An Environment for Studying the Impact of Spatialising Sonified Graphs on Data Comprehension

Rameshsharma Ramloll and Stephen Brewster

Department of Computing Science, University of Glasgow, Glasgow G12 8QQ,UK

Email: {ramesh,stephen}@dcs.gla.ac.uk

Abstract

We describe AudioCave, an environment for exploring the impact of spatialising sonified graphs on a set of numerical data comprehension tasks. Its design builds on findings regarding the effectiveness of sonified graphs for numerical data overview and discovery by visually impaired and blind students. We demonstrate its use as a test bed for comparing the approach of accessing a single sonified numerical datum at a time to one where multiple sonified numerical data can be accessed concurrently. Results from this experiment show that concurrent access facilitates the tackling of our set multivariate data comprehension tasks. AudioCave also demonstrates how the spatialisation of the sonified graphs provides opportunities for sharing the representation. We present two experiments investigating users solving set data comprehension tasks collaboratively by sharing the data representation.

Keywords

Visually impaired users, sonified graphs, spatial sound, data overview and discovery, gesture tracking, collaboration, awareness, and empirical studies

1. Introduction

AudioCave is a test bed for the study of spatialised sonified graphs in an immersive audio environment. Lessons learnt from experiments in this environment will inform the design of future applications targeting the data visualisation needs of visually impaired and blind users. In this paper, we describe the environment and refer to a number of experiments it has enabled us to perform together with the results obtained.

The first significant study for representing data by non-speech sound for blind and visually impaired users shows that such an approach can help the understanding of line graphs in this target group [1]. Another independent set of experiment evidence reveals that there is a close correspondence between the perception of auditory and visual graphs with regards to gross differences in function shape, as well as slope and level (height) perception [2, 3]. More recent evaluations studying real-world data comprehension tasks have confirmed the benefits of providing facilities to access numerical data through non-speech sounds [4]. However, most evaluations in the literature tend to focus on data comprehension tasks involving access to a single sonified numerical datum at a time. AudioCave allows us to investigate the impact of accessing multiple sonified numerical data concurrently on multivariate data comprehension tasks. We begin this investigation with an example minimal scenario requiring access to the non-speech representations of three variables concurrently.

2. An example soundscape in AudioCave

We illustrate how three sonified graphs can be placed at the corners of an equilateral triangle (25 cm sides) on a 2D surface so that each corner of the triangle produces a stream of pitches describing a relevant function. Suffice to mention that the choice of the spatial configuration of the sound sources and their number is incidental because our investigation needs a starting point. In this paper, we have not evaluated the effect of allowing users to create their own soundscapes by placing and moving the sound sources according to the requirements of data comprehension tasks. This is an important aspect we plan to investigate in the future. For simplicity, we assume that these streams are looped. In Figure 1, for example, one stream falls in pitch to a low minimum and picks up again to where it started to represent the U-shaped graph (A), another increases steadily in pitch to represent a monotonically increasing graph (B), and the last stream increases, decreases and then increases again in pitch to represent a sine graph (C). The three sources are rendered relative to a virtual observer located (1) on a user's extended index finger and (2) looking in its direction. The position and orientation of this observer are determined by an electromagnetic sensor mounted on the said index finger. The user hears three sonified graphs in the same way she would if she were to position herself, looking in the same direction as her index finger, at the centre of a similar triangle but bigger (25 m sides) with speakers (facing upwards) located at its corners and playing the relevant sonified graphs.

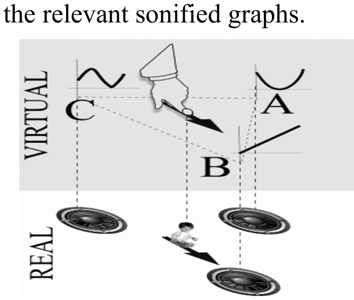

Figure 1 Virtual to real mapping in the AudioCave soundscape

3. Numerical Data to Pitch Mapping

Our sonified graphs use a simple equation (1) to map each numerical value of the data to the appropriate MIDI pitch[5, 6]. Pitch has been found to be useful for representing numerical magnitude.

$$MIDI\_Pitch = \lfloor 40 + 80 * (value) \div (high - low) \rfloor \quad (1),$$

where $high$ refers to the maximum value, low refers to the minimum value of the numerical data and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

We exclude the first 40 MIDI values. This is based on earlier findings showing the difficulty in discriminating between low pitches. We also do not use the 7 highest MIDI values because our sound card could not produce them to a satisfactory quality and discrimination between high pitches is also known to be poor. Equation 1 also gives a better linear relationship between the perceived pitch and corresponding number. The sonified graphs in Figure 1 are the results of mapping three series of numbers that describe three functions according to equation (1).

