A GENERIC EVALUATION MODEL FOR AUDITORY FEEDBACK IN COMPLEX VISUAL SEARCHES

Timothy Neate[1 *], Norberto Degara[2], Andy Hunt[3], and Frederik Nagel[2]

[1]Department of Computer Science, Swansea University, UK
[2]EAR Group, Fraunhofer IIS, Germany
[3]Department of Electronics, University of York, UK

ABSTRACT

This paper proposes a method of evaluating the effect of auditory display techniques on a complex visual search task. The approach uses a pre-existing visual search task (conjunction search) to create a standardized model for audio, and non-audio assisted visual search tasks. A pre-existing auditory display technique is evaluated to test the system. Using randomly generated images, participants were asked to undertake a series of visual search tasks of set complexities, with and without audio. It was shown that using the auditory feedback improved the participant's visual search times considerably, with statistically significant results. Additionally, it was shown that there was a larger difference between audio and non-audio when the complexity of the images was increased. The same auditory display techniques were then applied to an example of a real complex visual search task, the results of which imply a significant improvement in visual search efficiency when using auditory feedback.

1. INTRODUCTION

Tasks that require extensive searches of complex visual fields often take time and practice, and can also cause visual fatigue. Due to the ubiquity and power of modern computing devices, research areas such as medical imaging are transferring their visual searches to computers [1]. This allows for modern image processing, and interaction techniques, to be used in tandem with the human visual senses to explore larger-than-screen images, to let the user look for potential areas of interest.

The current standard for guiding a user towards a specific area of interest in an image is to use visual cues, for example – highlighting a specific on-screen area [2], or using arrows to point to an off-screen area [3]. However, these visual methods come with some inherent problems. The extra visual information can distract the attention of the user, and requires additional visual and cognitive processing, which may result in the user missing an important graphical feature. Also these visual cues can cover potentially useful information on the screen – obstructing the user's line of sight. In some other cases it is not possible, or technically difficult, to add visual cues, such as when using an electronic microscope or in an operating/surgery room.

A potential solution to the problems associated with providing visual feedback is to use alternative modalities (such as auditory, or tactile feedback) to extend the visual domain. When undertaking complex visual tasks, the provision of auditory assistance has been shown to improve performance [4] [5] [6]. Unused capacities in other senses can take over, offering additional data to be processed by the brain. An example of the use of sound to augment the analysis of images is the examination of cervical smear slides for cancerous cells [7] [8] – a task in which a clinician must search tens of thousands of cells per slide by viewing it with their microscope, looking for suspect (potentially cancerous) cells as they go. Some of the approaches described in Edwards et al's papers proved to be relatively successful in applying auditory feedback to assist a clinician when undertaking this extensive search task, such as Podvoiskis' approach [1], which showed that subjects were able to identify and classify specific types of cell from their auditory response alone.

Despite the previous work, no standard model of evaluating audio-assisted visual searches has been defined. In work such as [8] the focus was on a very specific task, and therefore it is hard to consider the techniques for other contexts. Additionally, approaches such as [6] focused on visual searches that were too simple, and therefore the effect of the auditory feedback could not be properly judged when assisting a visual search. Although work has recently been done into standardising some specific tasks in the field of auditory display [9], to the best of the authors' knowledge, no paper proposes a standardised method to evaluate audio-assisted visual searches.

The aim of this paper is to evaluate the usefulness of a single auditory feedback method, compared to no sound, in a series of complex visual searches. This is done with the aim of defining a standard method of evaluating auditory display techniques, such that the community can evaluate new approaches against the 'benchmark system' described in this paper.

The remainder of this paper discusses a series of experiments undertaken to test auditory display techniques developed to assist users in complex visual searches. Some background on the key auditory display concepts is given, along with an overview of the type of visual search tasks undertaken in Section 2. Next the proposed system is discussed in more depth, along with the experimental techniques in Section 3 and results in Section 4.

*Correspondence: `tdjneate@gmail.com`

[1]Using a cursor to interact a visual field and attain an auditory response based on its colour range.

Finally, Section 5 presents the conclusions of our experiments.

2. PROPOSED AUDITORY DISPLAY METHOD FOR COMPLEX IMAGES SEARCHES

This section describes the different elements of the proposed auditory display. Section 2.1 gives an overview of the proposed display system. Section 2.2 discusses the graphical data processing, and the sound mapping and interaction design are described in Sections 2.3 and 2.4.

