluded that D/HOH individuals rely heavily on paralinguistic information expressed from facial expressions and gestures therefore overlays such as captions and graphics



should never interfere with access to this information. It also concluded that the use of graphics, icons or animations seemed to have potential for use in captioning emotive sound information, music and sound effects but that these devices must be used carefully.

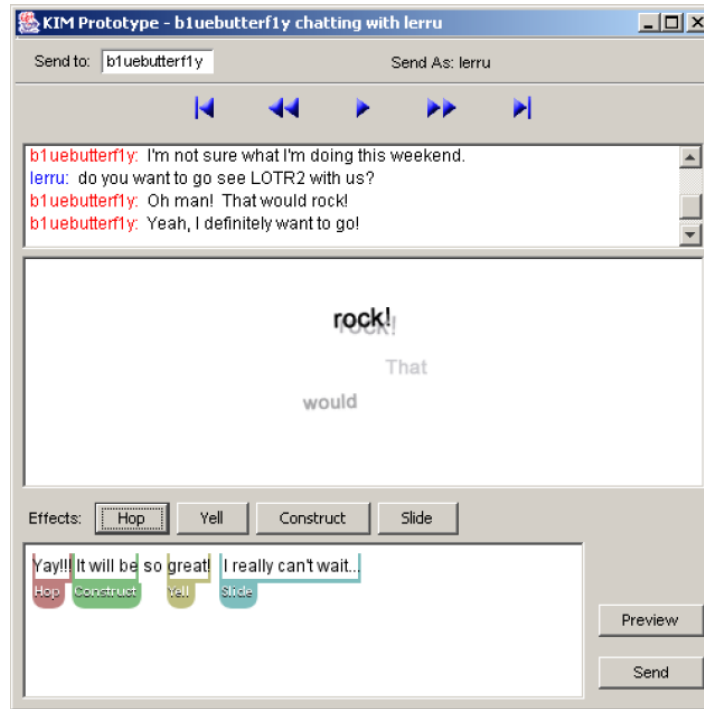
### **2.3.1 Animated Text / Kinetic text**

Animated text or kinetic typography emerged recently as an alternative way to express emotion, mood, and tone of voice. [25] examined the relationship between properties of animation and emotion, asserting that kinetic typographic parameters such as position and size can correspond to prosodic features of voice. Animated text and kinetic typography are also often used in title sequences of films and television to convey emotion. A notable example is the movie, *Se7en*, that uses trembling letters with a deteriorated, scratchy typeface design to convey a sense of terror in its titling sequence [26].

Kinetic typography was explored by [6] to evaluate its impact on instant messaging communication. Researchers created the Kinetic Instant Messenger (KIM) as shown in Figure 7 that integrated kinetic typography with instant messaging. As the kinetic typography message is played, it is also added to the conversation log in regular text. KIM provides users with four different animation effects:

1. Hop: text jumps up from and returns to the bottom of the screen.
2. Yell: text zooms in quickly and shakes.
3. Construct: individual letters rotate and slowly converge in the middle of the screen.
4. Slide: text scrolls horizontally across the screen, fading in and then out as it moves.

The authors reported that kinetic text has the ability to add dramatic meaning to the way in which emotions are conveyed [6].



**Figure 7. KIM displays incoming messages and replays messages in the main conversation [6]**

Another study using animated text in a chat system was performed by [27] and it explores the impact of animated text when used to express affect in online communication. This system estimated the affective state of a user by gathering data from physiological sensors and manually specified animation tags. This state was then presented to another user as animated text. Galvanic skin response (GSR) measures were used to indicate arousal level and animation tags were used to assess whether the emotion was positive or negative (valence); the combination of arousal and valence was then used to predict the user's emotion using [28] model of emotion.

The twenty different types of animation were implemented as shown in Figure 8. The user could then select an example or specify their emotional state directly through a tag embedded in a text message. For example, "<happy> I am happy!" presents "I am happy!" with happy motion. User testing with six participants showed that there was a good correlation between GSR data and user-reported tension. The authors indicated that GSR can be used to determine changes in mental tension in real time during an online conversation. The results also

suggested that emotional information might be able to increase the subject's involvement in the conversation.



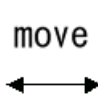







Happy↻	Drop↻	Move↻	Jump↻	List↻
				
Question↻	Attention↻	Sad↻	Chat↻	Tremble↻
				

Figure 8. Examples of animations used in [27]

Both of these studies seemed to demonstrate that kinetic text can enhance the ability of text to convey emotion without further descriptive wording [6, 27], however neither attempted to determine which aspects of the animations excited particular emotions. These studies provided informed this thesis particularly the findings that demonstrated that animations can add emotional elements to textual messages.

In 2006, [7] conducted a study where animated captions characterized emotions contained in music, speech and sound effects (as shown in Figure 9) and compared them with regular closed captions for the same content. The study reported that HOH participants responded positively to the moving captions as they provided improved access to the emotive information contained in the content.



**Figure 9 Kinetic text used in [7]**

## **2.4 Emotions**

Human emotion is the result of a combined processing of audio and visual cues [29]. Emotions affect the way we communicate every day. It is difficult to determine how many emotions there are or describe all different kind of emotions we use when we communicate, but [30] proposed a psychological model of emotion suggesting that all emotions can be reduced to a set of five to eight primitive emotions. These primitive emotions are sadness, anger, happiness, fear, surprise, disgust, anticipation and acceptance.

### **2.4.1 Emotions in sound and music**

Music has a strong presence in culture. It is a method of expressing and conveying cultural information and knowledge [31] that is universal.

It has been reported that music evokes different responses in the individual:

- Emotional responses [32] are an important medium for conveying cultural information.
- Evoking autobiographical memories.

- relaxation [33] as an escape for stress and anxiety; and
- pleasure [34]

Music also, often accompanies other stimuli. For example most televisions shows and films contain information and content in auditory form and when mixed with visual cues create the entertainment value of the presentation.

Another use of sounds and music can also be to create a sense of irony or comedy as this auditory approach can be so powerful that it will carry long-term cultural significance [35].

#### **2.4.2 Use of EC to provide emotions through music**

[36] considered a different approach that used an early version of EnACT to communicate the emotional information of a song through animated lyrics (see Figure 10) [36]. Participants in this study were presented with two songs using animated lyrics. Participants were asked to rate the understanding of the animated text. Overall, there was a positive reaction to the animated lyrics of the songs. Participants were also able to identify the videos presented to them as songs even though there was a serious attempt to mask the fact that the stimuli were songs. Participants also expressed the desire to have other visuals to accompany the animated lyrics to obtain the full attention of the viewers. Using animated lyrics or EC as music demonstrated that the moving text did not interfere with the readability and overall understanding of those lyrics. More importantly it showed that the participants were able to understand the animations of each of the emotions presented in the videos.

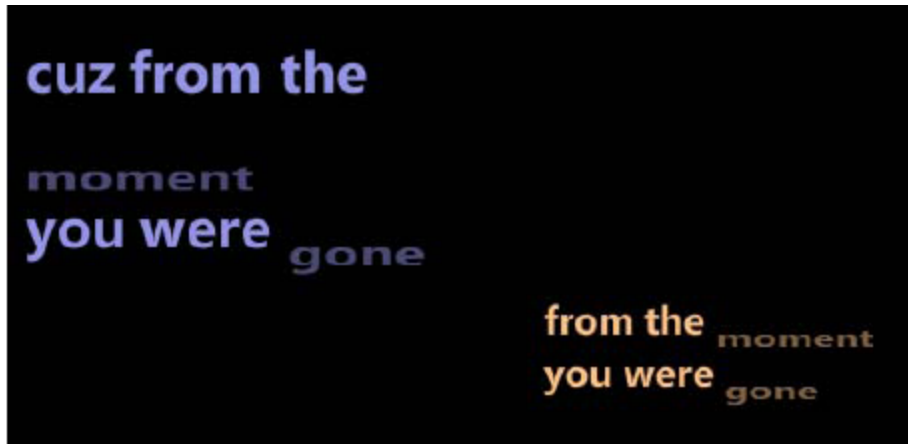


Figure 10. EC showing lead singer upper left and background singer on the bottom right

The study done by [36] and [7] showed that is potential for the use of EC to caption video and music in order to deliver some of the emotional content of it.

Music, sound effects, and speech prosody are important creative elements of a TV or film. Expressing them in an alternative modality requires a new way of thinking about what those sounds represent and how they can best be converted to a visual equivalent and still maintain the original meaning and creative impact on audiences.

In this chapter provided the history of CC, its problems and alternatives that could help on providing an alternative way of delivering information. Extensive research has been done regarding the use of animation to express emotions in text, however not a lot of research has being done regarding animated caption to improve CC. The research done at Ryerson University by [3], [5], [7] and [36] to the date of writing are one of the very few studies that have tested alternative methods for conventional CC in North America. The work provided by [7] inspired the development of EnACT once the animations for the basic emotions (sad, fear, anger and fear) were tested and provided positive results. [36] evaluated the animations that EnACT was able to

display providing also positive results as participants were able to understand the emotional content that the animations were displaying.

This thesis discusses the usability of the EnACT System and the potential to be used by Professional Captionists as a potential add-on or plug-in to existing captioning software tools out there.

## **Chapter III – Methodology and Implementation**

In this chapter, the technologies used in this thesis including a historical perspective on EnACT detailing the software's current functionality and my contributions will be presented. As parts of EnACT were developed prior to this thesis, it is important that I outline my contributions in the development cycle.

### **3.1 Target Users**

My research is focused on the usability and use of EC by the target users for EnACT. Target users fall into two groups:

- 1) I have termed the first group “Amateur Captionists” to describe users who have little to no training in any form of captioning, but have the desire to add captions to their or other's online video materials. These users have basic to advanced computer use knowledge. They may have some experience with simple video editing tools such as Windows Moviemaker™ or iMovie™, however they are not considered to be as proficient as professional video editors, video content producers or captionists.
- 2) The second group is termed “Professional Captionists” to describe users who create captions for television, film or video content industries (both online and broadcast) as paid employment. These individuals often work for third party post-production services for or broadcasters. These users would be considered as the main primary users of EnACT for this study since EnACT was created initially as an add-on to their existing captioning software to create animated captions.



### **3.2 Research Questions**

The main goal of this study was to test the usability EnACT and receive feedback therefore the following research questions were formulated to address my research focus:

- What are the usability and improvement outcomes of working with EnACT?
- What is the impact on the captioning process and users as a result of EC?
- Is EC technology feasible?

### **3.3 Study Design**

EnACT was created and developed as an add-on or plug-in to existing captioning tools so Professional Captionists are able create EC, however EnACT also provides basic captioning functionalities such as editing of dialogues, choosing the location of the caption on the screen and editing the time for each caption to be displayed on screen so it can be used by Amateur Captionists. For this to be possible, EnACT and its goal to create EC needs to be simple but effective with users with extensive to none captioning experience. To evaluate ease of use, ease of learning, flexibility and feasibility of EnACT as a captioning add-on tool to create EC a usability study was created.

Usability as defined in ISO 9241-11 is, “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” ([37], pp. 1). EnACT was designed as an add-on or a plug-in to existing captioning tools where the main users would be Professional Captionists, however, EnACT was also designed to be intuitive enough so any user regardless of any computer, education or professional background could create EC with it.

The study was accomplished using a conventional usability method of administering a pre-questionnaire questionnaire to gather demographic data such age, sex and computer experience as well as the participant's level of completed education, followed by a training sessions where EnACT was introduced and a general overview on how it worked, after this the participant was given three tasks to complete on his/her own. Finally, the participants were given a post-study questionnaire collecting information regarding their experience with EnACT. In addition to this the Professional Captionists were engaged in a discussion to explain their experience and comment on any possible change that the software might need. This was important since Professional Captionist are considered to be the primary target of this software. The Amateur Captionists were encouraged in providing with comments but not required to do so once they finished completing their post-questionnaire. ([see Appendix D](#)).

Twelve Amateur and three Professional captionists were included as participants, however, as Professional Captionists were difficult to recruit and because they were considered the primary user, the methodology was modified to include a case study methodology [49] for the Professional Captionist participants. More detailed information about the processes, opinions and considerations from Professional Captionists was collected using detailed interview techniques. Ethics approval was provided by the Ryerson Ethics Board (see Appendix B for the ethics approval letter).

All participants were recruited using a variety of techniques including creating a blog that specified the nature of my study, the location of where the study could take place and the compensation each participant would receive. I also joined social media sites such as Twitter and LinkedIn that allowed me to search for Professional Captionists and contact them directly. I used

email to contact my professional network and ask for their assistance in reaching Professional Captionists.

Once the pre-study questionnaire was completed, participants were provided with an introduction to EnACT where I gave them background information of CC and its problems, and then how EC could be used as an alternative solution. After the introduction all the participants were encouraged to informally browse through the functionalities and ask questions as needed.

Participants were then asked to work through a set of thirteen training tasks that consisted of step-by-step instructions of processes that familiarized them with the basic functions of EnACT (see Appendix D for a list of training). The training tasks included requiring the participant to load the video file to be marked-up with the EC, adjust the script and assign words within the script one of the four basic emotions and intensities given as functionalities within EnACT. This training usually lasted no more than 10-15 minutes and let the participant become more comfortable with the software.

During training, participants were also introduced to the talk aloud protocol [38]. Briefly, the talk aloud protocol involves the participant speaking out loud their thoughts on the action and activities they are engaged in as they work through the study tasks. Talk aloud protocol was chosen because of its ability to capture data that may provide insight into the real-time thoughts and opinions of EnACT as the participants worked through the tasks. Because users were asked to describe what they are doing at the time that they faced the task in the study, talk aloud protocol was able to capture richer data that can be compared against the screen records of their user behaviour with the program and to examine if the intended user work flow was encouraged by the program design.

Participants were then asked to begin the study tasks using EnACT and were provided with printed copies of the three study tasks. Time restrictions were not given and users were encouraged to take as long as they required to complete the tasks accurately and in full. Once the tasks were completed, participants were asked to complete a post-study questionnaire (see Section 3.4.2 for the details of the specific questions and Appendix D for a copy of the questionnaire).

As mentioned before, this study involved fifteen participants where twelve were Amateur Captionists and three were Professional Captionists (twelve males and three females in total). Ages ranged from 18 to 59 with 11 participants in the 18-29 range, one in the 30 to 39 range, two in the 40-49 range and one in the 50-59 range. The educational background varied amongst participants (two with graduate education, eight with undergraduate education, one with college education and four with high school education). They were required to have general computer experience which includes familiarity with basic text editors or multimedia players such as Windows Movie Player.

### **3.3.1 Usability study with Amateur Captionists**

The study for Amateur Captionists was designed to investigate the usability of the EnACT software by that target group. The location of the study was flexible and dependent on participant availability since most of them wanted to participate after work hours and I wanted participants to be comfortable when doing the study; therefore studies occurred in a number of different locations including some participant's workplaces. The remaining studies occurred at Center for Learning Technologies (CLT) at Ryerson University (TRSM 3-174).

### **3.3.1.1 Tasks**

Participants were asked to complete three tasks in total, ranging from low to high difficulty levels. Each task was designed based on the difficulty level assigned by myself. Tasks were designed to be completed in succession, with the following task built on learning from the previous one. This was done so that the later analysis of data from the participants would show what functionalities were more challenging to use than others.

The three tasks for the usability study were as follows for both participant groups:

The first task required the participant to load a video script file in .rtf format; select and mark-up five words in the Script Editor Area (SEA) with emotions (see Appendix F for the complete study procedure). The steps that the participant had to follow to complete this task can be seen in Appendix F

The second task required the participant to load a video file and its corresponding .rtf script file. The participant was then asked to assign emotions to words within the script, assign these selected words with appropriate time information that determined when the captions should appear and disappear on screen. The steps that the participant had to follow in this task can also be seen in Appendix F

The third task involved the participant loading the project created in Task 2 and performing changes to the project. Participants were asked to make changes to the previously assigned emotions and the times assigned to the duration of emotions appearance onscreen. The steps that the participant had to complete can be seen in Appendix F:

Once the tasks were completed, participants were asked to complete a post-study questionnaire.

### **3.3.2 Case study with Professional Captionists**

The objective of this component of study was to carry out an in-depth examination of the processes that a Professional Captionist would employ with EnACT in their normal captioning work.

Participants for this group were difficult to recruit due to scheduling difficulties because captionists were under very tight deadlines to deliver captioned materials to their employers. Conflict of interest with their current employers was also cited as a barrier to participation in the study. For example, one participant had to ask permission from her employer before participating in this study. The location of the study was also problematic as many captionists work from home which at times was away from the city. Some of the Professional Captionists that were contacted were also hesitant to participate due to their unfounded fears that any kind of digital automation might threaten their job viability. In the end, three Professional Captionists agreed to participate in the study.

#### ***3.3.2.1 Tasks***

The Professional Captionists completed the same tasks performed by the amateur captionists. In addition, after completing the post-study questionnaire they participated in a detailed discussion where they analysed and made comparisons of EnACT to the functionalities of their current captioning software and speculated as to how EnACT would perform in conjunction with those tools.

### **3.3 Equipment**

As the location of the study varied with each participant, the study setup needed to be mobile.

With this in mind, the equipment used to complete the study included:

- Two laptops with EnACT installed. Only one laptop was used during the study, however the second laptop was used as a back-up in the event that the first laptop failed. The specifications for the laptops can be found in Appendix J.
- One pair of headphones.
- One microphone.
- CamStudio, an open source screen recording program.
- Information and consent form (see Appendix K)
- Pre and post study questionnaires

The specifications of the laptops used to run EnACT can be seen in Appendix J:

Participants required headphones to listen to the audio that was present during the playback of the video file used during the study. Good quality audio was important in performing the study as one of the tasks required participants to synchronize the EC animations in time with the audio.

A microphone was required to record the voice of each participant from the talk aloud protocol. To record the screen actions and voice, CamStudio [39] was used.

### **3.4 Data Collection and Analysis**

All studies were conducted over a four month period in various locations. Qualitative and quantitative data were collected for studies with both participant groups. To obtain the quantitative data, questionnaires were. To collect the qualitative data, participants were asked to talk about their thoughts out loud as they worked through the three tasks. The interactions on

screen and the participant's verbalizations were recorded using CamStudio. The Professional Captionists were also asked to engage in a discussion at the end of their post-study questionnaire to consider their experience with EnACT and its limitations. They were encouraged to also make suggestions, provide their ideas for improvements, discuss what they would like to see in the future for the program and discuss whether they would be willing to use EnACT in their captioning work. Finally, written notes were taken by the researcher as a potential source of clarification during the data analysis if necessary.

Once the data was collected, descriptive analyses such as frequency analysis were used to analyze the quantitative data and a thematic analysis was used to analyze the qualitative data.

### **3.4.2 Usability questionnaire**

The post-study questionnaire contained eight questions regarding the usability of the software (see Appendix D for a copy of the questionnaire). The first question asked participants to rate the level of difficulty using a Likert scale for each of the three tasks in the study where 1 represented "Very Difficult" and 5 represented "Very Easy".

