

Do Localised Auditory Cues in Group Drawing Environments matter?

Rameshsharma Ramloll

Co-operative Systems Engineering Group, Lancaster University, LA1 4YR Lancaster, UK
ramloll@comp.lancs.ac.uk
Phone: +44 1524 59-3041

John Mariani

Co-operative Systems Engineering Group, Lancaster University, LA1 4YR Lancaster, UK
jam@comp.lancs.ac.uk
Phone: +44 1524 59-3041

Abstract

In this paper, we present the design rationale for a group drawing tool exploiting localised auditory cues to describe user activities. Our hypothesis is that these cues are important for two reasons. Firstly, they make participants aware of the details of execution of peer activities. This is especially significant when these activities are out of visual focus. Secondly, they convey intentional information among participants. The latter has been found to influence significantly inter-participant conversations during real world collaborative drawing activities. Our approach for adding sounds to the group drawing environment involves associating localised auditory messages to the palette, tools, primitive drawing objects and cursors representing metaphoric hands or points of gaze. These mappings give rise to dynamic soundscapes describing operations being or intended to be performed. We discuss the auditory authoring and browsing techniques involved in our group drawing environment together with their implications for the design of future collaborative environments involving auditory ecologies.

Keywords Collaborative drawing, intentionality, awareness, gestures, auditory localisation.

1 Introduction

A number of studies on collaborative drawing in the Computer Support for Co-operative Work (CSCW) community indicate that what is drawn on a shared workspace is the result of a complex process between participants [11,17]. When designing virtual environments to support this activity, it is important to provide as far as possible, affordances to cater for the process as a whole instead of simply, its output. A common observation that reiterates this view is that even though shared white boards have been around for a while, they are not often used mainly because of their inherent inability for transmitting rich gestures which are important for conveying intentionality information. Tang found that in the real world,

'hand gestures play a prominent role in all workspace activities (about 35 percent of all actions). Participants gesture to enact ideas, to signal turn-taking, to focus the attention of the group and to reference objects on the worksurface'[17].

This suggests that in addition to representing the results of any drawing activity on the shared workspace, information about who has the intention of carrying it out and how it is being or likely to be performed, also have to be supplemented. In the real world, the visual medium is the dominant medium through which information about gestures is accessed. The intentions of a participant can usually be derived by observing the tools she selects and her gesturing, such as finger pointing, head rotations, or gaze shifts, either for rehearsing any drawing activity or for indicating areas of interest. However, in virtual collaborative drawing environments, such visual cues are difficult to capture especially when only the worksurface is represented. A number of research efforts have focused on the visual aspects of gesturing such as,

'how to display a gesture on screen? If there is to be a gesture icon how large it must be? How should it move?'[10]

We are however interested in how the auditory medium can be exploited to express gestures when they are out of visual focus. In addition, providing appropriate auditory information may help to reinforce the functionality of gestures themselves. In our opinion, gesturing is not always a strictly visual phenomenon but can also be a combined visual and auditory one, and sometimes, even a purely auditory one. In this view, for example, the increasing or decreasing volume of footsteps may be interpreted as an auditory gesture indicating a possible increase or decrease in the interest of the person, who is doing the walk, about the observer. Another real world example is as follows. In a real world collaborative drawing environment, a participant may be aware that one of her peers wants to draw in some specific area, if she hears the latter shifting away objects covering the spot where she intends to make her contribution.

In this paper, we focus on the design of auditory awareness widgets [8,9] and explain how they may be used to improve interactions between participants engaged in virtual collaborative drawing sessions. We mention in passing that our drawing tool is built as part of an effort to get an insight into the design of auditory interfaces for

collaborative environments in general. Thus, the group drawing environment also suits our purpose by providing us with the opportunity for experimenting with various sound mappings describing peer activities. The lessons learnt from this exercise will be useful when building the auditory interface of our multi-user desktop, Moksha, which is described elsewhere [15]. Even though research about the use of sounds in collaborative environments have been reported earlier [7,8], environments allowing designers to get an insight into these developments through hands-on experimentation are mostly within the reach of the research labs where they originated. The easiest way for us to study auditory soundscape design is therefore to create and interact with a minimal collaborative environment involving an auditory ecology. Hopefully, this process will allow us to learn about the intricacies of soundscape design which is critical for the success of our future multi-user desktop environment. The similarity between the group drawing environment and Moksha is that in both cases, activities or operations are performed on shared objects, namely, graphics primitives in the former and desktop elements, such as file, application or device icons, in the later. Representing the location and nature of these activities in such a way that they can be ubiquitously perceived by participants is a general goal we wish to pursue.

