AUDIO FLIPBOARD: A SPATIAL AUDIO DISPLAY EXPLOITING SIMULTANEITY IN THE PRESENTATION OF A COLLECTION OF ORGANISED MEDIA ARTICLES

J Sinker University of Salford, Salford, UK
B Shirley University of Salford, Salford. UK

1 INTRODUCTION

Over the last twenty years personal media devices have become increasingly common place, evolving from large cumbersome devices capable of comparably low level functions and storage, to modern devices that are truly pocket sized or smaller with high level functionality and storage capacity. Currently mobile devices such as phones and tablets offer a wide range of functions; from making phone calls, playing music, and browsing the internet, to displaying literature as an alternative to a book, but all devices utilise a visual display system and subsequently face the same problems. Visual displays are limited by the screen size of the device, as devices grow smaller and smaller developers are faced with a 'bottleneck' in the rate of data presentation. Visual displays are rendered virtually inoperable when a user's visual attention is occupied by another task, such as driving, this redundancy is also more than significant when considering the visually impaired who may be entirely incapable of operating a device based on visual information alone.

An audio display offers an alternative means of data presentation that is not limited by device size, more specifically a spatial audio display can be designed to take advantage of the human auditory system's ability to monitor multiple concurrent data streams for subject specifics. Often referred to as the 'cocktail party effect'[1], when placed in a busy environment a person is able to selectively focus their auditory attention on a single conversation whilst also 'scanning' other conversations for words of interest. Moving away from the analogy, when presented with multiple concurrent audio data steams a person can actively 'select' a source of interest whilst spontaneously monitoring other sources for salient events.

This paper presents 'Audio Flipboard' a spatial audio display and interface for browsing multiple articles of auditory media; borrowing from the visual display centric idea of the iOS application 'Flipboard', in which a user's favourite feeds and news sources are collated automatically by the app and displayed in a dynamic 'magazine' format. The user is presented with multiple aggregated audio data streams, an assortment of user selected content from sources such as internet radio stations and cloud libraries, spatialised laterally around the head in a circular pattern; the display can be rotated around to bring the stream of interest to the focal point of the display, i.e., directly in front of the user. A simple touch controller is utilised for navigation with a rotary control of display orientation and a vertical slider control for 'focus' that adjusts the attenuation level imposed on the streams radiating outward from the current focal point.

The 'Audio Flipboard' is a proof of concept of an audio display designed with little or no need for an accompanying visual display; a user should be able to navigate the display with- out ever having to look at the controller. Such a design can be implemented for use in a situation in which a user's visual attention is occupied already, such as while driving, but also provides a possibility of greater accessibility of modern devices.

2 DESIGN

2.1 Initial Design Phase Summary

The initial design phase consisted of the construction of the core system architecture within Pure Data, an open source graphical programming language developed by Miller Puckette[2], and the iOS application touchOSC[3]. A set of simple subjective tests were devised to attempt to illustrate correlation between number of audible sources in the display and user confusion.

Subjects were asked to 'locate' varying numbers of sources using an early version of the system architecture. This served firstly to verify the function of the developed system as a binaural audio display, and secondly to identify an optimum number of concurrently audible sources to be considered in further design refinements of the system.