There were fourteen sub-questions for participants to assess:

- The level of difficulty loading the script file in the software.
- Assigning emotions to the words within the script.
- Adjusting the intensities of the emotions selected for words within the script.
- Saving a new project.
- Locating and opening a saved project.
- Loading a video file in the software.
- Adjusting the text size of the script as it appeared in the SEA.



- Changing the default colours for the emotions.
- Adjusting the font type of the script as it appeared in the SEA.
- Viewing the changes made in the SEA.
- Viewing the enhanced captions in the video file from the software interface.
- Reading the text of the captions as it played in the video file from the software interface.
- Changing the emotions assigned to words within the script file from a previous version of the project.

These questions were important to the study because they captured ratings of intuitiveness and ease of use of the software. These data were also used to identify problems and issues that were experienced by participants when using the software and in assessing the potential as a possible plug-in to existing captioning tools or stand-alone tool for amateurs.

In the second question, participants were asked about their opinions of the layout of specific interface objects displayed within the interface using a five-point Likert scale where 1 was “Very Poor” and 5 was “Excellent”.

The specific interface objects that were assessed for layout were:

- The script file is loaded into the software.
- The drop-down menu that allows the user to select the emotions.
- Functionality that allows the user to specify the intensities of the emotions.
- Option menu which contained functionalities such as changing the font and the colour of the emotions in the SEA.
- Functionality where the video file is loaded into the software.

These questions were important in determining whether the interface layout fit with the common user interactions that were required for captioning and to understand the way in which participants perform their own captioning tasks and any interaction habits or expectations that they developed from their own experience.

Question three asked participants to rate their confidence in successfully marking-up captions with emotions in the future. This question provided data that assessed if the user was able to learn the software in sufficient time to be confident in using it in the future without assistance. The question used a five-point Likert scale where 1 was “Very Confident” and 5 was “Not at all confident”.

A final forced choice question asked participants about their interest in using EnACT to caption their own work in the future. This question followed the same format as the rest of the questions using a five-point Likert scale where 1 was “Very Comfortable” and 5 was “Not Comfortable at all”.

Four open-ended questions were also added to this study to allow participants to elaborate on their experience and provide more in-depth responses than those provided in the forced-choice section of the questionnaire. Participants were asked about what they thought were the easiest task/s to perform with EnACT, their understanding of EnACT’s limitations, suggestions for improvement and any additional comments about the software.

For the Professional Captionists, once they were done completing their post-questionnaire, were engaged in a discussion where they explained how the experience was and if they would like to see any changes to the software UI or engine. Amateur Captionists were encouraged to do this but not required to since they were not considered to be the primary target.

### **3.4.3 Use Cases**

Professional Captionists were considered to be the primary target in these study since EnACT was originally developed to be a plug-in or add-on to an existing captioning tool therefore the participation of these three participants were considered as uses cases to analyze their experience while completing the tasks given.

A use case in the study would begin with the captionist opening a script file, in a RichTextFile (RTF) format and its associated movie file of any format. The text and movie files are then automatically processed separately by the script parser and the video encoder modules and displayed in the respective windows of the interface. After the user is done applying the desires emotions to the script, the EnACT parser will parse the script into the speakers and dialogue from the file while the movie encoder transforms and encodes the movie file into a flash video file (.flv). A user is able to preview their work by clicking on the “Show Preview” button to examine the attributes assigned in the editor area. The EnACT engine renders the text animation with the video file which is then displayed in the preview window of the software.

Once the tasks were completed, the Professional Captionists were engaged in a discussion to explain and talk about their experience with EnACT and any possible change or addition to the system’s UI or Engine.

### **3.5 System Description/Design**

In this section a description of the software, EnACT, is provided along with its development history and an outline of my specific contributions to it.

Prior to EnACT, the primary method of creating animated text for a video file is through specialized software tools such as Apple’s LiveType™ or Adobe After Effects™. “Adobe After

Effects offers extensive video editing features including the creation of vector graphics, working with 2D and 3D elements, editing with multiple cameras and manipulating key-frame values” [40]. Apple’s LiveType™ is part of Final Cut Studio™ and is primarily used to create animated title sequences for video projects. It includes functionality that uses fonts, textures, objects, templates and effects that can animate the titles [41]. Both of these software tools are capable of creating animated text, however, they are both intended for use by graphic designers with specialized design skills and not for text-based captioning. These professional design tools can be difficult to learn and use for novice users.

EnACT was designed to create and embed simplified animated text for time-based media (e.g., video and animated graphics) specifically for use by non-graphics experts [42]. It uses only four specific animations that relate to four basic emotions as outlined by Ekman [30] and a limited set of intensity modifiers for the emotions along with time in/out and position functions.

EnACT was created to complement existing captioning tools as either a plug-in or add-on rather than operate as an autonomous, full-featured captioning tool, however, basic functionalities were added such as: controls to mark the start and end times for the timing of the captions, SEA to edit captions/dialogue, options for caption placement in one of the nine possible locations on video area, and the ability to adjust text and font styles.

### **3.5.1 EnACT: Version 0.5**

An early prototype of EnACT was created by Quoc Vy in 2008 [42] (see Figure 11 for a system diagram). This limited version of the system was created to demonstrate the potential functions for a tool that could support animated captions. There were many deficiencies and functional limitations that needed to be resolved before it could be evaluated by users. For

example, to view a video file with the enhanced captions, the user would be required to manually find the location of the EnACT Engine.swf file to generate their video to view. The editing of the captions “begin” and “end” times (which set the times for a caption to appear on the screen) were mouse dependent and not accessible for keyboard users.

For my thesis, I maintained the basic system design skeleton from this version and added:

- A create project wizard functionality for each user to create their own project.
- A preview button on the UI so users can mark-up the captions then test and view their changes throughout the mark-up process.
- Video format conversion of any video to a flash format using ffmpeg, a command library tool to convert multimedia file formats.
- Fixed and improved major bugs in the code.

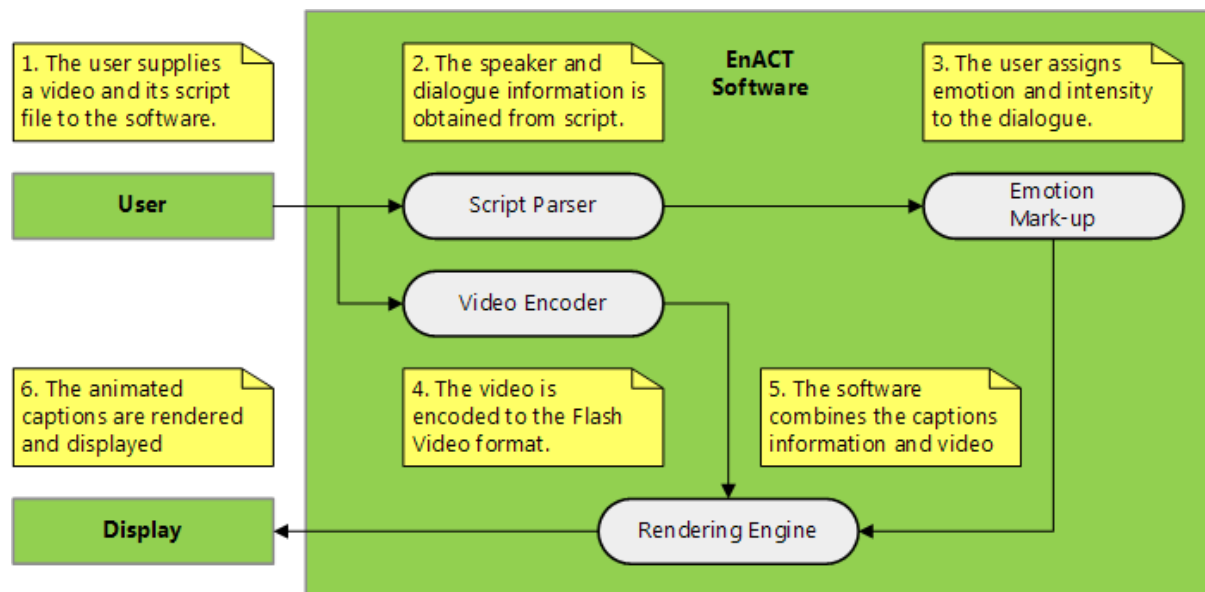


Figure 11. System Design for EnACT [42]

## **3.6 Development Platform**

The EnACT software system uses several important technologies: the C#.NET framework, Adobe Flash and the Extensible Markup Language.

### **3.6.1 C#.NET Framework 3.5**

The C# development environment was chosen as the primary language because it was an Object Oriented (OO) and “type-safe programming language derived from C and C++” [43]. EnACT relies on a wide variety of media file formats, and for this reason a higher level programming language was used because of the many existing, specialized libraries available for media manipulation in C#. Using the existing libraries in a high level environment, rather than build them independently in a lower level language, proved to be beneficial in the development cycle.

For future development of the software, the .NET framework also allows for a potential transition to a web application as tentative next step in the evolution of EnACT. As media formats evolve the development time is expected to be reduced because the .NET framework is maintained and updated on a regular basis by Microsoft.

The Integrated Development Environment (IDE) used to code the UI, the script parser, video encoder; mark-up of emotions and their corresponding intensities was Visual Studio 2008 (VS 2008). The UI and console application features include:

- UI design with drag and drop graphical elements.
- Syntax highlighting and automatic syntax checks.
- Comprehensive de-bugging tools.

### 3.6.2 Adobe Flash and ActionScript 2.0

A computer-based, Internet infrastructure was chosen for creating EC because of its flexibility and accessibility compared with the limiting broadcast standard of EIA608 and the lack of display and decoding hardware for CEA708 signals (see Chapter 2 for a discussion of caption standards). Adobe Flash was selected as an authoring application because it is an Internet/web-optimized tool designed for creating and displaying rich media content, particularly animation as well as having a good reputation as an easy to use prototyping tool. The Adobe Flash player was known to be “installed in 98.7% of internet-enabled desktop computers in the mature market of the US, Canada, UK, Germany, France and Japan” [44]. The Adobe Flash player was also available free for download to anyone and does not restrict users in any location, thereby making EnACT more accessible to our target audience.

Apple has chosen not to support Flash in their latest mobile devices (iPads and iPhones), however, they are continuing to provide support for their laptops and desktop machines. This means that animated text content cannot be viewed on Apple’s mobile devices. Android and Blackberry phones, however, do support Flash [45] and therefore, EnACT will produce animated content for some mobile devices. We expect that as new players become universally accepted, EnACT can be modified to produce playable files in any new format.

ActionScript, originally developed by Macromedia [42], is a simple but powerful Object-oriented (OO) scripting language used in Flash to add interactivity to applications.

Flash and ActionScript were used together to create the *EnACT Engine*. Flash was chosen to render and display the EC, and ActionScript was used to retrieve information from the Extensible Markup Language (XML) file created by the EnACT UI. The XML file contained

data specifying the mark-ups assigned by the user to each enhanced caption e.g., the emotion and intensities to use for each word, the location of the caption on screen and the video to create the animations. An example file can be found in Appendix H.

Once an XML file is created, the EnACT Engine renders and displays animated captions and outputs this data as a .swf file which can be played in any web browser or computer with a Flash player installed.

### **3.6.3 Extensible Markup Language (XML)**

XML is a standardized markup language used to represent and store data in an organized and retrievable format. XML models data as a tree of elements that contain character data and has attributes composed of name-value pairs [46].

XML is an independent transformable file format that was chosen as the primary communication between the C#.NET and Adobe Flash platforms to render the animated captions. XML introduces a flexible environment to share data and variable values between both platforms and thus allows them to be independent from each other. Data was captured while the user marks up elements within the script. These were stored within the XML file using descriptors assigned by developers. In comparison to relational databases, XML is more portable. If a database approach was used more support at a developer level would be required for the user to ensure that the correct database is installed in the client computer.

To populate the XML file, the UI of EnACT gathers speaker identification (ID), and the dialogue of the speaker values from the mark-ups selected by the user (e.g., words selected within the text script). . The dialogue is then parsed into words with the specific emotion and intensity values assigned. If there is a word that has not been manually marked-up, the word is



automatically assigned the default “no emotion” value with zero intensity (See Appendix H for a sample file). The xml file then contains:

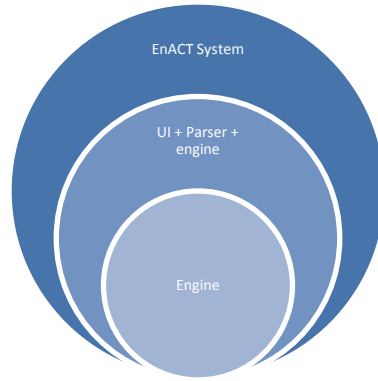
- Timing attributes where the caption is told when to appear and disappear.
- The speaker ID showing the name assigned in the script indicating who is speaking in the dialogue.
- Location of the caption to appear on the video.
- Alignment of the caption (left, centre or right justified).
- For each word it contains:
  - Emotion type, ranging numerically from no-emotion, happy, sad, fear, anger (0,1,2,3,4,5)
  - Intensity value, ranging from no-intensity, low, medium and high (0,1,2,3)

### **3.7 History of EnACT development**

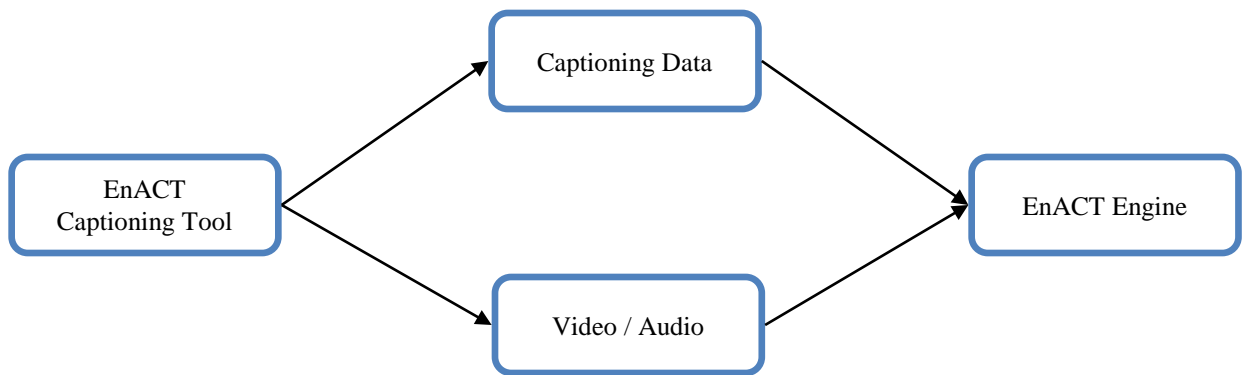
EnACT has been in existence since February 2006. Since that time, it has evolved considerably. I was involved with the research team in 2006 as an advisor, however, did not make any major contributions until I began my Master’s work in late 2008. In this section I will briefly explain the major milestones in the development of EnACT and highlight my contributions to the project.

#### **3.7.1 EnACT Engine - initial Versions (1.0 and 2.0)**

The EnACT Engine is the rendering engine for EC and is a component of the EnACT system as seen in Figures 12 and 13. It is used by the EnACT software during the process of creating EC according to the values assigned in the XML document.



**Figure 12. Relationship of the different EnACT system components.**



**Figure 13. The EnACT captioning tool is divided into two major components that are needed for the EnACT engine to render the EC**

Development of the EnACT Engine involved the use of Adobe Flash, ActionScript 2.0 and XML beginning February 2006 and was finalized August 2008.

In my Masters work I did not make adjustments to the EnACT Engine. However, at the time of writing the animations rendered by the Engine are undergoing improvement by other students and ActionScript 2.0 is being converted to ActionScript 3.0.

### **3.7.2 EnACT Editor Prototypes**

The purpose of the EnACT Editor is to allow users to create the EC and assign them to dialogue at specific times throughout the video. A UI is provided to users so they are able to easily manipulate the EC. The output of this work is an XML file and a playable video with embedded enhanced captions.

Three prototypes have been developed for the EnACT editor, each informing the next in the development process. My thesis work is based on the last iteration.

#### ***3.7.2.1 EnACT editor: Version 1.0 (September 2006 – December 2007)***

The first EnACT prototype was initially developed by Qiong (Jane) Zhang with the assistance of Richard Hunt who designed the interface in the September 2006. This version was extended by Qiong Zhang and Jorge Mori until December 2007 (see Figure 14 for screen shot of this first prototype). It was created using Visual Studio 2005 with C#.NET framework in a Windows XP machine environment.

This interface consisted of a main menu, four text boxes on the left; on the right side, a player component and mark-up tools for the dialogue and text.

This prototype included functionality for:

- Four types of emotions (Happy, Sad, Anger, Fear)
- Three levels of intensities (low, medium, high)
- Time –in and Time-out for the captions on the screen
- Ability to select captioning placement on the screen
- Alignment of text

- Speaker ID information.
- Ability to change the background and foreground color of the captions
- Option to edit the font and font size.
- A place-holder preview window was designed for and appeared in the interface however; this was not a functioning feature for the user.