At this point, it is appropriate to mention a word about our choice of a vector based drawing environment rather than a bitmap based one. Such a drawing environment provides a wealth of opportunities for the construction of complex and abstract soundscapes as it involves basic graphics primitives such as lines, circles, squares etc. and drawing operations that only exist in the electronic medium. The sound mapping for a bitmap drawing environment while being more straightforward to implement, is more likely to attract a sound mapping that naturally reflects corresponding real world sounds. Therefore, we chose a vector drawing environment rather than a bitmap drawing one because we need to learn more about abstract soundscape design and certainly how participants cope during collaborative tasks in the resulting fairly complex abstract soundscapes.

2 Overview of the group drawing environment

2.1 The analogous drawing scenario in the real world

We first describe the real world collaborative drawing scenario (Figure 1) that we attempt to approximate in the electronic medium. It is fairly restrictive but can be quickly implemented and is good enough to produce an environment reasonably rich in participant-participant interactions, which is much needed for our experiments. In this environment, only one drawing operation can be performed at a time with the rest of the participants being observers. Here, Marash is said to be holding the floor and performing the drawing operation while his peers are mere observers. All participants can control their visual focus on the shared workspace. The drawing operation is said to be asynchronous, as only one person can draw at a time. On the other hand, the observation operation is synchronous as all participants are in control of their own private focus of vision concurrently. They thus have concurrent independent areas of visual focus or views of the same workspace. The participants have some awareness of the current holder of the drawing tool and the focus of vision of peers. For a participant to be able to start a drawing operation, she has to make a request for the tool from the current user who may in turn give it to her, or she also has the option to grab the drawing tool. The social rules the participant decide to follow to determine whose turn it is to do a drawing operation determines the floor control policy they adopt.

In the corresponding approximation of the group drawing environment in the electronic medium, each participant has a particular view of the common canvas as shown in Figure 2. Each participant can scroll to any desired location at will but is able to perform drawing operations only if he has floor control. Holding the floor in our group drawing environment is equivalent to holding the unique drawing tool in the real world scenario. In Figure 2, Marash has floor control and is selecting a 'cloud' object (object A in Figure 2) in his view. Provision is made to allow each participant to be aware of the position of each other's window, together with the nature and location of activities performed on the common canvas through cues in the auditory medium.

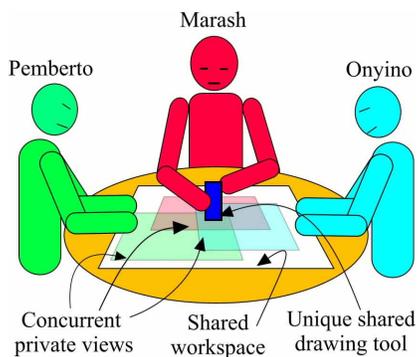


Figure 1 Group drawing in the 'real'

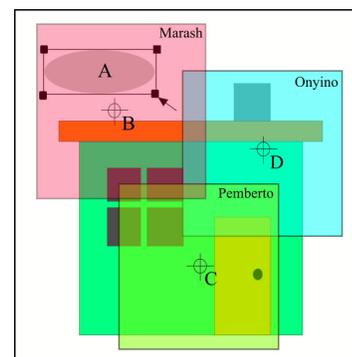


Figure 2 Group drawing in the 'virtual'

2.2 A user view of our group drawing environment

2.2.1 Floor control

Before being able to perform a drawing operation, the participant has to get hold of the floor. Clicking on the 'get floor' button of the floor manager window (Figure 3) informs his client to make a request for the floor. If the peer currently holding the floor is not currently doing any drawing operation, the 'get floor' request succeeds and the participant has floor control. When the participant gets the floor, events describing drawing operations from his client are broadcast to peer clients. However, events about the position of a client's window are always broadcast to peers irrespective of who is actually holding the floor.

2.2.2 Group drawing environment functionality and sound mapping strategy

For a user currently holding the floor, clicking on a button in section (a) allows her to draw simple objects such as squares, circles and straight lines, type text, select objects, delete them, move them around, duplicate them and move the common canvas. She is also free to work at the zooming level of her choice by clicking the appropriate scaling button from section(c). Our sound mapping approach makes use extensively of parallel auditory messages to describe complex situations on the shared canvas[3][4]. Table 1 describes the sound mapping strategy used in our collaborative auditory drawing environment. The representative sounds are grouped into 5 categories namely Primitive, Adjective, Participant, Action and Status sounds. Clicking on a primitive graphics object button such as a circle causes a 'Primitive' sound which is a short .wav file describing a base loop to be played. A similar design option has been described elsewhere [5]. If the object to be drawn is a filled circle, then the 'Primitive' sound is mixed with the 'filled' adjective which is another sound. Status sounds are triggered to describe the state of progress and details of an operation. Each colour of the palette is associated with a voice message informing participants of the colour being picked. Short voice messages are also associated with actions select, move, delete and duplicate. The selection status is described by causing the selection operation to trigger the source associated with the selected object and localising it with the later. For example, selecting a circle will trigger a 'circle' sound localised with the position of the circle. Selecting different objects at once will cause all the relevant representative sounds to be triggered and localised with the corresponding objects. This feature is presumably useful to inform peers of the details of the selection operation and of the presence of hidden selected objects. Users can also use the selecting operation as a way to create placeholders or to guide peers to a particular place.

