

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

Ratan Bahadur Thapa
Oslo and Akershus University
College of Applied Sciences
Oslo, Norway
s300681@stud.hioa.no

Mexhid Ferati
Oslo and Akershus University
College of Applied Sciences
Oslo, Norway
mexhid.ferati@hioa.no

G. Anthony Giannoumis
Oslo and Akershus University
College of Applied Sciences
Oslo, Norway
gagian@hioa.no

ABSTRACT

Screen-reader users access images on the Web using alternative text delivered via synthetic speech. However, research shows that this is a tedious and unsatisfying experience for blind users, because text-to-speech applications lack expressiveness. This paper, poses an alternative approach using an experiment that compares audemes, a type of non-speech sounds, with alternative text delivered using synthetic speech. In a pilot study with fourteen sighted users, findings show that audemes perform better across many areas. Specifically, audemes required lower mental and temporal demands and led to less effort and frustration and better task performance. Moreover, participants recognized audemes with higher accuracy and lower errors. Audemes were also perceived as more engaging compared to alternative text delivered using synthetic speech. Additionally, audemes were found to be richer in delivering information. This study suggests that non-speech sounds could substitute or complement alternative text when describing images on the Web.

Categories and Subject Descriptors

H.5.2 [User Interfaces] Auditory (non-speech) feedback.

General Terms

Human factors, design.

Keywords

Screen-reader; non-speech sounds; alternative text; Web image; accessibility.

1. INTRODUCTION

Images on the Web are used ubiquitously to deliver information. However, perceiving images has been a constant challenge for visual impaired persons who typically use assistive technologies to access the Web. For example, they often use text-to-speech (TTS) software, such as screen readers [1], which translate textual information into synthetic speech. To effectively translate images into synthetic speech, Web developers must include textual descriptions for images. However, many images on the Web lack such a description [2].

SAMPLE: Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '10, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/12345.67890>

The Web Content Accessibility Guidelines (WCAG) [3, 4], an industry standard, requires text descriptions for images, known as *alt text*. Alt text is an easy and relatively proven method to make images accessible to the blind people [5, 6]. Recent studies have attempted to develop new approaches that automatically include alt text when images on the Web lack such information [7]. A usability study with disabled people, however, found that including alternative text often is insufficient to guarantee the accessibility of websites [7]. Another study reports that alt text should be meaningful and easy to perceive in order to convey the appropriate message to the user [8].

WCAG 2.0 suggests that alt text should be as short as possible. This may undermine the quality of alt text by limiting the amount of information, and may affect the style of writing in a way that inhibits comprehension. Automatic accessibility checkers raise warnings if alternate text is too short or too long. Moreover, tools that check and verify website accessibility levels currently ignore the quality of alternative text due to the lack of a standardized methodology. Thus, the lack of appropriate and meaningful ways to represent Web images remains an obstacle for screen-reader users and others approaching the Web with non-visual assistive medium.

For such scenarios, this article argues that non-speech sounds provide a possible solution. Non-speech sounds have been used to encode messages and convey information using audio – e.g., to inform people about their medication or about upcoming events [9, 10]. Moreover, in some cases, reactions to audio stimuli have been found to be faster than visual stimuli [11]. For example, audible emergency alert signals [12] minimize reaction time for safety measures. The study found that sound can offer an intuitive form of the information it presents. This could allow people to understand and memorize information in different ways. Additionally, research shows that blind people compared to sighted individuals have higher levels of acoustic ability [13]. They can process auditory stimuli faster than sighted people [14]. For example, a study with visually impaired students demonstrated that their accuracy at identifying the location of sound source is superior to that of sighted students [15].

When comparing non-speech sounds with synthetic speech, the latter has the same drawback as a serial medium such as text. Text may use many words to describe simple information due to its lack of expressiveness [16]. Users must listen to all the words to comprehend a message. Using non-speech sounds, messages can be composed in shorter forms and therefore can be heard more rapidly. Moreover, 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

[17]. Similarly, research has also found that recognition accuracy decreases significantly with the increased presentation rate [18]. It requires practice even for highly skilled visually impaired persons.

Considering the differences between non-speech sounds and synthetic speech generated from screen-readers, this article evaluates their performance when used to describe an image on the Web. In the next section, this article looks at existing studies addressing this issue and the type of non-speech sounds commonly used in user interfaces. It then proceeds by describing the details of the experiment that was used to compare quantitatively and qualitatively non-speech sounds and speech. Afterwards, this article analyses the results and discusses some of the limitations of the study. Finally, this article concludes by highlighting useful directions for future work.

2. RELATED WORK

Non-speech sounds have been used for several decades to represent information in a computer user interface. Several types have been designed depending on the specifics of the information intended to be delivered. The two most used non-speech sounds to represent brief objects or information on a user interface are earcons [21] and auditory icons [22]. To illustrate, an earcon is the abstract sound we hear when receiving an email, while an auditory icon is the sound of a crumpling piece of paper when deleting a file.

Auditory icons are suitable for user interactions, alerts, and helpful for navigation [23]. Its true application is based on the direct representation of an associated concept. It is, however, very difficult to accurately classify or create an auditory icon for every word or concept. Considering that auditory icons are based on the natural sound an object makes, there is an intuitive link between the sound and the object or concept that it represents. In other words, it leverages the knowledge people have of natural events and uses the same sounds to represent an object or event in a user interface.

Earcons are generated from abstract synthetic tones to create an auditory message [22]. Typically, there is no natural association between the sound and the object or event that it represents. Because of this quality, they are easier to use and create, but might be frustrating and difficult for users to learn and remember since their association with an actual object or event is arbitrary. In a study using the Mathtalk system [24], earcons were used to indicate structural delimiters and provide an abstract overview of the entire equation. It was found that cognitive effort required to decode each pattern detracted users from processing the mathematical content.