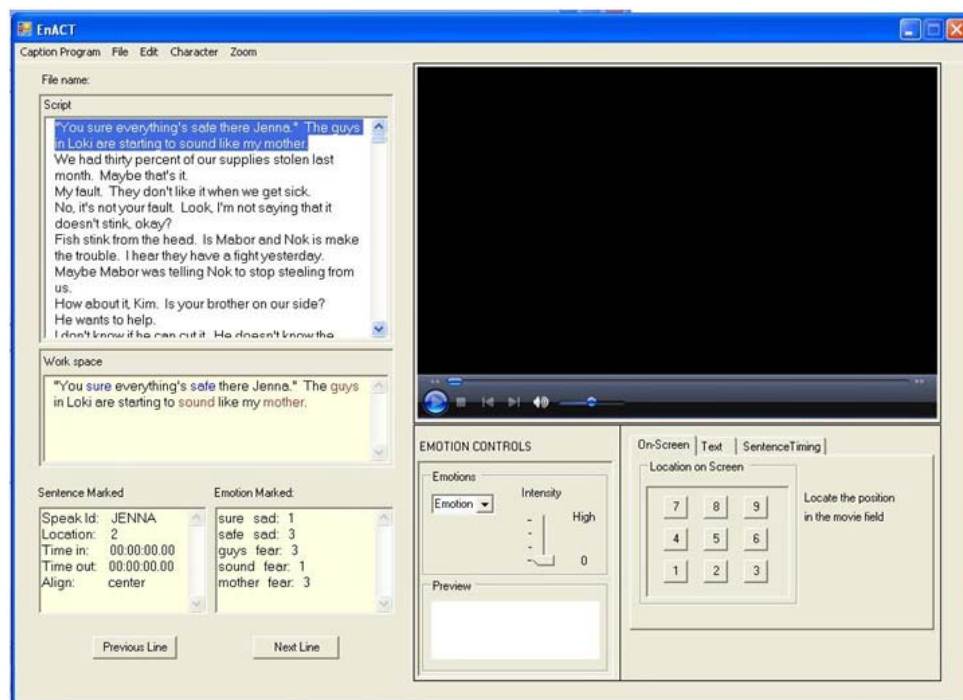
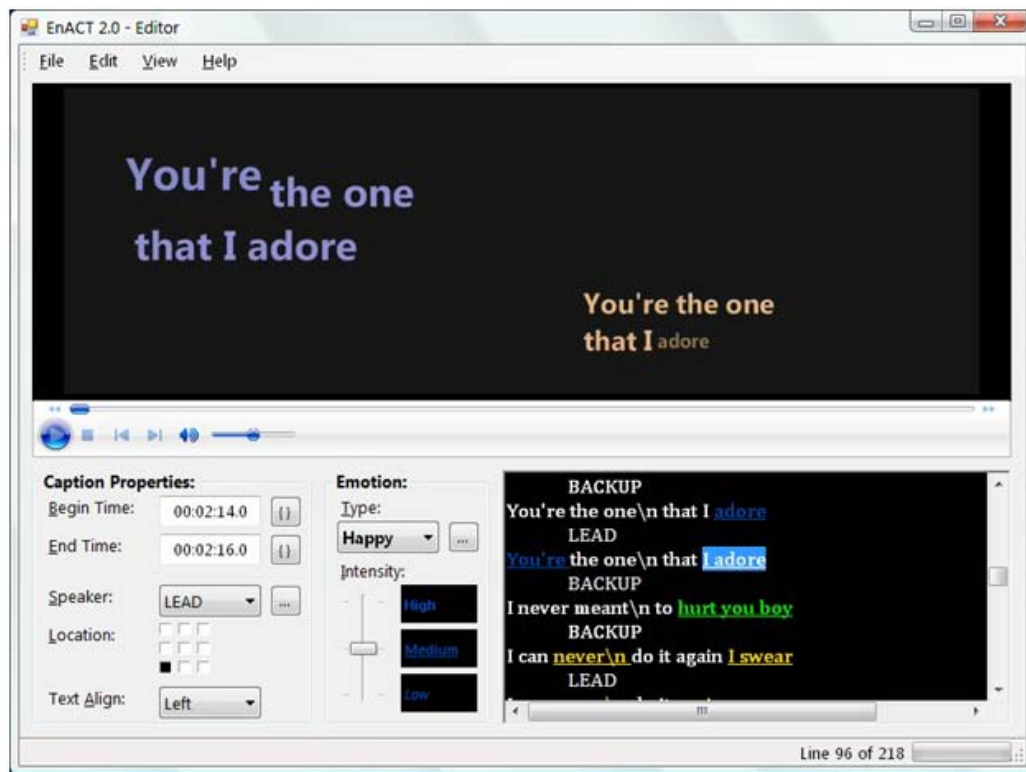


Figure 14. The First EnACT prototype developed by Zhang, Hunt and Mori, 2006

### 3.7.2.2 EnACT Editor: Version 2 (December 2007 – August 2008)

The next iteration of the EnACT editor involved a complete redesign of the UI to reduce visual clutter and organization of the main UI elements so that they more intuitive for the user. As shown in Figure 15 a larger video window was added and text unnecessary for core user tasks was removed. In addition, the caption editing flow was stream-lined so that it occurred in three

main panels: a Caption properties panel, an Emotion panel and a Workspace panel. These panels were organized so that users could work through the assignment of timing, emotions and intensities while referring to the text/script. Visual indicators of emotion and intensities were also added along with a global settings viewer (see [42] for further description of this version of the interface).



**Figure 15. Interface elements of the EnACT system**

#### 3.7.2.2.1 Issues with EnACT Version 2

A major limitation with this iteration was the lack of preview functionality within the work area. Users were able to view the video file, edit and mark-up the text of the captioned dialogue; however, they were not able to preview the EC within the video. This disrupted the user's work flow by requiring them to run the EnACT engine manually to generate a preview.

This cumbersome step meant that for a user to compare minor changes in their work they would spend more time generating the preview than altering their work. This discouraged users from checking and saving their work regularly. This action was not intuitive and the software design did not assist users in optimizing their workflow, making the process of creating EC tedious and frustrating. Furthermore, the interface favoured mouse users and forced keyboard users to interrupt their workflow. This was a problem as it was not efficient since marking up captions from the UI was a slow process and it was faster to go to the XML file directly and edit the captions from there.

A second issue was that EnACT could not load the entire dialogue from a script into the system as shown in Figure 16 is a screenshot of EnACT version 2 and the dialogues.xml file containing the marked up dialogue from the script file once it has being parsed by the EnACT System. The script contains four dialogues but only two appear in the SEA.

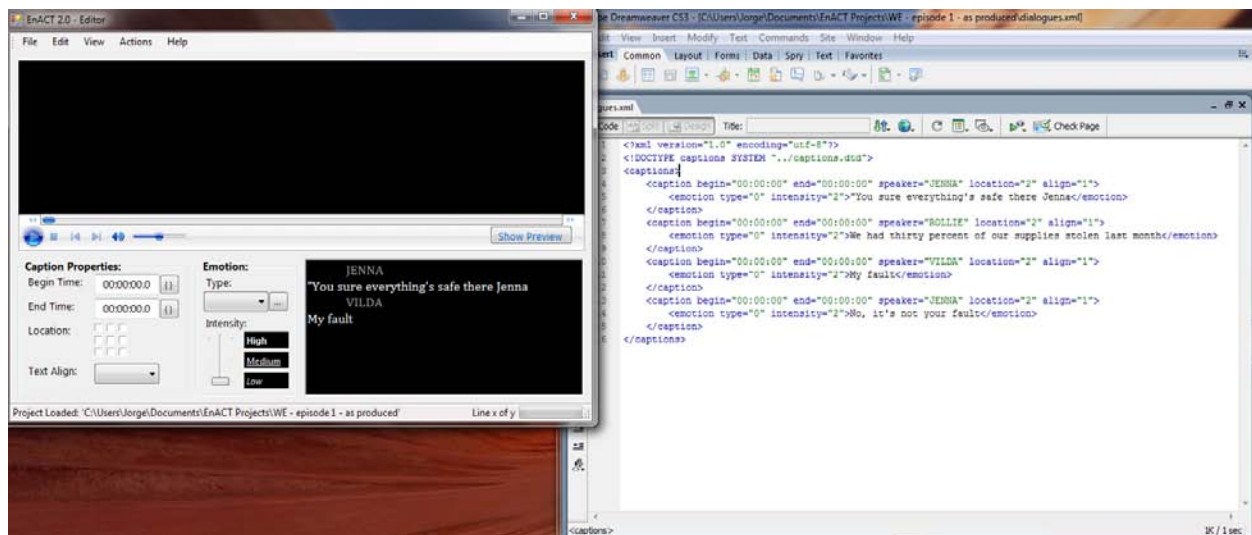


Figure 16. The script contains four dialogues but only 2 appear on the SEA

Another third major issue was that every time the user wanted to save her/his work, EnACT would record the incorrect information. In the file dialogues.xml, each emotion was as a

numerical value assigned according to the emotion selected for that word (Happy = 1, Sad = 2, Anger = 3, Fear = 4) in the “emotion type” variable.

When the user saved the project the value will get changed and replace it with a “-1” as shown in the code sample in Figure 17.

```
<caption begin="00:00:20" end="00:00:25" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="-1" intensity="0">Carlo,</emotion>
    <emotion type="-1" intensity="0">blow</emotion>
    <emotion type="-1" intensity="0">into</emotion>
    <emotion type="0" intensity="0">the</emotion>
    <emotion type="0" intensity="0">tube</emotion>
</caption>
```

**Figure 17. Example code of error in the “emotion type” variable**

Additional issues that are identified with this version of the interface are summarized in Table 1.

**Table 1. Problems of EnACT 2.0**

<b>Problems</b>	<b>Why it is a problem</b>
No Home project folder when software run for first time on a new machine	This is the main folder where all the projects created by the users are stored; not having this folder crashes the program as it does not know where to locate the main files.
Only one project was created and used.	There was no multi-user project creation meaning that in order to create a new EC video, the previous version had to be deleted.
A bug stopped the program from loading the entire script and it some of lines of the dialogue were missing.	This problem was a very critical and important issue as the parser of the script movie file was not parsing the dialogue correctly giving the wrong speakers the wrong dialogue and missing some of the dialogue
Problems with timing buttons	The buttons were created to assist the user to select either a "Begin" or "End" caption time so the user would not need to insert it manually. The main problem here was that the user had to manually click the button with only the mouse in order to give the captions their time attributes. Professional captionists usually perform all captioning tasks with only keyboard shortcuts. Requiring mouse use would interfere with their normal way of working.
Problem with the Save and Save As button	When the EC caption dialogue file was saved, it would save the wrong emotion and intensities attributes to the selected words
No Preview Button	The user could not see their work as they marked up the words.
Lack of keyboard shortcuts	Captionists perform most of their using the keyboard. Forcing them to use a mouse with an EnACT add-on to their regular captioning software would interfere with their workflow.



### 3.8 EnACT Editor Version 3 (September 2008 – Present)

While EnACT Editor Version 2 was a major advance from Version 1.0, there was still a considerable amount of original work to be carried out and limitations to overcome. In this section, I will explain new additions to the interface and the new workflow model. An example of EnACT Version 3 is shown in Figure 18.

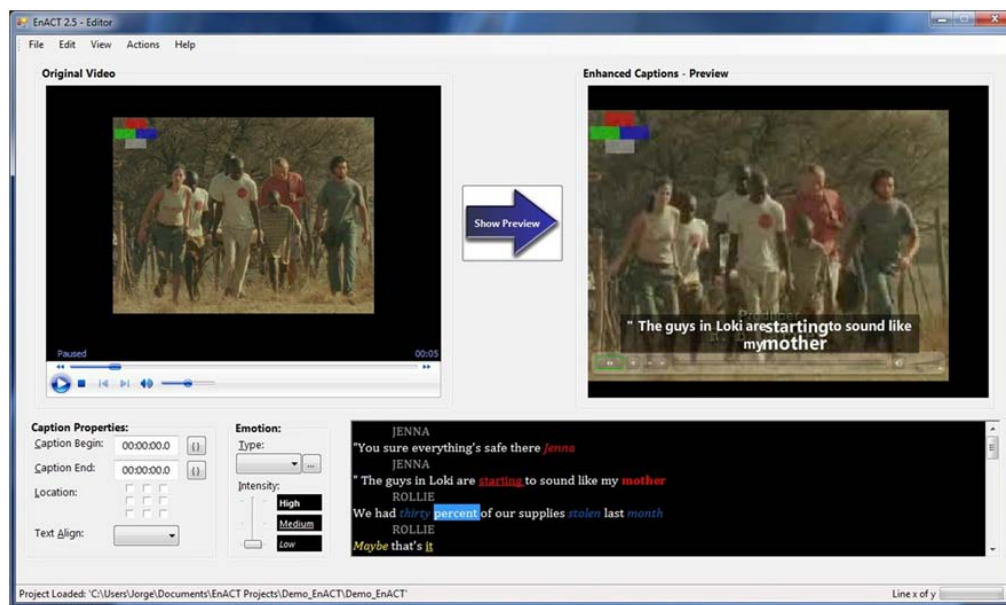


Figure 18. Screenshot of EnACT Editor, Version 3

#### 3.8.1 Resolutions implemented in EnACT 3.0

- 1) Loading uncompleted scripts: Examining the version 2.0 code the method for parsing the dialogue and displaying into the SEA was called `parseDialogue()` method (see Figure 19 for the pseudo code and Appendix I for the complete method code for EnACT).

```

parseDialogue()
{
    create list of emotions

    Access the xml dialogue file

    while the program is reading the file
    {
        if nodetype is not an element
            continue
        if the name of the node is not "emotion"
            break

        obtain emotion type
        obtain emotion intensity
        obtain text

        add to the list of emotions
    }
}

```

**Figure 19. A code sample from the parseDialogue method**

It was found that the problem of skipping dialogue elements was the use of the statement “`continue`” as shown in Figure 19; the `continue` statement starts a new iteration when a condition is met and therefore skips some dialogue lines. The solution for this was to remove the related “if” statement as shown in Figure 20.

```

parseDialogue()
{
    create list of emotions

    Access the xml dialogue file

    while the program is reading the file
    {
        if the name of the node is not "emotion"
            break

        obtain emotion type
        obtain emotion intensity
        obtain text

        add to the list of emotions
    }
}

```

**Figure 20. The bug fix in the parseDialog method**

- 2) Incorrect assignment of emotion type: When saving the project a “-1” was added to the type values inside the emotions xml element. Examining the code, the problem arose in the writeDialogues(string path) method. This method was in charge of updating the xml file with the new values for emotions, intensities and text of each xml element. I discovered that the “-1” was appearing because the emotion type was classified as “unknown” type. (see Figure 21 for pseudo code and Appendix I for full source code).

```

writeDialogues(){
    Initialize the number of lines of the richtextbox

    for all the lines in the text box
        case of selecting a speaker:
            obtain name
            if caption has changed
                remove all changes to captions
            else
                skip captions
            break

        case of selecting a dialogue:
            split the word in the sentence
            for all the words in the sentence
                highlight the selected word

                get the selected word
                get the emotion type
                get the intensity

                add the emotion to caption struct
                break

    open dialogues.xml for writing
    write the word, emotion type and intensity in dialogues.xml
}

```

**Figure 21. WriteDialogue method in pseudo code**

The pseudo code in Figure 22 provided the solution to this problem.

```
writeDialogues(){
    Initialize the number of lines of the richtextbox

    for all the lines in the text box
        case of selecting a speaker:
            obtain name
            remove all changes to captions
            skip captions and move to the next line
            break

        case of selecting a dialogue:
            split the word in the sentence

            for all the words in the sentence
                highlight the selected word

                get the selected word
                get the emotion type
                get the intensity

                if the emotion type is unknown
                    the emotion type is selected as none

                add the emotion to caption struct
            break

    open dialogues.xml for writing
    write the word, emotion type and intensity in dialogues.xml
}
```

**Figure 22. WriteDialogue bug fix**

Table 2 describes the remaining problems that were fixed in the developing of EnACT 3.0.

**Table 2. Problems and Solutions that were solved between Version 2 to Version 3**

<b>Problem</b>	<b>Solution</b>
No Home project folder when software runs fresh on a new machine	The program does folder and file checks before it creates the home folder for EnACT projects
User only able to create and update one project file.	Added functionality for the user to create multiple project files. Users are now able to select the saved file destination and within this directory path, folders now contain all elements related to the corresponding project.
Unable to create multiple project files.	
Problem with the Save and Save As button	This problem was fixed once the bug from the problem above was fixed.
Lack of keyboard shortcuts	Keyboard shortcuts were added for video controls and add begin caption time and add end caption time so users could use keyboard or mouse.

- 3) Preview functionality was missing: In EnACT Version 3 the original video panel was resized and placed on the left side of the screen. This location was chosen because in Western cultures a majority of people will look at the top, right corner of a page first consistent with the way in which a page is read; left to right [47]. It was also important to place this screen in a ‘before’ position so that a preview screen could be placed next to it to imply that it would be the ‘after’ video. The original and preview screens were then connected with a “Preview” button. The video containing the EC was then displayed and viewed in this preview window without the user having to locate appropriate files and manually generate a preview, load a new window, or change to a different player

application. The captionist can then make adjustments accordingly without having to interrupt their workflow or exit EnACT.

EnACT can accept any movie format which is then converted to a flash movie file (FLV) using the ffmpeg library in C# [48] when the user presses on the preview button. The code that processes the conversion is included in this thesis in Appendix B

- 4) Creating a new project: Figure 23 shows the main window that appears when a new project is to be created. The user is prompted to provide a script, and movie or TV file as separate documents. A script file is normally a text file narrating the movement, actions, expressions and dialogue of the characters.

Previous versions of EnACT used a “one-project-for-all” idea, meaning that the tool created enhanced captions for a single project without supporting multiple project versions. My approach allowed for multiple project designs so that each user can create and save their own project without overwriting other projects they may have created in the past. The user can also save different versions of the same file, in case they prefer to save their work in this way. Overall, this implementation was made to provide more flexibility to the user in the number of projects they could handle simultaneously as well as giving them control over the way they manage their files and/or projects.

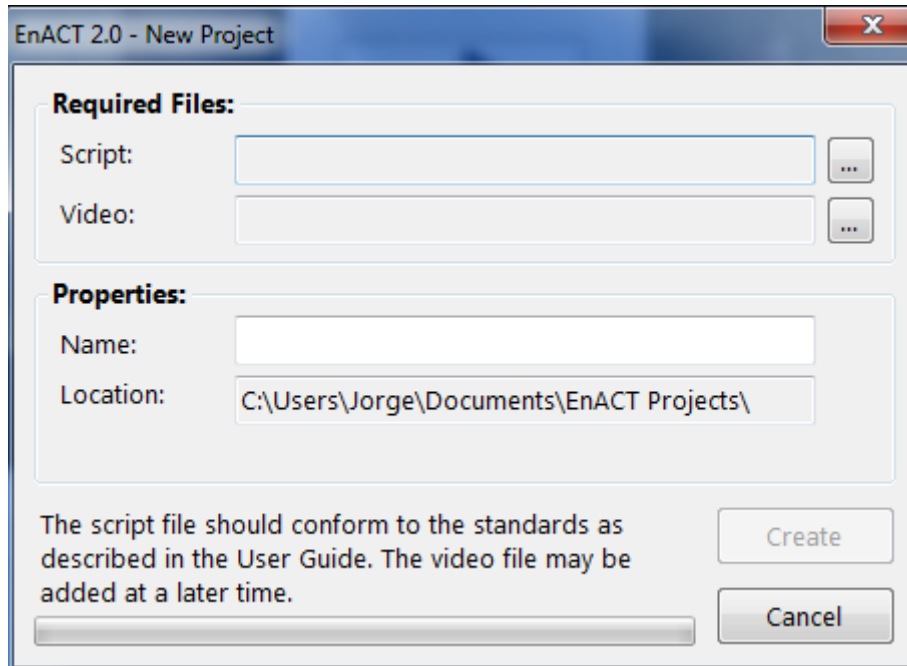


Figure 23. Creating a new project in EnACT

## 3.8.2 EnACT Script Properties

### 3.8.2.1 Script Editor Area (SEA)

The purpose of the script editor area is to display speaker names and their associated dialogue parsed from the original script file. It is here where the user can select the word(s) to which enhanced captions can be added and assigned emotions and intensity values. In the script or work area, the colours and fonts presented for each emotion are intended to display a relative difference between each other; they are not intended to represent any particular emotion or meaning. All colours and fonts can be customized through the option menu (see Figure 24), accessed through the main menu. These changes are only applied in the work area editor and will not affect the enhanced captions that appear within the video.