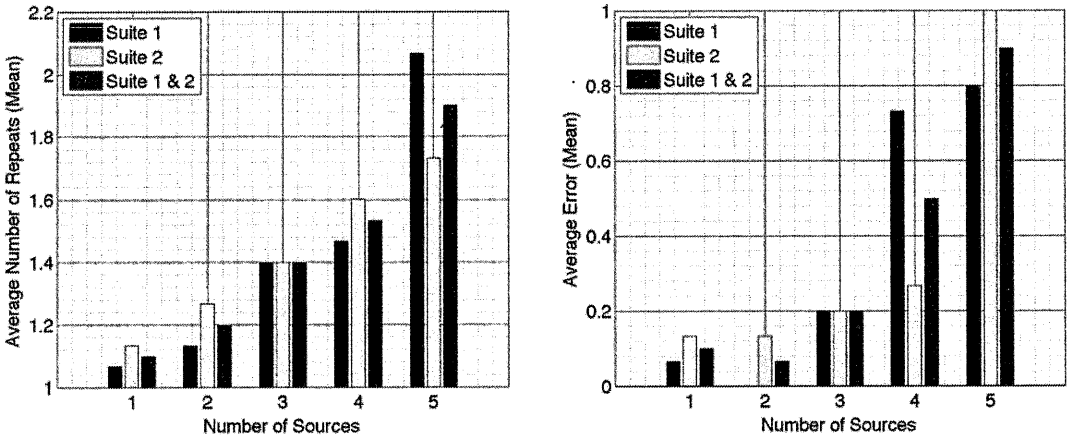

Figure 1, Average Number of Test Repeats and Average Number of Perceived Sources Error

The results in *Figure 1* represent a collection of 15 participants, and clearly illustrate that the 'difficulty' increases significantly with the introduction of more than 3 concurrent sources. These findings are congruent with findings of other authors[4].

2.2 Secondary Design Phase

2.2.1 Aim

The aim of the secondary design phase was to utilise the preliminary test data already acquired to develop and refine the proof of concept application for Audio Flipboard. Concluding with a second round of subjective testing in order to determine the relative merits and demerits of the system from the perspective of potential users. The secondary design phase testing also serves to attempt to determine the best source distribution regarding the organisation of speech and music sources.

2.2.2 System Architecture

The secondary design incorporates direct user control via touchOSC on the iPad; drawing certain parameters from the findings of phase one subjective testing, and ultimately developing a scaleable system suitable for implementation at various levels of complexity and an open ended development process.

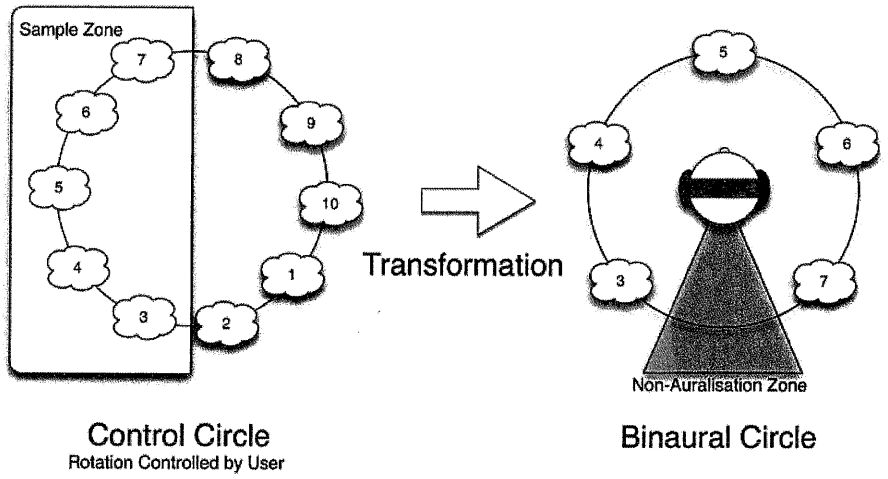

Figure 2, Secondary Design Phase Concept

Figure 2 illustrates the critical design concept underpinning the secondary design phase development: From the initial testing conducted, it was clear that five concurrent sources represents or even steps over a margin of confusion or difficulty for the listener, subsequently it was decided that five sources should be available for concurrent playback on the system, but only at one extreme of the 'focus' setting, the other extreme resulting in the auralisation of a single source directly in front of the listener, this 'focus' system has been identified in previous works as strongly recommended in the design of an audio display[1], and is explained in greater detail later in this section. With the imposition of a maximum of five sources spatialised about the listener a subsystem needed to be implemented such that the system could be used to browse a collection of more than five files or streams.

