INTERACTIVE SPATIAL AUDITORY DISPLAY OF GRAPHICAL DATA

Timothy Neate & Andy Hunt

Audio Lab
University of York
York
UK

tdjneate@gmail.com
andy.hunt@york.ac.uk

ABSTRACT

This paper describes a series of experiments designed to test auditory display techniques for the interactive sonification of graphical features on a tablet computer. The aim of these experiments was to evaluate new sonic methods for finding graphical features on a tablet computer screen for both regular size, and extended (larger than physical screen size) displays. A series of tests was designed to evaluate the techniques, and determine the effectiveness of binaural information in a series of goal-oriented searching tasks. The results show that the use of binaural audio gives faster location times, allows better searching techniques, and yields an improvement in locational ability when using auditory and visual information in tandem.

1. INTRODUCTION

In an era where displays are smaller, and screen real-estate is limited, the Human-Computer Interaction community is continuously exploring new approaches to tackle the challenge of fitting more content on less screen space. In the field of sonification, several approaches have been tested in an effort to expand screen displays into the auditory domain. With increased audio processing capabilities and interaction modes on smaller and more portable devices, the field of auditory display is becoming a forerunner in presenting additional information by means of multimodality.

One approach used to extend the visual domain is to place all non-essential information into a spatialized auditory field, with different zones of priority [1]. Approaches such as these allow the user to concentrate on information that is of high importance first, and then deal with information that is of less importance.

Other methodologies simply present all the visual information as it is, with a direct mapping into a raw auditory form [2]. Techniques such as this have been found to be highly successful in enabling those with visual impairments to gain a better idea of their surroundings, but tend to require extensive training on behalf of the user [3]. An alternate approach to this is to filter the data that we seek in the visual domain, before transforming it into the auditory domain [4]. This approach favours a goal-oriented searching task, where the user already knows what they are looking for, but does not fare well when representing raw image data.

This paper describes a goal-oriented approach to enhancing visual representation by means of interactive spatial auditory display. It is found that the approach described can aid users in locating graphical features on displays of different sizes with minimal, or even no visual cues. The work's novelty is derived from binaurally sonifying the relationship between the interaction of the user and the graphical features on a tablet computer, as well as the techniques developed to handle this interaction. As far as the authors know, this type of interaction to feature mapping has not been implemented with spatialized audio before.

This paper first covers the relevant background work in each of the relevant topics of this paper. It then goes on to discuss the implementation of the auditory display and interaction techniques that were developed. The next three sections then discuss the test's setup, the methods of user testing, and the results of the tests. This is done such that the testing procedure can be reproduced or scrutinised. A discussion section then outlines the results' implications, and then to finalise the paper a conclusions and further work section summarises the paper and discusses potential further work in the area.

2. BACKGROUND

This study brings together three main areas of research in the auditory display community – the sonification of graphical data, spatial audio sonification, and new interfaces for auditory display. The following three subsections give a background on each of these areas.

2.1. The Sonification of Graphical Data

In recent years there have been several methods devised for the transformation of graphical data into the auditory domain. Generally, these take two approaches: transforming all of the graphical data into a complex auditory field that envelops the listener holistically [2] [5], and a goal-driven exploratory methodology where the graphical data is first filtered – the user being left only with the specific features required [4] [6] [7].

Meijer's approach [2] involves scanning a video feed, mapping frequency to the height of a pixel in the display, and mapping the brightness of the pixel to the amplitude of a sinusoidal oscillator. This results in a highly reactive, complex auditory field that is best used to describe complex images or videos. Approaches such as this require the user to learn the mappings over extended periods of time. On the other hand, Bologna's work [8] endeavours to filter specific colours and only transform them into the auditory domain, resulting in a system that is easier to use than Meijer's, but can only provide limited goal-oriented information.

2.2. Spatial Audio Sonification

Spatialized audio has been used numerous times by those wishing to transform information meaningfully into the auditory domain [6] [9] [7] [1]. It is a highly suitable method to use for representing a physical direction in Cartesian space because of our innate ability to determine the location of a spatialized source within 11.8 degrees [10, pg. 39].

Binaural audio, or the notion of portraying 3D sound over headphones, has become an invaluable tool in the auditory display community for representing spatiality [7] [6]. It allows for effective, cheap, and portable 3D audio – meaning that we can present complex auditory fields outside of the lab environment.