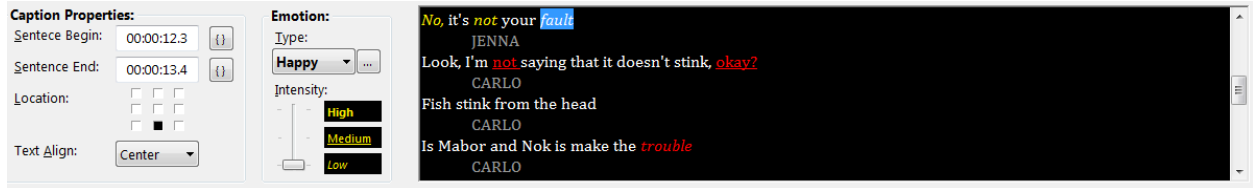


Figure 24. Editing and selecting emotions and intensities in EnACT

### 3.8.2.2 Colour options

Figure 25 shows the options for colour choices to differentiate the emotions that are used; the default colours are shown in Table 3. To change the colors for the default emotions, the user can click on one of the colour boxes corresponding to each emotion and change to a colour of his/her choosing. The same can be done by selecting the font that the user wants to use. These changes only happen in the script editor area (see Figure 25).

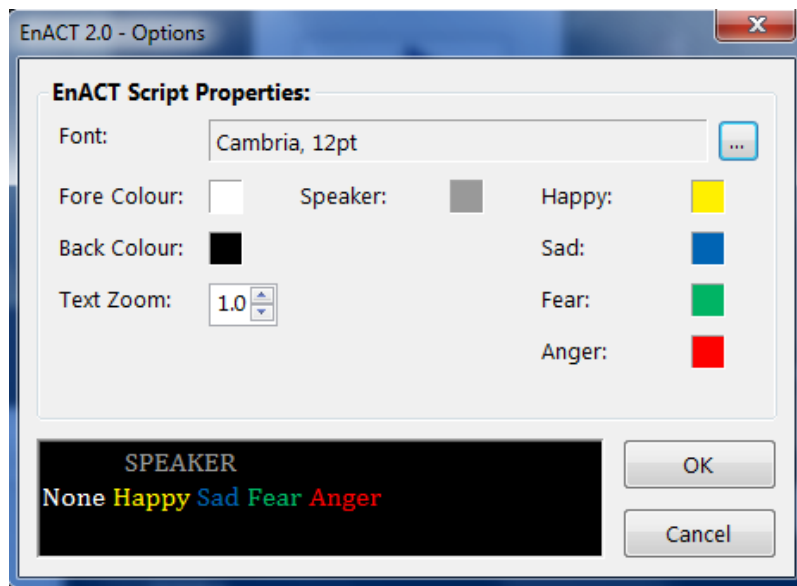


Figure 25 EnACT - Script Properties

**Table 3. Default colours for emotions in EnACT**

Emotion	Colour
None	White
Happy	Yellow
Sad	Blue
Fear	Green
Anger	Red

EnACT underwent a development process and different prototypes were constructed. The initial structure and underlying framework was designed in 2008 by Q. Vy as an undergraduate research assistant. The software was only partially complete and had not undergone any user evaluation. My Master's thesis consisted of completing EnACT, adding new functionality as outlined in this chapter and then carrying out a series of user evaluations with both target user groups.

## Chapter IV - Evaluation

### 4.1 Usability

Results from the usability study questionnaire can be grouped into four distinct groups;

- The first group entails a set of questions that rate the difficulty of performing aspects of each task
- The second group is made up of a set of questions that ask participants to rate the location of chosen elements in the UI of EnACT.
- The third group includes one question that gauges the confidence and one question that gauges the comfort level of participants when using EnACT.
- The last group consists of comments from the participants.

For analysis, the responses gained from the first category using a Likert rating scale were condensed from five to three categories into positive, negative and neutral as we did not meet the assumptions of the chi-square test for a 5-point Likert scale. However, [49] suggest that the condensation of Likert scale categories has no effect on the statistical outcome and is permissible in data analysis. The two positive categories were grouped together as one category and coded as “3” and the two negative categories were grouped together and coded “1” in the questionnaires scale. By grouping the data in a three-point Likert scale the assumptions of chi-square were met.

A chi-square analysis was performed on all questions within the first three categories. There were five significant chi-square results; Table 1 shows the significant results to the alpha error probability level of 0.05.

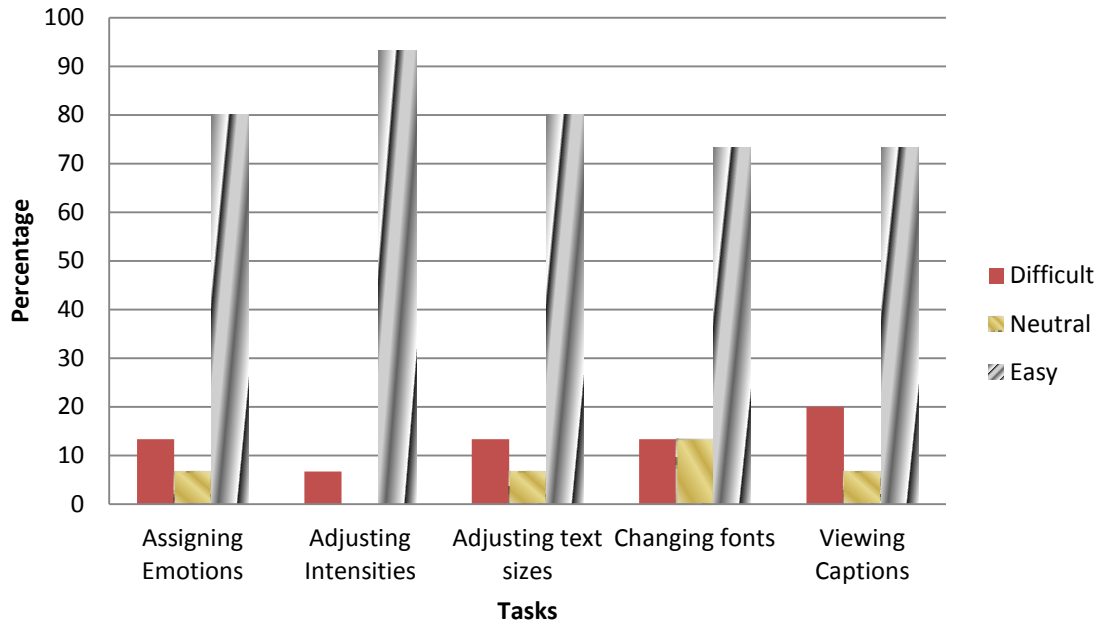
**Table 4. Chi-square table for ratings in the difficulty of task attempted where 1 = Difficult and 3 = Easy**

Tasks	X <sup>2</sup>	Df	Mean	Standard Deviation
Assigning emotions	14.80	2	2.67	0.72
Adjusting intensities	11.28	1	2.87	0.51
Adjusting text sizes	14.80	2	2.67	0.72
Changing fonts	18.87	3	2.47	0.99
Viewing Captions	11.20	2	2.53	0.83

Figure 26 shows the results of the participants in specific tasks during the study. Twelve out of fifteen (80%) participants reported that the task of marking-up the script with emotions was ‘Easy’ (M = 2.67, SD = 0.72). From these results, it appears that the functionality of selecting an emotion from a dropdown box or by using the right-click function on a word did not inhibit the user from performing this action.

This positive result appears to be consistent with results from the task of assigning intensities to the marked-up words. Fourteen of the fifteen participants (93%) rated the task of selecting an intensity for emotions as easy (M = 2.87, SD = 0.51).

Participants also had the task of adjusting the text size and the font in the SEA. Twelve participants (80%) reported that adjusting the text size was ‘Easy’ and useful (M=2.67, SD=0.72) and eleven participants (73%) found the task that required changing the fonts was ‘Easy’ (M=2.47, SD=0.99). When reporting their experience viewing the EC in the preview window, eleven participants (73%) rated the task as ‘Easy’ (M=2.53, SD=0.83). The result from this question is important as the preview function allows users to preview their EC and try different emotions, intensities and other effects with the script they are marking-up.



**Figure 26. The study showed positive feedback during specific tasks assigned to the participants**

The EnACT UI was designed with simplicity in mind. Controls were designed to be intuitive to the user and therefore reduce the required training time to produce enhanced captions quickly.

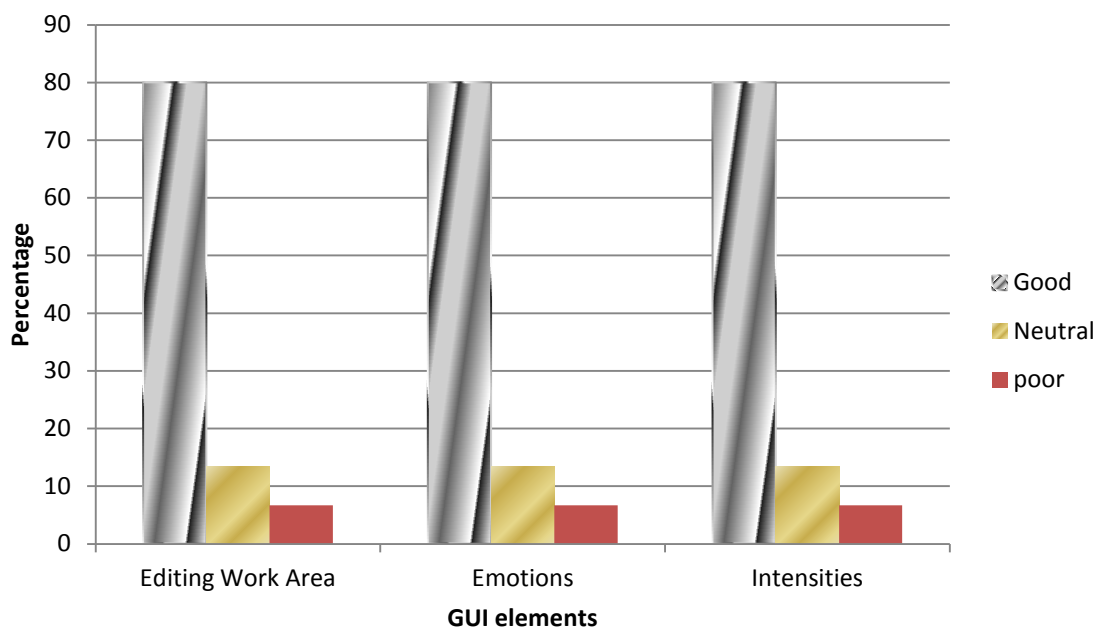
For the group of questions related to location of UI elements, there were three significant chi-square results with  $p < 0.05$  (see Table 5). The frequency of responses can be seen in Figure 27.

**Table 5. Chi-square table for the rate of opinions of the location of elements where 1=Poor and 3=Good**

Location	X <sup>2</sup>	Df	Mean	Standard Deviation
SEA	14.80	2	2.73	0.59
Emotions	14.80	2	2.73	0.59
Intensities	14.80	2	2.73	0.59

There were twelve participants (80%) rated the location of the script editor area as ‘Good’ ( $M=2.73$   $SD = 0.59$ ). One participant suggested that “adding tabs for each scene to avoid scrolling through an entire script” would be helpful. Seven participants offered suggestions for improving the UI design including building a “larger script display and an option for auto-scroll” and “positioning the script display between the caption properties panel and the emotions panel.”

Twelve participants (80%) rated the position of the emotion and intensities as ‘Good’ ( $M=2.73$   $SD=0.59$  for both factors). These two elements were placed next to each other in the UI, which was one possible explanation for the similar ratings from participants.



**Figure 27. Experience regarding the GUI**

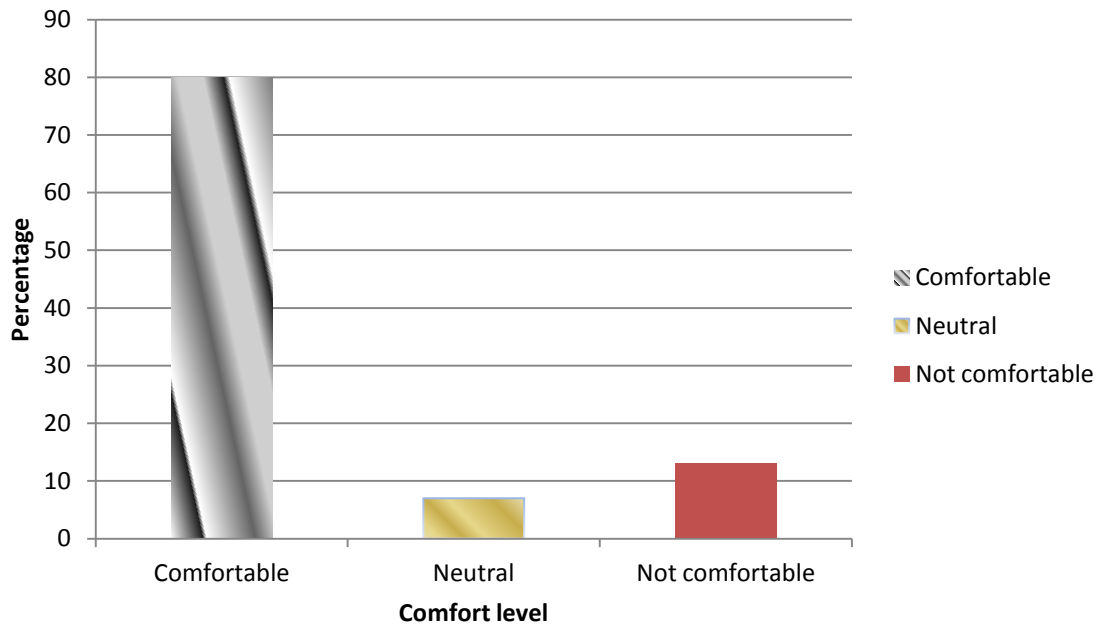
A chi-square analysis was performed to compare the responses from participants rating of their confidence and comfort level when using EnACT. As shown in Table 6, there were two significant chi-square results. Fourteen participants (93%) were confident about using the software on their own ( $M=2.93$ ,  $SD =0.26$ ). This result is important as it supports the aim of the

EnACT's design; to create a software design that reduces training time for the user and was easy to use. The results indicate that it was possible for a user to become confident enough to use EnACT without much help within a short period of time.

**Table 6. Chi-square results of the confidence rating from participants from using EnACT and participant's comfort rating when using the software where 1 = low and 3 = high.**

	<b>X</b>	<b>df</b>	<b>Mean</b>	<b>Standard Deviation</b>
<b>Confidence in using EnACT</b>	11.26	1	2.93	0.26
<b>Feeling comfortable using EnACT</b>	14.80	2	2.67	0.72

Similarly, twelve participants (80%) rated their comfort level when using EnACT for captioning a movie or TV show as 'High' (M=2.67, SD=0.72). See Figure 28 for the frequency of different responses. These results indicate that the task of marking-up captions with emotions and corresponding intensities using EnACT is easy and comfortable to accomplish within the limits of the application.



**Figure 28. Participants rating of the comfort level when using EnACT to caption a movie file**

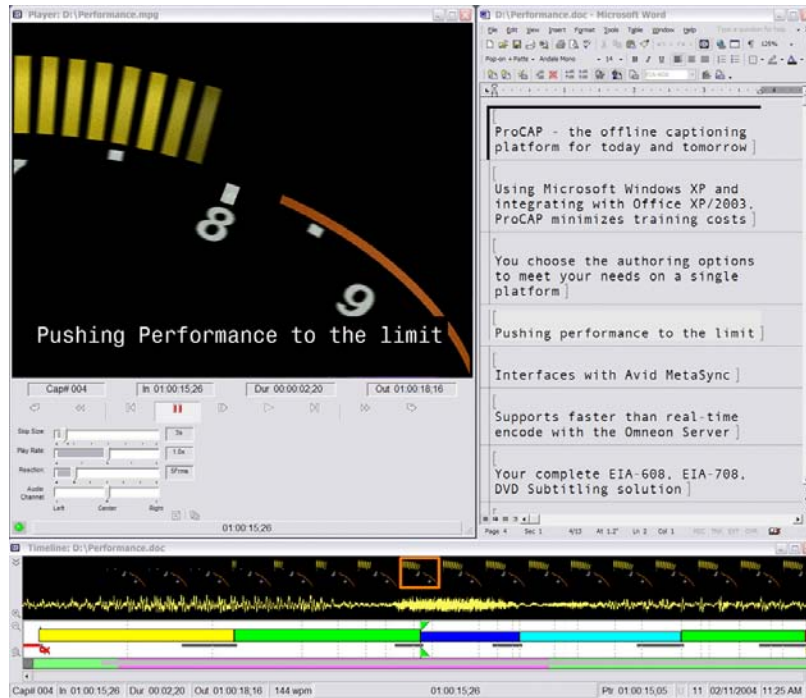
## **4.2 Case study**

An important component of understanding user experience for this software was to gather thoughts and opinions of Professional Captionists using EnACT. In this study, three case studies were conducted to examine the research questions as specified in section 3.4.3

### **4.2.1 Participant 1**

G is a female in the age group of 50-59 with a bachelor's degree and who has been employed as a closed captionist professional for the past fifteen years. She is currently employed by the Canadian broadcaster and works with a software captioning tool named ProCap as her primary closed-captioning software tool (See Figure 29 for a screenshot of ProCap).





**Figure 29. Screenshot of the dashboard of the professional captioning tool ProCap**

The interview with G lasted for approximately 60 minutes. The first 15 minutes of the study was used as an introduction to EnACT, 27 minutes were used as time for the participant to test the software and 33 minutes to complete the questionnaire and for discussion.

After working through the training tasks, G commented on the UI of the software. She made multiple comments and comparisons regarding the captioning software that she uses and EnACT. The recommendation that she insisted upon the most was to include a larger script working area in the dashboard design because she wanted to see more of the dialogue at one time while she was working with the SEA.

G experienced difficulties setting and editing the start and end times for each of the marked-up captions within the script. At minute 11, she became frustrated. She mentioned out loud that she thought the process was “annoying” since every EC required a time input that had to be entered manually. This was something G considered to be “really tedious” and also “time

consuming”. She made a comparison with the way her current captioning software, ProCap, treats the timing of captions in an automated manner = “The out-point of any caption is always going to be the in-point of the next caption – “So those two numbers need to be one and the same. So that’s how our software operates.” G also reported that these time codes are essential to a captionists role. As they are an established process in the production of captions, captioning software needs to represent the time codes associated with the captions in the script in more detail to assist the captionist This could be achieved by adding milliseconds to the video player in a format hh:mm:ss.ms where: hh: Hours, mm: Minutes, ss: seconds, ms: milliseconds.

G also compared how she is able to separate long sentences into two lines if they contained more than three or four words.

G experienced difficulty understanding the mark-up process related to setting the times to create EC. She thought that at least one word needed to be marked-up with an emotion in order to set an in and out-points. However, it is not mandatory for the user to select an emotion for a word in every line in order to add in-points and out-points to the script since if words are not assigned an emotion they are tagged as “no emotion” by default. G found this functionality design confusing as she thought that she had to mark-up each word within the script before the in-point could be added or edited.