By adding a layer of abstraction between the user control input and the 'binaural circle' such a system is realised. For the purposes of development and testing a total number of sources of ten was selected, in further development this number could be expanded almost limitlessly. The 'control circle' consists of a number of sources distributed evenly about a ring, the user has control of one of the sources position on the ring, all other sources are kept at a constant offset in relation to the controlled source. A portion of the 'control circle' containing five sources at any one time, in this case exactly half or the circle, is sampled and re-mapped to the 'binaural circle'. In this case the transformation, or effectively sample rate, is simply a multiplication by a factor of 2; it is important to note that for this transformation to remain simple, the position, size, and effective 'focal' position of the sample zone must be given special consideration.

The 'earplug ' object, which provides the spatialisation cues within Pure Data, takes arguments of angle of azimuth between 0 and 360 degrees, in which 0/360 represents directly in front of the user. Therefore the post transformation angle calculated from the sample zone of the control circle must not exceed the bounds of 0 and 360, i.e., the sample zone should extend from 0 to 180 on the control circle. However this means that the focal point of the display, 0° on the binaural circle is mapped to 0° on the control circle, meaning that sources 'pop' in and out of the binaural circle directly in front the user. This was overcome by taking advantage of the inherent ambiguity between the user control scheme and the position of sources on the control circle; by including a 90° offset in the transformation function. The subsequent transform function output ranges from -90° to +90°, which is wrapped to 0° to 360° using a modulo operator.

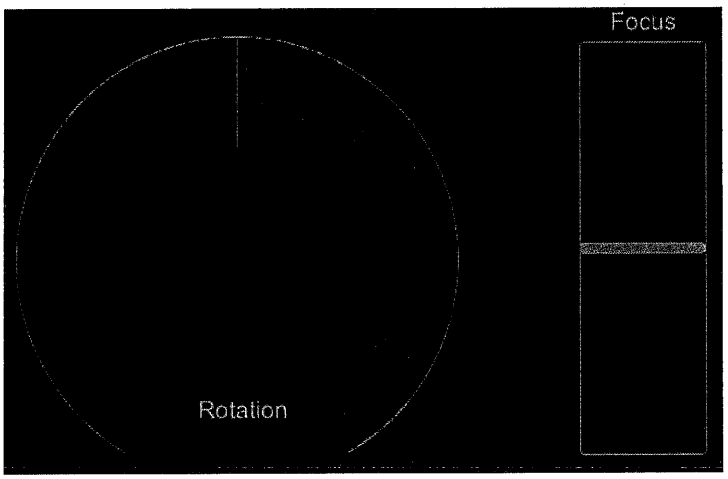

Figure 4, touchOSC Controller for Audio Flipboard

The design of the control was intentionally simple, as the design of the application uses as little visual stimuli as possible, and the control surface should be operable without drawing the user's visual attention for any more than a quick glance. The choice of a rotary fader and a vertical fader to represent rotation and focus respectively attempts to take advantage of common design paradigms that occur in day to day life[5]; if the user were to imagine themselves at eye level with the surface, the controls loosely describe turning to observe a visual stimulus and narrowing or widening one's field of view, to take in less or more of one's surroundings.

Both the rotary fader and vertical fader output numerical data ranging from 0 to 1, for the rotation control this value is scaled to represent the value 0 to 360 then offset by 180 so that the focal point of the display coincides with the default central position of the controller seen in *Figure 4*, and for the focus control the range is mapped to 0.2 to 8.2. The value is fed into an expression that determines each source's relative gain based on it's angle of azimuth, adapted from 'The Amblr'[4]. The gain expression gives a response similar to a cardioid polar pattern, the 'sharpness' of which can be altered with the manipulation of a single constant.

The equation as given by Stewart and Sandler[4]:

$$g(\psi) = e^{\frac{-\psi^2}{4}}$$

Can be re-written as:

$$g(\psi, f) = e^{\frac{-\psi^2}{f}}$$

Where 'ψ' is the angle of azimuth and 'f' is the value of the 'focus' control.

Altering the focus parameter between the range of 0.2 to 8.2 alters the polar pattern of the angular gain from a narrow single source focus to a much broader multi source focus. This facility allows potential users to 'tune' the display to a state with a number of sources of relative amplitude that they are most comfortable with, a feature specifically identified as missing in the Sonic Browser[6].