However, an image is typically a complex entity and difficult to represent using auditory icons and earcons. When communicating complex content, such as large, continuous data, sonification is used [25]. Several studies demonstrate the efficiency of this method for blind users [26-29]. For example, to communicate visual content created from a camera device, an audio representation is generated by mapping the continuous frequency of horizontal and vertical dimensions [30]. Similarly, audible methods of continuous data representation are used in navigation systems to assist blind people in navigating their surrounding [31, 32].

Sounds have been also used for communicating the shapes of objects by modifying pitch and intensity to correspond with the object's shape and size [33]. This helped users with visual impairments to easily follow the sound reference and recognize the object. Similarly, a study reports that a speech interface, which allows users to explore and identify graphs and tables, resulted in a significant decrease in subjective workload, temporal demand, and number of errors compared to haptic interface [34]. Other studies with blind people examined converting image to sound via method

of edge detection [35], and converting extracted image information into a haptic environment [36].

The sonification process, however, requires a reasonably large amount of digital data that can be translated into a sound by modifying the frequency, pitch and intensity. Images found on the Web, rarely have characteristics for such straightforward data-to-sound translation. Although it is possible, for instance to translate the Red, Green, and Blue values of image pixels into sound, such information will not be useful to a screen-reader user. Instead, the image should be represented as an auditory form that will make sense for the user.

Audemes provide a novel category of non-speech sounds that can be used to represent complex content. They were initially invented and tested with blind and visually impaired users to convey thematic content [37]. They are similar to auditory icons, but are semantically more flexible than other non-speech sounds. Because of this flexibility, they can represent more complex content compared to auditory icons. The meaning of audemes is typically generated by concatenating sounds, and although meanings are not completely open and arbitrary, they start broad and then narrow to additional sound cues, which merge into a single meaning [38].

The process of meaning generation of an audeme first starts with identifying the cause of the sound, and then following the reference. For example, "neighing of a horse" could be used to initially identify the animal. Afterwards, the established link is referenced to further focus the meaning of the audeme, which could be "horse riding", "horse polo", "horse racing", or other events related to the subject. The ability of audemes to generate meaning makes them potentially a suitable medium to describe images.

The design of audemes is based on empirical knowledge, which often results in the creation of sounds derived from the personal preference of the sound designer. It could contain sounds alone or in combination from natural events or abstract, musical tones. A study shows that audemes significantly improve and increase the recognition of concepts for blind and visually impaired participants [39]. Similarly, audemes for content navigation were tested on a touch-screen interface and they were found to be easy to learn, memorable and navigable for visually impaired teenagers [40, 41]. In terms of information retaining, audemes were found helpful in reducing memory erosion and even after five months, the content was better remembered with audemes than without them [37, 38]. Audemes were found to have potential in scientific applications including gaming and productivity as well as in education as a tool for better memory retention [42].

Considering these qualities of audemes, this article developed an experiment to evaluate their performance and suitability for describing images. Additionally, since audemes are generated from sounds that people identify from their experience, this article anticipates cognitive effort will be lower when using audemes.

3. METHOD

3.1 Experimental Setting and Participants

This paper uses both qualitative and quantitative data. Quantitative data include the workload perceived by participants when recognizing test images, and usability ratings of the test prototype. Qualitative data covers participants' comments while performing tasks, which was triangulated with their post-test reactions and feedback. Observation data sheets comprised other aspects of user data and trends, such as time spent on each prototype and attempts made on each test image during the experiment phase.

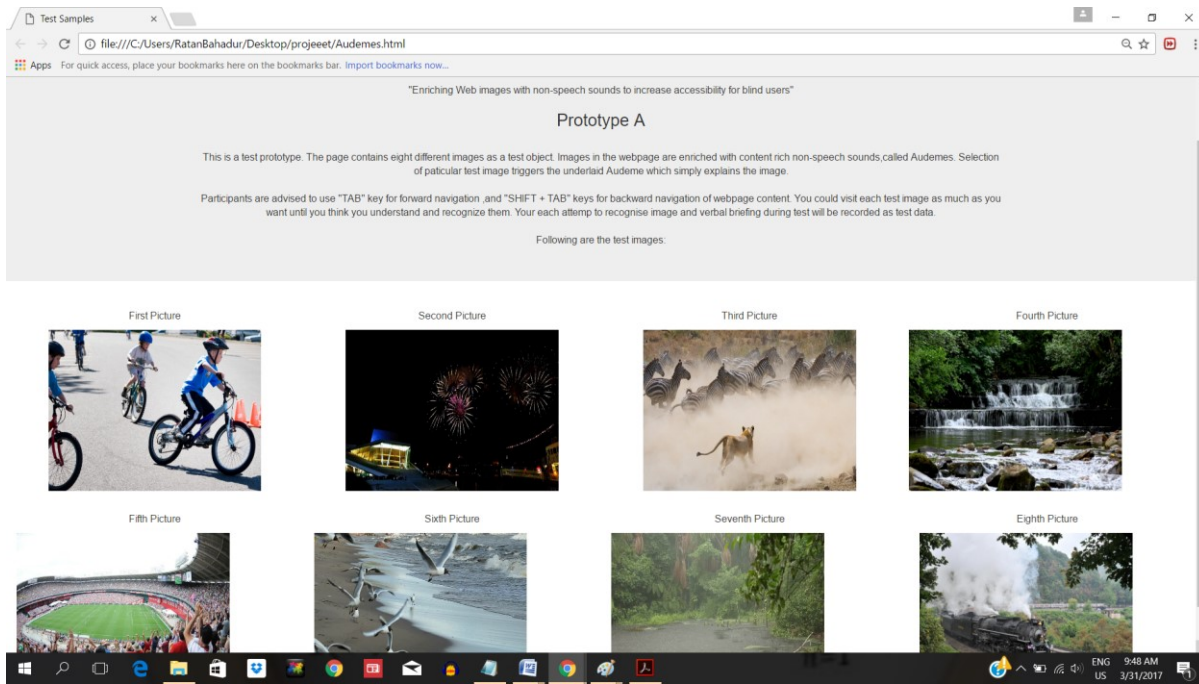


Figure 1. Screenshot of prototype “A” using images with audemes.