The limited number of emotions was another frustration cited by G throughout her experience as she felt limited and unable to do her job properly as working with the current number of four emotions was “tedious” and difficult to choose for the captionist.

#### **4.2.2 Participant 2**

J. is a male closed-captionist professional between the age of 18-29 and has a college diploma. He is currently employed by a digital media video post production house located in Toronto.

The study lasted approximately 60 minutes. The introduction was 15 minutes long and 25 minutes was spent by the participant testing EnACT followed by a discussion and questionnaire that took 20 minutes.

During the first minute into completing the study tasks J described dealing with the timing of the captions as being uncomfortable. He compared the software that he uses at his work with how he deals with the timing with this software. J made a suggestion to improve this functionality by using “one set of time code for each line (which) makes it a lot easier.” Each time J marked-up the script he would only insert the in-point and not set the out-point because he was unaccustomed to having to set it with his work software. This created confusion and also frustration for him when using the EnACT.

J also pointed out that the software that he uses for captioning had error-checking capabilities to prevent human errors such as using the wrong timing (when the in-point time is larger than the out-point time) or characters entered into the system that cannot be technically displayed within the media.

J suggested that when editing or changing the in-points and out-points of a marked-up part of the script the software should also select the corresponding frame in the video without the EC that appears in the top left area of the EnACT dashboard. That way, the user would know exactly where in the script and video he/she would be adjusting.

Another problem that J encountered throughout all of the tasks was the set of four emotions and three intensities to represent emotions in the video. After previewing the EC, he was concerned about the way that the viewer would understand the animations for each of the four emotions. He thought that the EC would not be understandable as he described “it will be hard to get used to, I think, for some people just because they’re trying to read along and some of them shake and like they’re more confused about why its shaking unless it gets explained to them beforehand which I know if you just have a pamphlet you’d have to send out”. J was particularly concerned about the way that the fear and angry emotions were represented by the shaking animation. He believed that this could lead to confusion and distraction for the audience.

During the use of EnACT, it was noticed that J’s attention was fixed on adjusting the emotions and intensities of the marked-up parts of the script. He explained that he was spending more time adjusting the intensities to represent the meaning of the dialogue in the video as accurately as possible. A pattern emerged while he was marking-up the words: first he set all of the in-point and out-point times for the script, and then he focused on the marking-up the script. It seemed that J enjoyed adjusting the settings for the marked-up script and previewing the video once the timing of each line of the script was complete.

Overall, J commented that EnACT was a new and exciting project that has not been done before. After working with EnACT and creating EC J reported feeling confident enough to create EC in the future without assistance.

### **4.2.3 Participant 3**

T is a female Professional Captionist, between the age of 18-29 and her highest level of education completed is a high school education. The study lasted approximately 40 minutes in

total. 15 minutes was spent introducing T to the study, 13 minutes was spent by the participant to finish the testing tasks and finally, 12 minutes was spent by T. to provide her feedback with the questionnaire.

Since the beginning of the training tasks T did not appear to have any problem learning and understanding the design and functionalities of EnACT. T completed all the testing tasks faster than the other Professional Captionists who participated in the study and also completed the tasks without requesting assistant or asking questions.

The functionality of the timing for the in-points and out-points was not a problem for T as it was for the other Professional Captionists who participated in this study. T was able to set times for each caption with ease if compared to any other part of the testing tasks, and did not make any comments during the study or in her questionnaire that would indicate that she had issues with this functionality.

T's suggestions focused on the number of emotions provided in this version of EnACT. She suggested that it would be useful to create a larger set of emotions for EnACT in future, as she thought that some of the videos could be more accurately described with different emotion words such as sarcasm.

Once T previewed the EC that she had created in the testing tasks, the audio recording and verbal interview at the end of the study captured her comments that expressed how impressed and surprised she was to create EC with EnACT.

### **4.3 Discussion**

In Sections 4.1 and 4.2 of this chapter the data from the user and case studies were presented and analyzed in two sections as follows:

- **Usability (See Section 4.1):** Amateur Captionists were asked to caption a short video and then were asked about their experience with the software by completing a questionnaire. Table 4 and Figure 26 present the results with significant difference about the difficulty of tasks given to the participants. Table 5 and Figure 27 report results with significance difference about the Graphical User Interface design, Figure 28 also presents results that show the participant's rating of their comfort level and confidence level if they were to use EnACT in the future.
- **Case Studies (See Section 4.2):** Professional Captionists were asked to test EnACT and provide in-depth commentary and analysis that describes how EnACT could fit into their everyday captioning tasks.

#### 4.3.1 Working with Emotions

Screen recorded data showed that all participants selected an emotion before selecting an emotional intensity. Table 4 and Figure 26 showed that of these two actions, selecting intensities of the emotions was rated as easy for all participants.

In the comments users reported that selecting an emotion from the given set (sad, anger, fear and happy) for the captions was a more challenging task than selecting an intensity for each emotion. One participant said that they “found it hard to choose what emotion goes with certain text”. This comment was amongst much of the similar feedback obtained from both groups of participants, indicating that, irrespective of their captioning needs, all participants believed that the selection of the emotions limited their ability to complete the testing tasks in the study.

[29] defines human emotion perception as “the result of a joint processing of audio and visual cues”. There is a wide variety of possible descriptive labels that could be used to interpret

an emotional state as detected by the captionist in this scenario. The limited choice of descriptive emotion labels may increase the cognitive effort required by a user to interpret all the emotional cues from a video and then to use their judgment to best fit this interpretation into a label provided. This becomes particularly problematic in complex scenarios where the captionist must understand when comedic devices such as sarcasm are being used and may require additional thought and consideration by some captionists to label this appropriately when creating EC

As a consequence of the difficulty posed by the limited set of emotions, participants commented that they did not believe they were equipped to make an accurate judgment of the emotions as they appeared within the video. This led to many suggestions for future versions of EnACT to provide a larger set of emotions for users. Whilst this point is valid as it could theoretically reduce cognitive load required by the user, the article on page 537 of [50] reports that “decision making tends to adapt the decision strategy and information processing to the type of decision aids available in such a way as to maintain a low level of effort expenditure”. Providing a larger set of emotions in the next version of EnACT could have a negative impact on the confidence level of the user when assigning emotions to words as it could do the opposite of what they are asking for by increasing the effort expenditure. The inclusion of a greater set of emotions may also have an impact on the design of the user interface and because of this implication may add further pressure to the user in mastering the software as [51] mentions, “the single greatest deterrent from getting started with a program is the amount of complexity new users must face in the very beginning.” EnACT’s interface is designed so that Amateur Captionists can get started quickly and Professional Captionists could quickly adapt to EnACT as an add-on or plug-in to their existing captioning software. By creating the program with the intention of reducing the learning curve for new users, Professional Captionists can learn

advanced functionalities and keyboard shortcuts easily so as not to disrupt their established workflow.

Further research is required to determine the optimal number of emotions and intensities for EnACT that will assist Professional and Amateur Captionists to be efficient in their work processes without overloading them with interface complexities or cognitive load. For this experiment, a basic set of four emotions (happy, anger, sad and fear) as specified in [52] were used rather than the eight emotions (sadness, anger, happiness, fear, surprise, disgust, anticipation and acceptance) reported in [30]. Further support was found to confirm the decision to use four basic emotions for EnACT. Psychological models of emotion proposed by [53] and [30] suggest that “all emotions can be reduced to a set of five to eight primitive emotions - sadness, anger, happiness, fear, surprise, disgust, anticipation and acceptance”. However, Acton [53] reports that in more of 50% of his studies; sadness, anger, happiness and fear are common denominators.

The limited set of emotions was also chosen to focus on examining the process of marking-up the script and creating EC, rather than focus on the process of interpreting and selecting the best emotion for the video. Complicating the decision making process with an extended set of emotions may have interrupted the user work flow by placing a heavy cognitive load on only one portion of the testing tasks and could have skewed the results accordingly.

Furthermore, having unique animations for a large set of emotions was not feasible because obtaining the correct animations for a new set of emotions would require further experimentation and analysis of the artistic and psychological processes (similar to what was reported in Rashid article). The animations generated would then need to go through another



round of evaluation with users prior to it being added to the EnACT system. This is possible for future works but it was not the focus of this study since I was testing the usability and feasibility of the software.

#### **4.3.2 EnACT User Interface**

All participants reported being satisfied with the location of the SEA (Table 5), however, some participants suggested that the size of the SEA could be problematic if they were to perform the same task with a longer script. In particular, a longer script would require more physical manipulation scrolling to navigate the small script viewing area by the user. At present, only four to six lines of script appear in the SEA on the interface. This may also impose a higher cognitive load on the user to remember and locate areas in the script for their own reference, without the ability to skip ahead quickly. The scrolling action is very mouse dependant and since Professional Captionists are primarily keyboard users it could become a very frustrating and time consuming task.

One participant offered one potential improvement to this, suggesting that tabs could be used to index each scene of the script, organizing a larger script into more manageable parts. Other participants suggested an increase in the size of the SEA; however, they did not give any indication as to how the other interface elements would be affected if this change were to take place.

Overall, participants rated the location of the emotion and intensity mark-up functions positively. Due to the close functional relationship between an emotion and the intensities, both elements were placed in close proximity to each other which may explain the similar ratings for both elements. Additional functionality was given to users to perform the same action by right-

clicking on a word in the script, where they would be presented with a graphical display of a menu of emotions and intensity levels that can be selected, as shown in Figure 30. All participants were aware of the ability to mark-up the script with the right-click functionality to select emotions as it was described during the introduction of the software at the beginning of the study. Most of the Amateur Captionists used the right-click function in the SEA to mark-up their script, whilst only some of the Professional Captionists used the right-click function to change an already marked-up part of the script from one emotion to another, as required in the testing tasks.

Based on my observations, Professional Captionists would be more inclined to use the keyboard rather than the mouse because they are habituated to keyboard use for captioning. This could explain why using the right click functionality would become an unusual action to perform for them. For Amateur Captionists who are also regular computer users, using the right click to assign the emotions and intensities is a more familiar task since it is available in more commonly available software such as Microsoft Word™.

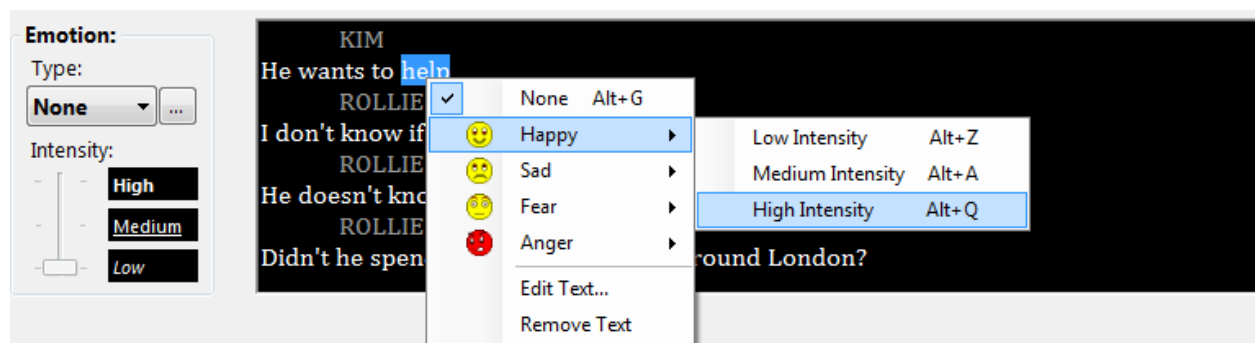


Figure 30. An alternative way to mark-up the script with emotions and intensities uses the right-click

#### 4.3.3 Confidence and Comfort Level using EnACT

Table 6 shows the reported confidence level of participants when selecting and adding emotions to words in the script and also their reported comfort level when using EnACT.

Both groups of participants rated their confidence level in using EnACT as high ( $M=2.93$ , where the highest is 3). Fourteen out of fifteen participants thought that selecting a word(s) to assign an emotional value was simple, but not necessarily an easy semantic task. These responses somewhat contrast with comments many participants made throughout the completion of the tasks. Participants noted that the set of emotions provided by EnACT was too limited in order to provide an accurate representation of what was occurring within the video. Having this limited set of emotions may have frustrated the participants as they expected a bigger set to choose emotions. This expectation may explain the substantially greater number of negative comments around the emotion labels from users when performing the tasks. Whilst the introduction to the study outlined that the testing tasks would not judge each user based on the accuracy of the emotions selected in their study, three Amateur Captionists and all of the Professional Captionists still reported that they were limited in their ability to assign emotions to the script. As many of the tasks required each participant to repeat the process of assigning emotions multiple times, the assumption in the study design was that participants would become accustomed to the process and understand that the emotions assigned would not have to be an accurate representation of the emotional content in the video. This provided the participants with a more sense of comfort and confidence in using the software, as the more they used it, the easier and faster they could adjust and/or change the emotions from words.

#### **4.3.4 Participant suggestions and opinions on EnACT**

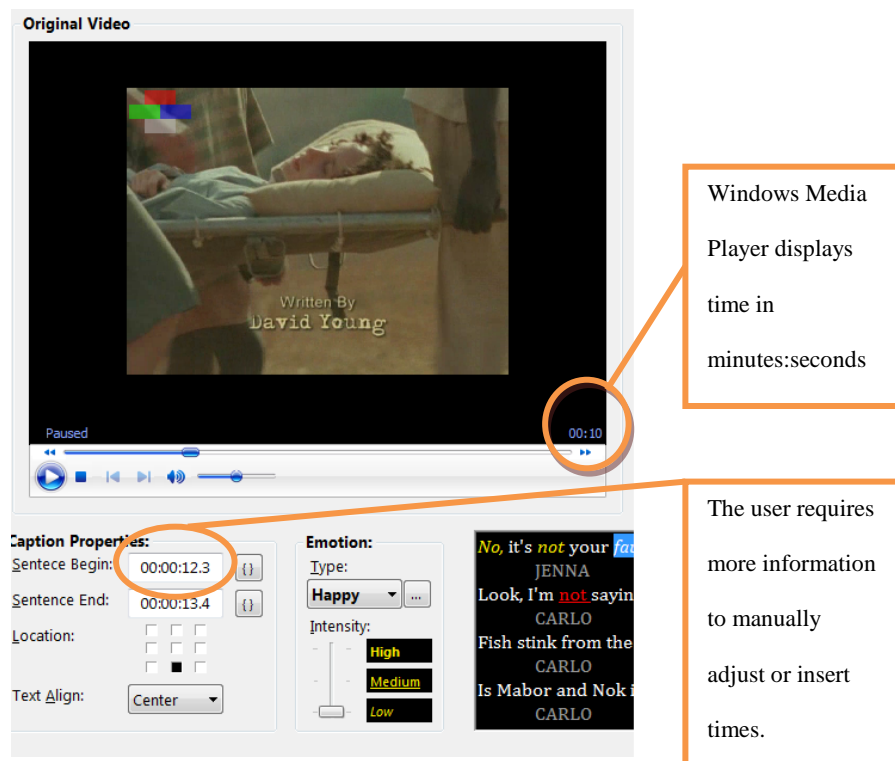
The overall reaction towards EnACT was positive and participants said that the use of the software to create EC was a “well thought of concept” and a “very good idea”. The additional functionality of the right-click feature to assign emotions and intensities was well received by participants, especially the Amateur Captionists, as another way to mark-up the script. Both

groups reported that this was useful particularly when they wanted to change emotions and intensities of previously marked-up parts of the script. Whilst this feedback was positive, participants reported negatively in the questionnaire that initially the task of assigning only one of four emotions to the script limited their choices when marking-up words as explained in Section 4.3.1

One emotion that participants suggested adding to the existing set of emotions was sarcasm. This is because participants found that the video used in the evaluation tasks contained “a couple dialogues to contain sarcasm in their voice.” Participants were confused about how they should represent this with the four emotions given. Adding animated text for other emotions is possible, however, “sarcasm is a very complicated emotion that can be difficult to understand and interpret” [54]. According to [55] “sarcasm is conveyed by slower tempo, lower pitch level and greater intensity than non-sarcasm”. Understanding and accurately representing additional emotions to the existing set goes beyond the scope of this thesis, since it focuses on the understanding of user interaction with EnACT’s interface. A more in-depth investigation is required into the visual representation of more complex and sophisticated emotions such as sarcasm.

Participants reported that they found assigning and altering the timing for their EC “tedious” and “difficult”. By observing the screen recording captured, it can be seen that many participants experienced difficulty synchronizing their captions to the corresponding dialogues in the video. The video player in the top left hand section of the interface is responsible for playing and controlling the original video file to be captioned. This video player contains a control bar that displays basic information of the time that has lapsed as the video plays, as shown in Figure

31. The time information is displayed as hh:mm:ss where hh represent the hours, mm the minutes and ss the seconds.



**Figure 31. Windows Media Player does not display the time in the same format that is required for input in the EnACT interface to set the timing for EC.**

This feedback about timing issues is valuable and can be used to further develop the capabilities of EnACT however, it is important to remember that EnACT is created with the intention of becoming an add-on to captioning software for Professional Captionists rather than a standalone application. The issues with setting the timing of EC could be overcome by the existing timing functionality in professional captioning software that would be available for EnACT to build upon.

Further development will still be required to EnACT however, to ensure that the timing functionality is more user-friendly to Amateur Captionist users. The timing assignment to EC

would need to be made more intuitive to reduce difficulties in their workflow. A potential solution to this problem would be to create a time display in this same video player window in the format of hh:mm:ss.ms where ms are milliseconds or to use an alternative media player that is capable of displaying the time in the same format and also allow the user to move the frames. Further development of EnACT could explore the creation of a custom media player that incorporates all or some of these elements, or replacing the current media player with an existing player that has the desired functionalities. Alternatively a time marker could be included where the user could right-click on a frame and the time would be automatically entered in the EnACT time input fields with the corresponding timing. Another possible solution to this issue could be to implement some speech recognition algorithms to delineate speech from non-speech spaces and so more automatic processing of time.

Participants suggested and requested more control of video playback in the media player, particularly with setting the timing for the EC. Participants suggested that a new functionality should be added that would allow users to “move the video back/forward frame by frame”. Another suggestion from a participant requested that functionality be added to the software so that once the video plays it also automatically follows the script with a marker that will show the current dialogue that is being spoken, within the SEA. The text, time and video would then be synchronized together and easier to control. During one interview, one of the participants used a virtual interface sketching software known as Balsamiq [56] to demonstrate his ideas as shown in Figure 32.