3.1. Audio Stream Characteristics

Figure 2 describes some non-spatial characteristics of auditory streams we consider to achieve satisfactory stream segregation [7]. An approximate temporal onset and decay difference of 20 ms between corresponding sonified numerical data (i.e. of the same index in each data series that describe a given function) is maintained. After a number of trials we settle for an approximate delay of 50 ms between successive tones within a stream so that the tonal sequence can be presented at a speed which does not compromise the comprehensibility of the streams. We do not allow subjects to vary these delays to facilitate the data analysis of our experimental data.

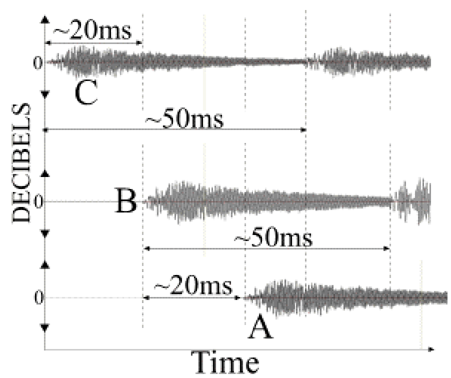

Figure 2 Amplitude characteristics of audio streams

3.2. Browsing Soundscapes and Navigating Sonified Graphs

The position and orientation of the virtual observer on the index finger, determine how the different sources of the soundscape are brought into focus. For example in Figure 3, at position 1, the sonified graph B will be heard more loudly. At position 2, the user will be able to hear stream B on her left and stream C on her right with a third fainter stream A in between the two possibly behind. At position 3, A and C will appear from the left and right respectively, with B faintly in the middle possibly in front.

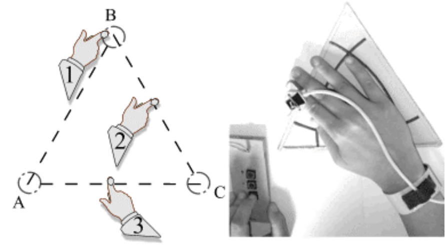

Figure 3 Positioning sensor for controlling data access

AudioCave does not only allow users to listen to looped sonified graphs. It also enables users to produce and control the streams using a keypad (Figure 3, right). The keypad contains three keys each having a specific functionality. Any key press generates three localised pitches, one for each function, at an interval of ~20 ms. All the functions in our data set each contain 100 values which when sonified are mapped to a series of 100 pitches. The top and bottom keys allow the three series to be ascended and descended respectively and generate a 'hissing' noise at the boundaries of the data sets. The middle key plays the three pitches in the current position within the series. Sounds are triggered during key press and stopped during key release.

3.3. System software and hardware

AudioCave consists of three primary software modules: (1) the data to pitch conversion module, (2) the streamed audio source-positioning module and (3) the user behaviour data-tracking module. The data to audio conversion module handles the mapping of numbers to pitch and allows a user to navigate through the data using a keypad. The streamed source-positioning module allows virtual sources to be placed in space to create the needed soundscape. It also deals with the tracking of the 6 degree-of-freedom electromagnetic sensors required to determine the position and orientation of observers within the soundscape. The behaviour-data tracking module captures the timed position and orientation of a user both within the soundscape and the sonified graph continuously.

Briefly, the AudioCave hardware (Figure 4) comprises of the following. The Lake® CP4 is an off the shelf audio convolution tool which together with its Multiscape™ application facilitate the creation of a spatial 3D sound environment populated by multiple streamed sources and observers. AudioCave uses the software hooks provided by Multiscape™ to control the virtual observers and sources within the soundscape. AudioCave's data to pitch conversion module controls three external daisy-chained synthesisers to produce the streamed audio required for each function. The Huron™ rack is an 8 input-8output connector panel for the incoming and outgoing XLR cables.

The Polhemus® FASTRAK® is an electromagnetic six-degree-of-freedom tracking instrument. It computes the position (X, Y, and Z Cartesian coordinates) and orientation (azimuth, elevation, and roll) of a sensor as it moves through space. The system utilises a single transmitter and can accept data from up to four sensors. AudioCave uses these tracked data to control the position and orientation of sources and observers in the virtual audio environment. The scenario described earlier can be constructed by moving a tracked sensor towards each corner of the triangle and dropping a sound source representing a function at that location. A sensor mounted on the index finger and coupled to an observer allows the browsing of the soundscape. In the single user setting, a virtual observer represents the user in the soundscape. In the multi-user setting, a virtual observer coupled with a virtual source carrying the speech stream of the user represents the latter in the soundscape.

