Visual Reaction to Subtitling in Television and Media

Ross Bryant
Abross@clemson.edu

Jill Constante
Constaj@clemson.edu

Mike Manna
Mannam@clemson.edu

Tony Serechia
Serecha@clemson.edu

Tawanna Starks
Stawann@clemson.edu

ABSTRACT

One of the greater influences in technology today is subtitling, the ability to provide verbatim textual translation of the spoken language in a television program or film [1]. By offering subtitling in media today, physically impaired individuals or those struggling with language barriers are given the opportunity to enjoy the same standard entertainment as others. Subtitling is generally enabled in an attempt to aid an individual's comprehension of media. An interesting question arises: Are subtitles a beneficial tool for educating or can they be distracting individuals from focusing on the visual aspects of a television show or movie? Using the latest eye tracking technology, it is possible to detect and record eye fixation points. An eye tracker is an input device that delivers real time coordinates of individual's eye movements [4]. We are experimenting to determine whether subjects spend more time reading subtitles (text) as opposed to concentrating on the visual images of a video. The specific usage of an eye tracker for this experiment is to provide data on participants when exposed to video clips with and without subtitling present.

Keywords: Eye tracking, subtitle, movie, audio

Introduction

Eye movement studies have been increasing over the years and continue growing as analysis helps researchers gain insight on human behavior [2]. Eye movement results are used and applied regularly. We have chosen to research eye behavior involving video images containing textual information combined with audio because there is little or no research in this area especially involving actual real-time eye tracking equipment.

1.1 Subtitling

Subtitling was first introduced to the public in 1947 by Emerson Romero, a deaf man and former actor. He was able to splice text between frames like the old silent movies. There was a segment of action as well as a segment of text. There was no synchronization between the action and the text segments. Romero's efforts were appreciated by the deaf community and two years later J. Arthur Rank, a movie mogul in the 40's and 50's, provided a captioned feature-length film at a movie house in London. These subtitles were etched onto pieces of glass and required a well trained operator to swap the glass frames as the movie played. The experiment was never reproduced, however the deaf community again showed great interest in the emergence of subtitling. Around the same time, a company in Belgium devised a new method of captioning. This method involved printing captions directly onto a master copy of film and using this to make open-captioned reproductions for distribution. This method is similar to the method used in open captioned movies today [5].

1.3 Objective

We have designed an experiment to obtain coordinates of human eye fixations when exposed to animated videos. We compare results of participants where videos contain textual representations to those without textual translation. Audio will be enabled in both videos. When movies or films are subtitled, the textual translation is typically displayed to all audiences whether or not they require it. With the newest generation of impressive movies out today like the "The Matrix", "Shark Tales," "Lord of the Rings", etc., we are interested in determining if individuals focus attention to subtitles even when they are fully capable of enjoying the movie without them. Normally, most television and media provide an option to turn subtitling on or off. An example scenario could be: a deaf individual wishes to enjoy a movie with subtitling turned on. Assuming the impaired individual watches a movie with his/her (non-deaf) friend, we will use our experimental results to explain whether or not the non-deaf individual is potentially distracted by the presence of subtitles or if they are ignored altogether since they are unnecessary for successful understanding of the film. Our experiment was devised to determine if the majority of subjects' eye movements follow the subtitles in media when audio is available. Based on the experimental results we obtain we can conclude

and generalize the effects on other similar events. One popular interest is advertising. Exactly how much time do individuals spend reading fine print on television advertisements compared to more obvious visual stimulants? This experiment could also be helpful in determining if humans notice banners at the bottom of television concerning weather updates or other informative reports. Ticker reports of athletic scores might be distracting, on the other hand sports fan may wish to read them.

