Title of Work:

Using non-speech sounds to increase web image accessibility for screen-reader users

Conference:

35th ACM International Conference on the Design of Communication

Rationale to ensure venue quality:

The conference is a conference run by the ACM, which is an immediate indicator of relatively high quality. Specifically, this conference is run by the Special Interest Group on the Design of Communication (SIGDOC) of the ACM. Additionally, the conference is sponsored by several well-known and prominent American Universities and Technology Companies, including UC Davis, Arizona State, Texas Tech University, UNT, and Adobe. Finally, given that this is the 35th iteration of this conference, it is safe to say that it is well established and well-recognized.

Problem Statement:

One of the most difficult things to effectively represent via a screen reader for vision impaired individuals is images. Traditionally, alternative text can be defined that describes the image to the user, but this can be a "tedious and unsatisfying experience for blind users" because text-to-speech generally lacks any form of expressiveness and because alternative text is frequently written in a manner that can make it a distraction and a hinderance to the rest of the experience. One method that could make the experience more natural and better convey the information about an image is to make use of audemes, a type of non-speech sound, to convey information about an image. Audemes can offer a more engaging and information-rich experience to visually impaired users when navigating web pages with a screen reader.

Paper Synopsis:

Images have become an integral and ubiquitous part of conveying information through web interfaces. However, not all users can easily perceive information in this manner. Perceiving images has frequently been a challenge for visually impaired users. The most common accessibility technology available to these users – screen readers and text-to-speech software – often struggle to effectively communicate image-based information. First, screen readers require that web developers annotate all image tags with a textual description of the image. It is not a given that developers will include this annotation, and when they do, it is often poorly optimized for conveying information about the image in that it can is too short, too long, or not sufficiently descriptive. Tools to automatically check the quality of alternative text are not currently viable because there is no defined accessibility standard for alternative text quality.

One potential alternative that has been successful in other, similar capacities is non-speech sounds. Non-speech sounds have been used to encode message or convey audio notification, based on the idea that humans can instinctively react or "train" themselves to react in a certain way to a certain sound. For example, audible emergency alert signals minimize reaction time in response to an emergency. This is because synthetic speech suffers some of the same drawbacks as text- it relies on the use of many words to describe a concept due to lack of expressiveness and users must listen to all words to fully comprehend the message. Non-speech sounds can use shorter forms to communicate more rapidly. Also, an experiment investigating memory load of natural sounds compared to synthetic speech showed that synthetic speech "puts a heavier load on short-term memory for young and old adults," and "recognition accuracy decreases significantly with the increased presentation rate."

To that end, Thapa et al. investigated the viability of audemes in place of alternative text as a screen reader-friendly image annotation. They gathered 14 sighted users that had at least a bachelor's or master's degree in a relevant technology field and had performed a comparative study between each user. Two web pages were developed, each with eight similar images. The first had audemes descriptions and the second had alternative text annotations. In order to simulate a blind use, the screen was disabled so participants could only navigate by sound. A simplified interface was set up to account for a lack of familiarity with screen reader technology, where the users navigated forward with "tab" and backwards with "shift + tab." The audemes were designed to be short, non-speech sound symbols under four seconds long, comprised of various combinations of sound effects, which include "natural or artificial context, abstract sounds, and music excerpts."

After exploring both sites, each user completed a paper version of the NASA Task Load Index (TLX) to evaluate the workload they perceived in identifying each image. After this, the System Usability Scale was administered for each site. The results indicated that the participants experienced significantly less workload in identifying test images with audemes compared to alternative text. Not only was there was a lower mental and temporal demand required to recognize images with audemes, there was a significant increase in task performance and reduction in frustration while as well. These results quantitively indicate that audemes offer superior accessibility. Qualitatively, users remarked that the synthetic speech for the alternative text was "stressful" and "caused them to lose concentration and miss descriptions," while the audemes were "pleasant" and not irritating when visiting an image multiple times.

Future Work:

The evaluation done was very comprehensive, so I don't see much room for future work in that regard. However, I would like to conduct these tests in a more realistic scenario to potentially offer more useful feedback, using actual vision impaired users in a regular usage scenario, rather than sighted users visiting just pages with images for comparison.