While blind participants would provide more valid data on the experiences of persons with disabilities, recruitment was a challenge, and so this paper opted to conduct a pilot test with fourteen sighted users (males=12, females=2). These participants were randomly recruited based on their knowledge and experience in web design and accessibility. Most of them (N=9) were students of the Master program in universal design of information and communication technology at Oslo and Akershus University College of Applied Sciences in Norway, which guaranteed general knowledge of Web accessibility and screen readers. Participants had at least either bachelor or master university degree and solid computer and internet browsing skills. The age of the participants ranged from 20 to 39 years. None of them identified as a person with a visual impairment. All participants were initially briefed and provided informed consent. Most of participants (N=12) had not previously used screen readers; however, they were briefly trained to use them before the test.

3.2 Test Prototypes

In order to investigate the significance of audemes as a means to represent Web images, a comparative study with alt text was conducted. Two web pages were developed as test prototypes. Prototype “A” contained eight images enriched with audemes as their description, and prototype “B” contained eight similar images with alt text. The structure of the prototype pages was divided into two sections. The first section contained information about the prototype and instructions on how to use it, whereas the second section listed the images as test objects. Each prototype had the same design and layout with information organized in the same order. Images were organized in two rows in a set of four images per row. The design of the test prototypes was kept the same to capture only participants’ reflections on the difference between delivering images with audemes or alt text. Prototype “A” is shown on Figure 1, while both prototypes are publicly available.¹

3.3 Navigation

To accommodate for participants’ lack of experience with screen readers, the prototypes were built to be navigable entirely by keyboard. To simulate lack of vision, the computer screen was turned off, so the participants could only navigate by sound. The navigation was simplified, so the participants only had to use two keys; Tab key to sequentially move forward through the images with an associated audeme or alt text, and Shift+Tab to move backwards. In addition, the mouse and touchpad were also disabled. This setting is depicted in Figure 2.

In order to achieve this simplified navigation, some JavaScript coding was used. The script played the audeme audio file when the user navigated over the test image using the Tab key. In the prototype using alt text, the screen reader simply read out the textual description behind the test image.

3.4 Audeme Design

The design of audemes is subjective and based on the designer’s consideration. They heavily rely on meaning derived from semiotics structure and an intuitive link between sounds and natural events. Audemes are defined as short, non-speech sound symbols, under seven seconds, comprised of various combinations of sound effects, which include natural or artificial context, abstract sounds, and music excerpts [38].

Audemes were developed following a guideline from a previous study that showed audemes composed of two to five individual sounds lasting three to seven seconds improved encoding and long-term memory [19]. Additionally, for this experiment, audemes lasted less than four seconds in order to match the speech duration of alt text. Table 1 shows the images that were used and their corresponding audeme names. Table 1 also includes links to the

¹ <https://audemes.000webhostapp.com/ratan/>








corresponding audemes. Table 2 lists the images and corresponding alt text used in Prototype B.

Table 1. Test content of prototype “A”: Images enriched with audemes.

Test Image	Audemes
 <p>Source: e2sport</p>	 <p>kids_cycling_audeme.wav</p>
 <p>Source: visitoslo</p>	 <p>new_year_audeme.wav</p>
 <p>Source: staticflickr</p>	 <p>lion_zebra_audeme.wav</p>
 <p>Source: yting</p>	 <p>river_flow_audeme.wav</p>
 <p>Source: ussoccerplayers</p>	 <p>stadium_soccer_audeme.wav</p>
 <p>Source: clubedafotografia</p>	 <p>seagull_beach_audeme.wav</p>
 <p>Source: picdn</p>	 <p>rain_forest_audeme.wav</p>

 <p>Source: railpictures</p>	 <p>steam_train_audeme.wav</p>
--	---

Table 2. Test content of prototype “B”: Images enriched with alt text descriptions.

Test Image	Alt Text
 <p>Source: thebetterindia</p>	Students are playing basketball in a tournament organized by project KHEL.
 <p>Source: mundy.assets.d3r</p>	Tourists are enjoying the Amazon rainforest river ferry expedition by boat.
 <p>Source: wallpaperscraft.ru</p>	Flocks of penguins are jumping in ice snow water in Antarctica.
 <p>Source: googleusercontent</p>	Debris of residential area of 2011 Tsunami destruction in Japan.
 <p>Source: userscontent2.emaze</p>	Typical wild African elephant family herd at a watering hole.
 <p>Source: pbs.twimg.com</p>	Arial view of hundreds of sheep leaving the stall.
 <p>Source: k37.kn3</p>	A group of wielder working in a bridge construction.



Picture of a space rocket taking off from launch station.

Source: rock-palace

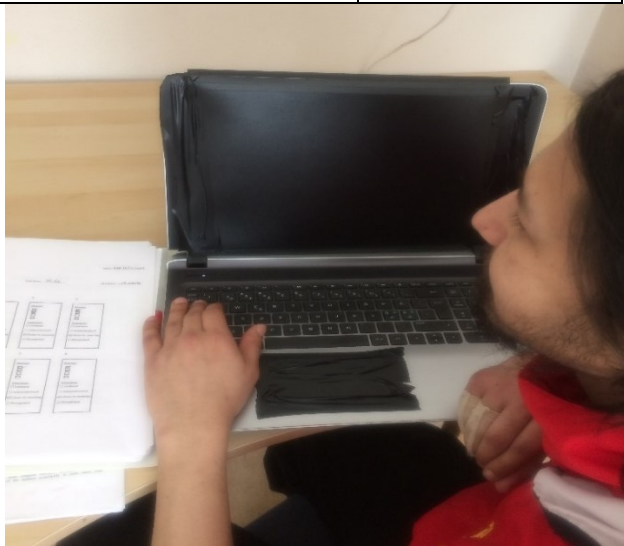


Figure 2. Participant accessing test prototype with NVDA.