Our ultimate goal is to show a short scene from a movie and establish a formal conclusion on whether the average person reads subtitles even if they are needless since participants have audio available. From the experiment we can conclude whether subtitling is a distraction to individuals when not used as a reference. If subject's eyes do not show significant fixations on subtitles we can generalize the results to other examples suggesting individuals only read athletic event scores scrolling across the bottom of the television screen if they are truly interested opposing the idea it is an unavoidable distraction. We expect eye fixations to significantly vary between the viewing of the two videos. Furthermore, we believe both male and female subjects will considerably focus their eye movements toward the subtitled portion of the video and expect eye fixations to centralize around the upper portion of the video where animation is located when subtitling is disabled.

2. Methodology

The goal of the experiment is to determine if subtitles have significant affects on human eye movement. The movie presented was a small scene from the animated film "Finding Nemo" [3]. Having a constant of audio and placing emphasis on the visual aspects helps in determining if subtitles are a distraction or enhancement of an understanding of a movie.

2.1 Apparatus

The major equipment used includes a Tobii Eye Tracker shown in Figure 1, one Linux PC, and one Windows PC. The Tobii Eye Tracker is a monitoring device consisting of a camera and infrared LED which detects eye movements and provides coordinates of human eye movement patterns. Currently it is the most accurate eye tracker available [4]. Average accuracy over a set of individuals has been tested to 0.5 degrees using standard accuracy measurement principles for eye-trackers. It is capable of producing a resolution up to 1280 X 1024 pixels. In normal conditions the Tobii can produce a frame rate up to 50 Hz [4]. The Linux PC hosted a program written in C++ specifically designed for our experiment. The program divided an MPEG video stream into PPM image files. These

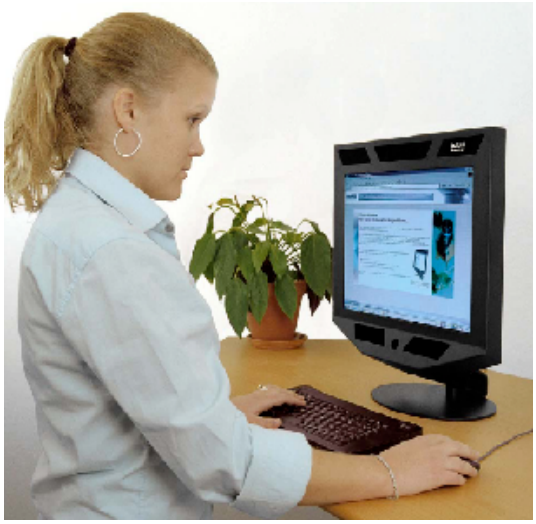

Figure 1: Tobii Eye Tracking station [4]

image files were synchronized with the audio which was extracted using 'ffmpeg'. The GUI was designed using GLUT, while OPENGL handled the PPM images and used SDL to handle the audio. The program was also designed to manage the data collection. The Windows PC was responsible for the physical collection of data from the Tobii Eye Tracker, and forwarded all recorded data points to the Linux machine. Once the Linux machine received input, our program then saved the data allowing us to perform our experimental analysis generating useful statistics.

2.2 Stimulus

Stimulus included two versions of the same animated video. We showed two sample videos both in English language with English subtitling. Figures 2 and 3 show sample screenshots of each video subject groups viewed. One video contained subtitles (shown in Figure 2) while they were eliminated in the other clip (shown in Figure 3).

Figure 2: Video image with subtitling

Figure 3: Video image without subtitling

2.3 Questionnaire

A 13 part questionnaire was issued following the experiment to achieve maximum experimentation. The questionnaire results were used for two reasons. 1) Provide background information to help with the diagnosis of each subject individually. It provided information including age, vision ability/accuracy, and if the participant had seen the video before. These results were useful for helping us accurately examine final results so we could address any possible biases or limitations. 2) Results of the questionnaire were compared to results of the subject's eye fixation points confirming their tendencies. The questionnaire evaluated our subject's recollection on the contents of the video displayed. Results showed whether subjects recalled more information portrayed through textual information or by graphical representation. The questionnaire included eight questions based solely on graphical representation and three questions pertaining to dialogue of the movie. Two remaining questions were for informational purposes. Figure 4 below shows a subject completing his questionnaire.