It is also useful to group the representative sounds according to the representative sound location strategy (Table 2). The representative sounds of activities of the owner of the drawing client are collocated with the pointer (e), the metaphoric hand of the user. However the sound sources representing the participants themselves are collocated with their respective listeners. The localised representative sound sources are rendered relative to a sound sensor or listener which is positioned at (f) and described by a graphical widget to remind peers of its location. By default, the sound sensor (f) is positioned at the centre of the visible portion of the shared canvas. Thus, just by listening, it is hoped that participants will have some awareness of the occurrence and location of given operations even if they are out of visual focus. Associating sources to the sensors also allow the positions of peer windows on the canvas to be deduced. This is an attempt to provide for mutual awareness in our group drawing environment.

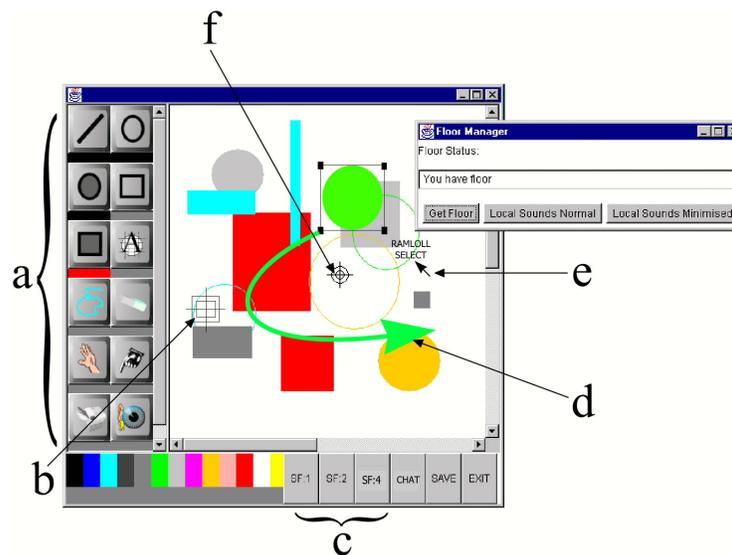


Figure 3: snapshot of the shared auditory tool

Primitive	Adjective	Participant	Action	Status
Straight-line		Marash	Select object(s)	Progress
Circle	Filled	Pemberto	Move object(s)	Start
Square		Onyino	Delete object(s)	End
Text		Ramloll	Duplicate object(s)	
			Select color	

Table 1: Sound mapping strategy to describe drawing operations

Representative sounds	Primitive	Adjective	Action	Status	Participant
Localisation coupling	Coupled with position of graphics objects and/or mouse pointer				Coupled with the position of participant listener

Table 2: Sound mapping strategy to describe location of drawing operations and window positions

2.2.3 Look, feel and sound of the group drawing environment

The following scenario describes the 'look, sound and feel' of the interface for two participants. The snapshot in Figure 3 is that of participant Ramloll's client. He has selected a filled circle and is moving it along the path described by arrow (d). Ramloll's listener (f) is positioned at the centre of the view area. The other participant, Marash, has a listener (b). Ramloll is informed of the presence of Marash by hearing the sound source associated with his listener (b). Thus if the intensity of this sound increases from the left, Ramloll knows that Marash has growing interest in viewing what he is doing. Marash on the other hand, by listening to the sounds can infer what object has been selected and what operation is being performed on it. In this case, he is able to infer where the selected filled circle is being dragged. He is also presumably aware of the viewpoint of Ramloll by listening to the sound source associated with Ramloll's listener (f). The 'local sounds normal' and 'local sounds minimised' buttons on the floor manager window provides participants with the ability to shut down sounds due to local operations. This may help them to focus on sounds associated with peer listeners.

3 Implementing the group drawing application

3.1 Soundscape construction

The sound rendering is based on Intel's 3D RSX [16]. To keep matters simple, all sources are point sources and their ranges are kept uniform. We believe that it may be a good idea to provide participants with the facility for changing the sound ranges dynamically in order to create soundscapes that suit their ability for the interpretation of sonic events. However, this facility is not provided in the current prototype. The auditory messages associated with the localised sources are short .wav files mostly downloaded from the WWW. Some effort is spent into selecting sounds that do not mask each other when played synchronously, and that in our opinion, are easily distinguishable. In the current prototype, as mentioned earlier, the listener is fixed at the centre of the client window. The effect of the sound rendering is mainly in terms of left and right panning with volume changes according to distance. Auditory browsing experiments in similar environments have been carried out earlier [13].