2.3. Interfaces for Auditory Display

As technology has become more powerful, its design has become more suited to our interactions. In the area of interactive sonification, we always try to develop for the best platforms we can at the time. From the first modern personal computers, up until a few years ago, this has almost exclusively involved interaction by mouse or similar PC peripheral. Touch interaction has not been sophisticated enough to become a viable portable platform for development until a little over 3 years ago, with the rebirth of tablet computer – Apple's iPad.

Now that we can use an extensive array of different interaction techniques to explore data, there are fewer limits in the world of interactive sonification – the human-computer interaction loop has become stronger, and there is less cognitive strain on behalf of the user – as they are free to think about what they are interacting with, and concern themselves less with how to operate the system.

3. PROPOSED SYSTEM

The system developed allows the user to experience an image in the auditory domain. When they interact with the image on an iPad by touching the screen they experience auditory feedback that indicates features around their point of touch – their colour represented by different sounds. They are then able to locate these features by moving their touch location around the screen and using the various auditory parameters:

- **Binaural panning** to describe its direction
- a **pulse train** to describe its distance; and
- an **alert** when they find it.

Additionally, other parameters have been developed to assist users when locating multiple features, or when searching on extended displays. The implementation of the experimental system can be broken down into three main parts – the location of the image feature, the user interaction, and the auditory feedback. This system is depicted in Figure 1.

3.1. Touch to Image Feature Calculation

The image-processing algorithm used in this implementation finds the average Cartesian point of a specific colour by summing the positions of pixels' 'x' and 'y' coordinates within a specific threshold set by the user. Using iOS touch delegate methods it is then possible to track the user's touch. Once this has been tracked, it is possible to find the vector between the user's touch, and the image feature, as outlined in Figure 2 and Equation 1 , where 'N' is

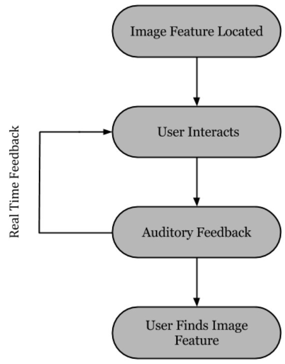

Figure 1: Interactive Graphical sonification system

the number of pixels within the filter in the image, 'P' is the filtered pixel's coordinate in the respective direction, and 't' signifies a touch point by the user. The letter 'd' denotes the dimension this algorithm travels through – this will be either 'x' (left to right), or 'y' (top to bottom). The end result of this algorithm is an integer for both dimensions that represents the average Cartesian coordinate of the pixels filtered.

$$\Delta_d = \frac{\sum_{P_{d=0}}^{N_d} P_d}{N_d} - t_d \tag{1}$$

The angle (Θ) of the vector is then determined using Equation 2, and the magnitude (M) with Equation 3.

$$\Theta = \frac{180}{\pi} \arctan\left(\frac{\Delta x}{\Delta y}\right) - \frac{\pi}{2} \tag{2}$$

$$M = \sqrt{\Delta x^2 + \Delta y^2} \tag{3}$$

As Equation 3 calculates the angle of the vector from '12 o'clock' for the user's touch, this allows for the audio processing system to project sources around the listener, with their finger as the 'central point'. Equation 3 is used to determine the magnitude (M) of the vector. This system is described in Figure 2.

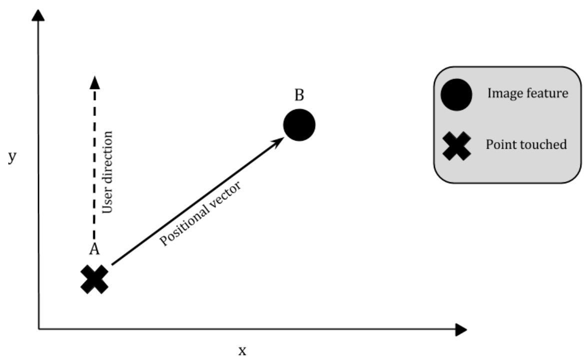

Figure 2: Vector between touch-point and image feature

3.2. Multi-touch Interaction

There needs to be a clear differentiation between when users want to activate the sound mapping, and when they need to move the virtual image around the screen display. After some tests with a small group of individuals it was established that one-finger touches should be used to activate and update the sound engine, and two fingers should be used to move the virtual image around the display. By doing this, it not only offers a clear differentiation between the two techniques, but also allows for simultaneous use of both techniques.

