An Evaluation of a Prototype Sensory Substitution App for Visually Impaired People

HELEN PETRIE[1], ALISTAIR BUCHANAN[2] AND ANNA BRAMWELL-DICKS[3]

[1] Department of Computer Science, University of York, UK
e-mail: helen.petrie@york.ac.uk
[2] Cadscan Limited, Chester, UK
e-mail: ab@spiral-scratch.com
[3] Department of Theatre, Film and Television, University of York, UK
e-mail: anna.bramwell-dicks@york.ac.uk

Abstract

There are 40 million blind people worldwide and over 120 million people with significant visual impairments. Their use of mobile smart phones and tablets fits the same profile as the general population; most younger people with sight loss are very conversant with modern technologies, and increasingly older visually impaired people are too. Smartphones and tablets now have accessibility features built in, using a gesture and audio interface designed specifically for visually impaired people to replace the tapping on visual icons and swiping used by sighted people. These features provide easy interaction with applications but are limited to text, buttons and switches, and provide no information about visual content such as maps or camera imagery. This study took a user-centred design approach to develop a prototype for an app that could enable visually impaired users to interact naturally with visual information presented on the screen of a mobile device using touch and sound together, a form of what is often known as sensory substitution. The user scans the information presented on screen by running their finger over a two-dimensional array which generates a set of audio signals that represent the image's individual pixels. Pitch, volume and pan were used to present different pixel values and locations conveying different types of visual information to the user. Four totally blind participants evaluated the prototype. They were shown three images, including simple 2D and more complex 3D images. All the participants found it easy to interpret the information in the images, although the more complex images would need speech tags as well as the sound information to make the images understandable. All participants thought the app had major potential for presenting image information to blind people, particularly in the contexts of education, culture and personal navigation.

1 Introduction

Approaches for presenting visual information to blind and visually impaired people has been an important area of research for many years. Many of the existing approaches require tactile images to be created, which is an expensive and time-consuming process. Visually impaired people interact with tactile images by scanning their fingers over the image as a form of sensory substitution. The user uses their senses of proprioception and touch together to convert positioning and depth information they have received through touching the tactile image into a mental model of the image's content. However, tactile images are expensive to print, and if someone wants to access a lot of visual information, then they would require a lot of tactile images to be printed. This project focused on whether it would be possible to combine the user's sense of proprioception with their sense of hearing instead of using their sense of touch to interpret the content of 2D and 3D images through the development of a prototype for the SoundTact app.

A visual sensory substitution device (SSD) typically takes visual information and converts it to an alternative mode allowing a visually impaired person to receive the information using their senses of touch or hearing instead of vision. The vOICe system (Meijer, 1992) is an example of a SSD that takes visual information from an image and converts it to an auditory signal. With the development of relatively cheap touch screen technology in tablet form, SSDs have extended so that instead of converting visual information into a single alternative modality, the information can be dual-coded into data that is transmitted to the user's senses of both touch and hearing together, rather than one or the other. For example, in the EdgeSonic project the user interacts with an image displayed on a tablet using their sense of touch and hearing to interpret its content (Yoshida et al. 2011) and similarly with exploring maps (Delogu et al. 2010). However, much of this research that has incorporated a multi-modal approach within the SSD does not seem to provide evidence of having undertaken a user-centred design approach to the development of the interactions, sounds and interface. Instead, these studies report on blind and visually impaired users being brought in to evaluate the success of the final system, rather than being incorporated at the design stage. Our approach, with this SoundTact project, has been to involve potential users of the system from the very early stages of the design process in order to respond to their needs and develop appropriate requirements for all aspects of the system.

2 Design and Development of the SoundTact App

We used an agile approach to develop the SoundTact app, since we had little understanding of how well the interface would be received, or how the more general aspects of the system when operating on an iPad would affect usage.