3.2 Event distribution

The group drawing application is implemented in Java. Whenever an operation is performed at the interface of a given client an event is generated. The event distribution strategy between the different client applications is based on that of the Java Collaborator Toolkit [1]. The floor control mechanism of the toolkit has been changed to add some element of synchronicity in the interaction. In particular, the positions of the listeners are always broadcast to peer clients irrespective of who holds the floor. Only events describing drawing operations and produced by the client holding the floor are allowed to trigger corresponding behaviours in remote peer clients. Since the behaviour that an event triggers also depends, in our case, on the origin of the event, we modify the Java Collaborator packet format by including an additional field in the data packet (see Table 3). This field describes the client where that particular event is produced. The collaborative event mediator on receiving an event will examine the packet and

trigger appropriate behaviours according to an inference engine. These modifications are necessary because the Java Collaborator Toolkit was mainly designed for a WhatYouSeeIsWhatISee environment. Our case involves a relaxed WYSIWIS (more accurately WhatYouSenseIsWhatISense) environment as the behaviour of the clients are client specific thereby allowing participants to have relatively independent views of the common workspace.

Each client has a buffer used to store the latest listener position of every other client so that every client is 'aware' of the listener position of their peers at all times. The application can be re-implemented to support full duplex interaction but since our drawing environment is an object oriented one, managing the synchronous construction of objects is more involved and will be tackled at a later stage. For now, we are mainly interested in creating a minimal environment in order to learn about the effects of adding localised sounds to a collaborative drawing environment. The collaborative event mediator (Figure 4) contains inference rules that are triggered to give rise appropriate behaviours in every client. For example if an event from Marash describing his window position reaches a peer client, the client's collaborative event mediator will trigger the appropriate sound that represents participant Marash and will update the position of his representative source in the current environment (See Figure 4). How events resulting from the drawing operations produce appropriate interface behaviours in peer clients is described in detail elsewhere[1].

length
App id
type
Window id
Component id
Client id
Event id
x
y
key
Click count
arg

Table 3 Modified collaborator packet format

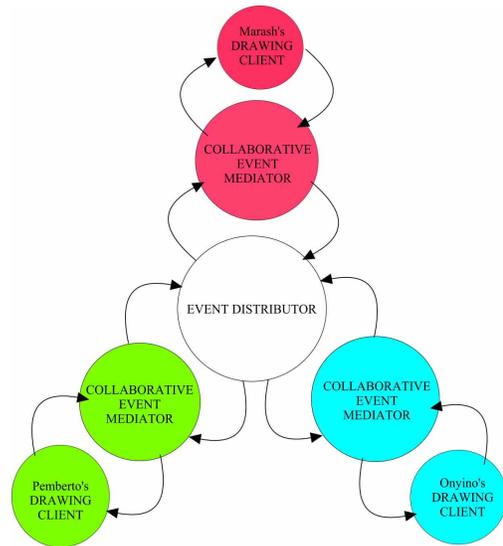


Figure 4: The flow of events in the group drawing environment

4 Experimenting with the group drawing environment

4.1 Hypothesis

We hypothesise that adding localised representative sounds to the virtual environment allows real time location and type awareness of activities which are out of visual focus. Our sound mapping strategy allows peers to be aware of the location, initiators and progress status of activities occurring on the common canvas. Associating a sound source with each participant allows the later to be aware of each other's location and to determine whether he or she is being observed and by whom. Such an awareness is deemed to free the participants from having to literally keep a close watch on peer activities in order to keep abreast with the current state of affairs in the group drawing environment. We believe this is likely to increase fruitful interactions between participants.

4.2 Testing procedure

4.2.1 Experiment Design

Figure 5 shows a typical arrangement involving two participants. Each has access to the soundscape through a pair of headphones to prevent sound interference as both participants are in the same room. They are allowed to communicate with each other using a text chat application and draw on the common canvas using the mouse. Note that the use of the chat tool is not mediated by any floor control mechanism. Even though our implementation is able to support many clients at one time, the groups taking part in the experiments vary from two to three mainly because of the availability of participants at the same time. It is worth mentioning here that increasing the number of participants also tends to increase the delay effects which hinder interactions between participants. Also, the floor

control policy we have adopted tends to encourage some participants to remain idle for long periods and this problem becomes more apparent as the number of participants increase. In order to increase the likelihood that the participants will be working out the view of their peer fairly frequently, the window of each client is restricted to a quarter of the area of the canvas which as such occupies the full screen.

