Chalk Sounds: The Effects of Dynamic Synthesized Audio on Workspace Awareness in Distributed Groupware

Carl Gutwin, Oliver Schneider, Robert Xiao
Department of Computer Science
University of Saskatchewan
110 Science Place, Saskatoon, S7N 5C9, Canada
carl.gutwin, oliver.schneider, robert.xiao @usask.ca

Stephen Brewster
School of Computing Science
University of Glasgow
Glasgow, G12 8QQ, UK
stephen.brewster@glasgow.ac.uk

ABSTRACT
Awareness of other people's activity is an important part of shared-workspace collaboration, and is typically supported using visual awareness displays such as radar views. These visual presentations are limited in that the user must be able to see and attend to the view in order to gather awareness information. Using audio to convey awareness information does not suffer from these limitations, and previous research has shown that audio can provide valuable awareness in distributed settings. In this paper we evaluate the effectiveness of synthesized dynamic audio information, both on its own and as an adjunct to a visual radar view. We developed a granular-synthesis engine that produces realistic chalk sounds for off-screen activity in a groupware workspace, and tested the audio awareness in two ways. First, we measured people's ability to identify off-screen activities using only sound, and found that people are almost as accurate with synthesized sounds as with real sounds. Second, we tested dynamic audio awareness in a realistic groupware scenario, and found that adding audio to a radar view significantly improved awareness of off-screen activities in situations where it was difficult to see or attend to the visual display. Our work provides new empirical evidence about the value of dynamic synthesized audio in distributed groupware.

Author Keywords
Awareness, distributed groupware, procedural audio.

ACM Classification Keywords
H.5.3 [Information Interfaces and Presentation]: *CSCW*.

General Terms
Design, Human Factors, Experimentation.

INTRODUCTION
Workspace awareness – the up-to-the-moment knowledge of who is in a shared space, where they are, and what they are doing – is an important factor in smooth and natural collaborative work [17]. Awareness is more difficult to maintain in distributed settings than during face-to-face work; to overcome this limitation, several kinds of awareness displays have been designed that provide information about collaborators. For example, researchers have investigated participant lists, multi-user scrollbars, duplicate views of what others can see, and radar views (miniatures of the entire workspace overlaid with representations of other people in the session) [29].

Most awareness displays are visual, and therefore suffer from three limitations. First, displays must be visible in order to be useful, but many scenarios (e.g., both small-screen and large-screen settings) make it difficult for users to see the awareness information. Second, visual information about activities may be difficult to see if the action is small or if the workspace is cluttered. Third, the observer must attend to the awareness display in order to see it, but as tasks become more demanding, it becomes more difficult to notice changes in the display.

Visual presentation, however, is not the only option for awareness information. Audio information is a natural part of shared activity in the real world, and previous research has shown that non-speech audio can successfully be used to help maintain awareness in groupware systems (e.g., [3,5,7,8,11,13,20]). Audio has several advantages that can overcome the drawbacks of visual awareness displays: audio takes no space and does not need a location on the screen; audio is not affected by workspace clutter; audio can be processed without requiring visual attention; and audio can be used in parallel with visual information. In addition, audio can be effective for several types of awareness information, such as *whether* activity is occurring right now, *when* actions start and stop, *where* actions are happening, the *type* of activity, and the *qualities* of the action (e.g., lines drawn slowly or quickly).

In this paper we evaluate the use of audio awareness as an enhancement for visual awareness displays. We explored two main questions: can dynamic audio generated with granular synthesis adequately convey information about the type and character of off-screen activities; and can the addition of audio information improve on a visual awareness display, particularly in situations where visual presentations are difficult to see or attend. We evaluated the quality of the synthetic sound by asking people to

CSCW 2011, March 19–23, 2011, Hangzhou, China.

determine the shape, speed, and pressure of off-screen drawings, using the synthesized sounds and matched recordings of real chalk sounds. We found that people were almost as accurate with synthesized audio as with real sounds. We then compared visual and audio awareness in a real groupware system where participants had to carry out an individual task and also stay aware of off-screen activity. We tested situations with different levels of workspace clutter, different radar-view sizes, and different degrees of visual attention required by the individual task. We found that the addition of audio awareness information significantly improved people's ability to stay aware of off-screen events, compared to the visual radar view alone. Participant preferences mirrored the performance results: people universally preferred the condition which combined the radar view with the synthesized audio.

This work makes two main contributions. We show that previous results in audio awareness can be extended to produce dynamic synthesized sounds that accurately reflect the type and quality of digital actions. Second, we provide empirical evidence that dynamic synthesized audio can significantly improve awareness of off-screen actions, particularly when a visual display becomes difficult to see or attend to. Our results suggest that audio awareness should be used more often in shared-workspace groupware.