Figure 4: Subject completing questionnaire

2.4 Subjects

Subjects consist of twelve volunteer college aged students, six males and six females, to prevent any age or sex discrepancies. We experimented on twenty total subjects, but discarded results of eight subjects due to error in the data their eye movements produced. Students were selected from the undergraduate population at Clemson University who met specific requirements (1) Must be between 18 – 24 years of age (2) Must speak/comprehend English fluently (3) Must not have severe hearing or visual impairments (4) Must not have prior knowledge of this experiment. Each student was randomly assigned to one of two condition groups. It is very possible males and females would have different results. We compared the results separately to determine if there exists a difference between sexes. Subjects were divided into two groups. One group saw the video with subtitles and the other saw the video minus the subtitles. Each group was given the same questionnaire to determine which aspects of the video the participants concentrated on.

2.5 Procedure

1. **Pre-test:** Subjects were instructed to read and sign an informed consent form detailing the experiment objectives. Subjects were then read an introduction to our experiment. This introduced the subject to the Tobii Eye Tracker informing them we will first calibrate their eyes so the eye tracker can adjust to their individual settings. They were warned to keep their head as still as possible to prevent corrupt data calculations. They were next told a series of collapsing dots would appear on the screen one at a time and to watch them until they disappeared not anticipating where the next one will arrive. Finally, subjects were told they would view a short video clip and to watch it as if it were appearing on television.

Subjects were next placed at least 50 cm away from the Tobii eye tracker [4]. The calibration program was launched to determine the exact eye movement of the right and left eyes of the subject. A total of 16 yellow dots were displayed for calibration. This allowed subjects to get familiar with the equipment before beginning the experiment. On completion of the calibration phase, volume on speakers was turned up to ensure subject could effectively hear the audio portion of the movie clip.

2. **Test:** Before performance of this phase of the experiment, the pre-test must have been performed and done correctly. The testing procedure was divided into two phases:
   a. With Subtitling: The visual and audio cues were presented to the subject. While viewing and

listening to a specific time scene of the animated film "Finding Nemo" with the subtitling feature turned on, the subject's eye movement was recorded.

b. Without Subtitling: The same procedure with subtitles was performed with the subtitling feature disabled during the time period the same scene was shown to a different set of subjects while their eye movement was recorded.

3. **Post-Test:** After successfully completing the experiment, the subjects were instructed to complete a questionnaire in relation to what was seen, heard, or even read during the duration of the experiment. All questions were related to visual and audio aspects that could possibly cause a distraction to very obvious things in the scene of the film shown.

3. Results

After reviewing and analyzing the data collected on our subjects, we can see subjects who were exposed to subtitles had slightly larger statistical values in terms of fixations and percentages, as well as a slight performance increase on the questionnaire. Contrary to our thoughts, the subjects with subtitles actually performed better on the visual aspects of the questionnaire, but significantly worse on the audio-related questions. Since the quizzes focused more on visual aspects of the film, the results obtained cannot be directly interpreted. Table 1 reports the findings of fixations we obtained from our experiment.

Subjects	Subtitles Yes/No	Total Number Points		Total Number Fixations	
		Inside region	Outside region	Inside Region	Outside region
Female 1	No	228	1279	6	44
Female 2	No	0	1620	0	41
Female 3	No	140	1060	5	44
Male 1	No	43	1537	2	58
Male 2	No	16	1539	0	48
Male 3	No	164	1345	4	31
Female 4	Yes	49	1433	1	52
Female 5	Yes	57	1512	2	33
Female 6	Yes	30	1494	1	55
Male 4	Yes	964	644	63	32
Male 5	Yes	157	688	13	35
Male 6	Yes	0	1600	0	40