Additionally, a sound-mapping parameter was used for when a user extends the bounds of an image. It was decided that a 'boing' sound should be used for this. A simple oscillator with harmonics was panned binaurally, dependent on which side the user had scrolled too much on – its frequency used to represent how much they had scrolled in excess of the scrollable screen.

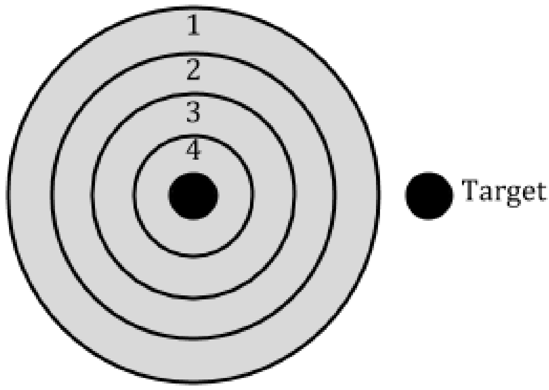

Figure 3: Example of four proximity zones

3.3. Auditory Display Mappings

The vectors calculated are used to drive the audio engine, which was written in Csound, and developed for iOS using the Csound-iOS API developed by Steven Li and Victor Lazzarini [11]. This allows the flexible audio processing capabilities of Csound to be combined with the touch interaction of the iOS platform. Several parameter mapping sonification methods were developed in Csound to provide interactive auditory feedback to the user. These are described below:

Pulse train – Pulse trains have been used to represent distance through sound with great success [12]. The decision was made to increase the pulsing as the users touch got closer to the image feature as it complements the instant real-time feedback we are familiar with when interacting with real-world systems – the closer the touch, the more frequent the pulse, therefore the faster the feedback to the user as they get closer. From this, it is possible for the user to make fine adjustments of position towards the shape faster, without having to wait for the next pulse. Several proximity zones were devised, as shown in Figure 3. Each proximity zone used a different speed of pulse train – the higher the proximity zone, the faster the pulse train. On an image the same size as the iPad screen, typically 9 proximity zones were used. For larger images, the number of proximity zones was scaled up accordingly so that the mappings remained consistent.

The sound used for the pulse depends on the colour of the visual object being represented. For the purposes of the experiment, four main synthesizers were built – a noise-based synth for BLACK, and three subtractive synths tones differing in pitch to represent the colours RED, BLUE, and GREEN. It would be possible to scale this up to more synthesizers for additional colour mappings.

Binaural Panning – The user was placed in the auditory field by touching the interface. This allowed them to them move through the auditory field with the various image features appearing around them, panned binaurally. The HRTFs, made by Bill Gardner and Keith Martin at the MIT Media Lab [13], were used as they are high quality measurements made using a KEMAR (binaural dummy head) designed according to the mean anatomical size of the population, therefore resulting in a good average head [14]. The Csound opcode 'hrtfmove2'[15] was used to interpolate the source (pulse train) around the listener as it offers good quality imaging, with minimal processing.

Alert Sound – A simple alert sound was used when the user ran their finger over specific areas to indicate that they have found something. This alert sound was ideal when the participants of the experiment were undertaking goal-oriented tasks as it provided them with some element of closure.

Volume – it was noted in the initial designs that when there were multiple sources of sound i.e., multiple image features detected, that the sounds often clashed. Therefore, a volume parameter was developed. This used the proximity zones described previously, but instead of controlling the speed of the pulse train, the volume parameter was used to control the perceived amplitude of sound emanating from the image source, increasing as the user travels closer to it.

4. EXPERIMENTAL SETUP

To evaluate the effectiveness of the techniques developed, the participants were divided into two groups; a 'control' group – B, and a 'treatment' group – A. Group B were provided with some auditory display parameters to find a specific feature in the image. Group A, however, had additional parameters – with the aim of testing whether the techniques developed affected the performance (determined by speed of location) of the participants undertaking the tasks.