RELATED WORK

Audio has a rich history in human-computer interaction, and several areas of previous work provide the foundation for our studies. Here we briefly review research in awareness, audio feedback, audio for groupware systems, and sound synthesis.

Group Awareness

Group awareness in collaborative work – "an understanding of the activities of others, which provides a context for your own activity" [10] – is now a common topic in CSCW (e.g., [1,5,18,23]). There are many ways in which people maintain awareness of their collaborators – general understanding of who is around in an environment, noticing body posture and gestures to manage turn-taking in conversation, or tracking another person's actions for tightly-coordinated actions. People's actions in a shared workspace are of particular interest – that is, the understanding of who is in the space, where they are working, and what they are doing [17].

Face-to-face settings provide rich awareness information, but much of this is lost in distributed groupware. The awareness problem is particularly acute when people can work in different parts of the shared space (i.e., relaxed-WYSIWIS groupware [30]), effectively preventing collaborators from seeing what others are doing. To address this problem, researchers have proposed awareness displays that restore some of the lost information: participant lists [1], multi-user scrollbars [19], duplicates of others' views [16], or miniatures of the entire space with representations of people overlaid on top – also called *radar views* [29]. Awareness displays can provide useful information to collaborators in distributed workspaces (e.g., [18,19]). However, most awareness displays present information visually, which has the drawbacks described above.

Audio Feedback

One way to provide awareness without requiring visual attention is to use sound. Audio feedback has been extensively studied in HCI (e.g., [3,6,12,14,28,31,32]), and researchers have looked both at symbolic sound (e.g., a sound played to indicate a particular event) [6], and dynamic sound, which represents a continuous action such as dragging an icon across a desktop [12]. For example, Gaver's Sonic Finder provided dynamic parameterized sounds for actions in the Macintosh Finder: different file sizes produced differently-pitched sounds when selected, and actions like file transfers produced a continuous filling-up sound that indicated the progress of the action [12]. Gaver investigated synthesis techniques for a variety of sounds that could be used in these environments, including scrapes, impacts, and breaking sounds [15].

Other researchers have used continuous sound feedback in a variety of settings, such as haptic environments where sound is used to increase the sense of immersion [2,21]. These more recent systems often use *granular synthesis* [4,27] to produce sound, a technique that allows for a wider range of sounds than earlier approaches (see below).

Audio Awareness in Groupware

Groupware researchers have explored a variety of non-visual awareness modalities for distributed settings (e.g., haptic feedback [22,23]), but a majority of this work has looked at audio as a way to maintain awareness. First, transmitting real-world non-speech audio has been investigated as a way to provide awareness of activity at another location (e.g., the Thunderwire audio awareness system [20]). These systems, however, do not provide much awareness of computer-based activities, since these actions often do not produce characteristic real-world sounds.

Second, several systems have included symbolic audio as a way to indicate events occurring at a remote location (e.g., EAR [15], which represented sounds from an event server for a broad community; GroupDesign [5], which duplicated local sound feedback for remote participants; or ShareMon [8], which played sounds to indicate background activity such as file sharing). Observations and studies of these systems showed that sound helped distributed groups maintain a sense of what collaborators were doing, and helped people to coordinate shared activity (such as meetings or social events [11]).

Third, some systems have used sound as a more direct representation of specific activities. ARKola [13] used continuous sound feedback to help distributed pairs operate a simulated bottling plant. The sounds reflected the bottling machine's state (e.g., the conveyor speed), and participants reported that they made use of the audio feedback to understand what their partner was doing, and to adjust their own behavior [13]. Cohen's 'Out To Lunch' system [9] created a parameterized soundscape to indicate a group's keystrokes, mouse clicks, and mouse movement; however, the sounds were representative of the rates of activity, rather than literal reflections of specific actions. Audio Draw, based on the ENO sound server [3], indicated actions such as selection, dragging, and resizing with sound; some of these were dynamic and parameterized to the activity (e.g., the scraping sound used with dragging was tied to the speed of the mouse). Another drawing editor [25] played different sounds for different participants, and spatialized the sound in the 2D environment to help with location awareness. Finally, several commercial multi-player games provide sound cues that are parameterized based on the actions of other people (e.g., vehicle sounds that are based on speed); in addition, sounds in 3D game worlds are also often spatialized.

These systems have made it clear that sound can be a valuable resource in helping people maintain several different types of group awareness. However, early groupware systems were limited in the range of synthesized sounds that they could produce (and the ways that these could be parameterized to dynamic activity), and no studies assess whether audio information can improve on common visual awareness displays such as the radar view. This is the question that we address in our studies below; before turning to these experiments, however, we review work in sound synthesis, and in particular, the technique of granular synthesis that we used in our work.

Dynamic Parametric Sound Synthesis

The goal for dynamic audio awareness is sound that reflects the continuous characteristics of an action – e.g., a sound that changes based on the speed and pressure of an input stroke with a pencil on paper. There are many possible ways of synthesizing sound in this fashion; three main approaches are additive-subtractive synthesis, mathematical modeling, and granular synthesis.