4.2.2 Familiarisation with the group drawing environment and task allocation

Each participant is explained the sound mapping strategy and asked to listen to different sounds and to remember as much as possible their associated meanings. Every participant is allowed to get familiar with the interface and the sound mapping for half an hour before the start of the experiments. They are then asked to collaboratively produce a simple representation of an object such as a car, house or a scene of their choice. Twelve participants take part in the experiments. The order of the experiments with sound and without sound (the control) is alternated so as to minimise any learning effects. After each task the participants are administered the NASA Task Load Index [12] questionnaires which are used to determine the subjective workload of a set task. They are also asked to rate the intentionality of the group drawing environment under the auditory and visual conditions. The intentionality of a common environment defines its ability to convey information about the interests, wants, needs and the details of ongoing activities of participants mainly through observation of their gestures. The intentionality variable rating ranges from 0 to 20 with 0 equivalent to poor and 20 equivalent to good.

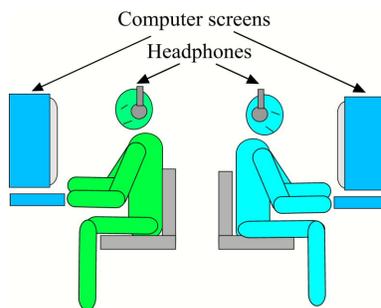


Figure 5: A typical experimental set-up

5 Evaluation of our group drawing environment

5.1 Participant feedback

5.1.1 Comparing subjective workload under the auditory and visual conditions

Figure 6 shows the average scores of participants for the workload categories in the NASA TLX. Note that we are more interested in the ratings for individual workload categories than in the overall workload value that can be determined. Also, it is worth mentioning that an increase in each rating increases the overall workload and that the performance category varies from high to low in the NASA TLX. Contrary to our expectation, the results do not show a dramatic decrease in subjective workload in the environment under the auditory condition even though a marginal decrease is noted. A majority of participants report that the localised sounds inform them that 'something' is happening roughly to their left or right but are unable to interpret the soundscape quickly and easily enough to allow it to be exploited. Thus, the sounds are considered more distractive and annoying than useful. None of the participants make use of the sound associated with the listener of each client to infer the location of their peers when the listeners of the later are hidden. Since the soundscape is not interpreted easily, it does not have a dramatic positive effect on the subjective workload experienced by participants. The soundscape may even offset any benefits that may have been obtained by imposing a cognitive load on participants trying to interpret it in real time.

5.1.2 Comparing subjective intentionality under the auditory and visual conditions

Figure 7 shows the average scores of participants for the intentionality category. In this case, there is a marked difference in the subjective intentionality rating for the auditory condition as compared to the visual one. Many participants report that the feeling of objects having a life of their own is diminished in the auditory environment. They seem to 'know' when an object is going to be drawn, or when a tool is going to be picked for e.g. by hearing quick movements of sources towards the left. Some participants feel more comfortable using the text chat tool in the auditory condition. They explain that hearing activity sounds in the background provides them with the information that somebody has the floor and they may as well communicate using the text chat concurrently. Under the visual condition, it is not as easy, as they have to often check whether they can get floor by clicking on the 'get floor button' of the floor manager. Thus, under the auditory condition, they are also more likely to be successful in getting the floor quickly as listening for activities provide the participants with information that allows them to decide about the right opportunities to make a floor request.

5.1.3 Complaints and suggestions by participants

Some participants complain about the choice of sounds which are considered not to be distinguishable enough. This trend was not general though. Some participants argue for the need to be able to refer to the sounds verbally as it was difficult for them to describe certain sounds to others especially in situations where they need to ask others if they have heard a particular sound and how to interpret it. None of the participants are able to use Intel's RSX to determine directly elevation information about the sources. They are only able to do so by paying attention to changes in volume when there is a relative motion between the source to be located and listener. A few participants suggest that there are too many redundant sounds. For example, they find that the sound associated with a tool when it is being moved about before an operation is performed to be annoying and prefer to listen to progress sounds alone, or just a simple common sound for mouse motions. Again participants are strongly divided on this issue as some argue that the sound associated with the tool and localised with the mouse pointer does help them to have an idea of what is to come.