For some of the experiments, the participants needed to be visually restricted so that they could not see the iPad screen, and instead only operated by touch and sonic feedback. Typically, blindfolds are used for this purpose. However, blindfolds can often be considered unethical, and may cause distress in the user. Therefore a visual restrictive device, similar to the device used by Fernström [16], was used. A simple visual restrictor (shown in Figure 4) was created out of a cardboard box and some cloth to ensure that the user could not see through it. The iPad could then be placed in the box and the user could freely interact with it.

Each participant was given a small training session at the beginning of Test 1 and Test 2. The training allowed them to see an example of the tests they were about to undertake such that they knew the relevant auditory and interaction parameters to complete the remaining tests. They were encouraged to practice until they felt that they could find the required image features without visual cues.

Figure 4: Makeshift visual restriction device

5. USER TESTING

Willing participants were asked to undertake a series of tests to examine the techniques developed. There were two main types of test – Test 1 and Test 2:

- **Test 1** focused on using auditory display techniques to portray images the size of a standard iPad screen; and
- **Test 2** involved extending the size of the image – using large scrollable images displayed on an iPad.

The demographics are first discussed such that an impression of the sample can be gained, then the rest of this section describes how each individual test was run, states its results, and discusses any significant findings.

In total 18 participants undertook the experiment – nine in each group, with an average age of 25.8 (standard deviation = 4.3). The group included people of British, Chinese, Dutch, Greek, American, Belgian, and Russian nationalities. In all, 12 were male, and six were female. The majority of the participants (13 out of 18) were from the Department of Electronics (University of York), predominantly in the Audio Lab, and the remaining five were from the Department of Computer Science. Due to the large number of people from the Audio Lab, the subject set included a relatively large number of musicians – 13 out of 18.

Some questions were asked specific to sound perception, binaural audio, and familiarity with tablet devices. It was found that three out of 18 participants claimed to have some form of sound-to-colour synaesthesia (the perception of one sense in the form of another), and all but one of the participants had experienced binaural audio before. When played a short binaural sample and asked to identify where they believed the source to be coming from, 15 participants said they knew exactly where it was at all times, and the remaining three said that they knew where it was most of the time. With regards to tablet computer/smartphone experience, 16 people owned devices, and the other two had some experience with them.

5.1. Results

This section will describe the details of each test and then go on to state the results.

5.1.1. Test 1.1: finding a black dot [with/without binaural]

In this test the user was tasked with finding a black dot on a screen by means of sound alone. Group A was given the pulse train, alert sound, and the binaural panning parameters. Group B were stripped of the binaural panning parameter and were provided with only mono audio, and thus acted as a control group so that the effect of the binaural parameter could be evaluated.

In this test all participants were able to find the black dot using the auditory feedback. Group A (who used the binaural mapping parameter) succeeded in finding the dot with a mean time of 15.6 seconds (blue line in Figure 5), with a standard deviation of 14.7. Group B (who did not use the binaural mapping parameter) were able to find the dot with a mean time of 18.4 seconds (red line in Figure 5), and a standard deviation of 10.7. The null hypothesis for Test 1.1 was that there would be no difference in times between Group A and B when searching for the dot – the binaural audio would make no difference. The alternative hypothesis was that using binaural audio would speed up the time they took to find the dot. The results were tested using a t-test for two independent samples, attaining a p value of 0.675 – suggesting low levels of confidence in the results.

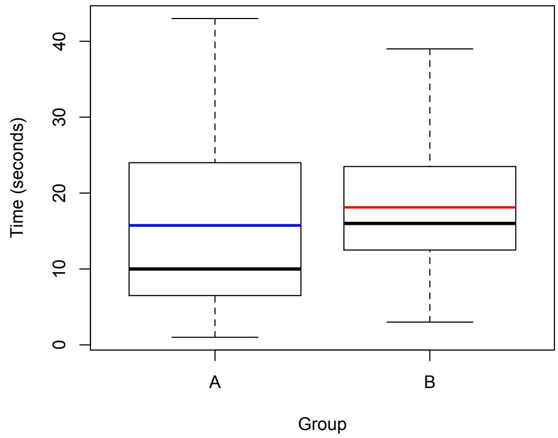

Figure 5: Boxplot for Test 1.1

5.1.2. Test 1.2: three coloured dots [Group B not told colour mappings]

For this test both Group A and Group B were asked to locate three coloured dots on a screen using the pulse train, alert sound, and binaural panning parameters. Group B, however, were not told the colour mappings, which were as follows: Red = high-pitched sound, Green = middle-pitched sound, Blue = low-pitched sound. The main aim of this test was to determine whether we have some preconceptions about how colour relates to sound. In addition we hoped to get qualitative information about how subjects coped with more than one acoustic target.