Additive and subtractive synthesis. These techniques take basic waveforms (e.g., sine waves, or various forms of noise waves) and combine and filter them to produce sound that approximates the desired activity sound [14,15,26]. For example, different kinds of scraping sounds can be built from filter banks that model the different resonant modes of particular materials [14]; similarly, basic chalk sounds can be simulated using white noise and a flanger filter. The main drawback to this technique is that it requires expertise in creating realistic sounds from the basic building blocks.

Mathematical modeling. Several researchers have investigated ways of generating sound by modeling the physical materials of the source objects, and their interactions with the environment [26]. Although these models can produce highly realistic sounds, they can also be extremely complex, difficult to build, and computationally expensive.

Granular synthesis. In this method, very small pieces of a source sound (i.e., 'grains' of 50ms or less) are used to build up a real-time dynamic sound (e.g., [2,4,21,27]). The source sounds are recorded from actual actions in the real world (e.g., pencil strokes on paper); to synthesize different sounds from the source, many grains are overlaid on one another, and are played at different speeds, volumes, and with different envelopes. This allows a high degree of control over the resultant sound, although the output is always strongly related to the source sound used to create the grains. Granular synthesis can reflect different elements of the sound such as different kinds of strokes, different pressures, and different stroke speeds, if these qualities are part of the set of input samples [2,21].

Granular synthesis has the advantage over other techniques that no artificial models of the action sound are required; the designer needs only to record a representative set of action sounds with the materials and tools of interest. These real-world data can then be used to synthesize parameterized audio, a process that can be done inexpensively, and can be repeated for the several types of sound that may occur in a groupware system.

GRANULAR SYNTHESIS FOR CHALK SOUND

We developed a synthesis system for creating chalk sounds that are parameterized by the speed and pressure of an input stylus on a tablet. The system is made up of several recorded chalk sound files, and a granular synthesis engine.

Sound files of chalk strokes were recorded at a real chalkboard, using a Shure SM58 microphone; sounds were 44.1 KHz 16-bit samples. Seven files were recorded: one for each combination of light/heavy pressure and fast/slow speed; and three recordings of the chalk hitting the chalkboard (for heavy, medium, and light pressure). These recordings were used as input for the granular synthesis engine. The source recordings were divided into 'grains' of 400 samples. To generate a chalk sound, the engine played 32 simultaneous grains.

A pressure-sensitive pen and tablet provide input (current pen speed and pressure, and location in the workspace) to the chalk engine. Generated sound is parameterized through interpolation: source sounds are recorded at different speeds and pressures, and the input parameters indicate how to select samples from the source. Individual grains are selected by randomly choosing a sample index within the recording. After a grain is selected, it is attenuated and stereo-panned according to the input parameters, and finally enveloped. Enveloping applies an increasing-then-decreasing amplitude mask to the grain in order to avoid clicks or other artifacts during playback. In addition to the dynamically-chosen grains, we also play a pre-recorded sound at the beginning of each stroke for the initial sound of the chalk touching the board. This sound is parameterized based on the user's initial stroke pressure.

The system was built using Python, with the enveloping implemented as a C routine. The engine will run on any sound hardware capable of supporting at least 32 simultaneous voices. There is very little latency in the generation system; in our test applications, there is no discernable lag between input and generation. Generation requires ≈6% of the CPU on an Intel Core 2 Duo processor.

In the next sections we report on two tests of our awareness audio. The first study focused on the quality of the sound produced by our synthesis engine, and the second focused on the effectiveness of audio awareness in groupware.

S1: HOW MUCH INFORMATION CAN SOUND CONVEY?

The goal of this study was to determine how well synthetic audio can convey information about activity to listeners, using real sounds as our performance baseline.

Methods

The study asked participants to interpret real and synthesized chalk sounds generated from drawings of different shapes at different pressures and speeds.

Participants, Procedure, and Task

Fourteen people (9 men and 5 women) were recruited from a local university. Participants ranged in age from 19 to 39 (mean 23.9), and all reported having no hearing impairments. Eight participants were regular online gamers.

The experimenter first played each of the sound types along with a graphical representation (see Figure 1), to familiarize participants with the shapes and categories. Participants also heard examples of each pressure level and drawing speed. The visual representations were then hidden, and participants were played a series of sounds; for each, they answered which shape, pressure, and speed they thought best described the sound. Participants were allowed to listen to the sound again (once only) if needed.