Table 1: Data calculations on fixations

3.1 Eye Behavior

The results from the subjects that viewed the movie with the subtitles had a greater amount of fixations when compared to those without the subtitles. The cause of this could be attributed to the added visual stimuli, confirming the notion that more fixations would occur. Although the ratios of fixations to points were similar, an advantage was noticed when viewing the results for subjects with subtitles. By setting a boundary around the subtitling region of both the non-subtitling and subtitling version of movie we were able to look at the number of fixations that fell within that region. This resulted in an average number of fixations for those subjects with subtitles were approximately four times as great as for those without. The comparison of 24% to 6% clearly shows, those subjects with subtitles present focused their gaze greatly within the boundary region previously established. Figure 5 and 6 below displays a comparison of the fixation points located inside and outside the subtitled region for all sets of subjects

Figure 5: Subtitled region with subtitling present

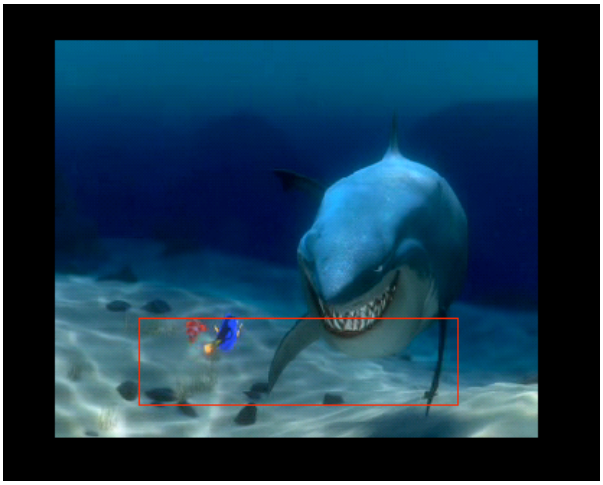

Figure 6: Subtitled region with subtitling disabled

3.2 Informational Assessment

The questionnaires given to the subjects produced results contrary to what was expected. The average scores on the questionnaires actually increased for

those with the subtitles. Acknowledging the fact subjects with extra visual stimuli would produce a greater percentage when answering the visual questions on the questionnaire in comparison with the other subject group of 95% to 83%. This can most likely be attributed to the fact that, presented with the extra visual stimuli on the screen, the subjects were forced to pay more attention to the happenings on the screen, and thus picked up on the visual aspects more than those other subjects. When taking into account the questions that focused on audio the subjects without the subtitles benefited greater in viewing the video in regards to the questions by 77% to 55%. The consequence for the subjects with subtitles is the lack of attention given to audio aspect of the video viewing.

4 Improvements

Upon completion of our experiment, we have identified possible limitations in our experiment that could be held accountable for data inaccuracies. Addressed below are several possible solutions for avoiding issues we encountered in our experiment. Researchers can consider these alternative approaches to prevent similar data inconsistencies from occurring in future implementations.

4.1 Program Implementation

Using the ffmpeg tool to extract the image files from the video clip added to the ease of use for our group. However, it increased the load time of the program and became a hog of system resources. Using another system would have eliminated the resource drain and load time. In addition, the video and audio portions of the clip were displayed using two separate systems (OpenGL and SDL respectively), having one system instead of two would have eliminated the use of a timer class which coordinated the two systems to work together.

4.2 Load time

Implementing the program took a great deal of time to load because of the usage of PPM image files. Having another method to use, the load time would reduce greatly. To effectively minimize the completion time of experiment we incorporated the load time in the pre-test portion of the experiment.

4.3 Movie Clip

Some recommended modifications to the movie clip to enhance the accuracy of the results obtained are movie quality, resolution, and movie selection.

4.3.1 Quality

The clip was extracted from a DVD using DVD Shrink. As a result of extraction there were minute tracking errors in the clip used, which caused the audio to “skip” in regards to the video. Multiple settings in the program were attempted to gain a better version of the scene, but were unsuccessful. Having a better version of the program could have eliminated any complex problems that existed for the subjects.