In the test, Group A, who were told the colour-sound mapping parameters, were able to get, on average, 2.45 out of the 3 dots correct. Group B, who were not told the colour-sound mapping parameters were able to get, on average, 2.34 dots correct. The null hypothesis was that by guessing at random, Group B would typically only be expected to get 1 out of 3 dots correct, and the alternative hypothesis was that they would guess more than 1 out of 3 of the dots correct. A Chi-Squared test was run to determine the odds of Group B getting this score by chance, and a confidence interval of p = 0.097 was attained – therefore showing relatively high confidence in the results.

5.1.3. Test 1.3: picture identification [both groups with same mappings]

For this test, users were asked to identify a simple picture (picture 3 in Figure 6) by interacting with it, and listening to its auditory response. They were asked to choose from four pictures (shown in Figure 6) the image they believed they had been interacting with. This test was designed to judge the success of the auditory display mappings when representing simple images.

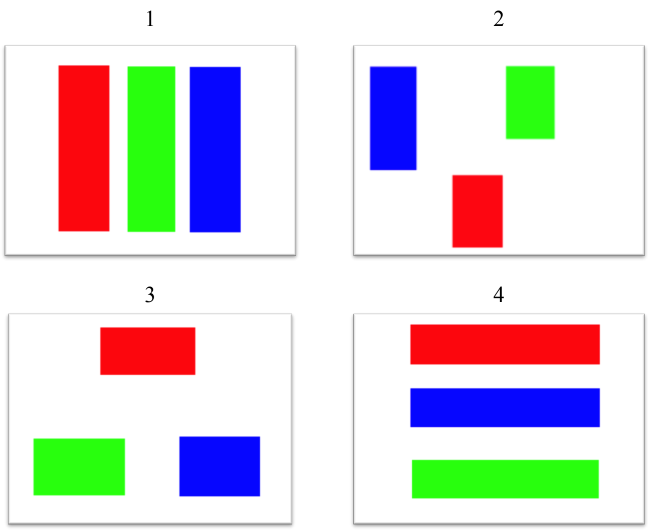

Figure 6: The four pictures the users were presented with (number 3 is being the one they were actually interacting with)

For this test the null hypothesis was that both groups would score the same as they would by guessing – each picture getting, on average, 4.5 users pick it. The alternative hypothesis was that the user would pick Picture 3, the correct answer, more than 4.5 times. In the test the participants chose Pictures 1 and 4 zero times, Picture 2 once, and Picture 3 (the correct picture) 17 times, in an average time of 27.34 seconds, and a standard deviation of 12.86. A Chi-Squared test was run to test the odds of this happening by chance – a confidence interval of $p = 4.562e^{-10}$ was attained – therefore suggesting very high confidence in the results.

5.1.4. Test 2.1: black dot in a large image [with/without binaural]

This following tests challenged the participants in tasking them with navigating a larger-than-display image. Test 2.1 was similar to Test 1.1 – the test where the users were tasked with finding a black dot using sound alone. However, the image used in this test was nine times larger than the iPad screen. The aim of this test was to evaluate the auditory display, and interaction techniques, when navigation around a larger image was involved.

For this test all but one (17/18) of the participants were able to find the black dot. The null hypothesis was that the additional binaural audio feedback provided to Group A would not speed up their performance, and that the mean times of the two groups would be the same. The alternative hypothesis was that Group A would be able to find the dot faster than Group B. In the experiment, Group A found the dot with an average time of 108.25 seconds (blue line in Figure 7) and a standard deviation of 68.16 and Group B 131.56 seconds (red line in Figure 7) and a standard deviation of 71.55. A t-test for two independent samples was used, attaining a p value of 0.5036, therefore suggesting relatively low confidence in the results.