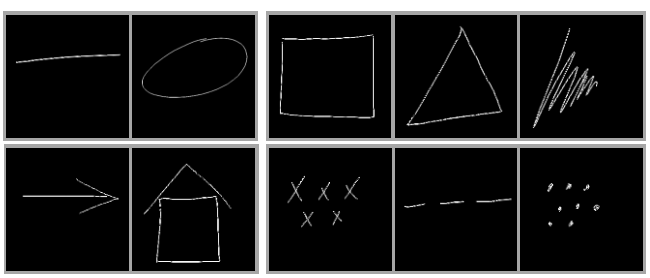

Figure 1. Shapes used in study 1: solid line, oval, square, triangle, scribble, arrow, house, Xs, dashed line, stipple.

Apparatus and Sounds

The study used custom software to present the sounds and record responses. Sounds were played on Logitech speakers placed on either side of the computer monitor, and the study was conducted in a quiet room.

Synthesized sounds were generated dynamically with the system described above, but used pre-recorded stroke input (therefore, all participants heard the same sounds). The real-world sounds were recorded at a real chalkboard using a MacBook Pro and an external microphone; each shape was recorded at each of the different speed and pressure levels.

The categories and specific shapes used in the study were:

- *One-line* shapes (solid line, oval): shapes that involve one line with no obvious corners.
- *Multi-line* shapes (square, triangle, scribble): shapes that use a single chalk stroke, but with multiple lines.
- *Two-stroke* shapes (arrow, house): shapes that involve two separate strokes.
- *Multi-stroke* shapes (Xs, dashed line, stipple): shapes that involve multiple separate strokes.

Study Design

We used a within-participants design with five factors:

- *Source* of the sound: real or synthesized
- *Category*: one-line, multi-line, two-stroke, multi-stroke
- *Shape*: one of ten different shapes (see Figure 1)
- *Drawing speed*: slow, medium, or fast
- *Pressure*: heavy, medium, or light.

There was one trial for each combination of these factors; trials were grouped by source (i.e., all sounds from one source were heard together), but trials for the other factors were randomly drawn from the pool. There were thus 120 trials for each participant.

Results

We organize our results in terms of participants' performance in interpreting the four qualities of the sound: shape category, specific shape, and pressure and speed.

Shape categories

Overall, participants were able to interpret shape categories very accurately, answering correctly 88% of the time (for synthesized, 85%; for real sound, 92%, Fig. 2). RM-ANOVA showed a significant effect of sound source ($F_{1,13}$=13.5, $p<0.005$), with accuracy higher for real-world sounds. There was an interaction between source and category ($F_{3,38}$=2.87, $p<0.05$): as can be seen in Figure 2, the difference was larger for two-stroke sounds.

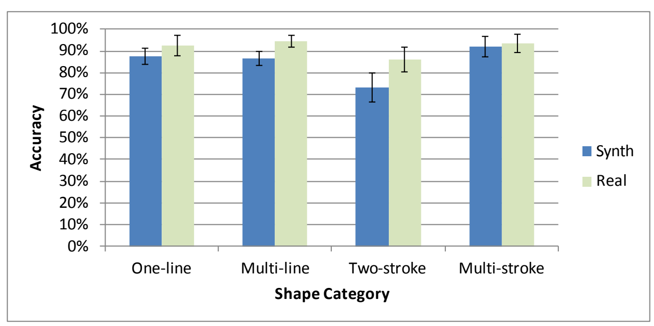

Figure 2. Accuracy (± std. err.) in determining category.

Specific shapes

Participants identified the shape 70% of the time (62% for synthesized, 77% for real, Fig. 3). RM-ANOVA showed a significant effect of sound source on accuracy ($F_{1,13}$=27.2, $p<0.001$), and significant differences between the shapes themselves ($F_{9,117}$=8.82, $p<0.001$). There was an interaction between source and shape ($F_{9,117}$=11.8, $p<0.001$).

The larger differences for oval and stipple shapes are possibly due to limitations in our playback of the initial contact sounds, and the fact that we did not model the varying sound of the chalk's tip on the chalkboard.

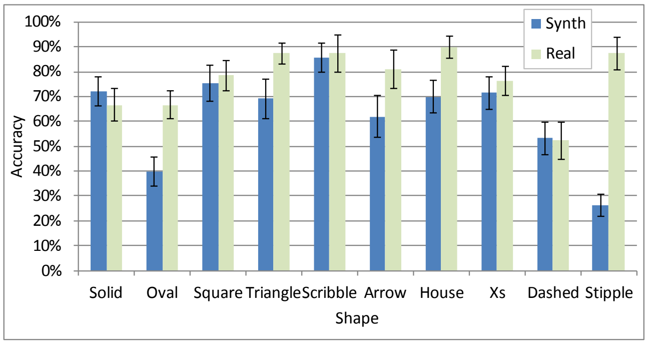

Figure 3. Accuracy ±s.e. in determining shape, by source.