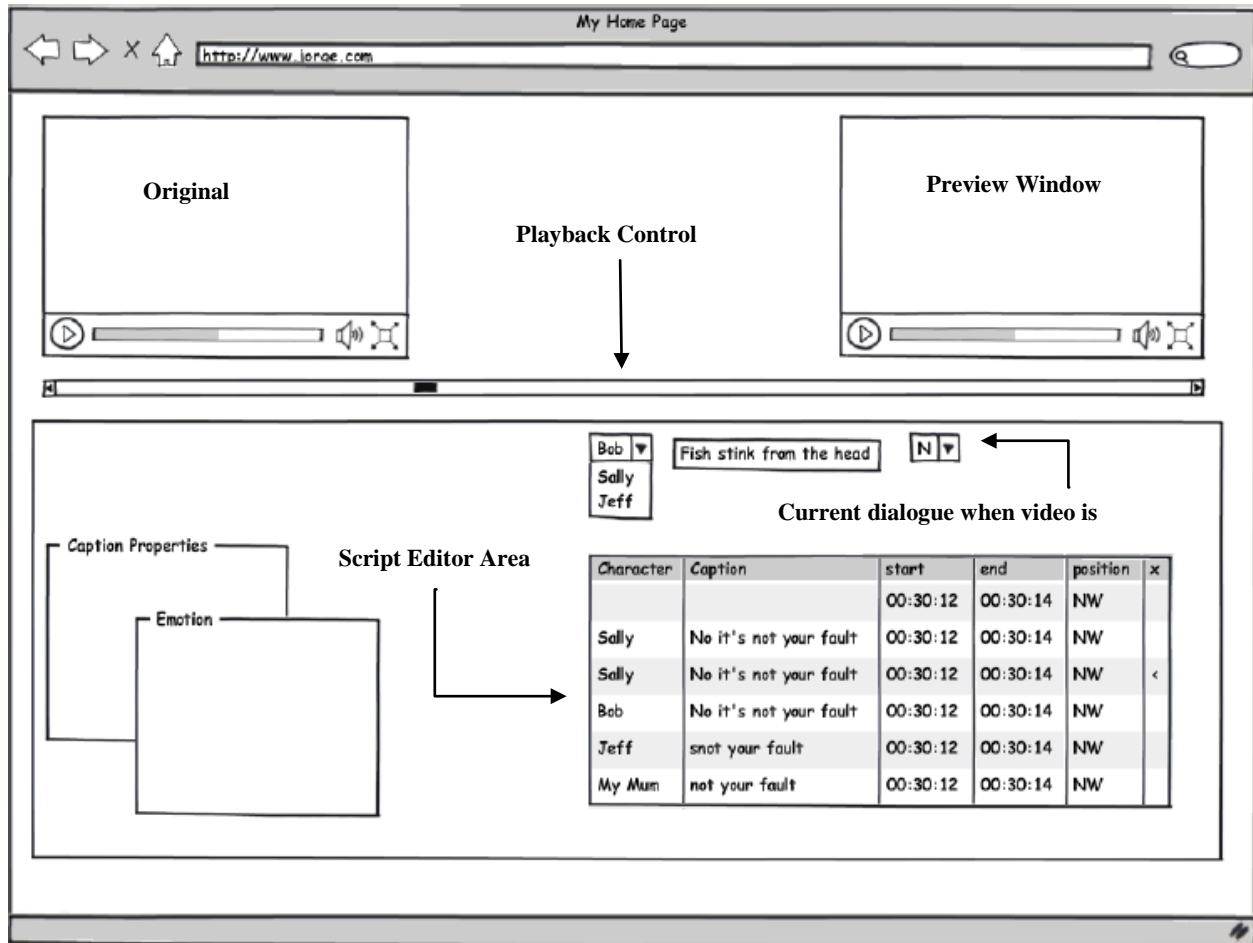


Figure 32. EnACT Version 3.0 redesigned by an Amateur Captionist participant based on his suggestions for improvement to the interface

As seen in Figure 32, the media player playback control is present at the bottom of the Original Video Window and also the Preview Video Window. These video progress bars will have playback control for both videos. In the lower portion of his redesign of the EnACT interface, the participant increased the size of the SEA and has re-imagined the RichTextBox as a dynamic table. This dynamic table would update or highlight text as it corresponds to the dialogue that is currently playing in both media players. This dynamic table would be divided into six columns that would contain the following fields:

- Character: This area will specify the speaker of the dialogue onscreen.

- Caption: This field will contain the written dialogue that is also being spoken and on both screens.
- Start: The time when the caption that corresponds to the dialogue is set to display.
- End: The time when the caption that corresponds to the dialogue is set to stop being displayed onscreen.
- Position: This field contains information that represents the location of the caption onscreen.
- X: This character acts as a marker to show the user what line of dialogue is being spoken in the video, while the video is playing.

This design could be an alternative to the current EnACT UI as it addresses some of small SEA and browsing through a big script issues identified by the participants with the current version of EnACT. This alternative version could connect the videos in the original and preview window with the new SEA, making the process of moving through the script potentially easier and more direct.

The participants of this study were impressed and comfortable using the software. Any problems participants encountered with setting the timing of the EC and assigning the emotions at the beginning of the experiment did not prevent them from creating EC. Some participants mentioned that after only a couple of minutes using the software they became more familiar with using EnACT. This is a positive result for EnACT in that it could benefit users in the use and creation of an alternative way to represent dialogue in video, and potentially enhance the entertainment experience of the audience.



#### **4.3.5 Limitations of the research**

The results obtained from this experiment were positive, however, there are a number of elements in the study that limited the results.

For the usability study fifty invitations were sent to Professional and Amateur Captionists. Of this number only fifteen participants agreed to participate in the study – three Professional Captionists and twelve Amateur Captionists. As a result of the low number of participants, most statistical analyses were not possible and interpretation of the results was limited. Scheduling time for the study with Professional Captionists was also challenging due to their demanding employment schedules, and as a result few were willing to discuss the possibility of testing EnACT. Many Professional Captionists were also difficult to find as fewer are employed in a fulltime capacity. This affected the timeline of the study as the integrity of results was heavily dependent on Professional Captionist participation. To overcome the constraints of participant schedules and the difficulty in finding suitable participants that were Professional Captionists, a different study could be designed that integrates the evaluation into a workplace setting and by also adding a higher reward upon completion of the study.

Another limitation in this study was that the cognitive workload of participants was not measured. This data would have provided further insight into the cognitive demands experienced by the user when creating EC. This could have been captured through alternative methods such as NASA TLX [57] or biometric measures such as galvanic skin response which do require self reports and could be more accurate representations of arousal or stress. This could have provided more balanced data to draw conclusions from and understand the user experience in more depth. For future study designs, a stationary study location would assist with the use of these

technologies as at this point in time they require professional set up and calibration to ensure the most accurate data is being captured.

Furthermore, this study required that Professional Captionists test EnACT with only a small portion of a real TV script. This scenario provided in the study did not take into account some of the actions they may have normally taken when dealing with a longer script so the results may differ if these participants were provided with a full script. A longer script may have provided them with an experience closer to what would be required if they were to use EnACT in a professional situation. Future studies should consider recreating a longitudinal task with a longer, complete script as this data could provide more reliable information that would have overcome any novelty effect that may have skewed the results of the study. More comprehensive data could be captured with a longer period of testing and could also account for the different learning curves among participants.

Finally, due to the time limit to complete the requirements for the Masters program, an integration of EnACT with a captioning tool currently in use in a professional setting was not possible. However, the results of this current study are encouraging and point to continuing with the EnACT project, making some of the modifications as suggested by participants and creating a tool that could be integrated with an existing captioning or video editing tool.

## **Chapter V - Conclusion, summary and future work**

### **5.1 Summary**

Little innovation has been done in Closed Captioning since its creation in the 1970's while the evolution in the television technology and film has increased dramatically. Research has shown that there was a need for more information to be displayed particularly for the non-speech information. Some research to address this issue was attempted in the past but had little success.

Past studies and research regarding improvements to CC led to the creation of Enhanced Captioning using animations to convey non-speech information such as emotions and related intensities. EC may improve the quality and enhance the entertainment value of a TV show or a movie by animating emotions and their intensities and adding them to the CC text, something that conventional CC cannot accomplish successfully with its static text. Digital television signals are becoming more prominent in the broadcasting industry, and the CEA-708 standards allow for the use of EC since the data bandwidth that it specifies allows the use of colour, animation and different fonts.

In this thesis a software tool to mark up words and create Enhanced Captions has been presented in detail. EnACT (Emotive and Affective Captioning Tool) is a markup captioning tool intended to be a plug-in or add-on to existing software captioning or video editing tools used in the industry but with enough basic captioning functionalities to use on its own. These functions include the ability to edit times for each caption, edit the dialogue in the SEA and choose the location where the captions should display on the screen. EnACT was designed to allow users to select four different emotions and three intensities and assign them to text in a movie or TV

script. EnACT then automatically renders those assignments into animated captions and then displays them as an overlay on the video on the screen. My contributions to EnACT included the fixing of major bugs in the software such as incomplete loading of the script into the SEA, creating and adding the preview window functionality by introducing the ffmpeg tool, allowing the functionality to create new project folders for new users and adding keyboard shortcuts when creating and editing the times for each caption.

Furthermore, this thesis provided an evaluation of the usability and functionality of EnACT with amateur and Professional Captionists. The main results of the studies showed that most participants rated the EnACT system as easy to use and EC as an alternative to current CC practices worth considering. However, most participants requested the addition of more emotions since only having four emotions was too restricting. They believed that they could not create the most accurate representation of the emotions with so few emotional labels. Other participants suggested increasing the size of the SEA so that it would be easier to see and handle longer scripts. Finally, participants suggested that the timing indicators for caption in and out points needed improvement because most users considered finding and adjusting the correct times a tedious and difficult task.

The task of marking up scripts with emotions for captions and understanding the output was shown to be feasible by Professional and Amateur Captionists.

Overall, EnACT 3.0 was an improvement to EnACT 2.0 as I was able to take EnACT and make it distributable and useable for Professional or Amateur Captionists to use it and create their own EC, by fixing major bugs such as loading incomplete dialogues to the SEA and adding more functionality such as creating new projects for the new users and the ability to convert any

video file into a flash file to integrate it with the EC and adding keyboard shortcuts as specified in section 3.8.1.

## **5.2 Future Research**

Although EnACT was reported to be a simple tool and the task of assigning emotions was also considered to be relatively straight forward and enjoyable for the participants, several improvements to the UI are required. These include creating a larger SEA that will provide a larger panel for the display of the TV or movie script so the user can navigate through the file with no problem, another issue that will require further research will be to include more emotions into the EnACT engine and then test with users to see if the animations to the new set represents the semantic meaning of the emotions presented.

The EnACT engine will also need to be improved since at the time of the writing of this thesis, the engine was written on ActionScript 2.0; an absolute version of ActionScript. The engine should be migrated to the newer version, ActionScript 3.0, in order to make it more maintainable and improve the animation of the emotions that EnACT contains. Considerations should also be given to convert the entire application into a more generic programming language such as C, C++ or Java in order for it to become more robust, mobile friendly and portable.

One of the next steps in the development of EnACT will be to work with an existing software captioning or video editing tool and try to assemble it as an add-on or plug-in. In addition, the research will introduce measures for cognitive workload such as the NASA TLX so that the impact of the evaluation tasks on workload can be estimated. .

Finally, another possible direction that EnACT could take would be to migrate part of its functionality to a web application. The output file created by EnACT is a flash file, so it can be

distributed or uploaded to multimedia websites such as YouTube, DailyMotion and more online communities.

## Appendices

### Appendix A: Definitions

This list contains the terms that are used in this thesis. These definitions were directly taken from [1] for more information and more definitions the source provides a wider list of definitions.

**deaf:** a medical term to those who little to no hearing. Also explains that it can be described as a “collective noun to refer to people who are medically deaf but who do not necessarily identify with the deaf community.”

**Deaf:** it is a sociologically term that involves people who are medically deaf or HOH and identifies themselves with the deaf community and their main communication technique is through sign language.

**Hard of Hearing:** refer to the people with mild to profound hearing loss; their main communication technique is speech.

**deafened (Also known as late-deafened.):**

This is both a medical and a sociological term referring to individuals who have become deaf later in life and who may not be able to identify with either the Deaf or the hard of hearing communities.

## Appendix B: Source Code

### Preview Button Code

```
// The preview feature of the program
private void preview()
{
    if (projectXML != null)
    {
        string newVideoPath =
            FileHelper.getPath(projectXML.PROJECT_FILE) +
            @"\video.flv";
        string replacePath =
            FileHelper.getPath(projectXML.PROJECT_FILE) +
            @"\video2.flv";
        string path = FileHelper.getPath(projectXML.PROJECT_FILE);
        string videoFiles =
            FileHelper.getPath(FileHelper.getPath(DEMO_PATH));
        string resources = path + @"\Resources";

        FileHelper.CreateDirectory(resources);

        updateSettingsFile();

        previewFlashVideo.LoadMovie(0, BASE_PATH +
            @"WE_demo\EnACT>LoadingMovie.flv");

        // If video does not need to be replaced, and has already been
        // converted
        // Files from the Resources folder that are needed to play will
        // be copied over to the project folder
        if (File.Exists(newVideoPath) && (!replaceVideo))
        {
            //Saving captions files
            SaveProject(false);

            //Copy files to video path to play

            //Copy files from Resources folder to main folder in order
            //to play
            copyImportantFiles(DEMO_PATH, path, @"\ClearOverAll.swf");
            copyImportantFiles(DEMO_PATH, resources, @"\settings.dtd");
            copyImportantFiles(videoFiles, path, @"\Settings.xml");
            copyImportantFiles(videoFiles, path, @"\Engine.swf");

            //Play video on preview panel
            startPreview("00:00:00");
            previewFlashVideo.LoadMovie(0, path + @"\Engine.swf");

            string oldVideo =
                (FileHelper.getPath(projectXML.PROJECT_FILE)) +
                @"\video2.flv";
            File.Delete(oldVideo);
        }
        // If user decides to change video in existing project, "video2"
```



```

// will be created
// Files will be copied over, and changes to the Settings.xml
// file will be made in order to load the new file.
else if (File.Exists(replacePath) && (replaceVideo))
{
    copyImportantFiles(DEMO_PATH, path, @"\ClearOverAll.swf");
    copyImportantFiles(DEMO_PATH, resources, @"\settings.dtd");
    copyImportantFiles(videoFiles, path, @"\Settings.xml");
    copyImportantFiles(videoFiles, path, @"\Engine.swf");

    changeSettings();

    startPreview("00:00:00");
    previewFlashVideo.LoadMovie(0, path + @"\Engine.swf");

    SaveProject(false);
}
else
{
    LoadingScreen converting = new LoadingScreen();
    converting.Show();
    convertVideo(projectXML.VIDEO_FILE);
    converting.Close();
    //Spaces in order to fit in the middle of the image
    BtnPreview.Text = "    Show Preview";
}
}
else
    MessageBox.Show("Please create or open an existing project first.", Application.ProductName + " - Error")
}

```

## Converting video code

```
//-----  
// This method uses the ffmpeg to convert the video  
//-----  
private void convertVideo(string videoPath)  
{  
    string converterPath =  
System.Environment.GetFolderPath(Environment.SpecialFolder.MyDocuments) +  
@"\EnACT Projects\ffmpeg\bin\";  
  
    string newVideoPath =  
FileHelper.getPath(projectXML.PROJECT_FILE);  
  
    Process proc = new Process();  
proc.EnableRaisingEvents = false;  
    //proc.StartInfo.FileName = @"C:\Users\Jorge\Documents\My  
Dropbox\Thesis\ffmpeg\bin\ffmpeg.exe";  
proc.StartInfo.FileName = converterPath + "ffmpeg.exe";  
  
    if (!File.Exists(proc.StartInfo.FileName))  
    {  
        MessageBox.Show("Software required not found, continue to  
find ffmpeg manually", "Converter not found!!");  
        if (findProgramDialog.ShowDialog() == DialogResult.OK)  
        {  
            proc.StartInfo.FileName = findProgramDialog.FileName;  
        }  
        else  
            return;  
    }  
  
    //proc.StartInfo.Arguments = "-i " + "\"" + videoPath + "\"" + "  
-ar 22050 -ab 32 -f flv -s 320x240 " + "\"" + FLASH_VIDEO_PATH + "\"";  
  
    if (!replaceVideo) // If replaceVideo flag is off, "video.flv"  
will be created.  
        proc.StartInfo.Arguments = "-i " + "\"" + videoPath + "\"" + "  
" -ar 22050 -ab 32 -sameq -f flv -s 320x240 " + "\"" + newVideoPath + "  
@"\video.flv";  
        else if (replaceVideo) // If replaceVideo flag is on,  
"video2.flv" will be created.  
        proc.StartInfo.Arguments = "-i " + "\"" + videoPath + "\"" + "  
" -ar 22050 -ab 32 -sameq -f flv -s 320x240 " + "\"" + newVideoPath + "  
@"\video2.flv";  
        proc.StartInfo.UseShellExecute = false;  
        proc.StartInfo.CreateNoWindow = false;  
        proc.StartInfo.RedirectStandardOutput = true;  
        proc.Start();  
        proc.WaitForExit(); //Comment this out if you dont want EnACT to  
freeze while conversion is happening  
        proc.Close();  
}
```

### Example of the dialogues.xml file

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE captions SYSTEM "../captions.dtd">
<captions>
  <caption begin="00:00:05.2" end="00:00:06.4" speaker="CARLO"
location="2" align="1">
    <emotion type="0" intensity="0">She's</emotion>
    <emotion type="0" intensity="0">going</emotion>
    <emotion type="0" intensity="0">to</emotion>
    <emotion type="0" intensity="0">be</emotion>
    <emotion type="2" intensity="3">okay?</emotion>
  </caption>
  <caption begin="00:00:07.0" end="00:00:08.1" speaker="RACHEL"
location="2" align="1">
    <emotion type="0" intensity="0">Yeah,</emotion>
    <emotion type="0" intensity="0">she</emotion>
    <emotion type="1" intensity="1">should</emotion>
    <emotion type="1" intensity="1">be</emotion>
  </caption>
  <caption begin="00:00:08.0" end="00:00:09.5" speaker="RACHEL"
location="2" align="1">
    <emotion type="0" intensity="0">We</emotion>
    <emotion type="0" intensity="0">just</emotion>
    <emotion type="0" intensity="0">have</emotion>
    <emotion type="0" intensity="0">to</emotion>
    <emotion type="0" intensity="0">make</emotion>
    <emotion type="2" intensity="1">sure</emotion>
  </caption>
  <caption begin="00:00:12.6" end="00:00:14.0" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="0" intensity="0">Her</emotion>
    <emotion type="0" intensity="0">heart</emotion>
    <emotion type="0" intensity="0">has</emotion>
    <emotion type="3" intensity="2">stopped</emotion>
  </caption>
  <caption begin="00:00:14.0" end="00:00:15.4" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="0" intensity="0">Carlo,</emotion>
    <emotion type="3" intensity="2">blow</emotion>
    <emotion type="0" intensity="0">into</emotion>
    <emotion type="0" intensity="0">the</emotion>
    <emotion type="3" intensity="2">tube</emotion>
  </caption>
  <caption begin="00:00:18.6" end="00:00:19.4" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="1" intensity="1">That's</emotion>
    <emotion type="1" intensity="1">it</emotion>
  </caption>
  <caption begin="00:00:21.4" end="00:00:22.5" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="0" intensity="0">Go</emotion>
    <emotion type="1" intensity="1">on</emotion>
  </caption>
  <caption begin="00:00:22.5" end="00:00:23.5" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="3" intensity="2">Every</emotion>
```