3.5 Procedure

Participants were briefed one week in advance about the conditions of the test when they received the consent form. On the testing day, they received a 30 minute training to gain some experience with non-visual access of Web content. First, participants familiarized themselves with the NVDA² (Non-Visual Desktop Access) screen-reader application. They were taught to use NVDA with eSpeak synthesizer and Ava, US-English Premium High vocalizer. Second, they were introduced to audemes with samples taken from an online audeme dictionary [20].

During the test, users were advised to access each test image as many times as they wanted until they thought they understood it. They were encouraged to speak out what they thought about the test images while accessing them. The number of attempts each participant made when accessing the images along with their comments during the test was recorded in an observation sheet.

After participants went through both prototypes, they completed a paper version of the NASA Task Load Index (TLX) [21] to assess the workload they perceived in identifying each test image. Two TLX measurement scales for each prototype were placed vertically right across the workload questions and participants were instructed to provide their workload ratings on the basis of comparison. Finally, the System Usability Scale [45] was administered for each prototype.³ Participants were advised to compare the prototypes while rating their usability.

4. ANALYSIS AND RESULTS

The examination of the data collected was initially conducted by ensuring its validity and suitability for statistical analysis. In all

² <https://www.nvaccess.org/>

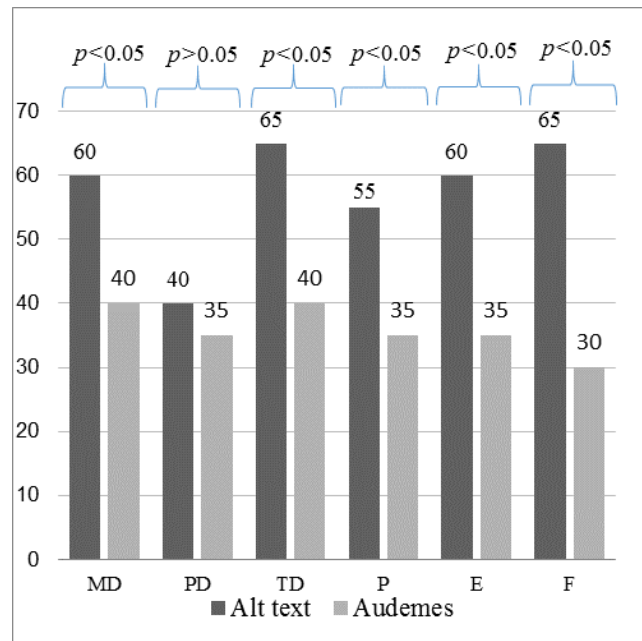


Figure 3. Paired t-test for each NASA-TLX subscales' workload (MD=mental demand, PD=physical demand, TD= temporal demand, P=performance, E=effort, F=frustration).

cases this paper reports the mean and standard deviation, while for the TLX data, and accuracy and error data, this paper also performed and report a paired t-test analysis to ensure significance. The TLX data analysis has been previously validated [46].

4.1 Results from the NASA-TLX

The overall findings indicate that participants experienced significantly less workload in identifying test images with audemes ($M=35.35$, $SD=18.46$) compared to alt text ($M=57.85$, $SD=8.92$), $t_{(13)}=7.87$, $p < 0.05$.

Looking at individual dimensions of the TLX, a significant decrease in mental demand was required to recognize web images with audemes ($M=39.28$, $SD=12.53$) over alt text ($M=61.78$, $SD=8.92$), $t_{(13)}=8.14$, $p < 0.05$. Similarly, there was a significant reduction in temporal demand with audemes ($M=38.75$, $SD=8.18$) over alt text ($M=63.92$, $SD=16.61$), $t_{(13)}=5.97$, $p < 0.05$. In terms of physical demand, however, no significant difference was seen when comparing audemes ($M=34.64$, $SD=13.93$) with alt text ($M=37.5$, $SD=9.53$), $t_{(13)}=0.86$, $p > 0.05$.

The results show that the task performance significantly increased when participants used the prototype with audemes ($M=32.85$, $SD=13.25$) compared to alt text ($M=55.71$, $SD=9.44$), $t_{(13)}=4.03$, $p < 0.05$. Linked to this, the results also show that there was significantly less effort required to recognize images with audemes ($M=36.07$, $SD=10.28$) compared to alt text ($M=60.71$, $SD=16.39$), $t_{(13)}=5.94$, $p < 0.05$. Moreover, there was a significant reduction in frustration while accessing web images with audemes ($M=27.14$, $SD=6.99$) compared to alt text ($M=63.92$, $SD=12.06$), $t_{(13)}=6.96$, $p < 0.05$. Figure 3 depicts these values.