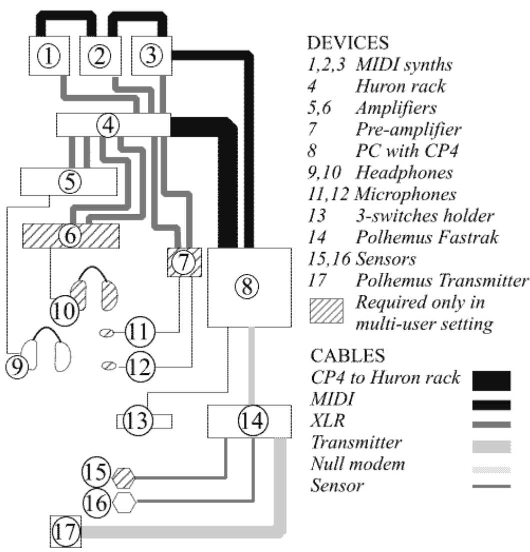

Figure 4 AudioCave hardware set-up

3.4. Evaluation methodology

Firstly, we compare the method of accessing sonified data sequentially to that in parallel for tackling multivariate data comprehension tasks. The multivariate data analysis task we focus on is the finding of intersection points between functions. While the latter is only one example of multivariate analysis, it shares common elements with other tasks in this category. Secondly, we explore the collaborative use of AudioCave to tackle another set of multivariate data analysis tasks under a number of conditions.

We make use of the NASA TLX [8] to determine the subjective workload of participants when tackling set tasks. The subjective workload (W) is the weighted average of 6 workload categories with possible scores ranging from 0 to 20. These as mental demand (M), physical demand (Ph), temporal demand (T), effort (E), performance (P) and frustration (F). In the multi-user experiments, we also record the speech exchanges between participants for conversation analysis.

All our subjects in the three experiments we present are blind folded. These do not represent our target users who are blind and visually impaired. However, our approach to test our methodology and system before repeating the experiments with our target users is both economical and practical. In addition, based on the results of previous researchers who carried out the same experiments with blindfolded and blind or visually impaired people [9], there are enough grounds to suppose that results obtained by the two groups will be similar.

4. Experiment 1: Identifying intersections

The goal of this experiment is to compare the effects of accessing sonified numerical data in parallel compared to doing the same sequentially when identifying intersection points between functions. An intersection point between sonified graphs is encountered when a close match between the pitch of each sonified datum at a given navigation position is observed.

4.1. Experiment Design

The hardware set up is as described in Figure 4. Three functions are randomly selected out of a set of 8 for the first two tasks and placed at the corners of a triangle (of sides equal to 25 cm and 2 mm thick) on a table. The blindfolded participant then browses the functions as specified by the first condition to locate intersection points. The timed location of the participant in the sonified graph is tracked throughout the task. The time of discovery and location of each intersection is also captured. After completing the intersection localisation task, participants are asked to recall the pitch change profiles and to select three best matching graphs from a set of eight. Another non-repeated set of functions is selected and the same procedure is carried out again. The NASA TLX is administered for the intersection localisation task under both conditions. The experiment is a balanced 2 condition within subject design. It lasts 1 hour 15 minutes with the first 20 minutes devoted to system familiarisation, training time and clarification of tasks.

4.2. Participants

The experiment involves 12 blindfolded participants (6 male, 6 female, 21 to 30 years old), none describing themselves as musicians.

4.3. Conditions

Serial (SER): The sensor to which the observer in the soundscape is coupled is fixed at the centre of the triangle pointing to its top corner. The participant uses one hand to select the function of interest by pressing a pre-specified key on the PC keyboard and uses the other hand to navigate the selected sonified graph using the keypad (Figure 4, device 13).

Parallel (PAR): The sensor to which the observer in the soundscape is coupled is mounted on the participant's index finger. The participant moves the latter around in the soundscape to focus on sonified graphs of interest while her free hand uses the keypad to navigate through them.

4.4. Hypotheses

H0: Participants locate more intersection points in the PAR condition than in the SER condition. H1: The subjective workload in the PAR condition is significantly less than that in the SER condition.

4.5. Data set construction

Eight data series representing 8 functions containing 100 integers between 0 and 100 (inclusive) constitute our data set. For additional insight, the data is presented in Figure 5.