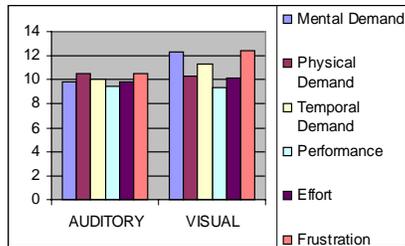


Figure 6: Comparing workload categories between auditory and visual conditions (Only one quarter of the canvas in view)

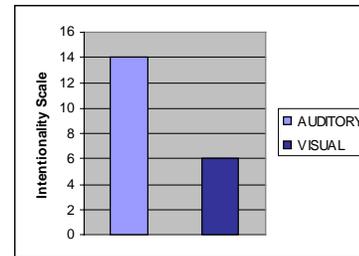


Figure 7: Effect of localised auditory cues on intentionality (Only one quarter of the canvas in view)

5.2 Our observations and discussion

While participants pass simple tests for the interpretation of complex sounds successfully during the sound mapping learning phase, they do not fare well under real work conditions, where interpretation of the soundscape is surprisingly poor except where voice messages are involved. Also abilities to interpret complex soundscapes seem to vary widely. Some participants are observed discouraging their peers to perform a given activity such as picking the same color, moving a selected object or deleting a selected object which illustrates the fact that intentionality information does get transferred. However these participants are in a minority.

In our opinion, learning the abstract sound mappings is not the only possible reason why we fail to get an overall marked decrease in workload under the auditory condition results. We suggest that another reason is that we are actually investigating the ability of participants to infer peer activities which are mostly at one extreme of the 'out of visual focus' condition, that is when the activities are completely out of view or hidden. The interpretation of sound sources under such conditions is more difficult just as is the case in similar real world situations. The ability to interpret complex soundscapes in real time is cognitively expensive and especially so when there is a lack of visual cues. When the soundscape loses its meaning it adds annoyance to the work environment and is very likely to increase the workload on the participants. We thus modify our experiment and look at the effect of sounds when all participants are exposed to the same view. Under the later condition, they are still able to control their own private focus of vision. However, the activities of their peers are not out of view even if they are not in focus. We postulate that peripheral visual cues can help participants to make sense of complex soundscapes in two ways. Firstly, the ability to match visible activities with sounds helps to learn the sound mapping. Secondly, there seems to be a dialectical relationship between what is seen and what is heard. How sounds are interpreted is influenced by the visual behaviour of the objects they are associated with. This aspect may be exploited to find ways of making complex abstract soundscapes more intelligible.

6. Do visual cues affect the ability to make sense of soundscapes?

6.1 Modifying experiment set-up so that each participant sees the whole canvas

The experiments are re-run but under slightly different conditions and with a different set of participants so as to minimise any learning effects. The drawing area is now that of the full canvas which occupies the full screen. In this series of experiments, all the participants are exposed to same view. The sounds associated with each participant's listener are turned off because they are not relevant in this particular case. Our aim is to evaluate our

claim that visual cues help to make complex abstract auditory soundscapes more intelligible.

6.2 Participant feedback

6.2.1 Comparing subjective workload under the auditory and visual conditions

Figure 8 shows the average scores of participants for the workload categories in the NASA TLX under the four different conditions. Only a few of the workload categories under the auditory conditions seem to be significantly less than their corresponding visual conditions. We are in fact unable to derive any concluding remarks based on the results obtained from the NASA TLX such as whether it is worth investing the effort to add localised sounds to decrease the workload on participants.

6.2.2 Comparing subjective intentionality under the auditory and visual conditions

Figure 9 shows the average scores of participants for the intentionality category under four different conditions. The change of partial to full view condition of the canvas seems to cause an increase in intentionality of the environment for both the visual and auditory conditions. However, there is a more prominent increase in the case of the auditory condition. These results seem to support our view that the presence of visual cues associated with the sonic cues does create a more comfortable environment as far as the interpretation of abstract soundscapes are concerned.

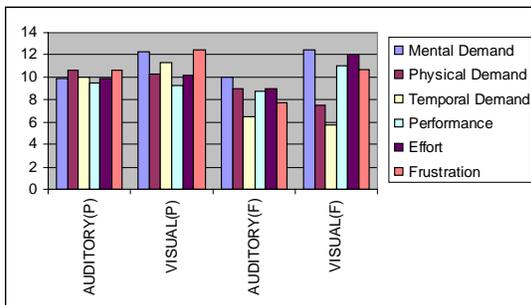


Figure 8: Comparing workload categories between auditory and visual conditions [Canvas in partial(P) and full (F) view]

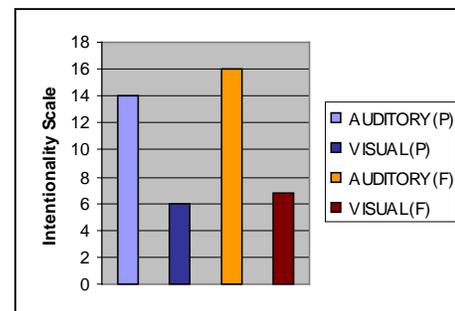


Figure 9: Effect of localised auditory cues on intentionality [Canvas in partial (P) and full view(F)]

7. Lessons learnt and implications for the design of collaborative soundscapes

The following summarises the feedback given by users who are involved in the experiments.