2.1. System description

A block diagram of the proposed system is shown in Figure 1. The system allows for a user to be assisted in a visual search task by providing sonic information about the target. When the user interacts with the image on the device (an iPad) an auditory response is produced. The sound is based on where in the image the touch occurred, and where the target feature is relative to this touch. Users receive auditory feedback, guiding them towards the target until they locate it.

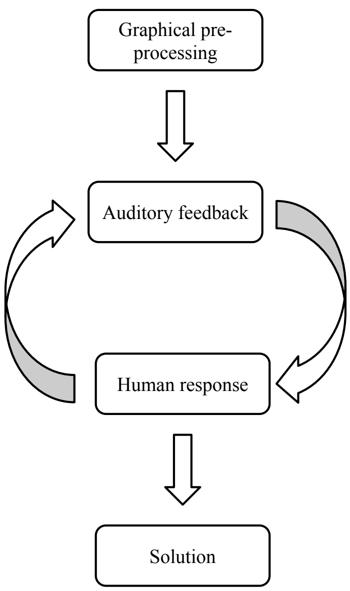

Figure 1: Block diagram of proposed system

2.2. Pre-processing of graphical data

When transforming graphical data into sound, researchers have typically made one of two decisions: 1) to transform all of the information into the auditory, creating a complex auditory output that the user must learn to decipher, or 2) a target-based approach where the user specifies the target information, and only this is transformed into the auditory domain.

An example of the first approach is Peter Meijer's system 'The vOICe' [10]. This system transforms visual information into its raw auditory form by mapping the brightness of pixels to the amplitude of oscillators. Approaches like this offer users a highly complex auditory result, which they must learn to decode. This approach has shown to be highly successful in aiding the visually impaired to perceive visual information [11] [12].

On the other hand, there is the second approach, based on targets. Approaches such as Bologna et al.'s 'See ColOr' project [13] [14] [15] [16] rely on methods that filter out the unwanted visual information (for example, specific colours), and transform only this goal-oriented information into sound. These approaches tend to benefit tasks where we know what the target looks like, and therefore can do some graphical pre-processing to isolate the required visual features and determine their positions. The main advantage of this method is that it offers a simpler auditory field that is easier for the user to learn.

In this paper, we assume that the coordinates of the targets are known, but it is assumed that some image pre-processing would be done beforehand dependant on the task at hand. The sonification mappings were developed with the interaction modes of modern tablet computers in mind – therefore the auditory feedback is acquired by touch interaction. By using information from the user's touch position, and combining it with the locational information of the target, it is possible to determine a 2D vector where the initial and terminal points are determined by the touching and target coordinates.

2.3. Sound mappings

The vector generated is used to drive the audio engine. A pulse train was used to represent the distance from the feature, and binaural panning was used to describe the direction. Inspired by the work of Yoshida et al. [17], it was decided that a pulse train could be used to represent the distance between the user's touch and the target feature – with the frequency of the pulses mapped inversely to the distance between the two. This means that the closer the touch to the target, the faster the rate of the pulse-train, allowing for more resolution, resulting in a tighter human-computer interaction loop as the user gets close to the target.

This pulse train is panned binaurally. The decision to use spatial audio was made because it allows us to easily represent a physical direction (in the image) in Cartesian space (in sound). This is possible because of our innate ability to perceive the location of a spatialized source to within 11.8 degrees [18]. In the auditory display community, binaural audio has been used to produce complex directional auditory fields over headphones [19] [20]. In our system the direction of the image feature was panned relative to the user's touch, as shown in Figure 2. The user is assumed to be using the iPad in landscape mode, and their direction is assumed to be directly in front of them ('up' on the device). The HRTFs used were made by Bill Gardner and Keith Martin of the MIT media lab [21]. They were found to be a good 'best fit' head as they were made on a KEMAR (binaural dummy head) which was designed according to the mean anatomical size of the population.