Pressure and speed

Overall, people were less accurate in determining pressure and speed (Figure 4): average accuracy was 45% for pressures, and 64% for speeds. RM-ANOVA showed that performance was significantly better with synthesized sounds than with real sounds (pressure, $F_{1,13}$=10.8, p<0.01; speed, $F_{1,13}$=10.2, p<0.01). However, there are two likely reasons for these results. It was more difficult to precisely control the pressure and speed for our real-world recordings, leading to more mismatches between the participant's answer and the 'true' value. Second, our categories of pressure and speed are relative (unlike shape), and so there is more variability in people's responses.

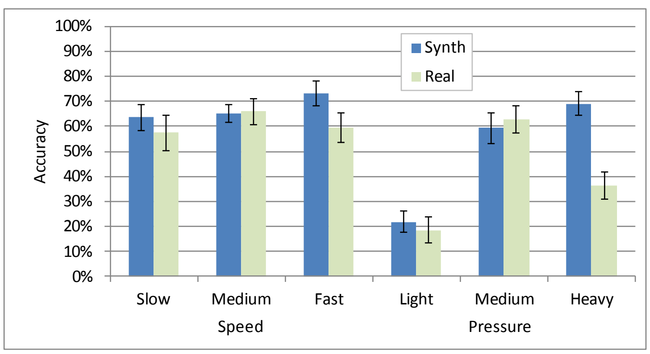

Figure 4. Accuracy ±s.e. in determining speed (left bars) and pressure (right bars).

Summary of Study One

The study showed that although real-world sounds can be interpreted more accurately than synthesized sounds, these differences are relatively small. For groupware awareness, the study shows that there is a large amount of information that can be conveyed in synthesized chalk sounds. Our next study considers whether this information can make a significant difference when added to a groupware system.

STUDY TWO: EFFECTIVENESS OF AUDIO AWARENESS

The main goal of the second study was to determine whether adding dynamic audio information to a visual awareness display would improve awareness of off-screen activity, compared to the visual-only view. We explore this comparison in three situations where visual displays have limitations: increasing workspace clutter, decreasing radar size, and increasing visual demand of the observer's task. Our hypothesis is that as visual awareness displays becomes more difficult to see or attend to, audio awareness will become more valuable.

A second goal of the study was to compare audio-only awareness presentations to a visual-only presentation, to see if there are situations where audio alone can be used effectively to support workspace awareness.

Methods

The study asked people to carry out an individual drawing task in a shared chalkboard application, and also keep track of off-screen activity, using one of several different awareness presentations. Off-screen actions were simulated for the study, but were based on pre-recorded traces of real drawing activity.

Participants, Procedure, Task, and Apparatus

Twelve participants (7 men and 5 women) were recruited from a local university; ages ranged from 19 to 34 (mean 25.4). All of the participants had normal vision and hearing. Four of the participants had experience with multi-player games, but none had seen the system used for the study, and none were participants in Study 1.

Participants were given a demonstration of the shared chalkboard system, were told about the simulated off-screen user, and were shown how to complete a task. Participants were instructed to maintain high accuracy on their individual drawing task but to also keep track of the off-screen actions. The system then presented six test trials in each of the experimental conditions. At the end of the session, participants completed a questionnaire asking them about their overall experiences and preferences. Participants were allowed to rest between conditions.

The individual task involved tracing a drawing in the main workspace, using the mouse to control the chalk (see Figure 5). Participants also had to keep track of off-screen activities – in each trial, one stroke or shape would be drawn in another part of the workspace. When the participant noticed that an off-screen action was occurring, they pressed the space bar as soon as the action was finished; this brought up a dialog where they could state where the off-screen action occurred (left or right), and what type of shape had been drawn. In all conditions that involved audio feedback, the participants' own drawing actions also produced audio (although these sounds were played at a lower volume).

The study was conducted using a custom-built groupware system developed in C# and the GT toolkit (hci.usask.ca/ /projects/GT). Procedural audio was generated using the synthesis engine described above. The experiment ran on a Windows 7 PC with a 1280x800 display; participants wore headphones for all conditions with audio feedback.

Study Factors and Conditions

The study examined one main factor (type of awareness presentation) and also looked for interactions with three secondary factors (workspace clutter, radar size, and attentional demand of the primary task).

Type of awareness presentation. The two main conditions for this factor were the radar view by itself, which showed a live miniature of the entire workspace including the off-screen regions (see Figure 5); and the augmented radar, which included synthesized audio awareness. For additional comparisons to explore the second goal of the study, we also tested two audio-only presentations: the synthesized audio on its own, and an abstract-audio condition that played simple symbolic sounds at the start and end of each off-screen action.

Workspace clutter. This secondary factor was used to look for interactions within the main comparison (between the radar and the radar+synthesized conditions). We tested three levels of clutter: none, sparse, and dense (0, 50, or 105 distracter shapes in the workspace, see Figure 5).

Radar view size. This secondary factor included small, medium, and large views (80x60, 240x180, or 400x300). The medium-sized radar is shown in Figure 5.