These results indicate overall better results from audemes compared to alt text. Participants commented that the synthetic

³ <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>

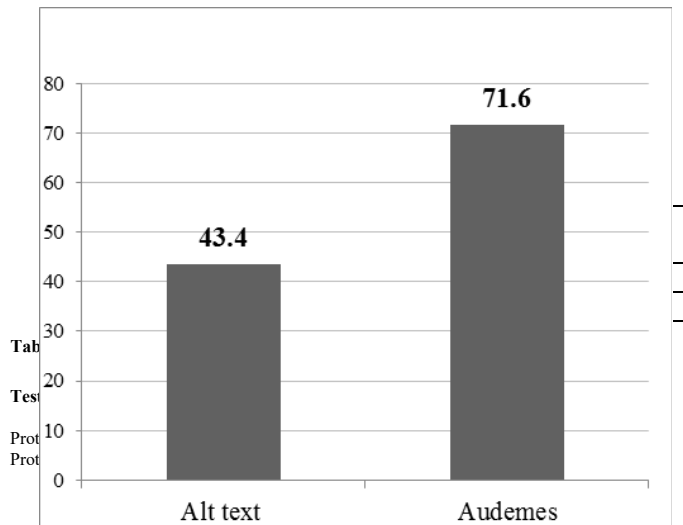


Figure 4. Mean SUS percentiles for each prototype provided by 14 participants.

speech from the NVDA screen-reader was stressful and caused them to lose concentration and miss the alt text descriptions. On the other hand, audemes were found to be more pleasant, and did not make participants feel distracted or irritated even when visiting an image multiple time.

4.2 System Usability Scale Results

In order to measure the usability level of the prototypes, the results from administered the System Usability Scale were calculated and compared. The results show that participants rated the prototype with audemes to be more usable compared to the prototype with alt text. This confirms the results from the previous section showing that participants perceived less subjective workload in identifying test images using audemes, which influenced their perception of usability.

Specifically, the usability ratings given by all participants shows that the prototype with audemes scored 71.6 in comparison to the prototype with alt text, which scored only 43.4 (Figure 4). Following the ranking analysis offered in this study, Table 3 lists the Prototype A as acceptable compared to B.

This difference between the two prototypes according to participants' comments was that audemes were perceived as more engaging than speech. However, the participants also commented that audemes were too short. They suggested that audemes should be longer in order to be more informational and easier to understand.

4.3 Results in Terms of Accuracy and Error

Part of the experiment aimed to measure image recognition accuracy. For each image, the results provide the level of accuracy using categories "Recognized" and "Closer to Meaning", and the errors using categories "Confused" and "Misunderstood". These responses were further adjusted with participants' feedback when images were shown to them. The results also provide the number of attempts each participant took to identify each image and the time they spent on each prototype.

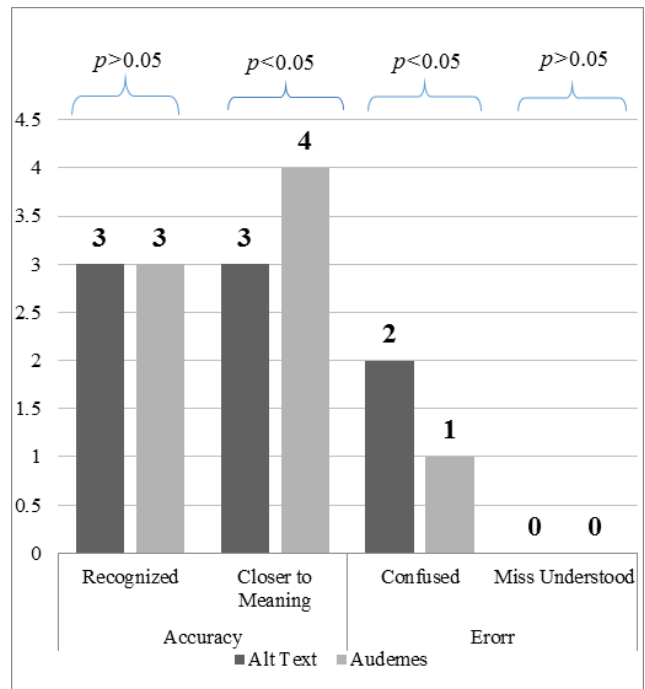


Figure 5. Accuracy and error rates means for image recognition.

The results show that participants' overall accuracy at identifying the images was significantly higher with audemes ($M=7$, $SD=0.78$) compared to alt text ($M=6$, $SD=0.81$), $t_{(13)}=4.03$, $p < 0.05$. In addition, participants recognized almost half of all images with the same accuracy, audemes ($M=3$, $SD=0.60$) compared to alt text ($M=3$, $SD=0.55$), $t_{(13)}=0$, $p > 0.05$. However, using audemes ($M=4$, $SD=0.74$), participants more accurately understood the meaning of the image than compared to alt text ($M=3$, $SD=0.67$), $t_{(13)}=4.76$, $p < 0.05$. In terms of error, in both conditions participants showed no signs of misunderstanding any of the images, however, they were significantly less confused when using audemes ($M=1$, $SD=0.82$) compared to alt text ($M=2$, $SD=0.92$), $t_{(13)}=2.80$, $p < 0.05$. These values are depicted in the Figure 5.

Participants commented that the confusion with alt text was caused by the unclear speech generated by the TTS and their inability to concentrate on it continuously. Similarly, they thought audemes were more closely associated with images. This indicates that compared to alt text, audemes may improve image recognition as well as reduce errors.

Figure 6 depicts the cumulative number of attempts participants made to recognize the images, and the time spent on each prototype. The results show significant differences for the latter measurement, namely, participants spent more time with audemes ($M=16$, $SD=2.44$) compared to alt text ($M=13.71$, $SD=1.38$), $t_{(13)}=4.94$, $p < 0.05$. In terms of attempts, the difference was not significant; audemes ($M=18.5$, $SD=3.03$) compared to alt text ($M=19$, $SD=2.26$), $t_{(13)}=0.07$, $p > 0.05$.

In relation to these results, participants commented that they perceived audemes as richer in terms of information, which helped them concentrate and visualize the images. Moreover, the results shows that four participants successfully identified all images using audemes, although it took them more attempts and more time than average. This indicates that audemes provided a richer experience

and helped participants immerse themselves, and perform better than alt text.