```

        <emotion type="0" intensity="0">five</emotion>
        <emotion type="3" intensity="2">seconds</emotion>
    </caption>
    <caption begin="00:00:26.4" end="00:00:27.9" speaker="RACHEL (CONT'D)"
location="2" align="1">
        <emotion type="4" intensity="1">Not</emotion>
        <emotion type="4" intensity="1">too</emotion>
        <emotion type="4" intensity="1">hard</emotion>
    </caption>
    <caption begin="00:00:31.7" end="00:00:32.6" speaker="CARLO"
location="2" align="1">
        <emotion type="0" intensity="0">How</emotion>
        <emotion type="3" intensity="2">long</emotion>
        <emotion type="0" intensity="0">do</emotion>
        <emotion type="0" intensity="0">we</emotion>
        <emotion type="0" intensity="0">do</emotion>
        <emotion type="3" intensity="2">this?</emotion>
    </caption>
    <caption begin="00:00:32.6" end="00:00:33.6" speaker="RACHEL"
location="2" align="1">
        <emotion type="0" intensity="0">As</emotion>
        <emotion type="3" intensity="2">long</emotion>
        <emotion type="0" intensity="0">as</emotion>
        <emotion type="0" intensity="0">it</emotion>
        <emotion type="3" intensity="2">takes</emotion>
    </caption>
    <caption begin="00:00:40.4" end="00:00:41.0" speaker="RACHEL (CONT'D)"
location="2" align="1">
        <emotion type="1" intensity="3">Heartbeat!</emotion>
    </caption>
    <caption begin="00:00:43.3" end="00:00:44.0" speaker="RACHEL (CONT'D)"
location="2" align="1">
        <emotion type="0" intensity="0">Her</emotion>
        <emotion type="1" intensity="2">breathing's</emotion>
        <emotion type="1" intensity="3">back</emotion>
    </caption>
    <caption begin="00:00:47.1" end="00:00:48.0" speaker="CARLO"
location="2" align="1">
        <emotion type="1" intensity="3">You</emotion>
        <emotion type="1" intensity="3">save</emotion>
        <emotion type="1" intensity="3">a</emotion>
        <emotion type="1" intensity="3">life!</emotion>
    </caption>
</captions>

```

## Appendix C: Ethics approval



To: Jorge Mori  
Computer Science  
Re: REB 2010-141: EnACT: Usability Study  
Date: September 2, 2010

Dear Jorge Mori,

The review of your protocol REB File REB 2010-141 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 2010-141) on future correspondence.

Congratulations and best of luck in conducting your research.

A handwritten signature in black ink, appearing to read "Nancy Walton".

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

## **Appendix D: Questionnaire**

### **Usability Questionnaire**

The purpose of this questionnaire is to understand how effective EnACT is to learn and use. This questionnaire should take about 15 minutes to complete the questionnaire. Remember that your participation in this study questionnaire is voluntary, you can choose to not to answer any of the questions provided. Thank you for your time and effort.

### **Demographics**

1. What is your hearing status?
  - a. Hearing
  - b. Cochlear implant
  - c. Hard of hearing
  - d. Deafed
  - e. Deaf
2. What is your gender?
  - a. Male
  - b. Female

3. What is your age?
- a. 18 – 29
  - b. 30 – 39
  - c. 40 – 49
  - d. 50 – 59
  - e. 60 +
4. What is your highest level of education completed?
- a. No formal education
  - b. Elementary school
  - c. High school
  - d. College (diploma, 2 or 3 years)
  - e. University (bachelor's degree, 4 or more years)
  - f. Graduate school
  - g. Prefer not to answer
5. How often do you use the computer per day?
- a. Never
  - b. Seldom
  - c. Sometimes
  - d. Often
  - e. Always

6. Please rate the how difficult you found the following tasks you attempted with EnACT

(please circle one number from 1 to 5 for your rating, or 0 if the task was not completed):

	Did not try	Very Difficult	Somewhat Difficult	Neutral / No opinion	Somewhat Easy	Very Easy
1. Loading in the script into the software	0	1	2	3	4	5
2. Assigning emotions to the words	0	1	2	3	4	5
3. Adjusting intensity of the emotions	0	1	2	3	4	5
4. Saving the project	0	1	2	3	4	5
5. Finding and opening a saved project	0	1	2	3	4	5
6. Loading a movie into the software	0	1	2	3	4	5
7. Adjusting the text size	0	1	2	3	4	5



8. Changing the colours for the different emotions	0	1	2	3	4	5
9. Changing to a different font	0	1	2	3	4	5
10. Seeing the changes you made in the text	0	1	2	3	4	5
11. Viewing the captions on the video	0	1	2	3	4	5
12. Using the video controls	0	1	2	3	4	5
13. Reading the text	0	1	2	3	4	5
14. Changing the emotion assignments	0	1	2	3	4	5

7. Rate your opinion on the location of the following elements in helping you use EnACT:

	Very poor	Poor	Neutral / No opinion	Good	Excellent
1. The location of the script	1	2	3	4	5
2. The location of emotions	1	2	3	4	5
3. The location of the intensities	1	2	3	4	5
4. The options menu	1	2	3	4	5
5. The location of the movie	1	2	3	4	5

8. Rate your confidence in being able to mark up captions with EnACT in the future without any assistance.

- a. Very confident
- b. Confident
- c. Neutral
- d. Not that confident
- e. Not at all confident

9. What did you find easiest to do with EnACT?

---

---

---

---

10. How comfortable would you feel if you were to use EnACT to caption?

- a. Very comfortable
- b. Comfortable
- c. Neither comfortable or not
- d. Not comfortable
- e. Not comfortable at all

11. What were the main limitations of EnACT that you found?

---

---

---

---

12. Do you have any suggestions that you think would make EnACT more effective for you to use?

---

---

---

---

---

13. Do you have any additional comments about your experience using EnACT?

---

---

---

---

## Appendix E: Training document

### Usability Study – EnACT Software

#### Goal & Methodology

The main goal of this usability study is to obtain initial feedback for EnACT. This study will consist of three cases where each case will have a different task varying in difficulty level.

#### Set of training tasks

- To create a new project click on File > New. The new project form window will become visible.
- Load the script file named: script\_training.rtf
- Load the movie file named: training\_video.avi
- Name the new project “Training\_*participantName*”.
- Click Create.
- Once the project is created you will be able to mark up words.
  - Change some words to a high, medium and low intensity with the emotions angry, happy, sad or fear.
    - **Note:** You can also right-click on the word to choose the emotion and intensity. This will also allow editing to any word selected.
  - Give a “begin” and “end” time to each dialogue.
    - **Note:** To make the process easier, drag across the progress bar in the movie and then click on the button next to the text field “begin time”. The time where your caption will appear in the video will become visible. Ensure the “end time” of your caption does not overlap with the “begin time” of the previous caption.

- Use the preview button to observe the enhanced captions.
  - **Note:** *EnACT will convert the original video into a flash file upon the first edit of the script for the video. Once you are ready to preview the marked up script, press “show preview” one more time to see your Enhanced captioned video.*
- Click on File > Save to save the project.
- Click on File > Close or close the window to exit the program.
- Run the software again.
- Click on File > Open and open the saved project called “Training\_participantName”.
- Click on View > Options and change the colour default settings of two or more emotions, the font type and size of the text.
- Mark up more words:
  - Change a word to a medium intensity with happy emotion.
  - Change a word to a low intensity with a angry or sad emotion.
- Save and Exit the program.

## Appendix F: Study Tasks

### Usability task 1

This case study requires you to load a movie script and mark up some of its words.

1. Create a new project.
2. Load the script rtf file “case\_study\_script”. **Note:** Do *not* load the movie
3. Create a new project called: <your\_name>\_case1.
4. Once the script is loaded and is visible, select five random words and mark them up with the emotion and an intensity that you believe they should have.
  - a. You should have at least 1 of each emotions (angry, sad, happy and fear).
  - b. You should have at least 1 of each intensity (high, medium, low).
5. Click on Save.
6. Exit the program.

### Usability task 2

This case study requires you to load the movie script and corresponding video file. You will be asked to mark up the script and edit the length of time for each captioning effect to show in the preview window.

1. Create a new project.
2. Load the script rtf file “case\_study\_script”.
3. Load the 45 seconds long video file “case\_study\_video”.
4. Name the project “<name>\_case2”
5. Watch the clip **once**.
6. Customize the default colour for the emotion “happy” to red and “anger” to yellow.

7. Mark up one or more words from the first ten lines of dialogue, using all of the emotions and intensities at least once.
8. For each of the words marked up, give the captioning effects a “begin” and “end” time.
9. Use the preview button to view your enhanced captions.
10. Save your progress.
11. Exit the program.

### **Usability task 3**

This case study requires you to load the video file and make changes to an existing project. You will be asked to make changes to the emotions and the length of time for each captioning effect.

1. Open the last project.
2. Make at least three changes to the emotions you previously marked up.
3. Adjust the “begin” and “end” time of each corresponding captioning effect.
4. Increase the font of the text.
5. Save your progress.
6. Exit the program.



## **Appendix G: Recruitment emails & Posters**

### **Captions haven't changed since 1970 – See what a difference Ryerson can make with your help.**

At the Ryerson Centre for Learning Technologies, we believe that captioning could better serve the deaf and hard of hearing. We are contacting you as we believe your captioning expertise is important to ensuring that our work fits in the expert captioning community.

Our research team has developed a captioning software tool, called EnACT, which allows captionists to create animated captions. We are running a study to gain initial feedback for this software tool and find new ways to improve viewer's experience in captioning. We have been exploring the use of enhanced and animated captions for the past four years and as a result have developed an animated caption tool called EnAct. This software allows captionists to tag text scripts with one of four different emotion types and intensity. The software then processes those tags into animations within the captions. We would like to know whether this tool is easy to learn and use before continuing our development work. Your feedback will be invaluable.

As part of the study, you will learn how to use EnACT. You will be asked to create enhanced captions for a video clip. To do this, you will use EnACT to indicate the emotions and their intensities on a text script using the mark up functions of EnACT. You will also be asked to make adjustments to captions, such as changing font styles and the visual tags attached to words. We will ask you to fill our short questionnaire, after the study. The study will take no longer than one hour of your time.

If you are interested in participating, please contact Jorge Mori at [jmori@ryerson.ca](mailto:jmori@ryerson.ca) to arrange an appointment that is convenient for you. Also, we can send you a consent form to participate, a formal description of the study, and a pre-meeting questionnaire ahead of time. As a thank-you for your participation, we will provide you with \$15 upon completion of the study.

The location will be at The Centre for Learning Technologies at Ryerson University

Location: 55 Dundas St. W. 9<sup>th</sup> floor, room 3-174.

Directions:

55 Dundas St. W is on the south west side of the Yonge and Dundas intersection. We are one building going west on the south side of Dundas Street (in the same building as the Canadian Tire and Best Buy on Dundas).

We will audio record the session. However, the audi006F will be used as a memory aid for the researchers only, and individuals will not be identified.

Jorge Mori

Ryerson University

350 Victoria St.

Toronto, Ont.

M5B 2K3

416.979.5000 ext. 2523

[jmori@ryerson.ca](mailto:jmori@ryerson.ca)

## Appendix H: Payment Receipts

This document acknowledges the participant of receiving \$15 for the feedback provided while being part of the Enhanced Captioning software study using EnACT (Emotive and Effective Captioning Tool) under the supervision of the main researcher Jorge Mori.

---

---

Date

Participant

---

Researcher

## Appendix I: Problems with EnACT 2.0 and Solutions implemented in EnACT 3.0

### Problem 1: Dialogues do not load properly in SEA

#### EnACT version 2.0 code snippet

```
private ArrayList parseDialogue () {
    ArrayList emotions = new ArrayList (10);

    emotion _emotion;

    while (reader.Read ()) {
        if (reader.NodeType != XmlNodeType.Element)
            continue;

        if (reader.Name.ToLower () != "emotion")
            break;

        _emotion = new emotion ();

        // Emotion
        try {
            _emotion.type = (Emotion) int.Parse
                (reader.GetAttribute ("type"));
            _emotion.intensity = (Intensity) int.Parse
                (reader.GetAttribute("intensity"));
        }
        catch (Exception) {
            _emotion.type = Emotion.None;
            _emotion.intensity = Intensity.None;
        }

        // Text
        try {
            _emotion.text = reader.ReadString ();
        }
        catch (Exception) {
            continue;
        }

        emotions.Add (_emotion);
    }

    emotions.TrimToSize ();
    return emotions;
}
```

#### EnACT version 3.0 code snippet

```
private ArrayList parseDialogue () {
    ArrayList emotions = new ArrayList (10);

    emotion _emotion;
```

```

while (reader.Read ()) {
    //if (reader.NodeType != XmlNodeType.Element)
    //    continue;

    if (reader.Name.ToLower() != "emotion")
        break;

    _emotion = new emotion ();

    // Emotion
    try {
        _emotion.type = (Emotion) int.Parse (reader.GetAttribute
            ("type"));
        _emotion.intensity = (Intensity) int.Parse
            (reader.GetAttribute ("intensity"));
    }
    catch (Exception) {
        _emotion.type = Emotion.None;
        _emotion.intensity = Intensity.None;
    }

    // Text
    try {
        _emotion.text = reader.ReadString ();
    }
    catch (Exception) {
        continue;
    }

    emotions.Add (_emotion);
}

emotions.TrimToSize ();

return emotions;
}

```

## Problem 2: Save button recording value of -1

### Dialogues.xml file created by EnACT 2.0 when saved

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE captions SYSTEM "../captions.dtd">
<captions>
  <caption begin="00:00:00" end="00:00:00" speaker="CARLO" location="2"
align="1">
    <emotion type="0" intensity="0">She's</emotion>
    <emotion type="2" intensity="2">going</emotion>
    <emotion type="0" intensity="0">to</emotion>
    <emotion type="0" intensity="0">be</emotion>
    <emotion type="0" intensity="0">okay?</emotion>
  </caption>
  <caption begin="00:00:00" end="00:00:00" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="-1" intensity="0">Carlo,</emotion>
    <emotion type="-1" intensity="0">blow</emotion>
    <emotion type="-1" intensity="0">into</emotion>
    <emotion type="0" intensity="0">the</emotion>
    <emotion type="0" intensity="0">tube</emotion>
  </caption>
  <caption begin="00:00:00" end="00:00:00" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="-1" intensity="0">Not</emotion>
    <emotion type="0" intensity="0">too</emotion>
    <emotion type="-1" intensity="0">hard</emotion>
  </caption>
  <caption begin="00:00:25.3" end="00:00:20.5" speaker="RACHEL (CONT'D)"
location="2" align="1">
    <emotion type="-1" intensity="0">Her</emotion>
    <emotion type="0" intensity="0">breathing's</emotion>
    <emotion type="-1" intensity="0">back</emotion>
  </caption>
</captions>
```

## Code EnACT 2.0

```
private void WriteDialogues (string path) {
    rtfScript.Visible = false; //rtfScript.UseWaitCursor = true;

    // Save Current Selection
    int SELECTION_START = rtfScript.SelectionStart;
    int SELECTION_LENGTH = rtfScript.SelectionLength;

    int length = rtfScript.Lines.Length;

    setProgressBar (0, length);

    bDisableEnACTFunctions = true;

    caption _caption = new caption ();

    int _start = 0;
    int _length;
    string[] _words;

    emotion _emotion = new emotion ();

    for (int i = 0; i < length; i++) {
        ProgressBar.PerformStep ();

        switch (i % 2) {
            case 0: // Speaker
                _start += rtfScript.Lines[i].Length;

                _caption = captionsXML.getCaption(i / 2);

                if (_caption.bDirty)
                    _caption.emotions.Clear (); // Reset Captions
                else
                    // Skip Captions
                    _start += rtfScript.Lines[++i].Length;
                break;
            case 1: // Captions
                _words = rtfScript.Lines[i].Split (' ');

                for (int j = 0; j < _words.Length; j++) {
                    _length = (j < _words.Length - 1) ?
                        _words[j].Length + 1 : _words[j].Length;
                    rtfScript.Select (_start, _length);

                    _emotion.text = _words[j];
                    _emotion.type =
                        getEmotionType(rtfScript.SelectionColor);
                    _emotion.intensity = getEmotionIntensity
                        (rtfScript.SelectionFont);

                    _caption.emotions.Add (_emotion);

                    _start += _length;
                }
            }
        }
    }
```

```

        _caption.emotions.TrimToSize ();
        break;
    }

    _start++; // Skip New Line
}

captionsXML.writeXML (FileHelper.GetFullPath (path, "dialogues.xml"));

bDisableEnACTFunctions = false;
bProjectDirty = false;

setProgressBar (0, 0);

// Restore Selection State
rtfScript.Select (SELECTION_START, SELECTION_LENGTH);

rtfScript.Visible = true; //rtfScript.UseWaitCursor = false;
}

```



## Code EnACT 3.0

```
private void WriteDialogues (string path) {
    rtfScript.Visible = false;

    // Save Current Selection
    int SELECTION_START = rtfScript.SelectionStart;
    int SELECTION_LENGTH = rtfScript.SelectionLength;

    // The number of lines in the richtextbox
    int length = rtfScript.Lines.Length;
    setProgressBar (0, length);

    bDisableEnACTFunctions = true;

    caption _caption = new caption ();

    int _start = 0;
    int _length;
    string[] _words;
    int mod;

    emotion _emotion = new emotion ();

    for (int i = 0; i < length; i++) {
        ProgressBar.PerformStep ();

        mod = i % 2;

        switch (mod)
        {
            case 0: // Speaker
                _caption = captionsXML.getCaption(i / 2);
                _caption.emotions.Clear(); // Reset Captions
                _start += rtfScript.Lines[i].Length;
                break;

            case 1: // Captions
                _words = rtfScript.Lines[i].Split(' ');

                for (int j = 0; j < _words.Length; j++)
                {
                    _length = (j < _words.Length - 1) ?
                        _words[j].Length + 1 : _words[j].Length;
                    rtfScript.Select(_start, _length);

                    _emotion.text = _words[j];
                    _emotion.type =
                        getEmotionType(rtfScript.SelectionColor);
                    _emotion.intensity =
                        getEmotionIntensity(rtfScript.SelectionFo
nt);

                    //Bug fix
                    if (_emotion.type == Emotion.Unknown)
                        _emotion.type = Emotion.None;
                }
            }
        }
    }
```

```

        _caption.emotions.Add(_emotion);

        _start += _length;
    }

    _caption.emotions.TrimToSize();
    break;
}

_start++; // Skip New Line
}

captionsXML.writeXML (FileHelper.getFullPath (path, "dialogues.xml"));

bDisableEnACTFunctions = false;
bProjectDirty = false;

setProgressBar (0, 0);

// Restore Selection State
rtfScript.Select(SELECTION_START, SELECTION_LENGTH);

rtfScript.Visible = true;
}

```

## **Appendix J: Computer Specifications:**

### Laptop 1:

- System Model: HP Pavilion Dv6000
- Operating System: Windows Vista Business Service Pack 2
- System type: X86-Based PC
- Memory: 2Gb RAM
- Processor: Intel(R) CPU T2250 @ 1.73GHz (2 CPUs)
- Storage: 120 HDD
- Graphics: Intel(R) GMA 950

### Laptop 2:

- Acer Aspire 7741G
- 4Gb DDR3 Memory
- 620GB HDD
- ATI Mobility Radeon HD

## **Appendix K: Consent Form**

**Project Title:** EnACT Usability Study

**Principal Investigators:** Jorge Mori, BSc. Ryerson University,

[jmori@ryerson.ca](mailto:jmori@ryerson.ca)

Deborah Fels, P.Eng., Ph.D., Ryerson University

[dfels@ryerson.ca](mailto:dfels@ryerson.ca)

### **Consent to Participate in Study from Subject**

#### **Information Form**

The purpose of this study is to obtain feedback for EnACT, a software tool used for creating animated captions. The result and data obtained from this study will be used in my thesis project, as it is part of my graduate program requirement. In order to do this, you will be provided with a short introduction on how the tool works and given about 5 minutes to practice with it or until you are comfortable with the EnACT interface. We then will ask you to create enhanced captions using the script and video for three different video clips using EnACT. You will do this by watching the clip, deciding which emotions and intensity the actors are trying to convey and assigning those emotions to words or phrases in the script (emotion tags). You will also be asked to make adjustments to the captions by changing font styles and the appearance of the emotion tags attached to words.