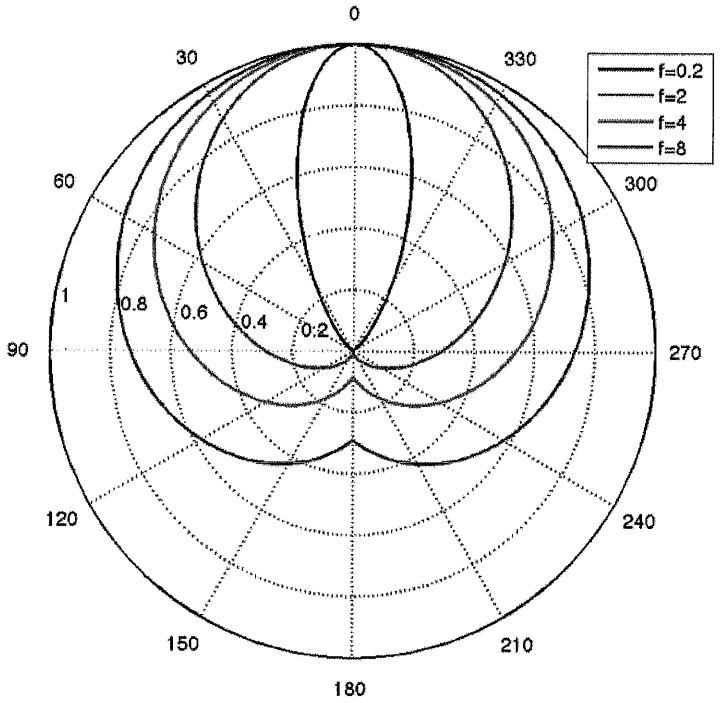

Figure 5, Polar Plots of the Gain Equation

Figure 5 demonstrates how the polar pattern response of the gain expressions changes with respect to the focus control value. The angular gain expression was introduced in order to better accentuate the 'focal point' of the display as being directly in front of the user, the concept also aims to reduce the effects of front-rear confusion and is recommended in previous works[7 4] .

2.2.3 Experimental Design

Ten students from the University of Salford were selected randomly, i.e., with no discrimination towards age or sex as participants for the experiment. The group consisted of eight males and two females between the ages of 20 and 26, eight of the participants study audio related courses and should be considered as somewhat trained critical listeners. Participants were not offered any compensation for taking part in the experiment. Each participant per- formed four short tests which although also serving to provide some objective data metrics, served primarily as an opportunity for the subjects to explore the display in a scenario that could easily be extrapolated to expected 'day to day' use of the Audio Flipboard system. The objective dependent variable of the experiment is the time taken to locate a given audio stream within the display; subjectively the purpose of the test is to identify the validity and usability of the system through feedback given by the subjects in the form of a short questionnaire and discussion with the experimenter.

2.2.4 Experimental Scenario

Each subject was led through the experiment by the experimenter working from a script where appropriate, in an attempt to remove any bias imparted by the experimenter. Before beginning the first test subjects were given an explanation of the control mechanisms and the display, then an opportunity to try navigating the display with a set of ten files not used in the experiment; this was done in order to negate any initial steepness in the learning curve or misunderstandings due to the explanation given.

For each test the subject was given a description of one of the ten audio streams in the display; five songs and five speech based broadcasts, given the song title, artist and genre description, or the broadcast title, broadcast station, and a contextual description of the broadcast respectively. Subjects were asked to locate the the audio file, line it up directly in front, draw up to full focus i.e., only that source audible, and finally press the submit but- ton, ending the test. This process of drawing full focus and aligning the target source with the focal point of the display was referred to as 'selection' by the experimenter, this word was chosen specifically in an attempt to aid the subject in extrapolating the experimental experience to practical use of the display as a consumer or user. Similar to the phase one experiment described in Section 5.1, the ten audio files selected for the test were deliberately chosen as suitably distinct, such that there was expected to be little to no confusion regarding multiple sources matching a single description given to the subjects. For each of the four tests the audio files were arranged differently amongst the sources: Five music then five speech in test one; Alternating one music one speech in test two; random permutations for tests three and four.