We began by developing a set of initial user requirements and/or design ideas that we wished to explore the suitability of. These ideas focused on two different situational contexts, one for at home use and the other for use during navigation. We developed use cases for both contexts, although we had some practical concerns about the suitability of the app for navigation if both hands would be required to use it as blind and visually impaired people will often need the use of one hand to hold their cane or dog's lead. These initial user requirements and design ideas related to the general interface (i.e. integration with VoiceOver), the type of audio cues that could be manipulated to convey information to the user (e.g. pitch, volume, timbre and stereo pan) and different options for the type of audio hardware (e.g. headphones, bonephones or through the device's integrated speaker hardware).

These initial set of user requirements and design ideas were incorporated into the development of the first version of the SoundTact app. This early stage prototype was used as a prompt for discussion in interviews with 4 totally blind participants. The prototype allowed us to demonstrate the concept behind the app which enabled us to further explore the user and system requirements with input from these prospective users.

This first version of the prototype included just the sphere image (Figure 1). When the user tapped the screen on the image with their finger a sine wave was generated and played to the user where the pitch of the since wave was mapped to the intensity of the image. A higher intensity

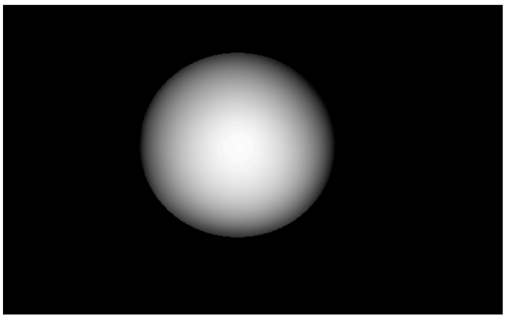

Figure 1: The sphere image

in the image indicated that this part of the image has less depth (i.e. is closer to the user) and was given a higher pitch.

These interviews expanded our initial set of requirements and identified a number of key features that the app would have to incorporate. These requirements included a need for a distinct background sound so that when the user had their finger over the black areas of the image (i.e. areas where there is no image information) it was not just silent, but instead a different type of sound would indicate this is an area of the image where there is no image information. Another key requirement that was identified during these interviews was a need for a clear and distinct edge sound so the user knows when their finger had reached the edge of the image's window. This feature is important because the edge of the image does not align with the edge of the device. Also, the participants felt that it was important the user could generate sounds by scrolling with their finger (i.e. by keeping their finger in contact with the screen at all times) instead of using a tapping motion to generate the sound as this would allow them to retain contextual information through proprioception.

Following this initial set of interviews, where we also discussed the different hardware options and potential use cases, we developed a second, more refined version of the app. This second implementation incorporated different audio to image feature mappings as well as more complex images.

A number of different 2D and 3D images were incorporated within the app. These images contain different levels of complexity of information, including 3D images of a sphere (Figure 1), a horse (Figure 3), a Buddha figurine (Figure 4). and a 2D map of a section of Campus West at the University of York (Figure 5).

For the 3D images, 3 different mappings of audio parameter to image feature were implemented. The first auditory mapping approach decreased the pitch of the sine wave as the depth value in the image became closer to the user. The second approach to mapping increased the amplitude (i.e. the volume) of the sine wave as the image feature had a smaller distance. While, the third mapping was a fusion sound that combined both pitch and amplitude manipulation, so that areas of the image that are closest to the user (i.e. have the shortest depth) were both at a lower pitch and louder. In the second version of the prototype a background sound was also included to highlight to the user when they were touching an area on screen that did not contain any image information. Further, a distinct edge sound was implemented to indicate to the user when they had reached the edge of the image with their finger.