It will take you about 60 minutes to finish the training and the three video clips. During the study you will be asked to talk out loud your thoughts on what you are doing. A screen recording program will record your voice and the computer screen. A researcher will be taking notes during your session on concerns or comments you may have as well as to record technical issues if they occur. You will also be asked to complete a questionnaire at the conclusion of the study. The first part of the questionnaire will contain demographics questions that will help the research team to classify the data obtained into correct data sets. The second part will contain questions to obtain feedback for EnACT.

#### **Confidentiality**

All raw data will be kept strictly confidential and kept in a locked cupboard or password protected server in the Centre for Learning Technologies at Ryerson University. However a summary of the data will be published in academic venues but no individual details will be identified in this summary. The information gathered from surveys will be strictly used for research and academic purposes with only the principal investigators having access to it. The database records will be stored for five years and then deleted from the server.

## **Risks and Discomforts**

The risks associated with participating in this study are minimal. You may experience some fatigue or frustration while creating the enhanced captions with the tool or from answering the questionnaires. However, you are able to take breaks at any time or stop participation in the study without penalty. You may also experience some discomfort with having your screen or voice recorded. In this case, you may choose not to participate in the study or alternatively you can record your opinions in writing and remain “off camera.”

## **Expected Benefits**

Individual participants will not receive any direct benefits; however, this study will benefit the general community of caption users. This study will test the user-friendliness of the Emotive and Affective Captioning tool (EnACT). We hope that this information may lead to improvements in closed captioning technologies and techniques. You will receive \$15 for your transportation costs and time.

## **Feedback**

A copy of any publications that arise from this research will be available to all members of the public through the Ryerson’s online publication system at <http://digitalcommons.ryerson.ca>.

## **Voluntary Nature of Participation:**

Participation in this study is entirely voluntary. If you do not wish to participate in this study it will not affect any current or future relations with Ryerson University or The Centre for Learning Technologies. If you choose to participate, you can stop the study at any time and for any reason without penalty. In addition, you may refuse to answer any questions or participate in any task at any point of the study without penalty.

## **Location of study**

The study will take place at Ryerson University in a usability room that is setup with the eye tracking and video recording equipment in the Ted Rogers School of Management building located at 55 Dundas St. W. You will be given the room number when the time of your participation is established.

## **Questions about the Study**

If you have any questions or concerns, about this study please feel free to contact Jorge Mori at [jmori@ryerson.ca](mailto:jmori@ryerson.ca) or Deborah Fels at [dfels@ryerson.ca](mailto:dfels@ryerson.ca). If you have any concerns or complaints about this study in regards to its ethical nature please contact the Research Ethics Board, c/o Office of the Vice President, Research and Innovation, Ryerson University, 350 Victoria St., Toronto, ON M5B 2K3, Tel: 416-979-5042

**Project Title:** EnACT Usability Study

**Principal Investigators:** Jorge Mori, BSc. Ryerson University,  
[jmori@ryerson.ca](mailto:jmori@ryerson.ca)  
Deborah Fels, P.Eng., Ph.D., Ryerson University  
[dfels@ryerson.ca](mailto:dfels@ryerson.ca)

### Consent Form to Participate in Study

I acknowledge that the research procedures described above have been explained to me and that any questions that I have asked have been answered to my satisfaction. I have been informed of that there may be a possible risk of psychological discomfort from having my screen and voice recorded or using the Emotive and Affective Captioning Tool, however, strategies are in place to reduce this risk.

I have been informed of the alternatives to participation in this study, including my right not to participate and the right to withdraw without penalty. I hereby consent to participate in the study and to be screen, video or audio recorded during the study. I have received a copy of the information sheet.

Signature of Participant: \_\_\_\_\_

Name of Participant (please print): \_\_\_\_\_

Date: \_\_\_\_\_

Agreed to be videotaped ☐ Agreed ☐ Disagreed

The details of this study were explained to me by:

Name of Investigator: \_\_\_\_\_

Date: \_\_\_\_\_

## Appendix L: Participants Scripts

Introduction to EnACT script used in the software:

FADE IN:

ACT ONE

EXT. BUSH -- DAY

Six BOYS carry two wounded FRIENDS on stretchers. A large helicopter passes low overhead. They pound over the ground, running for all they're worth.

EXT. AERIAL POV -- DAY

A series of shots as the plane flies over the African terrain.

INT. HELICOPTER -- DAY

The interior is noisy and jammed with cargo -- skids of boxes, oil drums. RACHEL, 24, is strapped into a jump seat. She's a nervous flier.

EXT. AERIAL POV -- DAY

RACHEL'S POV: A cluster of buildings on the ground. The corrugated roof of the largest building is emblazoned with the red and white ALLWORLD MEDICINE logo.

EXT. LANDING FIELD -- DAY

ROLLIE, JENNA and CARLO head for the strip, a group of PORTERS accompany them carrying VILDA on a stretcher. KIM, 14, pads alongside.

JENNA

"You sure everything's safe there Jenna." The guys in Loki are starting to sound like my mother.

ROLLIE

We had thirty percent of  
our supplies stolen last  
month. Maybe that's it.

VILDA

My fault. They don't  
like it when we get sick.

JENNA

No, it's not your fault.  
Look, I'm not saying that  
it doesn't stink, okay?

CARLO

Fish stink from the head.  
Is Mabor and Nok is make  
the trouble. I hear they  
have a fight yesterday.

JENNA

Maybe Mabor was telling  
Nok to stop stealing from  
us.

Jenna shoots Kim a look, is he listening? He's  
listening hard.

JENNA (CONT'D)

How about it, Kim. Is  
your brother on our side?

KIM

He wants to help.

ROLLIE

I don't know if he can  
cut it. He doesn't know  
the territory. Didn't he  
spend all last year  
hanging around London?



User tasks script used in the software:

FADE IN:

ACT ONE

CARLO

She's going to be okay?

RACHEL

Yeah, she should be. We  
just have to make sure.

RACHEL (CONT'D)

Her heart has stopped.  
Carlo, blow into the  
tube.

RACHEL (CONT'D)

That's it. Go on. Every  
five seconds.

RACHEL (CONT'D)

Not too hard.

CARLO

How long do we do this?

RACHEL

As long as it takes.

RACHEL (CONT'D)

Heartbeat!

RACHEL (CONT'D)

Her breathing's back.

CARLO

You save a life!

## References

- [1] Canadian Association of the Deaf. (2007) "Statistics on deaf Canadians," Retrieved on Oct 15, 2011 from [http://www.cad.ca/statistics\\_on\\_deaf\\_canadians.php](http://www.cad.ca/statistics_on_deaf_canadians.php).
- [2] R. Mitchell, "How many deaf people are there in the United States? Estimates from the Survey of Income and Program Participation," vol. 11, pp. 112-119, 2006.
- [3] D. I. Fels, C. Branje, D. G. Lee and M. Hornburg, "Emotive Captioning and access to Television", AMCIS 2005, 2330–2337.
- [4] S. Abrahamian. "EIA-608 and EIA-708 Closed Captioning". Last updated 2003, Accessed Jun 6<sup>th</sup>, 2011, pp. 4. 2003.
- [5] C. Silverman and D. I. Fels, "Emotive captioning in a digital world," in *Proceedings of the 8th International Conference on Computers Helping People with Special Needs*, 2002, pp. 292-294.
- [6] K. Bodine and M. Pignol, "Kinetic typography-based instant messaging," in *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, Ft. Lauderdale, Florida, USA, 2003, pp. 914-915.
- [7] R. Rashid, Q. Vy, R. Hunt and D. I. Fels, "Dancing with Words: Using Animated Text for Captioning," *Int. J. Hum. -Comput. Interact.*, vol. 24, pp. 505-519, 06, 2008.
- [8] Udo, J.-P., Fels, D.I., "Universal design on stage: Live audio description for theatrical performances," *Perspectives: Studies in Translatology*, vol. 18, pp. 189-203, 2010.
- [9] C. Stephanidis, "Adaptive Techniques for Universal Access," *User Modeling and User-Adapted Interaction*, vol. 11, pp. 159-179, March, 2001.
- [10] Q. Vy and D. Fels, "Using Placement and Name for Speaker Identification in Captioning," vol. 6179, pp. 247-254, 2010.
- [11] D. Sillman, "Line 21: Closed Captioning of Television Programs— A Progress Report: A Paper Presented at the 1978 Symposium on Research and Utilization of Educational Media for Teaching the Deaf," 2008.
- [12] H. H. a. D. E. Hsin-Chuan, "The Effects of Closed-Captioned Television on the Listening Comprehension of Intermediate English as a Second Language (ESL) Students," *J. Educ. Technol. Syst.*, vol. 28, pp. 75-96, 1999.
- [13] CRTC, "Broadcasting Public Notice CRTC 2007-54," (2007) retrieved on Jan 13<sup>th</sup>, 2012 from <http://www.crtc.gc.ca/eng/archive/2007/pb2007-54.htm>.

- [14] CRTC, "Broadcasting Notice of Consultation CRTC 2011-488," (15 August 2011) retrieved on Jan 13<sup>th</sup>, 2012 from <http://crtc.gc.ca/eng/archive/2011/2011-488.htm>.
- [15] CRTC, "Broadcasting and Telecom Regulatory Policy CRTC 2009-430," (2009) retrieved on Jan 13<sup>th</sup>, 2012 from <http://www.crtc.gc.ca/eng/archive/2009/2009-430.htm>.
- [16] CAB, "Follow up to broadcasting and telecom regulatory policy CRTC 2009-430, accessibility of telecommunications and broadcasting services (the accessibility policy) – working groups on quality of closed captioning," Canadian Association of Broadcasters, 2010.
- [17] Consumer Electronics Association, "CEA-708-D: Digital television (DTV) closed captioning," Consumer Electronics Association, 2008.
- [18] CPC closed captioning and Subtitling, "True 708 versus 608 captions," .
- [19] G. O. Crowther, "Adaptation of U.K. Teletext System for 525/60 Operation," *Consumer Electronics, IEEE Transactions on Adaptation of U.K. Teletext System for 525/60 Operation*, vol. CE-26, pp. 587, Aug., 1980.
- [20] K. Su and Y. Peng, "A method for teletext display" in *Computer Graphics, Imaging and Visualisation, 2006 International Conference on Computer Graphics*, 2006, pp. 231.
- [21] CAB, "Closed captioning standards and protocol for canadian english language television programming services," Canadian Association of Broadcasters, (2008), retrieved on Nov 15<sup>th</sup>, 2011 from <http://www.cab-acr.ca/english/social/captioning/captioning.pdf>.
- [22] A. B. Jordan, A. Albright, A. Branner and J. Sullivan, "The state of closed captioning services in the United States" pp. 1-47, 2003.
- [23] C. Silverman and D. Fels. Beyond captioning: The next frontier. in *Center On Disabilities Technology And Persons With Disabilities Conference 2001*, 2001 retrieved on Sept 14<sup>th</sup>, 2011 from <http://www.csun.edu/cod/conf/2001/proceedings/0217silverman.htm>.
- [24] A. Mehrabian, "Communication without words," in *Psychology Today* pp. 53-56, 1968.
- [25] J. Forlizzi, J. Lee and S. E. Hudson, "The kinedit system: Affective messages using dynamic texts," in *Proceedings of CHI 2003, (Ft Lauderdale, April, 2003)*, ACM, 2003, pp. 377-384.
- [26] D. Geffner, "First things first" (1997) in *Filmmaker Magazine* retrieved on Dec 15<sup>th</sup>, 2011 from <http://www.filmmakermagazine.com/issues/fall1997/firstthingsfirst.php> .
- [27] H. Wang, H. Prendinger and T. Igarashi, "Communicating emotions in online chat using physiological sensors and animated text," in *CHI '04: CHI '04 Extended Abstracts on Human Factors in Computing Systems*, Vienna, Austria, 2004, pp. 1171-1174.

- [28] C. Conati, R. Chabbal and H. Maclaren, "A Study on Using Biometric Sensors for Monitoring User Emotions in Educational Games," 2003.
- [29] E. Mower, Sungbok Lee, M. J. Mataric and S. Narayanan, "Joint-processing of audio-visual signals in human perception of conflicting synthetic character emotions," in *Multimedia and Expo, 2008 IEEE International Conference on*, 2008, pp. 961.
- [30] P. Ekman, "Basic Emotions," in *John Wiley & sons Ltd*, pp. 45-60, 2005.
- [31] D. W. Fourney and D. I. Fels, "Creating access to music through visualization," in *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, 2009, pp. 939.
- [32] T. Rose, *Black Noise: Rap Music and Black Culture in Contemporary America*. University Press of New England, 1994.
- [33] Knight, W.E.J., Rickard, N.S., "Relaxing Music Prevents Stress-Induced Increases in Subjective Anxiety, Systolic Blood Pressure, and Heart Rate in Healthy Males and Females," *J. Music Ther.*, vol. 38, pp. 254-272, 2001.
- [34] H. Kohut and S. Levarie, "On the enjoyment of listening to music", in *International Universities Press*, pp. 1-20, 1990.
- [35] Fourney, D., Fels, D., "Thanks for pointing that out. Making sarcasm accessible for all," in *Proceeding of the Human Factors and Ergonomics Society*, 2008, pp. 571-575.
- [36] J. Mori and D. I. Fels, "Seeing the music can animated lyrics provide access to the emotional content in music for people who are deaf or hard of hearing?" in *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, 2009, pp. 951-956.
- [37] T. Jokela, N. Iivari, J. Matero and M. Karukka, "The standard of user-centered design and the standard definition of usability: Analyzing ISO 13407 against ISO 9241-11," in *Proceedings of the Latin American Conference on Human-Computer Interaction*, Rio de Janeiro, Brazil, 2003, pp. 53-60.
- [38] S. Suh and T. Trabasso, "Inferences during reading: Converging Evidence from Discourse Analysis, Talk-Aloud Protocols and Recognition priming " in *Journal of Memory and Language*, vol. 32, pp. 279-300, 1993.
- [39] camstudio, "Camstudio Open source Free streaming video software," retrieved on August 14<sup>th</sup>, 2011 from <http://camstudio.org/>.
- [40] Adobe, "After Effects CS5.5 / Features", (2011), retrieved on May 25<sup>th</sup>, 2011, from <http://www.adobe.com/products/aftereffects/features.html>.

- [41] Apple INC, "LiveType 2 user manual." (2005), retrieved on Jun 2th, 2011 from [http://manuals.info.apple.com/en/livetype\\_2\\_user\\_manual.pdf](http://manuals.info.apple.com/en/livetype_2_user_manual.pdf).
- [42] Q. V. Vy, J. A. Mori, D. W. Fourney and D. I. Fels, "EnACT: A software tool for creating animated text captions," in *Proceedings of the 11th International Conference on Computers Helping People with Special Needs*, linz, Austria, 2008, pp. 609-616.
- [43] S. Reges, "Can C\# replace java in CS1 and CS2?" in *SIGCSE Bull*, vol. 34, pp. 4-8, June, 2002.
- [44] S. Reimers and N. Stewart, "Adobe Flash as a medium for online experimentation: A test of reaction time measurement capabilities", in *Behavior Research Methods*, vol. 39, pp. 365-370, 2007.
- [45] Adobe, "Flash Enabled Mobile Devices," (2011) retrieved on May 21st, 2011, from [http://www.adobe.com/flashplatform/certified\\_devices/](http://www.adobe.com/flashplatform/certified_devices/).
- [46] S. S. Chawathe, "Describing and Manipulating XML Data," in *IEEE Data(Base) Engineering Bulletin*, vol. 22, pp. 3-9, 1999.
- [47] J. A. Walker and S. Chaplin, "Visual Culture: An Introduction", in Manchester University Press, 1997.
- [48] S. Tomar, "Converting video formats with FFmpeg," in *Linux J.*, vol. 2006, pp. 10, Jun, 2006.
- [49] J. Jacoby and M. S. Matell, "Three point Likert scales are good enough," *J. Market. Res.*, vol. 8, pp. 495-500, 1971.
- [50] P. A. Todd and I. Benbasat, "The influence of decision aids on choice strategies under conditions of high cognitive load," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 24, pp. 537, apr, 1994.
- [51] D. Sharma and A. Gruchacz, "The Display Text Editor TED: A Case Study in the Design and Implementation of Display-Oriented Interactive Human Interfaces," *Communications, IEEE Transactions on*, vol. 30, pp. 111, jan, 1982.
- [52] R. Rashid, J. Aitken and D. I. Fels, "Expressing emotions using animated text captions," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4061 LNCS, pp. 24-31, 2006.
- [53] G. S. Acton, "Basic Emotions" (1998) Retrieved Oct 2<sup>nd</sup>, 2011, from <http://www.personalityresearch.org/basicemotions.html>.

- [54] Shamay-Tsoory, S.G., Tomer, R., Aharon-Peretz, J., "The neuroanatomical basis of understanding sarcasm and its relationship to social cognition," *Neuropsychology*, vol. 19, pp. 288-300, 2005.
- [55] P. Rockwell, "Lower, Slower, Louder: Vocal Cues of Sarcasm," *J. Psycholinguist. Res.*, vol. 29, pp. 483-495, 2000.
- [56] Balsamiq Studios, "Balsamiq Mockups", retrieved on Sept 5<sup>th</sup>, 2011 from <http://www.balsamiq.com/products/mockups>.
- [57] NASA Ames Research Center, "NASA TLX," retrieved on Dec 15<sup>th</sup>, 2011 from <http://humansystems.arc.nasa.gov/groups/TLX/index.html> .