Upon completion of the tests each user was asked to complete a short questionnaire with the experimenter, the design of which aimed to gather partially formal feedback from the subjects regarding the viability of the display in real applications.

Each subject performed the tests in a unique order generated using the 'randperm' function in MATLAB. This was done in order to avoid any false correlation between source orders and time taken, that could have in fact been due to the influence of a learning curve or perhaps cognitive fatigue occurring with the first or final tests respectively.

2.2.5 Experimental Measures

The 'data logging' of the controller inputs during each test offer multiple experimental measures: the time taken to complete each test from the length of the stored arrays; whether or not the subject was able to identify the correct source; and finally the manor in which the subject used the controls to navigate the display, although not easily quantified, this information provides further insight into the successes or failures of the intended usage in the design stages of the project.

The data collected from the questionnaire and discussion with the subjects will also provide a mixture of quantitative and qualitative data. Quantitive directly from the five level Likert scale[8], and qualitative from the discussion and general comments made throughout the questionnaire session.

3 RESULTS

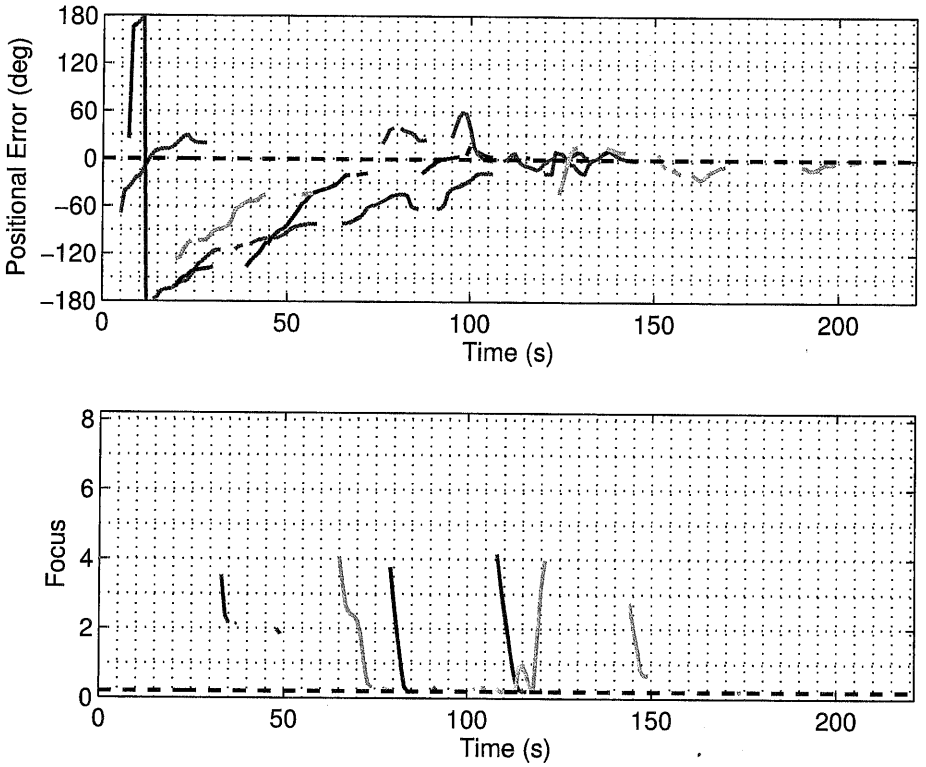

Figure 6, Subject One: Rotation Distance from Target and Focus

Figure 6 shows an example of the results regarding the subject's control of the device during the experiment, where matching colours of focus and positional error curves represent corresponding data sets from each of the four tests. The positional error shows the rotational distance from the current target source to the focal point of the display, essentially representing the subject's rotation control, wrapped to plus or minus 180°. The focus plot shows the value of the quotient of the exponent in the gain equation, for which the lower the value of the Y axis corresponded to the narrower focal width.

