Improved Captions for the Hearing Impaired

Samson B. Drews

Valerie Barr, Advisor

UNION COLLEGE

CS

Introduction and Background

Visual media is everywhere. Rapid advances in video accessibility are causing video to hold an increasingly significant role in our daily lives. This is exciting, but the hearing impaired audience have great difficulty in fully comprehending video content due to loss of key auditory information. Captioning is the solution for this, and has been since the 70's [1]. Different from subtitles, captions are meant to convey dialogue *and* sounds. However, currently, captions are generally illustrated statically, in a fixed region of the screen as in Figure 1 below.

Figure 1. Example of current captioning

This method delivers a limited amount of information and inhibits interpretation, especially for the hearing impaired. There are three main problems with current captioning:

1. Uncertainty as to which character is presently speaking.
2. Uncertainty of the pace that a given character is speaking.
3. Lack of conveying volume variation, aside from punctuation.

The existing approach has a long way to come before it is suitable for the hearing impaired audience. My research aims to enhance current captioning methods to give disabled individuals more equal opportunity to enjoy visual media.

Solution

To enhance viewing experience, compared to existing methods, this more dynamic approach to captioning will:

- Determine more suitable and obvious script locations depending on the location of the sound
- Highlight scripts in real time (like karaoke) to better convey pace
- Illustrate volume with variation of text size

The existing approach has a long way to come before it is suitable for the hearing impaired audience. Figure 2 offers a slightly more stylized approach that addresses the aforementioned issues.

Figure 2. Example of how captioning could better address its problems

Technologies

There are a vast amount of multimedia file types available and it was difficult to decide which to use. The technologies needed are straightforward and not difficult to use. All that is needed is WebVTT for the captions, HTML for the web pages, CSS for styling the pages/captions, PHP for updating the page, and little JavaScript for some nice animation.

Experiment Design and Results

In order to determine if contextual captioning is an improvement over the industry standard, user studies were conducted over the span of a week. The experiment was designed so that a participant was randomly assigned one of two videos and its corresponding survey using Google Forms. Each of the two videos and surveys were the exact same, the only difference was the captions. One video contained industry standard for captions and one had this paper's proposed contextual captioning. The below graph shows how easy it was for each participant to interpret certain aspects of improved captioning.
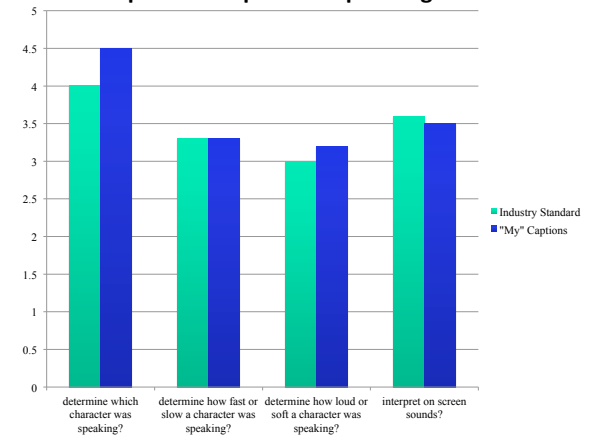
Evaluation and Future Work

From these results, it is difficult to determine if my independent variable, the type/form of captioning, had any significant impact on a given participant. As the bar graph above shows, contextual captions performed slightly better than the industry standard on the majority of the questions, but these results are not significant enough to be deterministic.

Moving forward, primary goals will be to add karaoke-style script highlighting and as much automation as possible. In terms of making sound-related words more dynamic, just as much emphasis needs to be put into styling as the word itself

References

[1] National Captioning Institute. History of closed captioning. 2016.

[3] W3C. Synchronized Multimedia Integration Language (SMIL 3.0).