Some extra parameters were developed to assist the user in locating the features – 1) an alert sound when they actually find the target, created by modulating a sinusoid at a high rate, and 2) a 'boing' sound to tell them when they had scrolled too far off the screen – a simple oscillator with harmonics, panned binaurally when a user extends the bounds of the screen. The sound engine was developed in Csound and integrated into iOS using the Csound-iOS API, developed by Steven Li and Victor

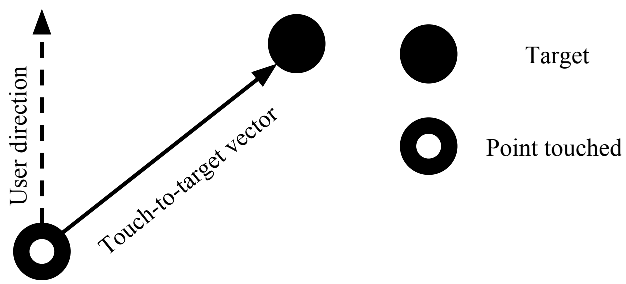

Figure 2: Touch-to-target vector that drives the audio engine

Lazzarini [22]. This allowed for the extensive audio processing capabilities of Csound to be used in tandem with the touch-based iOS environment.

2.4. Interaction design

The method of navigating the extended display is as follows. For each test the image was nine times the size of a conventional iPad screen (3^2), and interaction techniques had to be developed for the user to navigate this display. To ensure that moving around the screen was clearly differentiated from the interaction for sound feedback, two different methods of interaction were devised – two-fingered swiping-gestures for navigation, and one-finger touching for the sound feedback.

3. EXPERIMENTAL METHOD

3.1. Task definition

To assess the usefulness of using audio (by measuring the search time), the proposed system has been evaluated using a non-applied test. A number of users were set the task of finding a target shape in a conjunction-based search. A conjunction search is an example of a feature search in which the target is hidden in a visual field in which there are features that bear similarities to it – for example tasking someone with finding the anomalous shape (the blue circle) in Figure 3.

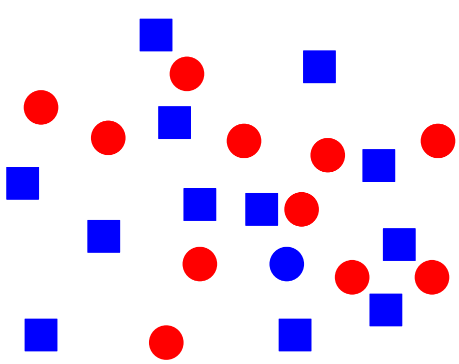

Figure 3: A simplified conjunction search

Treisman and Gelade's Feature-Integration Theory model [23] suggests that when we perceive a stimulus, we quickly register the features of the visual objects (such as size, colour and shape). However, the perception of the objects themselves (for example, a red square) takes more time – it requires a further level of cognitive processing to perceive the object as a whole. In general, it was found that the search for targets defined only by a difference in features results a linear relationship between the time to find the target, and the number of distractors [23]. This relationship allows for the complexity of the experiment to be judged, therefore enabling us to determine whether the auditory feedback aids us more or less for different levels of complexity.

The user, in each test, was tasked with finding a black triangle within a visual field of similar shapes – black squares and white triangles (as shown in Figure 4). The test aimed to judge two parameters – the effect of audio on the search task in general (audio vs. visual), and the effect of increasing the complexity of the image with audio and no audio.

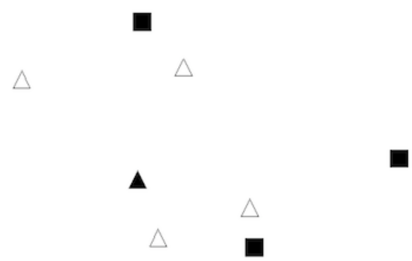

Figure 4: small section of conjunction search with target (black triangle) in view

Random images were generated in Matlab. Each shape (either a black square or a white triangle) was generated at a random position[2], the target (a black triangle) was then generated at a random position. The user was provided with a test script, and an iPad (iPad 4) to conduct the test on.

3.2. Test description

To evaluate the usefulness of audio feedback in terms of the complexity (density) of the image, two different tests were run. On the first test, at low density, the density of the image was 0.4 objects/mm. On the second test, high density, this was doubled to 0.8 objects/mm.

Test 1a ([low density] (no sound)) was an experiment in which the user was asked to complete the aforementioned conjunction search with a density of one shape per 2.5mm without sound. Test 1b ([low density] (sound)) then involved the users undertaking this task with sound, such that the effectiveness of the auditory feedback could be determined. Then, in Test 2 [high density], to judge the effect of increasing the number of distractors, the density was increased to one shape per 1.25mm. The effect of sound on this search was also judged by allowing the users to have just visual cues (Test 2a [high density] (no sound)) and to have the additional auditory cues (Test 2b [high density](sound)).