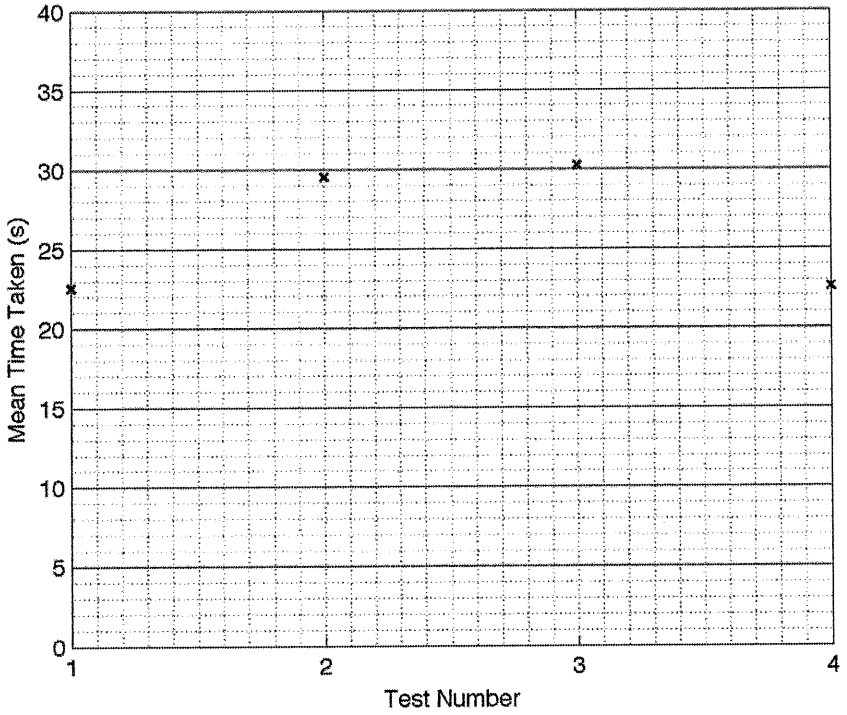

Figure 7, Mean Average Time Taken against Source Organisation

Figure 7 depicts the mean average time taken to complete each test, where test 1, 2, 3, and 4 refer to the test descriptions of source ordering: all speech then all music; alternating speech/music; and two random permutations respectively.

Test	Source Order									
1	1	2	3	4	5	6	7	8	9	10
2	1	10	2	9	3	8	4	7	5	6
3	2	7	3	10	4	1	9	5	6	8
4	8	2	3	7	5	1	9	4	10	6

Figure 8, Source Organisation by Test

The following table displays the data obtained from the Likert scale portion of the experiment feedback sessions, where 1-5 represent a negative to positive response respectively:

	Response					
Statement	1	2	3	4	5	Mean
1	0	0	0	3	7	4.7
2	0	0	0	2	8	4.8
3	0	0	0	1	9	4.9
4	0	0	0	3	7	4.7
5	0	0	0	1	9	4.9

Figure 9, Likert Scale Responses

The statements identified as 1 to 5 are as follows:

1. The design of the controls suitably represented the actions they performed.
2. It was easy to identify and 'select' a particular source of interest.
3. The controls were intuitive and easy to learn.
4. The controls were adequate to explore the display.
5. The display provided an effective means of exploring multiple audio streams.

All data processing for this section was performed in MATLAB.