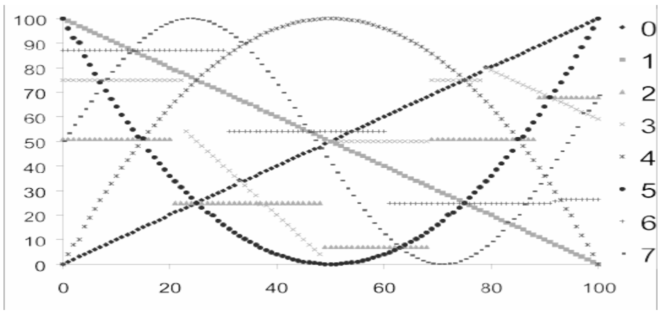

Figure 5 Plot of the 8 data series used in the experiment

Let us consider a data sonification mapping P which maps a number to a pitch and two continuous smooth functions f_1 and f_2. The design of the data representation constrains us to consider each function over the set of integers even though they may be defined over the set of real numbers. An intersection point between f_1 and f_2 occurs at integer x where $P(f_1(x)) = P(f_2(x))$. However, an intersection can occur between two consecutive integers x_1 and x_2. In addition, multiple consecutive intersection points can also be obtained. In order to simplify our intersection localisation task and facilitate the calculation of errors after task completion, we modified the data set so that the last two conditions did not occur. Other experiments will be carried out in the future to deal with situations where such constraints do not hold. In any case, we are so far primarily interested in whether our sonification approach narrows down the search space for our intersection point identification tasks and to keep the experiments as simple as possible.

4.6. Task definitions

A1. Find (0 to any number) instances where the three given functions have identical values. (8 minutes maximum)

A2. Select graphs representing the profile of pitch changes you encountered in the previous task from a list of 8 graphs (2 minutes maximum).

A3. Find (0 to any number) instances where any two of the three given functions have identical values. (8 minutes maximum)

A4. Select graphs representing the profile of pitch changes you encountered in the previous task from a list of 8 graphs. (2 minutes maximum)

4.7. Results

The average error is the average of the difference between the position of the actual intersection point and that of the discovered one. A discovered intersection point is deemed correct if the error is less than or equal to 2.

Questions	A1	A1 Error	A3	A3 Error
SER	3/12	1.2	6/12	1.5
PAR	9/12	1.4	16/12	1.3

Table 1 Average correct results for tasks A1 and A3

	SER	PAR	T_{11}	p
A2	30/12	31/12	-0.43	0.674
A4	31/12	29/12	1	0.338

Table 2 Average correct results for tasks A2 and A4

	M	Ph	T	E	P	F	W
3SER	14.6	7.3	12.2	10.5	9.3	9.8	11.9
3PAR	12.3	5.3	7.4	7.2	5.6	6.3	9.0
T_{11}	2.8	2.9	3.5	2.2	3.3	2.7	3.3
p	0.018	0.015	0.005	0.046	0.006	0.020	0.007

Table 3 Average TLX data for task A1

	M	Ph	T	E	P	F	W
2SER	13.3	6.0	10.3	9.1	8.6	8.4	10.8
2PAR	9.8	3.1	6.9	7.3	5.6	6.2	7.8
T_{11}	3.5	3.2	2.3	1.4	2.1	2.2	2.7
p	0.004	0.009	0.041	0.197	0.058	0.053	0.019

Table 4 Average TLX data for task A3

4.8. Discussions

Table 1 shows that there are more intersection points identified per participant for tasks A1 and A3 in the PAR condition than in the SER one. The average error in the SER condition is less than that in the PAR condition for task A1, but the average error is less in PAR condition for task A3. The average errors can be attributed to varying pitch discrimination abilities under various cognitive loads. However, they are small under both conditions. This illustrates that extensive musical abilities are not prerequisites to tackle set tasks. These results are encouraging especially because in a real application, participants will have the opportunity to narrow down their search space significantly before accessing the details in speech.

Table 2 shows that the majority of participants tackle these tasks successfully. There is no significant difference between the ability of participants to recall the pitch change profiles and match them with their sonified graph counterparts in the SER and PAR conditions. Therefore accessing sonified graphs in parallel does not appear to break down overview information about each sonified graph in AudioCave.

Table 3 shows that the overall workload and individual workload factors expressed by participants for finding an intersection point between three sonified graphs are significantly ($p < 0.05$) less in the PAR condition than in that the SER condition.

Table 4 shows overall workload is significantly ($p < 0.05$) lower in the PAR condition than in the SER condition when finding intersection points between pairs. However, this is not the case for effort, performance and frustration.

In general, the statistical results confirm hypotheses H0 and H1 which are also supported by post experiment comments by the participants. The result in table 4 is not as good as in table 3 because some participants find that having to listen to an extra stream, even if it is faint and far away can just be distracting noise to the task of finding an intersection between 2 functions. However, this view is not shared by others who do not find a third non-needed source so distracting that it hampers their current task. A number of participants report that the decrease in temporal and physical demands in the PAR condition for both A1 and A3 is distinctly noticeable. This result is also supported by the navigation trace of participants within each function. This is a plot of their position in each series of data against time and not presented here because of space considerations. The plot shows that participants tend to navigate the whole series of data probably to get a 'picture' of the whole sonified graph before starting to look for an intersection point. In the PAR condition, participants are observed to speed up in regions where the pitch differences are high and to slow down when the difference is low.