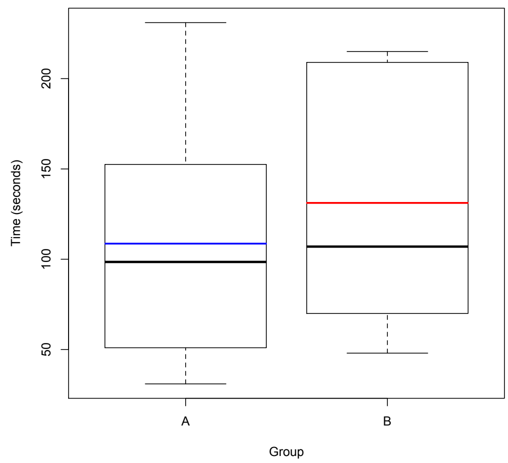

Figure 7: Boxplot for all participants in Test 2.1

5.1.5. Test 2.2: three coloured dots in a large image [with/without binaural]

In this test users were asked to navigate a large display and were tasked with locating three coloured dots. Again, the focus was on a comparison of binaural audio verses mono audio.

In this test, seven out of nine participants in Group A found all three dots, compared with four out of nine for Group B. The null hypothesis was that Group A, with the binaural audio, would perform the same as Group B, without the binaural audio. The alternative hypothesis was that Group A would be able to find the dots quicker, on average, than Group B. The average time to finish the test for all members of Group A was 242.8 seconds (blue line in Figure 8) with a standard deviation of 65.7, and for Group B it was 391.3 seconds (red line in Figure 8) with a standard deviation of 249.8. A t-test for two independent samples was used, attaining a p-value of 0.124, therefore suggesting some confidence in the results.

The average time for those who found all three dots in Group A was 254.85 seconds, with a standard deviation of 60.58. In Group B this was higher at 312 seconds, with a standard deviation of 255.16. A one tailed t-test for two independent samples

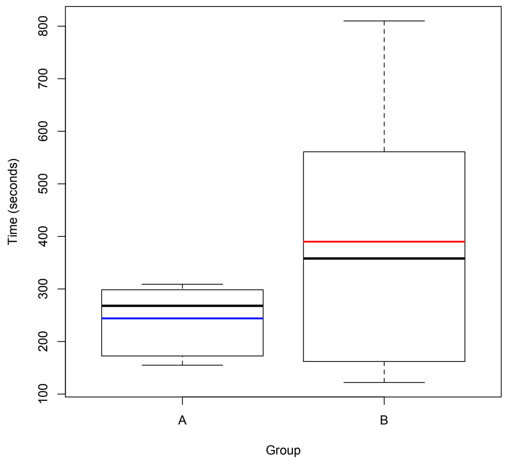

Figure 8: Boxplot for those who finished Test 2.2

produced a p value of 0.587 – suggesting relatively low confidence in the results.

5.1.6. Test 2.3 & Test 2.4: black dot in a large image [no visual restriction]

These tests involved a multimodal task using larger scrollable images – Test 2.3 featuring an image nine times the iPad screen, and Test 2.4 16 times the size. For these tests Group A were tasked with finding a black dot by means of using an auditory display, and Group B were tasked with finding the dot by means of visual cues alone. The aim of this test was to judge if the sonification techniques affected the performance of the user when searching.

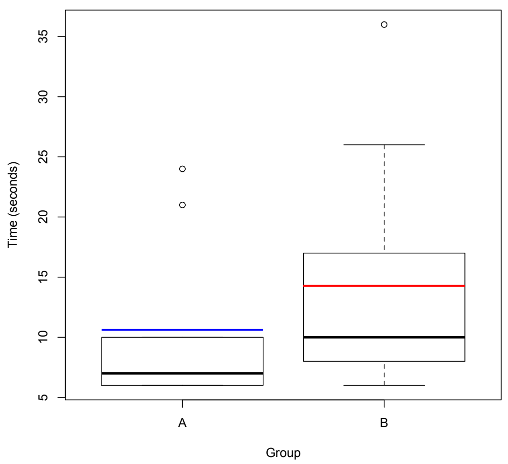

Figure 9: Boxplot for results of Test 2.3

In Test 2.3 the null hypothesis was that the group who were able to see the iPad screen and had auditory feedback (Group A) would perform the same as the group with visual cues alone (Group B). The alternative hypothesis was that Group A would be able to find the dot quicker. In the test, all participants found the dot – Group A attaining an average time of 10.6 seconds (blue line in Figure 9) and a standard deviation of 7, Group B attaining an average time of 14.3 seconds (red line in Figure 9) and a standard deviation of 10.3. A t-test for two independent samples was used, attaining a p value of 0.373 – showing relatively low levels of confidence.