[2]The Matlab script can be found in the supporting material with this paper. The positions are uniformly distributed within the image size.

The participants were asked to undertake each test in a certain order. This order was rotated between users to negate the training effect. 10 trials of each test were undertaken by the subject, resulting in a total of 40 tests per participant. The position of the target is different in each of these 40 tests, as discussed in Section 3.1.

A 'bonus' test was then introduced – a Where's Wally search task (more information on Where's Wally can be found at [24]). The users were asked to conduct a Where's Wally search on an extended display screen in the same manner they undertook the conjunction searches – by trying to find a specific target (Wally) within distractors (things that look like Wally), with and without auditory feedback. Two pictures were used for each participant, one with sound and one without. The picture with auditory feedback was alternated to account for the difference in complexities between the two pictures. As it was expected that some participants would not be familiar with Where's Wally, a small training session and an introduction to Where's Wally was given before the bonus tests.

During all tests, the conditions were the same for each participant – the volume on the iPad was set to the same level (volume level 8), and the room conditions (an acoustically treated listening booth at Fraunhofer IIS) remained consistent across all subjects. A high quality set of studio headphones (Beyerdynamic DT 990 Pro) was used, and the audio feed was taken directly out of the iPad's 3.5mm jack. Figure 5 denotes the set-up of the test.

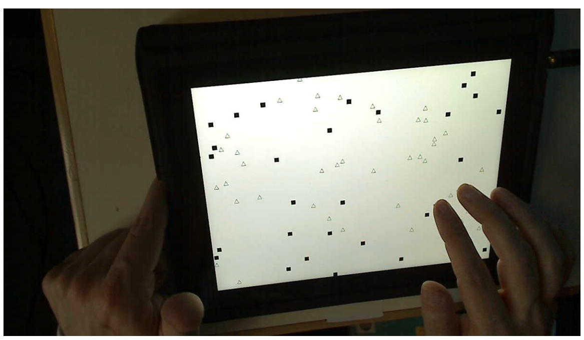

Figure 5: The test setup from the perspective of a webcam above the iPad

3.3. Evaluation

The auditory display proposed in this paper has been evaluated in terms of reaction time in the complex visual search described in Section 3.1. When the participant began the test, a timer was started on the device, and when they found the target the timer was stopped. This time was logged within the device in a scoreboard system. The times could then be noted down at a later date for analysis. In case of problems when logging the times, and to conduct further analysis, each test was filmed. A webcam was suspended above the user's hands as they interacted with the iPad, allowing for any nuances of the search task to be captured and stored for analysis later. A video showing the test being undertaken is publicly available [3]. It shows the user undertaking the visual search task with and without the auditory display techniques.

The times were then used to determine the effect of the auditory feedback on the task by looking at the means, Cohen's d effect size, and p-values based on an analysis of variance (ANOVA) test [25]. 95% confidences intervals (CI) for the mean values were calculated using bootstrapping [26]. It was expected that providing the participants with auditory feedback would help, as providing them with any information about the location of the target should give them an advantage. The main concern of this experiment was to find out if the techniques developed helped the participants locate the targets faster, and if so by how much, in hope that others would be able to compare these results to other types of auditory and non-auditory feedback.

An additional aim was to determine if, as the complexity of the visual search increased, the effect of the auditory feedback was greater. Treisman [23] suggests that as the number of distractors increases, there is a linear increase in the time it takes for participants to find the target in a conjunction search. It is of interest to see whether a similar effect occurs with or without auditory feedback, in the context of an extended display conjunction search. If this holds true we would expect to see a doubling in the time as the number of distractors is doubled due to this linear relationship, allowing us to quantify complexity in our model.

3.4. Participant demographics

16 participants took the experiment, all of whom were interns or researchers at Fraunhofer IIS apart from a visitor from the University of York. A small preliminary experiment and power analysis was done before this to ensure that this was enough participants for significant results. The average age of the participants was 29.9 years (standard deviation of 10.0 years). The group included people of German, British, Indian, Spanish, Australian, Italian, Bolivian, Venezuelan, and Chinese nationality. With regards to gender – 14 males took the experiment, and two females.