The majority of our users do not perceive genuinely out of the head sounds. This is not unexpected because a common Head Related Transfer Function [10] is used. However, they are all convinced of the generally correct behaviour of the soundscape when changing the position and orientation of their index finger. Determining approximately how far sources are towards the left or right seems easy. Finding whether a source is in front or in the back is problematic. This observation is supported by tracked data of the sensor mounted on the index finger of participants trying to locate a sound source. However, users develop strategies to localise sources quickly. For example, the front back discrimination problem is resolved by paying attention to the change in volume during vertical navigation. These observations are very similar to those obtained by other researchers [11] investigating simpler spatial audio environments mainly based on stereo panning.

5. Designing the collaborative AudioCave

Our experience with blind students in a classroom reveals that interactions with a data representation, even a non-visual one, is often punctuated by a need to discuss the representation with others. Similar situations arise when the teacher is trying to explain the nature of some data to these students. Auditory assistive applications accessed through headphones can cause isolation by obstructing speech exchanges during collaborations. In AudioCave, this shortcoming is addressed through the sharing of the virtual environment. The aim is to provide opportunities for auditory data representations to be shared in the same way their visual counterparts are. While researchers have spent considerable effort studying shared workspaces in the visual medium [12], we are not aware of similar work in the purely auditory medium.

Figure 4 presented earlier describes the hardware set-up of the multi-user version of AudioCave. The unidirectional microphones capture the speech of the participants to produce corresponding virtual speech sources within the shared soundscape. The additional sensor is used to track the position and orientation of the index finger of the other participant thereby catering for her 3D audio rendering needs. Our hardware as is can provide support for 4 participants in all. However, we do not consider scalability an immediate concern. The current goal is to study the opportunities presented by the spatialisation process for sharing the auditory representation.

5.1. Geometric transforms to bridge the physical separation

The receiver sensors mounted on the index finger of each participant, in addition to specifying the location and orientation of the virtual observers, also dictate where corresponding relevant speech sources are located. A simple strategy allows participants to share the same virtual workspace while operating in their own private physical workspace. This is achieved through geometric transforms applied to parameters tracked by sensors.

Transforming the coordinates tracked by the sensor (M) of one user using equation 2 produces the environment illustrated by Figure 6. In this case, the orientation of the soundscape with respect to each participant remains the same. We refer to this spatial relationship between the soundscape and participants as the WYHIWIH (pronounced Y-hee-wee, What You Hear Is What I Hear) configuration. This is similar in spirit to the WYSIWIS [13] (What You See Is What I See) configuration, a foundational abstraction for multi-user interfaces that expresses many of the characteristics of a chalkboard in face-to face meetings. In this configuration, the virtual representations of both participants are present in the circled region of Figure 6 (right hand coordinate frame of reference). Recall that azimuth is rotation about the z-axis.

$$\{x, y, azimuth\} \rightarrow \{2d - x, -y, 180 + azimuth\} \quad (2)$$

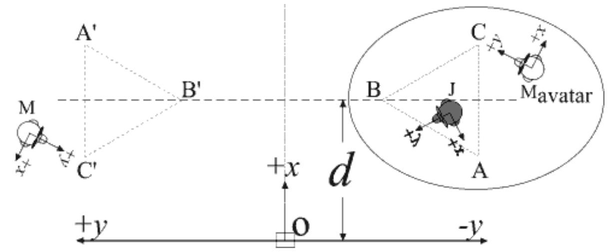

Figure 6 WYHIWIH configuration

Figure 7 illustrates the auditory perception of the soundscape by two participants in this configuration. Two participants M and J are sharing a soundscape consisting of three sonified graphs at A, B and C. Each participant has a receiver sensor mounted on his or her index finger. The raw coordinates from J are sent unmodified to the rendering engine while, the raw coordinates of M are transformed using equation 1 before they are sent to the rendering engine. The effect is that both M and J perceive they are sharing an identical soundscape and the behaviour of the avatar of their peers appears to be consistent. For example M will hear J's voice coming from the left and J will hear M's voice coming from the right.

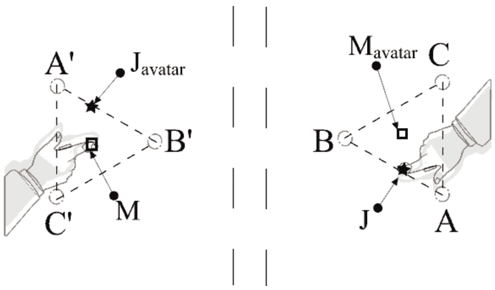

Figure 7 Controlling audio avatars in AudioCave

In line with the requirements of our next experiment design, we implement another soundscape configuration by transforming the coordinates tracked by the additional sensor for the peer user using equation 3 to produce the environment illustrated by Figure 8. We refer to this relationship between soundscape and participant as the relaxed-WYHIWIH configuration.