4.3.2 Resolution

The clip was displayed at a resolution of only 352 by 288 pixels. Therefore, the subject’s gaze was restricted to the boundary region, and because of the small resolution it affected the size of the boundary region. The smaller resolution makes the variation degree of error .5 become a larger factor in analyzing the results. Hence, a larger resolution would result in the subject’s eye movement to be greater, leading to more meaningful data. It ultimately would have made each fixation clearly distinct, resulting in readable data to use for analyzing.

4.3.3 Movie Selection

The movie was chosen based on its agreeable content and good combination of audio and video stimuli. However, it was a movie that a large portion of people had seen. If another movie had been chosen that the subjects had not already seen, forcing them to pay more attention to the movie and alleviate memory usage when answering the questionnaire. They would have had to pay more attention to the movie and could not use prior memory to answer the questionnaire.

4.4 Subjects

Realizing the complications involved in analyzing and gathering prominent data from each subject, it would be best to use a larger amount of subjects for this experimentation. This would lessen the effects of dead data and outliers.

4.4.1 Variety

Allowing a variety of viewing options for the movie could prevent the overwhelming majority of the subjects that were used for this experiment had seen the movie before. Pre-screening the subjects and only using those who have not seen the selected movie clip would alleviate previous viewing knowledge on the clip. This could possibly do away with the recollection efforts that some of the subjects may have embarked upon.

4.4.2 Corrected vision

A large number of the test subjects had some form of corrected vision. Those with glasses were asked to remove them and if they could not see, their data was not collected. Contacts were not assumed to have played any factor. It was unclear if the vision problems of the users translated into differences in the data collected. If it had been possible to ensure that all subjects had 20/20 vision, then such probabilities could have been eliminated.

4.5 Questionnaire

There were two apparent problems with the questionnaire. The quantity of questions and limited variety hindered our assessment of the subjects' complete retention.

4.5.1 Number of Questions

Keeping in mind that our subjects are college students and doing this on volunteer bases, making the test procedure quick and accurate was a positive attribute, a longer questionnaire would not be adequate. Hence, having a questionnaire focusing on quantity and quality could eliminate any outliers that may have thrown the averages off.

4.5.2 Variety

Creating a questionnaire which possessed a balanced variety of visual and audio questions could have helped in determining whether subtitles enhance the subject's knowledge of information pertaining to the movie clip.

5 Discussion and Conclusion

Based on the data we collected, we could see from the statistical analysis that the subtitling did lead to increased fixations in the bounded region. However, the questionnaire led to contrary conclusions. The questionnaire gave the impression that the subtitles actually helped better the overall scores of the subjects. Because the questionnaire was so visual-oriented, the increased overall grades cannot be directly linked to the subtitles. The subtitles instead created a negative effect, bettering the visual aspects but greatly reducing the auditory perceptions. Because of the relatively small number of subjects, and the closeness of the results, we cannot say that subtitles had either a negative or a positive effect on the subject's ability to view and recollect aspects of the movie clip. If future research is done, primarily with larger numbers of subjects and more distinct variables, rules could possibly be generalized.

References

[1] LIESBETH PYFERS: State of the art: Subtitling and Closed Captions in Multimedia Productions. http://www.equalaccess.nl/MultiReader/subtitling.htm, (2002).

[2] DUCHOWSKI ANDREW T.: *Eye Tracking Methodology: Theory and Practice.* Springer Verlag, London, UK, (2003).

[3] STANTON, ANDREW, UNKRICH, LEE: *Walt Disney*. Finding Nemo. (2003)

[4] TOBII TECHNOLOGY AB. *User Manual: Tobii Eye Tracker*. (2003).

[5] CAPTIONED MEDIA PROGRAM: *How Bird Hunting in North Carolina Saved Captioning*. http://www.cfv.org