Some test-oriented questions were asked. It was found that 12 participants owned touch-screen devices, one had some experience, and three claimed to have little experience of using them. No participants claimed to have no experience with touch-screen devices. The majority of the participants (11 out of 16) had both music and audio processing knowledge, two had only an audio background, and nobody claimed to have just a music background. However, three participants had no music or audio processing knowledge.

With regards to experience with 3D audio, seven people had good knowledge and experience with binaural audio, eight had listened to some before, and only one participant had not heard any before. To ensure that the participants did not have any issues with the visual element of the search task they were asked to self-report their visual abilities. All participants claimed to have either perfect, or corrected vision. Finally, the participants were asked to describe their visual search experience. One participant had good experience with visual searches, two had a little experience, and the remainder had no experience with visual searches, such as conjunction searches.

[3] https://dl.dropboxusercontent.com/u/30785467/Auditory_feedback_in_complex_visual_searches.wmv

Density	Audio	Mean time	CI
Low	no	21.57	$[18.40, 26.20]$
	yes	10.01	$[9.22, 11.25]$
High	no	36.28	$[31.50, 44.56]$
	yes	12.75	$[11.59, 14.28]$

Table 1: Mean reaction times and their corresponding 95% confidence intervals for each of the experiments (in seconds).

4. RESULTS

Analysed results of the tests are discussed in this section. We compare the differences between the tests, such that conclusions can be made about the success of the auditory feedback. The mean reaction times and their corresponding 95% confidence intervals for each of the experiments are outlined in Table 1 for quick reference and access.

4.1. Global effect of the audio feedback and density

Figure 6 shows a boxplot of the reaction time with and without audio feedback (namely, Sound/No Sound on the x-axis) and density (High/Low x-axis) factors. Mean values are shown as black circles on the box plot for each of the conditions. As we can see, the mean reaction time is lower when using audio feedback compared to the tests where no audio feedback is given. As we would expect, complexity increases the reaction time.

In order to assess the effects of the auditory feedback and the complexity of the visual search experiments, a 3-way ANOVA analysis of the reaction times has been performed. The density of the test image and the use of audio feedback are fix factors, whereas subjects were set as a random factor for the statistical analysis. Results of the ANOVA analysis show that the audio feedback and density factors had a statistically significant effect on the average reaction time ($p < 0.01$). In the following subsections we will analyze each of the individual factors, audio feedback and density, in more detail.

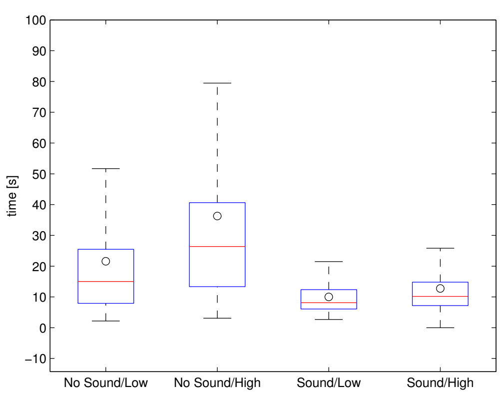

Figure 6: Boxplot describing the effect of the audio feedback and density on the conjunction search.

Subjects effect had a p-value of 0.51 meaning that the mean reaction time of participants do not show statically significant differences, and that no participant performed statistically better. This is something that we would expect since no participant was particularly skilled a priori in visual searches.

4.2. Audio feedback analysis

In Test 1 [low density] the participants were able to locate the target 100% of the time, with or without sound. Averaged subject times ranged from 8.11 to 40.27 seconds in the test with no sound, and from 5.45 to 12.23 seconds with sound. To judge the effect of auditory assistance on the visual search, a 2-way ANOVA analysis on the reaction time was done where the use of audio feedback was a fix factor and the subjects a random factor. On average, the participants managed to locate the target without sound in 21.57 seconds (CI $[18.40, 26.20]$). With the additional sound assistance, the participants were able to locate the feature in 10.01 seconds (CI $[9.22, 11.25]$). These results show statistically significant differences ($p < 0.01$). We found an effect size of 0.66, suggesting that the auditory feedback helps the participants improve their visual search a considerable amount.