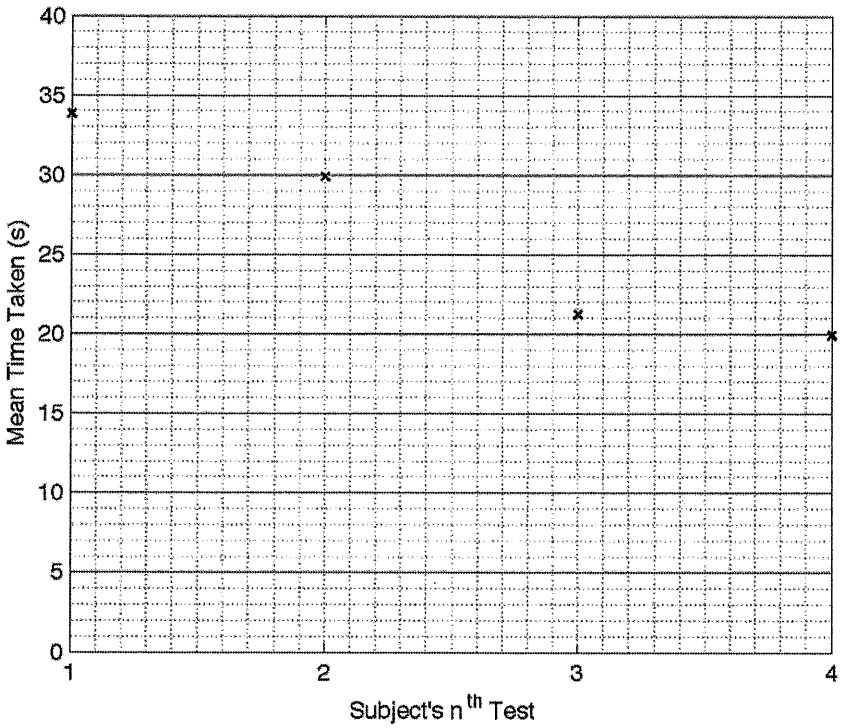

Figure 10, Number of Tests Taken against Mean Average Time Taken

4 DISCUSSION

Considering the manner in which the display was navigated, typically illustrated by *Figure 6*, it can be seen that subjects were able to 'hone in' on the target source with relative consistency, with all participants 'selecting' the correct source in all tests. Largely participants seemed reluctant to reduce the focus of the display below the default value, with two subjects in particular commenting explicitly on the increased difficulty of navigation when doing so during the feedback sessions. Interestingly all subjects seemed comfortable allowing the rotation control to roll over the ±180° point despite the visual 'break' in the otherwise circular control mechanism, due to this 'break' the audio display exhibited a noticeable 'hop' as the control value sent by touchOSC jumped as the user continued to scroll past the detection zone; this had the potential to cause an aural discomfort or perhaps even introduce an element of confusion into the experiment, however it was only mentioned by a single participant. This issue could be overcome by the introduction of a different control mechanism, ideally a 'closed' circle with the same behaviours as the current implementation; one participant did suggest the use of other control mechanisms such as a unified XY pad or perhaps iOS device accelerometer control.

Contrary to expectations, *Figure 7* shows that an alternating speech-music organisation of sources in fact led to an increase in the mean time taken to complete each test, with the all speech-all music organisation performing best of all four. However the variation in the reported mean averages are perhaps too small for a significant conclusion to be drawn, given a greater sample size with a greater number of test repetitions a more prevalent pattern may emerge.

Far more significant is the result illustrated by *Figure 10*, the mean average time taken to complete each test decreases noticeably as the subject completed each test in their unique order. Such a result is indicative of the subjective learning curve that comes with the display, seeming to level off after three tests, the data suggests that the display has a relatively shallow learning curve associated with it. This assumption is echoed in the feedback collected from the participants, with almost all participants choosing to comment on the ease of use of the display explicitly.

As shown in *Figure 9* user feedback was unanimously positive with a considerable tendency towards the highest value of the scale, the statements for which these responses were received were chosen with deliberate overlap in order to obtain more reliable results. The goal of the questionnaire was simply to establish as firmly as possible whether or not the Audio Flipboard display was of a valid design that subjects were comfortable in using, and could browse with ease and lack of confusion; the positive feedback on the Likert scale shows a strong tendency to support such statements and subsequently serves as sufficient 'proof of concept'.

The Likert scale results are well punctuated by the overall tone of the verbal open ended feedback obtained during the phase two experiment, with several participants commenting positively on the experience of using the display, several also stating their interest in utilising such a display if it were commercially available.

5 CONCLUSIONS

In conclusion the project can be considered largely successful; the Audio Filpboard display presents a novel system in which users are able to exploit simultaneity in listening effectively in browsing a collection of audio media.