$$\{x, y, azimuth\} \rightarrow \{x, y-l, azimuth\} \ldots\ldots\ldots(3)$$

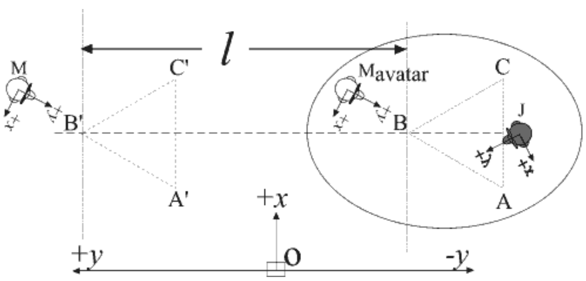

Figure 8 Relaxed-WYHIWIH configuration

Figure 9 describes two participants collaborating using AudioCave in the WYHIWIH configuration. The photograph has been augmented with information about the virtual sources representing the sonified graphs on the table. The next two experiments study data comprehension tasks tackled collaboratively under various conditions. We are in particular interested in investigating (1) whether our strategy of coupling the speech stream source of a participant with his or her observer leads to usability problems, (2) the impact of the choice of soundscape configuration on collaborative interactions between participants and (3) how collaboration actually takes place.

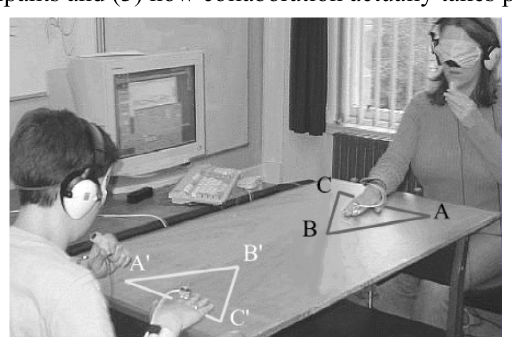

Figure 9 Discussing sonified graphs

5.2. Experiment 2: Collaborating in the WYHIWIH configuration

5.2.1. Participants

8 (4 male, 4 female, 21 to 31 years old) blindfolded participants take part in this experiment. There is only 1 person who describes herself as a musician in this group.

5.2.2. Experiment Design

The hardware set-up is as described in Figure 4. The experiment is carried out in a WYHIWIH configuration. In this experiment, the sources are placed on the bare table so that no tactile cues regarding the position of the sources are available. The sonified graphs are played continuously in loops. There is a gap of 2 seconds between loops. The experiment is a balanced 2 condition within subject design. It lasts 1-hour 20 minutes, with the first 20 minutes devoted to system familiarisation, training time and clarification of tasks.

5.2.3. Conditions

Dynamic Speech Source (DSS): In this condition, the speech of each participant is collocated with the position of the finger carrying the sensor.

Fixed Speech Source (FSS): In this condition, the speech and that of his or her peer appear to come from two adjacent and fixed points at the centre of the equilateral triangle in the immersive audio space.

5.2.4. Hypotheses

H3: The subjective workload of participants when tackling tasks in the DSS condition is significantly less than that in the FSS condition. H4: The number of errors in the DSS condition is less than in the FSS condition.

5.2.5. Training

Participants are required to play a hide and seek game to practise the localization of their peer in the immersive audio environment. This game involves each participant trying to find his or her peer counting 1 to 10 repeatedly until he or she is found. Once this situation is reached, roles are reversed and the game is started again.

5.2.6. Data set

The data set constituted of 800 values ranging from 0 to 100 to form 8 sonified graphs.

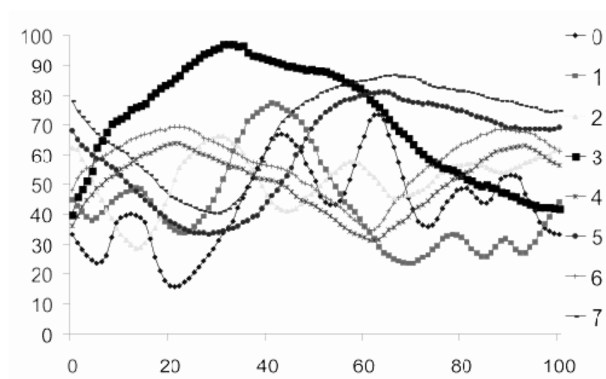

Figure 10 Data set used for collaborative data comprehension tasks

5.2.7. Task definitions

Participants are asked to collaborate and to reach a solution for the following set tasks:

B1. Which pair of sonified graphs have maxima and minima occurring concurrently for the whole portion of the sonified graphs presented? (8 minutes maximum)?

B2. Which of the three sonified graphs (if any) contain the absolute maximum? (8 minutes maximum)

B3. Which of the three sonified graphs (if any) contain the absolute minimum? (8 minutes maximum)

At the end of the experiment, the participants are subjected to the NASA TLX test to evaluate their subjective workloads under both conditions.