Attentional demand of individual task. The drawing task (described above) varied in attentional demand: in some conditions the sailboat template gradually moved, and participants had to watch the screen to keep their chalk on the template. We tested three levels of demand: none, slow, and fast (0, 40, or 80 pixels/second movement).

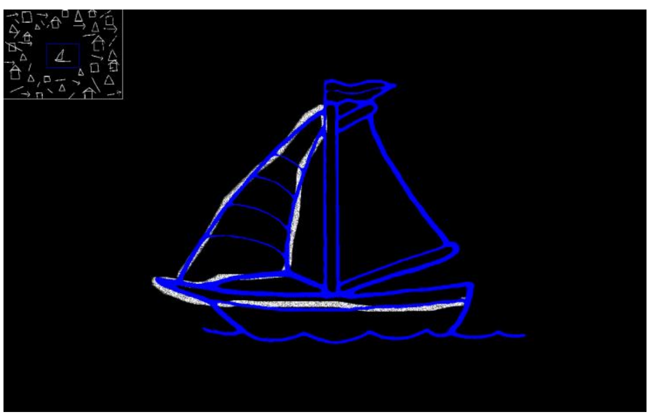

Figure 5. Experimental system with 1280x800 viewport centered in the workspace, with 240x180 radar (at top left) showing sparse workspace clutter. The main screen shows the sailboat template and a participant's drawing strokes.

Experimental Design

The study used a mixed factorial design based on a series of planned comparisons (see Table 1). To explore our main hypothesis, we compared Radar and Radar+Synthesized for each of the secondary factors (Clutter, Size, Attentional Demand). To explore the differences between audio-only and visual-only presentations, we compared the Radar to the two audio-only conditions in a subset of the secondary factors (see details below). The planned comparisons for these two investigations resulted in 24 total conditions.

Awareness presentation was rotated for each participant, so that each presentation was seen in the same position an equal number of times. Secondary factors were seen in sequence (e.g., small, medium, and large for radar size). There were six trials in each condition, meaning that there were 144 data points gathered in each session.

Three dependent measures were collected: accuracy in indicating *when* an off-screen action occurred; accuracy in determining *where* the action occurred; and accuracy in determining *what* type of shape had been drawn.

Table 1. Experimental factors and conditions

	Radar Only	Radar + Synthesized	Synth Only	Abstract Audio
Workspace Clutter	none, sparse, dense (size=med. no move)	none, sparse, dense (size=med., no move)	n/a	n/a
Radar Size	small, medium, large (no clutter; no move)	small, medium, large (no clutter; no move)	n/a	n/a
Attentional Demand (movement)	none, slow, fast (size=med, no clutter)	none, slow, fast (size=med, no clutter)	none, slow, fast	none, slow, fast

Results: Radar vs. Radar+Synthesized

To investigate our main hypothesis (assessing the value of adding audio information), we compared Radar to Radar+Synthesized for each of the secondary factors (Clutter, Radar Size, and Attentional Demand).

Radar vs. Radar+Synthesized: Effects of Clutter

Figures 6-8 summarize the performance of the Radar and Radar+Synthesized presentations, for participant response time (when), location accuracy (where), and shape accuracy (what). In these trials, radar size was always 240x180 pixels, with no movement of the template (see Table 1).

RM-ANOVA showed significant main effects of awareness presentation on all three dependent measures (response time, $F_{1,11}=14.8$, $p<0.005$; location accuracy, $F_{1,11}=12.6$, $p<0.01$; shape accuracy, $F_{1,11}=61.8$, $p<0.001$). In addition, there were interaction effects for all measures; as seen in Figures 6-8, the difference between Radar and Radar+Synthesized increased with additional clutter (response time, $F_{2,22}=10.8$, $p<0.001$; location accuracy, $F_{2,22}=7.9$, $p<0.005$; shape accuracy, $F_{2,22}=6.2$, $p<0.01$).

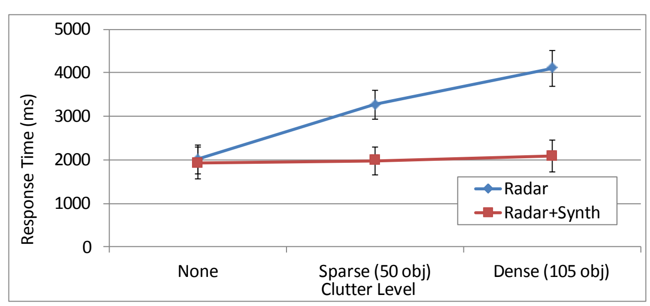

Figure 6. Response time (to notice an off-screen event) by presentation and clutter level. Error bars show ± std. error.