- The majority of participants report awareness of location of activities but less so for precisely what activity.
- Some participants report bad choice of auditory icons.
- Some participants do not agree with the localisation of sources strategy. They find a continuous source describing the current location of a tool annoying.
- There is a high preference for voice messages.
- Some participants argue for the need of easily describable and common rather than customised sounds in a collaborative setting.
- Visual cues improve the 'intelligibility' or interpretation of soundscapes and help participants to learn the sound mapping.

It is clear from our experiments that adding sounds to a collaborative environment is not a straightforward exercise. A large number of design and test iterations are probably needed before reaching the right sound mapping strategy in order to produce an 'intelligible' soundscape. Since users have different preferences for the appropriate auditory messages, it can be argued that there is a need to provide users with customisation facilities. However, we also point out that some participants prefer standard sounds that they can describe easily to each other. Finding the right balance between these competing requirements is challenging.

It is important for the environment involving complex soundscapes to provide some facility that will allow users to learn the sound mappings. This facility should at best be integrated in the environment rather than providing it separately. The effect of vision on the interpretation of complex soundscapes cannot also be ignored. Thus, when associating hidden objects with sounds, it may make sense to provide some visual feedback synchronised with the

sounds on the containing objects as well. This may help participants to find their sound source targets more easily.

8. Work in progress

Better audio rendering engines are certainly likely to improve soundscape comprehension. Unfortunately, commercial 3D audio rendering engines producing convincing externalised 3D sounds are currently hard to find. Such rendering engines often exploit synthesised Head Related Transfer Functions (HRTFs) which often are closely guarded secrets. However, the best externalisation of sound sources is more likely to be achieved, only if subject specific HRTFs are used [6]. How to achieve the latter in practice on a large scale is still an unresolved problem. Also, for the large majority of participants, the most convincing experience obtained from the rendering engine we are using is that for left and right panning while elevation localisation is poor or in-existent. In our experience with the version of the 3D sound API we are dealing with, the use of headphones instead of speakers also does not seem to make any difference.

We also make a few trial runs of our drawing environment with 4 speakers, two for left and right panning and the remaining for top down panning to create a naïve spatial sound environment (Figure 10). The top and bottom speakers are used to deal with the elevation localisation while the left and right speakers are used to deal with lateral localisation of sources. This is achieved by reserving a surrogate computer (b) to run a mirror of the drawing client at (a), but with the graphics environment reflected along the diagonal from the top left, to the bottom right of the screen in (a). The rendering engine on the surrogate computer is dedicated to the top and bottom speakers. To illustrate with an example, if a participant moves a source along a vertical line on her shared canvas in (a), this motion will appear to be along a horizontal line on the surrogate machine (b) which will trigger the appropriate panning responses for the top and bottom speakers. The net effect is that of a single rendering engine with a 4-speaker output. While this is an impractical design in a normal work environment, it is fairly adequate for our investigative purposes. In short, we are making opportunistic uses of aspects of the 3D sound API at our disposal that work convincingly to investigate our research problem. One problem with the 4-speaker approach though is that participants need to be geographically separate in order to avoid sound pollution which will make localisation difficult. However, sources above the listener and those below it can be distinguished more distinctly with this method. Our experience with this environment suggests that to create intelligible and useful soundscapes, it is not enough to be able to render convincingly 3D sounds. In particular, it is unlikely that an effective 3D audio renderer will compensate for an unsatisfactory approach towards sound mapping.

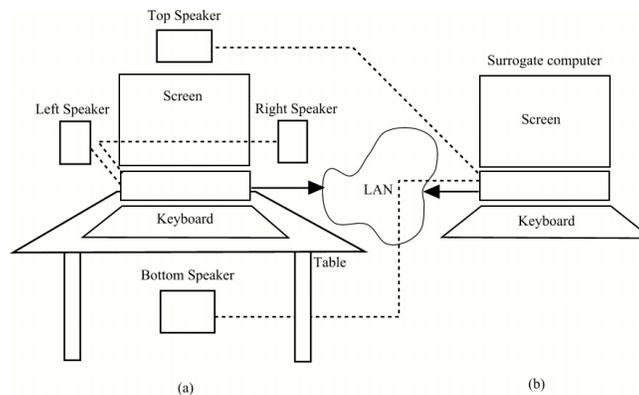


Figure 10: Naïve localisation strategy using 4 speakers

We are currently experimenting with new ways to browse a soundscape. In particular, the participant will have a direct coupling between his point of gaze and his listener which so far is fixed at the centre of the canvas. Elsewhere, the advantages of coupling the listener with the point of gaze which is tracked have been explained [14]. This is deemed to provide the user with more controlled access to localised auditory information. We also plan to study situations involving multiple shared drawing tools using auditory cues parametrised according to stacking order and visibility. Another idea worth investigating is whether it is helpful for a localised auditory chat tool to be integrated with the drawing environment. For example, a localised-streamed sound source of a participant's speech can be collocated with the mouse pointer. Thus, when she is conversing with her peers during the drawing activities her voice will appear to originate from her mouse pointer. Conversations such as the following 'if I move an object from here to there ...' will have a more forceful meaning in case the 'here' and 'there' sounds are heard at different locations.