5.2.8. Results

The fraction of incorrect answers for B1, B2 and B3 in FSS are 0/8, 1/8 and 1/8 respectively. The fraction of incorrect answers for B1, B2 and B3 in DSS are 0/8,0/8,1/8.

	M	Ph	T	E	P	F	W
DSS	9.1	2.6	8.1	7.8	2.2	6.1	7.0
FSS	12.5	4.1	11.9	11.0	4.1	9.2	9.8
T_7	-4.2	-1.6	-3.2	-2.3	-1.6	-4.1	-3.9
p	0.004	0.163	0.015	0.054	0.158	0.004	0.005

Table 5 Workload experienced in WYHIWIH

5.2.9. Discussions

Most participants complete the set tasks successfully within the 8 minutes allocated to them in both the DSS and FSS conditions. The error levels in both conditions are comparably low. Table 5 shows a significant decrease ($p<0.05$) in the mental, temporal, frustration and overall workload experienced by participants in the DSS condition. H3 is confirmed and H4 refuted.

5.3. Experiment 3: Collaborating in the relaxed WYHIWIH configuration

Experiment 3 is very similar to Experiment 2 except that it is carried out in the relaxed WYHIWIH configuration.

5.3.1. Participants

8 (4 male, 4 female, 22 to 30 years old) blind folded participants take part in this experiment. No one has described himself or herself as a musician in this group.

5.3.2. Experiment Design, Conditions, Training and Tasks

The experiment design and conditions are identical to that of Experiment 2. The training is identical to the one described in the preceding experiment. Participants are asked to collaborate and to reach a solution for the tasks: C1, C2, C3 which are identical to B1, B2 and B3.

5.3.3. Hypotheses

H5: The subjective workload of participants when tackling tasks in the DSS condition is significantly less than that in the FSS condition. H6: The number of errors in the DSS condition is less than in the FSS condition.

5.3.4. Results

The number of incorrect answers for C1, C2 and C3 under DSS are 1/8, 2/8 and 1/8 respectively. The fraction of incorrect answers for C1, C2 and C3 under FSS are 0/8,2/8,1/8 respectively.

	M	Ph	T	E	P	F	W
DSS	10.5	1.9	4.8	6.3	4.0	4.4	6.8
FSS	10.4	2.1	4.9	6.5	3.2	4.9	6.9
T_7	0.4	-1.0	-0.5	-0.8	1.1	-1.1	-0.1
p	0.731	0.351	0.598	0.451	0.320	0.316	0.941

Table 6 Workload experienced in relaxed-WYHIWIH

5.3.5. Discussions

Again, most participants are able to complete set tasks successfully. Table 6 shows no significant difference between the workload categories and the overall workload. Both H5 and H6 are refuted.

6. Meta-analysis of empirical results

The quantitative results suggest that the soundscape configuration has an effect on the significance of the workload difference between the DSS and FSS conditions. We cannot however explain why the average workload in the relaxed WYHIWIH configuration is less that that in the WYHIWIH configuration. Collaborative applications are known to be hard to evaluate using traditional hypothesis testing backed by statistical analyses. We therefore analyse the conversations of participants to gain additional insights in the sharing of the representation.

Observations of the collaborating participants show that they typically spend the beginning of the allocated time to visit the various sonified graphs independently. They then try to locate each other by using spatial audio cues. For example, they will quickly recognise that they are close to each other when perceiving a marked increase in the speech volume of their peer. Interactions within the audio space clearly show that participants are willing to and succeeded in talking about the relevant sonified graphs. Various groups exhibit a wide range of interactions within the audio space ranging from pairs who are frequently engaged in persistent meaningful conversations to those who talk to each other only on rare occasions mainly to check their answers. A small minority of users point out that they do not make much use of the spatial properties of the speech because they are more interested in the contents of the exchange. Our meta-analyses reveal the following main elements of the collaboration framework emerging in our shared audio space.

6.1. Preference for WYHIWIH configuration

Most participants find the WYHIWIH configuration more natural. They report that it allows them to give directions more easily and it is less complicated to describe the location of objects in the auditory environment.

6.2. Establishment of a common model of the audio space

Participants in the beginning of their collaboration often establish at the onset where the various sonified graphs are, confirm this information with their peer and make sure that they have a common model of the soundscape. This is typically achieved much faster in the WYHIWIH configuration.

6.3. Workspace awareness

In the shared audio space, awareness of peer location and current context are achieved by interpreting the spatial characteristics of peer speech. This information provided and exploited passively through the shared workspace, allows users to move smoothly between close and loose collaboration, and to assign and coordinate work dynamically[14, 15].