In Test 2.4 the null hypothesis was the same as in the previous test – that Group A, with the additional auditory feedback, would perform the same as Group B, who had no auditory feedback. The alternative hypothesis was that Group A would find the dot in less time than Group. The results show that Group A were able to find the dot with an average time of 12.4 seconds (blue line in Figure 10) and a standard deviation of 6, and Group B an average time of 15.2 seconds (red line in Figure 10) with a standard deviation of 11.1. Upon running a t-test for two independent samples, a p value of 0.519 was found – suggesting relatively low levels of confidence in the results.

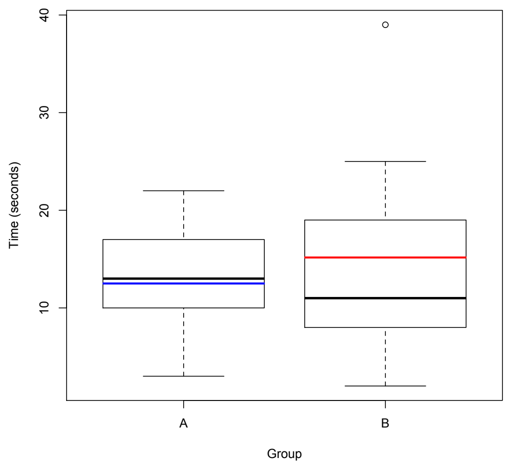

Figure 10: Boxplot for results of Test 2.4

6. DISCUSSION

It was evident from the results of Tests 1.1 (locating a black dot), 2.1 (locating a black dot on large screen), and 2.2 (locating three coloured dots on large screen) that the binaural audio allowed the participants to, on average, find the image features faster than with the other mappings alone – the pulse train, volume, 'boing' sound, and alert sound.

The results suggest that as the images became larger, and the tasks more complex Group A outperformed Group B more and more. In Test 1.1 (locating the black dot) Group A were able to locate the image feature 17.7% faster than Group B, and in Test 2.1 (locating a black dot within a large scrollable image), in which the image is 9 times larger, there was a 21.5% difference between the groups those with binaural audio appear to perform even better in comparison to the non-binaural group. The difference between the groups in Test 2.2 becomes larger – an (albeit small) increase of 0.92% between the two groups when an additional difficulty is added – this may suggest that as the tasks become more complex, the binaural audio helps more. However, the significance levels were not high enough for us to state this categorically.

Upon observation of the videos (all available on the link provided in the supporting material folder with this paper) it is possible to say, anecdotally, that the participants with the binaural audio (Group A) undertook more logical searching strategies, whereas the group without the binaural audio (Group B) normally undertook a more sporadic 'brute force' searching method, which would account for the larger spread in times, and therefore higher standard deviation, for Group B. However, it must be noted that even in Group A there was a large spread of times. We believe that this large variation between the participants, as well as the relatively low number of participants, has led to the lower levels of significance found.

Test 1.3 (image identification using binaural audio) indicated strongly that participants were able to detect a picture from sound alone using the techniques. The techniques excelled at allowing the participants to gain a quick overview of the graphical features on display with a very high success rate, and an insignificantly low probability of attaining the same results through guessing. The large variation in times can be attributed to the searching techniques of the individuals. From discussing the test with the participants after the experiment it became evident that some participants looked at the pictures beforehand and made a mental model of what they believed they were looking for. Meanwhile, others investigated the auditory response, and then by process of elimination chose their answer. The subjects with the faster times generally adhered to the first approach.

When comparing visual and audio cues, against visual cues alone in Tests 2.3 and 2.4 (black dot within progressively larger images with and without visual cues) the difference in searching techniques became even more evident. The relatively large standard deviations of Group B can be attributed to the sporadic searching patterns the group used. Sometimes a participant would randomly scroll around very quickly and get lucky, whilst others spent quite a while 'raster scanning' the image in a logical search. Most participants in Group A took a while to assess the local area, then gradually moved in a straight line towards the image feature, resulting in the higher mean times and smaller standard deviations.