The system received a strongly positive response from all experimental participants, suggest- ing that the audio display design was suitably ergonomic and has a reasonably high level of 'usefulness' in the opinions of the participants, representing a key portion of the likely target audience for such a system. It is not unreasonable to assume that the system architecture in it's current state provides a sufficient prototype platform that could be expanded into a wide range of other applications. A particular area of which such a system could be extremely valuable is, as eluded to previously is in the aid of the visually impaired.

Both the development and experimental elements of the project have proven Pure Data to be a somewhat ideal platform for system prototyping and implementation for a system of this caliber. The flexibility of the software allowed for a great deal of rapid trial and error development, and the availability of open source high level object externals greatly reduces the time taken to realise an idea or architecture that would be otherwise time consuming and comparatively complex in a more 'traditional' programming environment.

Although the experimental sample sizes were perhaps too small to conclude, beyond any shadow of a doubt, certain trends emergent from the data; the sample sizes are large enough and the feedback and data congruent enough to validate the Audio Flipboard as a viable audio display design.

6 FURTHER WORK

As it stands the Audio Flipboard system is functional, but it does not yet meet the requirements of a commercially useful standalone architecture. Immediate continuation of the works will see the exploration of running the DSP and control interface on a single mobile device; during the course of

the project several language libraries under the umbrella 'libpd' have been released. 'libpd' allows a Pure Data patch file to be placed in a 'wrapper' that another programming language, such as Objective C (iOS), can pass static media to and from. This has strong implications for the Audio Flipboard as the current system architecture may be adapted to work inside a single iOS application with relative ease.

It would be particularly interesting to attempt an implementation of the Audio Flipboard system incorporating head tracking, such that the display could 'follow' a user's head and subsequently strengthen the binaural image. It is possible that this addition could be realised through the use of an iOS device camera[9], given that the device is capable of the audio DSP it is likely able to cope with a visual parallel process also.

The current static nature of the audio files in the prototype system, although suitable for implementation on an MP3 device or similar, is ultimately limited in it's scope. A further step in the system design will be to incorporate a dynamic audio library capable of streaming content from online resources, be it internet radio stations or a personalised online library such as Spotify or Last.FM. The designation of content streams to in-system audio sources could be handled either within the Audio Flipboard application itself, through a desktop companion application, or even an internet browser plug-in.

It would be interesting to explore an additional spatial dimension, elevation, but this is not currently a priority. More appropriate for immediate attention is the inclusion of a means of HRTF optimisation, given a selection of different HRTF models the system could perform a simple 'best fit' test during a user's first interaction. The inclusion of such a subsystem would improve the spatialisation effects over a significantly wider user base.

Finally, further testing of the proof of concept design of the application would increase the validity of the conclusions of the project, but also give clearer insight into the apparent trends in the data captured during the project thus far.

7 REFERENCES

1. B. Arons, A review of the cocktail party effect, Journal of the American Voice I/O Society.
2. M. Puckette, Pure data: Another integrated computer music environment, Proceedings of the Second Intercollege Computer Music Concerts (1996).
3. hexler.net, http://hexler.net/, (2012).
4. R. Stewart and M.Sandler, The amblr: A mobile spatial audio music browser, IEEE International Conference on Multimedia and Expo, (2011).
5. W. Walker, S. Brewster, et al., Diary in the sky: A spatial audio display for a mobile calendar. 15th Annual Conference of the British HCI Group, (2001).
6. M. Fernström and E. Brazil, Sonic browser: An auditory tool for multimedia asset management, Proceedings of the 2001 International Conference on Auditory Display, (2001).
7. C. Schmandt and A. Mullins, Audiostreamer: Exploiting simultaneity for listening, CHI '95 Conference Companion on Human Factors in Computing Systems, (1995).
8. J. Preece, Y. Rogers, and H. Sharp, Interaction Design: beyond human-computer interaction, (John Whiley & Sons), (2002).
9. S. Basu, I. Esssa, and A. Pentland, Motion regularization for model-based head tracking, Proceedings of the 13th international Conference on Pattern Recognition, (1996)