5. DISCUSSION AND LIMITATIONS

This paper shows that compared to alt text, enriching images with audemes significantly decreases the workload for screen reader users. Moreover, participants found it easier to recognize images using audemes as the mental ability and activity required to perceive and recognize an image was diminished significantly. Audemes also offered a sense of enjoyment and less pressure, which resulted in less effort and increased performance. Participants also found it difficult to process speech generated from alt text. On the other hand, they found listening to audemes to be pleasant. This likely resulted in a lower subjective workload, which increased the perceived usability of the prototype.

This article argues that the improvement of audemes compared to alt text, is attributed to the ability of the medium to communicate context in addition to content. Content includes the identity and properties of the object or events in the image, while context carries additional information about the content, such as, non-verbal cues, emotions, and environmental information. This article argues that, alt text is appropriate for communicating the content, while audemes also communicate context. Participants found the contextual experience missing with alt text, in contrast to audemes, which provided rich informational cues about the image. This contextual information further helped clarify the content of the image. Consequently, users' accuracy when identifying the images was increased. This is an indication that audemes provide richer experiences by communicating information about the content and context. While the content of an image may be simple to include in alt-text, contextual information is more difficult to describe. This claim, however, is based on an initial observation, and thus should be further investigated.

Despite these results, this study, has several limitations. First, due to the difficulty in recruiting blind participants who are real screen reader users, the study simulated blindness with sighted participants, which influenced and introduced bias into the results. Blind people heavily depend on audio to substitute their lack of visual cues when interacting with the environment. This increases the effectiveness and efficiency with which they perceive and process auditory content compared to sighted people. Several studies report such claims, for example, compared to sighted people, blind people better utilize auditory information [14], they process auditory language stimuli faster [16], congenitally have enhanced processing of speech [22], and they show better perception of degraded speech with equivalent hearing conditions [23]. Hence, the outcomes of this study may differ with blind people. However, there are three significant results that offer a degree of confidence that similar outcomes could be found when tested with real screen reader users: (1) all participants unanimously favored audemes compared to alt text in terms of overall reduced workload, (2) the prototype with audemes showed higher levels of usability, and (3) audemes helped users achieve higher levels of image comprehension.

The second limitation of this study is the choice of the images included in this study, which make them suitable for comparison between audemes and alt text. However, images found on the Web vary and often depict content that is difficult to represent using non-speech sounds. For example, it would be difficult to develop non-speech sounds to communicate an image containing numerical information, such as, prices, serial numbers, codes, dates and times. Also, it is difficult to represent color, size, structure and texture of various objects. This indicates the need to develop a complex

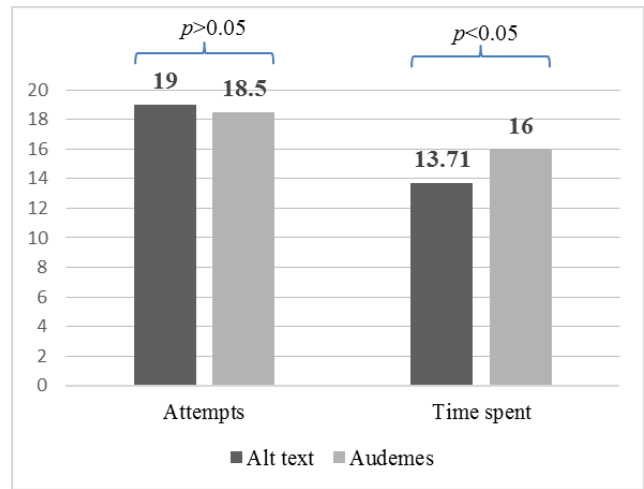


Figure 6. Number of attempts to recognize images and the time spent on each prototype.

audeme vocabulary and ontology that involves extensive training for the user. Additionally, although in this study we compared audemes and alt text in isolation, future research could test how and to what extent alt text and audemes complement one another.

Third, the length of the audemes for this study was set to four seconds to match the length of the speech representing the alt text. This length, however, might have influenced the results of the study, considering that other studies suggest audeme length to be up to seven seconds [38]. Moreover, the length of the audemes was a topic that was commented on by some of the participants.

6. CONCLUSION AND FUTURE WORK

In this study, we investigated the performance of audemes as compared to alt text for accessing images on the Web using screen readers. Overall, the findings indicated that audemes performed better across many dimensions. Specifically, they required lower mental and temporal demand, and resulted in less effort and frustration and better task performance. Moreover, audemes contributed to higher levels of accuracy and lower errors in image recognition. They were also perceived as more engaging compared to alt text delivered using synthetic speech. Additionally, audemes were found to deliver richer information by communicating the context in addition to the content of the image. These factors influenced the website's usability, as the prototype with audemes was rated more usable than the prototype with alt text.

For future work, this article suggests investigating audemes in terms of delivering content and context. Additionally, this article recommends measuring the effect of audeme length on the process of image recognition. Finally, this article suggests replicating the study using blind participants to increase validity.

7. REFERENCES

- [1] Kurniawan, S.H., Sutcliffe, A.G., Blenkhorn, P.L. and Shin, J.E., 2003. Investigating the usability of a screen reader and mental models of blind users in the Windows environment. *International Journal of Rehabilitation Research*, 26(2), pp.145-147.
- [2] Blanck, P., 2014. The struggle for web eQuality by persons with cognitive disabilities. *Behavioral Sciences & the Law*, 32(1), pp.4-32.