In Test 2 [high density] 100% of the participants were able to locate the target. However, as one would expect from a more complex visual searching task, the times were slower – the average times ranged from 14.42 to 70.82 seconds in the visual-alone task, and from 6.58 to 15.27 seconds in the visual and auditory task. The average reaction time without sound was 36.28 seconds (CI $[31.50, 44.56]$). The auditory feedback improved the time to 12.75 seconds (CI $[11.59, 14.28]$). A 2-way ANOVA analysis of the reaction times for the high density tests shows that the differences in time are significant (p < 0.01). A large effect size in the high density test of 0.83 was found. This effect size is larger than in the low density test, suggesting that audio feedback helps more in more complex searches.

4.3. Density analysis

We now analyse the density factor and compare reaction times in the case where no audio feedback is used. For this, a 2-way ANOVA of the reaction times in the tests where no audio feedback (Test 1a and Test 2a) is given. As before, the density is a fix factor and subjects are a random factor. As we would expect looking at the boxplot of Figure 6 and the mean reaction times and CIs reported in the subsections above, results show that subjects were significantly better in low complexity images (21.57 seconds) than in high complexity images (36.28 seconds).

The case where sound is used shows similar results. It is not surprising to verify that the mean reaction times are significantly better in low complexity images (10.01 seconds) than in high complexity images (12.75 seconds). The next section will statistically analyse the interaction effect between audio feedback and density.

4.4. Interaction effect

The 3-way ANOVA analysis of Section 4.1 also shows that there is a statistically significant interaction effect between the audio feedback and the density factors ($p < 0.01$). This suggests a deviation from the linear model assumed by ANOVA. Audio feedback does not help in the same way in the high and low density tests. Looking at the average reaction time in Figure 6, we can observe that the reduction in reaction time when using sound is larger in the high density test case, from 36.28 to 12.75, than in the low density case, from 21.57 to 10.01. As one might expect, the participants appeared to rely on the auditory cues more as the visual search tasks became more complex, often relying on the auditory cues even though the target was visible on the screen.

4.5. Applied test (Where's Wally test)

In the applied Where's Wally test, 14 out of 16 participants found Wally in the visual search, and everyone found Wally in the audio assisted visual search. There was a significant difference ($p < 0.05$) in the mean times in which the participants found Wally – those without sound found him, on average, in 70.26 seconds (standard deviation 55.43), and those with sound found him in 25.42 seconds (standard deviation 17.6).

5. CONCLUSIONS AND FURTHER WORK

The work shows that the additional auditory feedback can be used to reduce search time. In Test 1 [low density] there was a considerable difference between the participant's times with sound and without sound – those with sound completed the task in around half the time. And in Test 2 [high density] this gap grew even larger – they were able to locate the feature in around one third of the time it took the non-audio group. Results also showed a significant interaction effect between audio feedback and the complexity of the image. When compared with similar work [6] (a simple visual search without distractors using the same auditory methods) it is evident that the participants relied on auditory cues significantly more when the display's complexity was increased.

It is clear that the large difference in the means between the audio and non-audio groups sets a good baseline for others to develop better techniques. Further work should include developing methods that are able to improve the effectiveness of the users' search. With regards to reproducible research [27], the work discussed in this paper can be extended or verified easily as the Matlab code, Xcode example projects, and test scripts have been provided at the following link so that anyone may replicate these results, or improve the techniques developed.

`https://db.tt/D71PglVc`

Additionally, the methods discussed in this paper, or indeed any more effective methods that are developed from the testing procedure described in this paper should be applied to complex visual tasks such as the cervical cancer cell application discussed in the introduction. Perhaps the tasks themselves may mean that the techniques developed may require some additional work. However, the fundamental mapping techniques developed from the conjunction search task should provide a good baseline for any complex visual task. Moreover, as discussed in previous work [6], the methods can be applied to those with visual impairments, or those with their eyes engaged elsewhere, to navigate computer interfaces.

This paper has proposed a method for evaluating complex visual searches on extended display screens, with and without sound. By using the pre-existing conjunction search model, it allows us to evaluate the auditory display techniques that we have developed in a controlled manner – where we control the complexity of the search, in a scalable and repeatable manner.

6. ACKNOWLEDGEMENTS

The authors would like to thank all the participants who undertook this experiment, as well as the people at the EAR lab at Fraunhofer IIS, especially Nadja Schinkel-Bielefeld, Balaji Thoshkahna, David Worrall, and Christian Neukam, for their feedback and help running the experiments.