Figure 3: The Buddha image

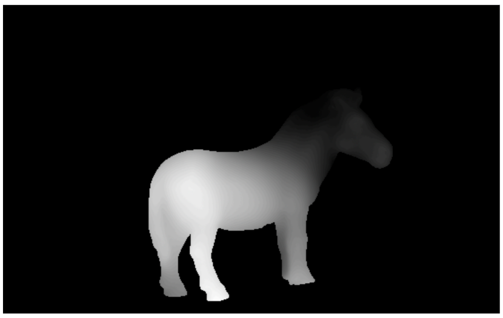

Figure 4: The horse image

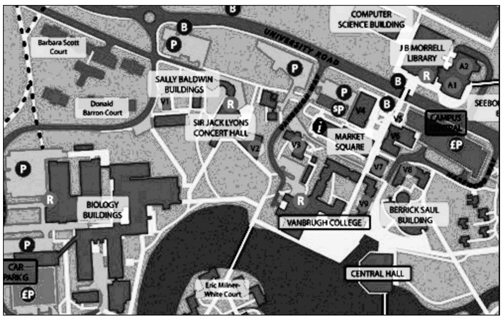

Figure 5: The University of York Campus West map image

3 Prototype Evaluation

Method

Four totally blind participants evaluated the second version of the SoundTact app. All participants were male, with ages ranging from 36 to 65 years. Two participants have been blind since birth and two lost their sight as adults. All participants were competent to expert users of assistive technology.

The participants were asked to interact with the different 2D and 3D images using the 3 different audio mapping approaches and provide feedback on the app. Figure 6 shows a visually impaired user interacting with the campus map image.

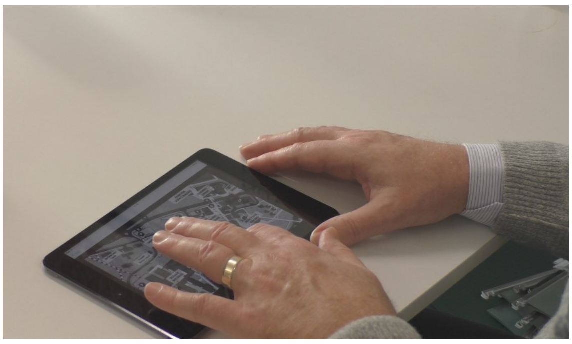

Figure 6: Visually impaired user testing map image

The first image used in the evaluation of the second version of the SoundTact app was the image of a sphere in the centre of the screen (Figure 1). This simple image was used to illustrate the basic features of the system – the edge sound, the background sound and the three main sound possibilities (volume change only, pitch change only, fusion of volume and pitch change). Participants were asked to try to identify what the image was, to trace the edge of the sphere and to find the point of the sphere that was closest to them. Participants were then asked about the different basic aspects of the system, particularly which of the three possibilities for the main sound manipulation they found to be most effective.

Next, the participants were asked to use the image of a statue of the Buddha in the centre of the screen (Figure 3). This is a much more complex image than the sphere, with a base and the legs of the Buddha closer to the viewer than the torso, and the head at a middle distance. Participants were asked to explore the image and identify which elements were closer to them. They were then told what the image was and given a guided tour of the image. Participants were then asked for their views about the use of the system for exploring these more complex images of 3D statues, possibly in an educational or cultural context.

The participants were then asked to interact with the image of a horse (Figure 4). Again, this is a complex image with the horse's bottom closest to the user and the head turned towards the right of the image. The participants were told that this was an image of a figurine of an animal and asked if they could identify what type of animal it was, and to find it's legs, face and body.

The final image used in the evaluation was a 2D image of a campus map of a small section of Campus West at the University of York (Figure 5). Participants were introduced to the different sounds created to the footpaths, roadways, buildings and the lake. A Wizard of Oz audio description was provided of building names and important features. Participants were asked to follow the footpath along the lake and explore the overall layout of the university campus. Participants were then asked for their views about the use of the system for exploring maps, both before making a journey and during the journey itself.

Finally, participants were asked for their views about how the system might be developed further and their overall opinion of the system.

Results

All the participants preferred the fusion sound in comparison to the pitch change only or the volume change only. Three participants felt it was easier to detect the changes in the fusion sound in comparison to the other two sound types. One participant felt there was little difference in detection of changes between the fusion and the pitch only sound, but that both of these were preferable in comparison to the volume only sound. One participant felt that the fusion sound was aesthetically more pleasing compared to the other two sound types.