- [3] Chisholm, W., Vanderheiden, G. and Jacobs, I., 2001. Web content accessibility guidelines 1.0. *Interactions*, 8(4), pp.35-54.
- [4] World Wide Web Consortium, 2008. Web content accessibility guidelines (WCAG) 2.0.
- [5] Evett, L. and Brown, D., 2005. Text formats and web design for visually impaired and dyslexic readers—Clear Text for All. *Interacting with computers*, 17(4), pp.453-472.
- [6] Slatin, J.M., The art of ALT: toward a more accessible Web. *Computers and Composition*, 2001. 18(1): p. 73-81.
- [7] Ferati, M. and Sulejmani, L., 2016, July. Automatic Adaptation Techniques to Increase the Web Accessibility for Blind Users. In *International Conference on Human-Computer Interaction* (pp. 30-36). Springer International Publishing.
- [8] Rømen, D. and Svanæs, D., 2012. Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users. *Universal Access in the Information Society*, 11(4), pp.375-385.
- [9] Esposa Jr, J.I., 2008. How to write a good alt text.
- [10] Buxton, W., 1989. Introduction to this special issue on nonspeech audio. *Human Computer Interaction*, 4(1), pp.1-9.
- [11] McGee-Lennon, M., Wolters, M., McLachlan, R., Brewster, S. and Hall, C., 2011, May. Name that tune: musicons as reminders in the home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2803-2806). ACM.
- [12] Bly, S., 1982. *Sound and computer information presentation* (No. UCRL-53282). Lawrence Livermore National Lab., CA (USA); California Univ., Davis (USA).
- [13] Papastavrou, J.D. and Lehto, M.R., 1996. Improving the effectiveness of warnings by increasing the appropriateness of their information content: some hypotheses about human compliance. *Safety science*, 21(3), pp.175-189.
- [14] Niemeyer, W. and Starlinger, I., 1981. Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness II. Central functions. *Audiology*, 20(6), pp.510-515.
- [15] Sánchez, J., Lumbreras, M. and Cernuzzi, L., 2001, March. Interactive virtual acoustic environments for blind children: computing, usability, and cognition. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems* (pp. 65-66). ACM.
- [16] Röder, B., Rösler, F. and Neville, H.J., 2000. Event-related potentials during auditory language processing in congenitally blind and sighted people. *Neuropsychologia*, 38(11), pp.1482-1502.
- [17] Doucet, M.E., Guillemot, J.P., Lassonde, M., Gagné, J.P., Leclerc, C. and Lepore, F., 2005. Blind subjects process auditory spectral cues more efficiently than sighted individuals. *Experimental brain research*, 160(2), pp.194-202.
- [18] Barker, P.G. and Manji, K.A., 1989. Pictorial dialogue methods. *International Journal of Man-Machine Studies*, 31(3), pp.323-347.
- [19] Smither, J.A.A., 1993. Short term memory demands in processing synthetic speech by old and young adults. *Behaviour & Information Technology*, 12(6), pp.330-335.
- [20] Slowiaczek, L.M. and Nusbaum, H.C., 1985. Effects of speech rate and pitch contour on the perception of synthetic speech. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 27(6), pp.701-712.
- [21] Blattner, M.M., Sumikawa, D.A. and Greenberg, R.M., 1989. Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), pp.11-44.
- [22] Gaver, W.W., 1986. Auditory icons: Using sound in computer interfaces. *Human-computer interaction*, 2(2), pp.167-177.
- [23] Gaver, W.W., 1997. Auditory interfaces. *Handbook of human-computer interaction*, 1, pp.1003-1041.
- [24] Edwards, A.D., 1997. Using sounds to convey complex information. In *Contributions to Psychological Acoustics: Results of the Seventh Oldenburg Symposium on Psychological Acoustics*, Oldenburg (pp. 341-358).
- [25] Kramer, G., 1993. *Auditory display: Sonification, audification, and auditory interfaces*. Perseus Publishing.
- [26] Roth, P., Petrucci, L.S., Assimacopoulos, A. and Pun, T., 2000. Audio-haptic internet browser and associated tools for blind users and visually impaired computer users. In: C. Germain and O. Lavielle and E. Grivel. *COST 254 Intelligent Terminals, Workshop on Friendly Exchanging Through the Net*. p. 57-62
- [27] Ramloll, R., Yu, W., Brewster, S., Riedel, B., Burton, M. and Dimigen, G., 2000, November. Constructing sonified haptic line graphs for the blind student: first steps. In *Proceedings of the fourth international ACM conference on Assistive technologies* (pp. 17-25). ACM.
- [28] Brown, L.M. and Brewster, S.A., 2003. Drawing by ear: Interpreting sonified line graphs. Georgia Institute of Technology.
- [29] Mansur, D.L., Blattner, M.M. and Joy, K.I., 1985. Sound graphs: A numerical data analysis method for the blind. *Journal of medical systems*, 9(3), pp.163-174.
- [30] Auvray, M., Hanne-ton, S. and O'Regan, J.K., 2007. Learning to perceive with a visuo—auditory substitution system: localisation and object recognition with 'The Voice'. *Perception*, 36(3), pp.416-430.
- [31] Wilson, J., Walker, B.N., Lindsay, J., Cambias, C. and Dellaert, F., 2007, October. Swan: System for wearable audio navigation. In *Wearable Computers, 2007 11th IEEE International Symposium on* (pp. 91-98). IEEE.
- [32] Walker, B.N. and Lindsay, J., 2006. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2), pp.265-278.
- [33] Sanchez, J., 2010, April. Recognizing shapes and gestures using sound as feedback. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3063-3068). ACM.
- [34] King, A.R. and Dr Supervisor Evans, 2007. *Re-presenting visual content for blind people*. University of Manchester.
- [35] Krishnan, K.G., Porkodi, C.M. and Kanimozhi, K., 2013, April. Image recognition for visually impaired people by sound. In *Communications and Signal Processing (ICCSP), 2013 International Conference on* (pp. 943-946). IEEE.