7. CONCLUSIONS AND FURTHER WORK

Several auditory display techniques were developed to allow for images of varying sizes to be explored by means of binaural interactive sonification. Gesture-based interaction modes were created to facilitate the exploration of these images on an iPad, whilst listening to an auditory representation. Seven tests were then carried out to deduce whether the techniques were improved by use of binaural audio.

The results from the tests showed that binaural audio could be used to improve our understanding of simple images of varying sizes. It is evident that the experiments could benefit from additional participants – the number used (eight in each group) was not enough to produce very significant results. It is recommended that if this experiment is replicated, additional participants should be recruited. To reproduce the tests described in this paper a folder including the test scripts, a document describing the technical setup, videos of the participant's performances (provided pending acceptance), and the code needed to reproduce the tests has been provided at the following Dropbox link:

`https://db.tt/eyVcfAFf`

Further work in this area should involve increasing the complexity of the images and the image processing algorithm. Similar methods could be used to explore more complex images where a user wishes to search for a specific colour, such as when scanning cervical cancer slides, or looking for objects in Deep Field space photography. Additionally, the binaural auditory display techniques could be used for numerous applications, not just the exploration of images. For example – improving immersion in computer interfaces or assisting those who are visually impaired, or have their eyes on other tasks, by extending the visual domain with spatial audio.

This paper has demonstrated the potential of binaural audio to provide real-time feedback to visually restricted or distracted users to improve the location of objects in the data being represented both on and off-screen.

8. REFERENCES

[1] David McGookin and Stephen Brewster, "Fishears - the design of a multimodal focus and context system," Tech. Rep., Glasgow, 2001.

[2] Peter Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, 1992.

[3] Amir Amedi, William Stern, Joan Camprodon, Felix Bermpohl, Lofti Merabet, Stephen Rotman, Christopher Hemond, Peter Meijer, and Alvaro Pascual-Leone, "Shape conveyed by visual-to-auditory subsitution activates the lateral occipital complex," *Nature Neuroscience*, vol. 10, no. 6, pp. 687–689, 2007.

[4] Guido Bologna, Benoît Deville, Pun Thierry, and M Vinckenbosch, "A perceptual interface for vision substitution in a color matching experiment," *Joint Conference for Neural Networks - IEE World Congress on Computational Intelligence*, pp. 1–6, 2008.

[5] Ryan McGee, Joshua Dickinson, and George Legrady, "Voice of sisyphus: An image sonification multimedia installation," *ICAD*, vol. 18, 2012.

[6] Wilko Heuten, Daniel Wichmann, and Susanne Boll, "Interactive 3d sonification for the exploration of city maps," Oslo, 2006, NordiCHI.

[7] Piotr Skulimowski Michal Bujacz and Pawel Strumillo, "Naviton: A prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired," *AES*, vol. 60, pp. 696 – 708, 2012.

[8] Guido Bologna, Benoît Deville, and Thierry Pun, "Sonification of color and depth in a mobility aid for blind people," *International Conference for Auditory Display*, vol. 16, pp. 9–13, 2010.

[9] Brian Katz, Emmanuel Rio, Lorenzo Picnali, and Olivier Warusfel, "The effect of spatialization in a data sonification exploration task," *ICAD*, vol. 14, pp. 1 – 7, 2008.

[10] Jens Blauert, *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*, Hirzer Ferlag, Massachusetts, 1997.

[11] Lazzarini, "The audio programming blog," 2012.

[12] Tsubasa Yoshida, Kris Kitani, Hideki Koike, Serge Belongie, and Kevin Scheli, "Edgesonic: Image feature sonification for the visually impaired," *Augmented Human Conference*, vol. 2, pp. Article 11, 2011.

[13] Bill Gardner and Keith Martin, "Hrtf measurements of a kemar dummy-head microphone," Tech. Rep., MIT Media Lab, 1994.

[14] Wen Zhang, Thushara Abhayapala, and Rodney Kennedy, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *Acoustical Society of America*, vol. 127, pp. 23472357, 2010.

[15] Brian Carty, "hrftmove, hrtfstat, hrtfmove2: Using the new hrtf opcodes," *Csound Journal*, vol. 9, 2008.

[16] Mikael Fernström, Liam Bannon, and Eoin Brazil, "An investigation of soft-button widgets using sound," *Le Journe de Design Sonore*, pp. 1–11, 2004.