7. REFERENCES

[1] B. Hemminger, A. Bauers, and J. Yang, "Comparison of navigation techniques for large digital images," *Journal of Digital Imaging*, vol. 21, pp. 13 – 38, 2008.

[2] C. Y. Ip and A. Varshney, "Saliency-assisted navigation of very large landscape images," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, 2011.

[3] S. Burigat, L. Chittaro, and S. Gabrielli, "Visualizing locations of off-screen objects on mobile devices: A comparative evaluation of three approaches," in *MobileHCI*, 2006.

[4] E. Haas, R. Pillalamarri, C. Stachowiak, and M. Lattin, "Audio cues to assist visual search in robotic system operator control unit displays," Army Research Laboratory, Tech. Rep., 2005.

[5] A. Podvoiskis, *MEng Project Report: Improving the Efficiency of Cervical Cell Sample Analysis*. University of York, 2004.

[6] T. Neate and A. Hunt, "Interactive spatial auditory display of graphical data," in *4th Interactive Sonification Workshop*, 2013.

[7] "Segmentation of biological cells for sonification," The University of York, York, Tech. Rep., 2008.

[8] A. Edwards, A. Hunt, J. Stammers, A. Podvoiskis, and R. Roseblade, "Sonification strategies for examination of biological cells," *ICAD*, vol. 16, pp. 193–200, 2010.

[9] N. Degara, T. Kuppanda, and F. Nagel, "The walking game: A framework for evaluating sonification methods in blind navigation," in *4th Interactive Sonification Workshop*, 2013.

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

[11] A. Amedi, W. Stern, J. Camprodon, F. Bermpohl, L. Merabet, S. Rotman, C. Hemond, P. Meijer, and A. Pascual-Leone, "Shape conveyed by visual-to-auditory subsitution activates the lateral occipital complex," *Nature Neuroscience*, vol. 10, no. 6, pp. 687–689, 2007.

[12] L. Merabet, D. Poggel, W. Stern, E. Bhatt, C. Hemond, S. Maguire, M. Peter, and A. Pascual-Leone, "Activation of visual cortex using crossmodal retinotopic mapping," Melbourne, 2008.

[13] G. Bologna, B. Deville, and P. Thierry, "Transforming 3d coloured pixels into musical instrument notes for vision subsitution," *Journal of Image and Video Processing*, p. 14, 2007.

[14] G. Bologna, B. Deville, P. Thierry, and M. Vinckenbosch, "A perceptual interface for vision substitution in a

color matching experiment," *Joint Conference for Neural Networks – IEEE World Congress on Computational Intelligence*, pp. 1–6, 2008.

[15] G. Bologna, B. Deville, and T. Pun, "Blind navigation along a sinuous path by means of the see color interface," *International Work-Conference on the Interplay Between Natural and Artificial Computation*, 2009.

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

[17] T. Yoshida, K. Kitani, H. Koike, S. Belongie, and K. Scheli, "Edgesonic: Image feature sonification for the visually impaired," *Augmented Human Conference*, vol. 2, p. Article 11, 2011.

[18] J. Blauert, *Spatial Hearing – Revised Edition: The Psychophysics of Human Sound Localization*. Hirzer Ferlag, 1997.

[19] W. Heuten, D. Wichmann, and S. Boll, "Interactive 3d sonification for the exploration of city maps." Oslo, NordiCHI, 2006.

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

[21] B. Gardner and K. Martin, "Hrtf measurements of a kemar dummy-head microphone," MIT Media Lab, Tech. Rep., 1994.

[22] Lazzarini, "The audio programming blog," 2012. [Online]. Available: http://audioprograming.wordpress.com/2012/04/03/csound-ios-sdk-released/

[23] A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, pp. 97 – 136, 1980.

[24] C. M. Distribution, "Where's wally official website," Website, January 2014. [Online]. Available: http://www.whereswally.co.uk/

[25] R. V. Hogg and J. Ledolter, *Engineering Statistics*. New York: MacMillan, 1987.

[26] A. Zoubir and D. Iskander, "Bootstrap methods in signal processing [from the guest editors]," *Signal Processing Magazine, IEEE*, vol. 24, no. 4, pp. 7–8, July 2007.

[27] P. Vandewalle, J. Kovacevic, and M. Vetterli, "Reproducible research in signal processing," *IEEE Signal Processing Magazine*, vol. 26, pp. 37 – 47, 2009.