All the participants liked the background sound, particularly in the fusion sound version. Two commented that it was effective in telling you that you were on the display, without creating a sound that drew the attention too much. All the participants liked the inclusion of an edge sound, although two of the participants thought that the particular sound chosen was a bit too intrusive. The participants who had taken part in the first evaluation were very pleased that they could now swipe the screen and hear the sounds continuously, rather than having to tap repeatedly.

All the participants found it easy to trace the edge of the sphere and to find the point closest to the viewer. The two participants who have been blind since birth were less confident about the actual shape of the object, as the pitch/volume change of the curvature of the object puzzled them. This is not surprising, given they would have only had access to tactile diagrams with very limited height differentials. Some training with real objects and their SoundTact images would be very useful for congenitally blind people to explain these concepts. The two participants who lost their sight as adults had no difficulty understanding the shape.

For the Buddha image, all the participants found it easy to trace the outline of the image and all immediately understood that the base of the statue was closest to the viewer. When told it was a statue of the Buddha, two of the participants (one adult blind, one congenitally blind) worked out the basic anatomy, one participant in detail. One of the congenitally blind participants thought that the lower pitch/volume head was "less detailed, not so in focus" as the higher pitch/volume areas, showing his lesser understanding of the visual world. Again, training in the system could remedy these issues.

The resolution of the map made it difficult to explore individual elements, this could easily be fixed by increasing the resolution and introducing scrolling to the display (suggested by one of the participants). Where necessary, participants were given some assistance in moving around the key features of the map.

Participants found it very interesting and helpful that the different elements of the map had different sounds. Quite accidently, the lake produced a very auditorily appropriate sound, a deep pitch with some reverb which sounded a bit like a swimming pool. In contrast, the footpaths produced a high pitched sound, which participants thought was appropriate as it was most salient and easy to follow. They thought that a more striking sound was needed for the roadway, as this would be a danger point for blind pedestrians.

Overall, all the participants thought the campus map had a great deal of potential. They thought that visually impaired students and visitors could study it before going out on campus, to give themselves an overview mental map of the campus. Students could also use the map to find routes to places that they did not visit regularly. Two of the participants thought they would take the map out with them, if the iPad could fit in a pocket and consult it using headphones. All the participants thought the map would need some simplification, verbal labels for buildings, some tweaking of the sounds and an "audio legend".

The basic principle of translating 2D (specific map feature information) or 3D information (depth) from images into sound has worked very well. All the participants in the evaluation found it easy to understand and thought it had major potential for presenting image information to visually impaired people.

All the participants thought the SoundTact system has major potential for presenting image information to blind people, for both the 2D and 3D information. They see the applications particularly in the educational, cultural and personal navigation areas.

One of the participants who works for a major organization of visually impaired people, suggested that the system could have a very useful application in the music education of visually impaired children. Visually impaired children are now taught as much as possible with their sighted peers and learn music with in a sighted context. However, the physical notation of music does not have any meaning to them. This participant thought that if it could be presenting using the SoundTact system, it would greatly help in learning music. He even suggested that one could press the notes and hear an appropriate sound.

4 Conclusion

This early stage work on the SoundTact project has taken a user-centred approach to the development of a prototype app that converts visual information to sound in real time when the user taps an area of an image. The concept has proven to be successful with blind users able to understand different features of both simple and complex 2D and 3D images. The research has identified a number of key user requirements for any such system, and demonstrates that there is clear potential for an app that exploits both touch and sound together to be used as an alternative to tactile images.

References

Meijer, P B L, 1992 "An experimental system for auditory image representations" *IEEE Transations on Biomedial Engineering* **39** 112 – 121

Yoshida, T., Kitani, K M, Koike, H., Belongie, S., Schlei, K., 2011 "EdgeSonic: Image Feature Sonification for the Visually Impaired" *Proceedings of Augmented Human Conference,* Odaiba, Tokyo

Delogu, F., Palmiero, M., Federici, S., Plaisant, C., Zhao, H., Belardinelli, O., 2010 "Non-visual exploration of geographic maps: Does sonification help?" *Disability and Rehabilitation: Assistive Technology,* **5(3)** 164 – 174.