- [36] Nikolakis, G., Moustakas, K., Tzovaras, D. and Strintzis, M.G., 2005. Haptic representation of images for the blind and the visually impaired. *CD-ROM Proc. HCI International*.
- [37] Ferati, M., Pfaff, M.S., Mannheimer, S. and Bolchini, D., 2012. Audemes at work: Investigating features of non-speech sounds to maximize content recognition. *International Journal of Human-Computer Studies*, 70(12), pp.936-966.
- [38] Mannheimer, S., Ferati, M., Bolchini, D. and Palakal, M., 2009. Educational sound symbols for the visually impaired. *Universal Access in Human-Computer Interaction. Addressing Diversity*, pp.106-115.
- [39] Mannheimer, S., Ferati, M., Huckleberry, D. and Palakal, M.J., 2009. Using Audemes as a Learning Medium for the Visually Impaired. In *HealthINF* (pp. 175-180).
- [40] Ferati, M., Mannheimer, S. and Bolchini, D., 2009, October. Acoustic interaction design through audemes: experiences with the blind. In *Proceedings of the 27th ACM international conference on Design of communication* (pp. 23-28). ACM.
- [41] Ferati, M., Mannheimer, S. and Bolchini, D., 2011, October. Usability evaluation of acoustic interfaces for the blind. In *Proceedings of the 29th ACM international conference on Design of communication* (pp. 9-16). ACM.
- [42] Meyer, C., Ahsan, S., Gupta, V., Parham, A., Patel, H. and Qureshi, M., 2014. Audemes: Exploring the Market Potential of a Sound Based Educational Tool.
- [43] Audeme dictionary. Available from: https://audemes.org/dictionary_2014/dictionary_2016.html.
- [44] Hart, S.G. and Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, pp.139-183.
- [45] Brooke, J., 1986. System usability scale (sus): A quick-and-dirty method of system evaluation user information. Digital equipment co ltd. *Reading, UK*.
- [46] Hart, S.G. and C.D. Wickens, *Workload assessment and prediction*, in *Manprint*. 1990, Springer. p. 257-296.
- [47] Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), pp.4-7.
- [48] Hugdahl, K., Ek, M., Takio, F., Rintee, T., Tuomainen, J., Haarala, C. and Hämäläinen, H., 2004. Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds. *Cognitive brain research*, 19(1), pp.28-32.
- [49] Gordon-Salant, S. and Friedman, S.A., 2011. Recognition of rapid speech by blind and sighted older adults. *Journal of Speech, Language, and Hearing Research*, 54(2), pp.622-631.

1. Bleicher, J. and J. Bleicher, *Contemporary hermeneutics: Hermeneutics as method, philosophy and critique*. 1980: Routledge & Kegan Paul London.
2. Blanck, P., *eQuality: The struggle for web accessibility by persons with cognitive disabilities*. 2014, New York: Cambridge University Press.
3. Chisholm, W.V. and G. Jacobs, *I.(editors)(1999) Web Content Accessibility Guidelines (WCAG)*. W3C Recommendation, 1999. **5**.
4. Consortium, W.W.W., *Web content accessibility guidelines (WCAG) 2.0*. 2008.
5. Evett, L. and D. Brown, *Text formats and web design for visually impaired and dyslexic readers—clear text for all*. *Interacting with computers*, 2005. **17**(4): p. 453-472.

6. Slatin, J.M., *The art of ALT: toward a more accessible Web*. *Computers and Composition*, 2001. **18**(1): p. 73-81.
7. Rømen, D. and D. Svanæs, *Validating WCAG versions 1.0 and 2.0 through usability testing with disabled users*. *Universal Access in the Information Society*, 2012. **11**(4): p. 375-385.
8. Esposa Jr, J.I., *How to write a good alt text*. 2008.
9. McGee-Lennon, M.R., M. Wolters, and T. McBryan, *Audio reminders in the home environment*. 2007.
10. McGee-Lennon, M., et al. *Name that tune: musicians as reminders in the home*. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2011. ACM.
11. Bly, S., *Sound and computer information presentation*. 1982, Lawrence Livermore National Lab., CA (USA); California Univ., Davis (USA).
12. Papastavrou, J.D. and M.R. Lehto, *Improving the effectiveness of warnings by increasing the appropriateness of their information content: some hypotheses about human compliance*. *Safety science*, 1996. **21**(3): p. 175-189.
13. Niemeyer, W. and I. Starlinger, *Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness. II. Central functions*. *Audiology*, 1981. **20**(6): p. 510-515.
14. Röder, B., F. Rösler, and H.J. Neville, *Event-related potentials during auditory language processing in congenitally blind and sighted people*. *Neuropsychologia*, 2000. **38**(11): p. 1482-1502.
15. Doucet, M.-E., et al., *Blind subjects process auditory spectral cues more efficiently than sighted individuals*. *Experimental brain research*, 2005. **160**(2): p. 194-202.

16. Barker, P. and K.A. Manji, *Pictorial dialogue methods*. International Journal of Man-Machine Studies, 1989. **31**(3): p. 323-347.
17. Smither, J.A.-A., *Short term memory demands in processing synthetic speech by old and young adults*. Behaviour & Information Technology, 1993. **12**(6): p. 330-335.
18. Slowiaczek, L.M. and H.C. Nusbaum, *Effects of speech rate and pitch contour on the perception of synthetic speech*. Human Factors: The Journal of the Human Factors and Ergonomics Society, 1985. **27**(6): p. 701-712.
19. Mannheimer, S., et al. *Educational sound symbols for the visually impaired*. in *International Conference on Universal Access in Human-Computer Interaction*. 2009. Springer.
20. Ferati, M. 2014; [Adumenes dictionary]. Available from: https://audemes.org/dictionary_2014/dictionary_2016.html.
21. Hart, S.G. and L.E. Staveland, *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*. Advances in psychology, 1988. **52**: p. 139-183.
22. Hugdahl, K., et al., *Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds*. Cognitive brain research, 2004. **19**(1): p. 28-32.
23. Gordon-Salant, S. and S.A. Friedman, *Recognition of rapid speech by blind and sighted older adults*. Journal of Speech, Language, and Hearing Research, 2011. **54**(2): p. 622-631.