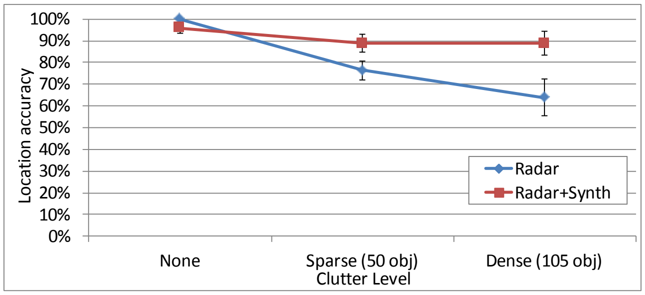

Figure 7. Accuracy in determining the location (left or right) of the off-screen event, ±s.e., by presentation and clutter level.

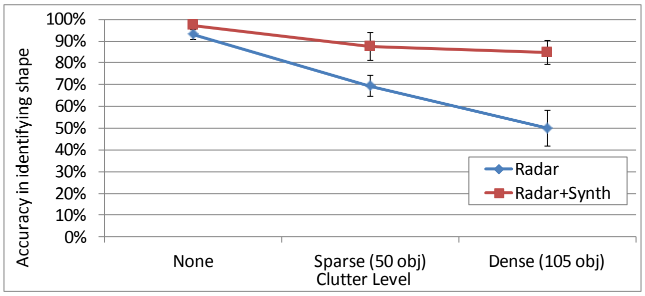

Figure 8. Accuracy in determining what shape was drawn, ±s.e., by awareness presentation and clutter level.

Radar vs. Radar+Synthesized: Effects of Radar Size

No differences were found on any measure when comparing Radar and Radar+Synthesized in terms of the size of the radar view (response time, $F_{1,11}$=1.02, p=0.33; location accuracy, $F_{1,11}$=2.10, p=0.17; shape accuracy, $F_{1,11}$=3.34, p=0.09). In addition, no interactions were found between the awareness presentation and the radar size (response time, $F_{2,22}$=1.7, p=0.21; location accuracy, $F_{2,22}$=1.88, p=0.18; shape accuracy, $F_{2,22}$=0.20, p=0.82).

As seen in Figure 9, the largest difference between the two awareness presentations was at the smallest radar size; this may suggest that even smaller sizes could lead to significant differences, but this is left for future work.

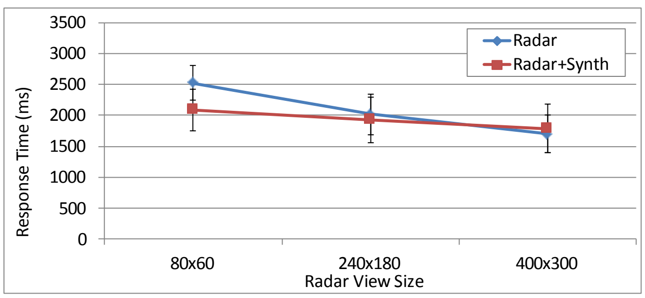

Figure 9. Response time ±s.e. by presentation and radar size.

Radar vs. Radar+Synth: Effects of Attentional Demand

RM-ANOVA showed a significant main effect of awareness presentation for response time ($F_{1,11}$=7.03, p<0.05), but not for location accuracy ($F_{1,11}$=0.05, p=0.82) or shape accuracy ($F_{1,11}$=1.73, p=0.21). In addition, no interaction effects were found: response time, $F_{2,22}$=1.82, p=0.19; location accuracy, $F_{2,22}$=0.94, p=0.41; shape accuracy, $F_{2,22}$=0.45, p=0.64. Figure 10 summarizes the response time results. Radar size for these trials was always 240x180 pixels, with no workspace clutter.

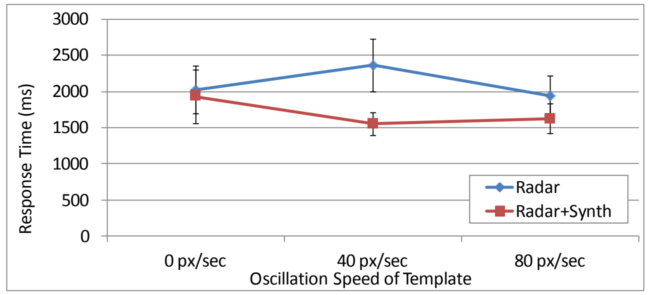

Figure 10. Response time ±s.e., by awareness presentation and amount of attentional demand (template oscillation).

There are substantial differences in response time between Radar and Radar+Synthesized for the slow oscillation condition, but less difference for fast oscillation. We suspect a training effect in this case, since participants always saw the slow condition before fast. In addition, observations suggest that participants in the Radar condition often ignored the movement of the template and drew lines from memory in order to spend more time looking at the radar view.

Results: Audio-only vs. visual presentations

To investigate how audio-only awareness presentations match up to the visual-only presentation, we compared the best and worst performance of the radar view alone, with the synthesized-only and abstract-only conditions. We analysed only response time and shape accuracy, since the abstract audio was not spatialized. The analysis shows that at its best, the radar view by itself outperforms the audio-only presentations, but that the synthesized audio is equal to or better than the radar when visual conditions are poor (e.g., the workspace is cluttered).