9 Conclusion

In this paper, we discuss the importance of adding localised sounds to describe activities in a group drawing environment. Our evaluation of this particular environment indicates that the addition of localised sounds did not

have a consistently positive effect on the workload experienced by users. This result seems counter intuitive especially in the light that most participants report an increase in the awareness of peer intentions and activities. It is probable that the difficulty in interpreting the sonic environment in real time may have actually tempered the benefits that we expected. The fact that participants did not use the sounds associated with peer listeners to infer who is observing them has also deep implications. This suggests that simply replacing a missing component in a given medium, such as gaze awareness which is in the visual medium, by an alternative source of information in a different medium, such as a localised sound to indicate a participant's approximate view point, requires a steep learning curve in order to be effective. However, we believe our work is still in an early stage and we plan to continue investigating our design approach by coupling the point of gaze with the listener and explore the pros and cons of such a design strategy. It is worth mentioning at this point that the benefits of adding localised sounds to a group drawing environment are likely to increase according to the degree of synchronous participant-participant interactions provided by the environment.

10 Acknowledgements

We thank all the participants that took part in our experiments for their inspirational comments and enthusiasm.

11 References

1. Abdel-Wahab H, Kvande B, Nanjangud S, Kim O and Favreau J. Using Java for Multimedia Collaborative Applications. In proceedings of the 3rd International Workshop on Protocols for Multimedia Systems (PROMS'96), October 1996, Madrid.
2. Beaudouin-Lafont M and Karsenty A. Transparency and awareness in a real-time groupware system. In UIST'92, Monterey, CA, Nov 15-18, 1992, 171-80.
3. Blattner M, Sumikawa D and Greenberg R. Earcons and icons: Their structure and common design principles. *Human Computer Interaction* 1989, 4(1): 11- 44.
4. Brewster S A. Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD Thesis, University of York, UK, 1994
5. Brewster S and Clarke C. The design and evaluation of a sonically-enhanced tool palette. In Proceedings of the Third International Conference on Auditory Display, Palo Alto, California, 1997.
6. Burgess D. Techniques for Low Cost Spatial Audio. In UIST'92, Nov 15-18, 1992, 53-59.
7. Cohen, J., Monitoring Background Activities. In Kramer, G. (ed.) *Auditory Display*. Addison-Wesley, 1994, 439-531.
8. Gaver W. Sound Support for Collaboration. In ECSCW'91, Amsterdam, 1991, 293-308.
9. Gutwin C and Greenberg S. Workspace awareness for groupware. CHI'96 Conference Companion, Vancouver, 1996, 208-209.
10. Hayne S, Pendergast M and Greenberg S. Gesturing through cursors: implementing multiple pointers in group support systems. In Greenberg S. et al. (ed.) *Groupware for real-time drawing, A designer's guide*. McGraw-Hill Book Company, 1995, 63 – 80.
11. Ishii, H. and Kobayashi, M. ClearBoard: a seamless medium for shared drawing and conversation with eye contact, in Proceedings of the ACM Conference on Human Factors in Computing Systems, ACM Press, 1992, 525-32.
12. NASA TLX information available at <http://www.dtic.dla.mil/iac/cseriac/tlx.html>.
13. Pitt I and Edwards A. Navigating the interface by sound. In G.D. Abowd and others (eds.) *A collection of papers on HCI*, University of York Department of Computer Science, Heslington, York, YO1 5DD, England.
14. Ramloll, R. and Mariani, J, Demonstrating gaze driven auditory browsing, submitted to ICAD' 98, Edinburgh 1998. Available at <http://www.comp.lancs.ac.uk/computing/users/ramloll/Demogaze.zip>
15. Ramloll, R. and Mariani, J. "Moksha: Exploring Ubiquity in Event Filtration-Control at the Multi-user Desktop", accepted for the WACC '99 (Work Activities Coordination and Collaboration Conference) February 22-25, 1999 Cathedral Hill Hotel, San Francisco, California, USA.
16. RSX Web Reference, Intel's 3D Realistic Sound Experience white, paper available at web link <http://www.intel.com/ial/rsx/wpaper/wpaper.htm>.
17. Tang J. C Findings from observational studies of collaborative work. *International Journal of Man-machine studies*, 1991 Vol. 34, No 2, 143-160.