6.4. Audio Deixis

Speech content in the FSS condition tends to be low in referential ambiguity (where reference is verbally explicit as well as deictically indicated by gesture, e.g. "Listen to A, I think it goes up, hangs in there for a while and then drops slowly"). Speech content in the DSS condition tends to be high in referential ambiguity (deictic reference alone, e.g. "Isn't this one rising all the time?"). We have demonstrated that deixis can occur in a purely auditory medium.

6.5. Divide and conquer strategy

Many participants adopt a divide and conquer approach to tackle the set tasks. This approach allows participants to distribute the cognitive load to make sense of the data. Gaver has reported a similar observation in the collaborative auditory soundscape of ARKola [16].

6.6. Iterative checking and confirmation

Participants often request their peer to check observations that they are not sure about. The iterative checks also raise participant confidence about their answers to the various set questions. A conflicting observation tends in all our cases to trigger rechecks until a consensus is reached.

6.7. Sharing of insights

Participants will often visit the sonified graphs in different orders and discover 'interesting' aspects of the data which they naturally share with or describe to their peer during conversations.

7. Conclusions

This paper describes the platform we used to investigate some advanced interactions with non-speech numerical data representations. We summarise here findings based on experiments carried out in AudioCave. Results from Experiment 1 in the AudioCave environment show that (1) providing access to three sonified numerical data concurrently facilitates tackling our intersection point localisation tasks and (2) this approach does not prevent the construction of trend information for each function. Results from Experiment 2 and 3 show that the spatialisation process offers rich opportunities for sharing the auditory data representation. The advantage of coupling the participant's speech source with his or her representative observer becomes more evident in the WYHIWIH configuration. The experiments will be repeated with blind and visually impaired students in order to compare the results.

8. References

1. Mansur, D.L., *Graphs in Sound: A Numerical Data Analysis Method for the Blind*, in *Computing Department*. 1975, University of California Davis: California. p. 65.
2. Flowers, J.H. and T.A. Hauer, *Musical versus visual graphs: Cross-modal equivalence in perception time series data.* Human Factors, 1995. **37**: p. 553-569.
3. Flowers, J.H., D.C. Buhman, and K.D. Turnage, *Cross-modal equivalence of visual and auditory scatterplots for exploring bivarate data samples.* Human Factors, 1997. **39**: p. 341-351.
4. Ramloll, R., S. Brewster, and W. Yu. *Using non-speech sounds to improve access to 2D tabular numerical information for visually impaired users*. in *IHM-HCI 2001*. 2001. Lilles, France.
5. Pollack, I. and L. Ficks, *Information of elementary multidimensional auditory displays.* Journal of the Acoustical Society of America, 1954. **26**: p. 155-158.
6. Brewster, S.A., P.C. Wright, and A.D.N. Edwards. *Experimentally derived guidelines for the creation of earcons*. in *HCI '95*. 1995. Huddersfield: Springer Verlag.
7. Yost, W.A. and G. Gourevitch, *Auditory Image Perception and Analysis: The basis for hearing.* Hearing Research, 1991. **56**: p. 8-18.
8. Hart, S.G. and C. Wickens, *Workload assessment and prediction.*, in *MANPRINT, an approach to systems integration*, H.R. Booher, Editor. 1990, Van Nostrand Reinhold: New York. p. 257-296.
9. Yu, W., *et al. Exploring computer-generated line graphs through virtual touch*. in *ISSPA 2001*. 2001. Kuala Lumpur: IEEE Catalog Number: 01EX467.
10. Begault, D.R. and E.M. Wenzel, *Headphone Localization of Speech Stimuli*, in *Proceedings of the Human Factors Society 35th Annual Meeting*. 1991. p. 82-86.
11. Pitt, I.J. and A.D.N. Edwards, *Navigating the Interface by Sound for Blind Users*, in *Proceedings of the HCI'91 Conference on People and Computers VI*. 1991. p. 373-383.
12. Ishii, H. and M. Kobayashi, *ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact*, in *Proceedings of ACM CHI'92 Conference on Human Factors in Computing Systems*. 1992. p. 525-532.
13. Stefik, M., *et al.*, *WYSIWIS Revised: Early Experiences with Multiuser Interfaces.* ACM Transactions on Office Information Systems, 1987. **5**(2): p. 147-167.
14. Dourish, P. and V. Bellotti, *Awareness and Coordination in Shared Workspaces*, in *Proceedings of ACM CSCW'92 Conference on Computer-Supported Cooperative Work*. 1992. p. 107-114.
15. Gutwin, C. and S. Greenberg, *Workspace Awareness for Groupware*, in *Proceedings of ACM CHI 96 Conference on Human Factors in Computing Systems*. 1996. p. 208-209.
16. Gaver, W.W., R.B. Smith, and T. O'Shea, *Effective Sounds in Complex Systems: The ARKola Simulation*, in *Proceedings of ACM CHI'91 Conference on Human Factors in Computing Systems*. 1991. p. 85-90.