For response time, RM-ANOVA showed a main effect of awareness presentation ($F_{3,33}$=13.35, p<0.001). A followup Tukey HSD test showed that Radar-best was significantly faster than all others, and Synthesized was significantly better than Radar-worst and Abstract sound (all p<0.05).

For shape accuracy, there was also a main effect ($F_{3,33}$=37.65, p<0.001). Follow-up analysis showed that Radar-best was better than all other conditions, and that Abstract-only was worse than all other conditions (p<0.05).

Results: Participant Preferences

At the end of the session, people were asked to rank the four presentation types in order of preference. Responses were very consistent, with an overwhelming preference for the condition that provided both visual and audio information. All twelve participants ranked the Radar+Synthesized presentation as their highest preference, and ten of twelve participants ranked the remaining conditions in the order Radar-only, Synthesized, Abstract.

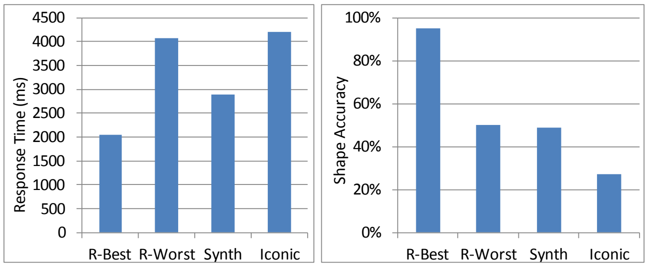

Figure 11. Response time and shape accuracy by awareness presentation (R-Best is a 240x180 radar view with no clutter; R-Worst is a 240x180 radar with dense clutter).

DISCUSSION

The two studies provide new empirical evidence that synthesized audio can add rich awareness information to visual presentations of off-screen activity. People are able to determine the types and qualities of actions represented by synthesized audio with a high degree of accuracy, and the addition of this auditory information significantly improves awareness over a radar view alone – particularly when there is workspace clutter or when the user's individual task demands their attention. The performance results are strongly echoed in participant preferences.

In the following paragraphs we discuss several issues arising from this work: how the audio information aided awareness, how these results will generalize to other groupware tasks, potential limits to the approach, and the main lessons for designers of groupware systems.

How and why does audio improve awareness?

One of the main problems with visual awareness displays is that they must be attended to in order to determine when off-screen events happen; and if the user misses the details of the event, they must determine what action has occurred by remembering and assessing the difference between current and previous states of the view.

Audio information about off-screen actions aids with both aspects of this problem. First, audio awareness frees people from having to constantly poll the radar view for new events; the guarantee of an audible notification of activity allows people to concentrate their visual attention on their individual task rather than the group awareness task. Audio information therefore acts as an event notification channel, telling people when they need to look at the radar view (this is a different kind of notification than seen in earlier systems such as ShareMon [7], in that audio and visual awareness tools are tightly coupled).

Second, the dynamic qualities of the synthetic audio also appear to be important, since they provide information about what type of off-screen activity is occurring (i.e., not just that something is happening, but some idea about what kind of action). Even though this information is not always highly accurate, it provides a rough understanding that can prime the user when they turn their attention to the visual radar – that is, it gives people information about what kinds of actions they should be looking for in the radar. Our second study suggests that this helped people maintain awareness particularly when it was difficult to determine what to look for in the visual display.

We did not test a combination of the radar view with abstract sound, but we suspect that it would also provide some improvement, since the sound icons would act as indicators of off-screen events, even if they did not provide information of the type of activity. However, early pilot testing with this combination suggested that people found the abstract sounds irritating in a drawing program; in future work we plan to explore this combination further.

Generalizing the results

Our experience with audio awareness suggests that our main findings can be generalized to other groupware systems and shared workspaces. First, there are many situations where group awareness is important, but where either seeing or attending to a visual awareness display can be difficult. Our study showed effects for attentional demand and workspace clutter, but there are also likely to be problems with other factors such as distance to the radar view (e.g., on large displays), other attention-demanding tasks (e.g., IM notifications), or screen-space constraints (e.g., on mobile devices). In addition, we believe that participants in our study were more focused on the awareness task (due to the nature of the experiment) than many groups will be in the real world – in these cases, audible awareness cues should be able to provide people with a general understanding of off-screen activity without them even having to look at the visual awareness display.

Second, synthesized audio awareness will be applicable to many other types of shared-workspace groupware, even though there are also limits to the way that real-world sounds can be applied in digital environments. Dynamic audio can clearly be useful in groupware systems where people use direct manipulation and spatial actions with a pointing device – activities such as dragging, pointing, sliding, selecting, drawing, handwriting, flicking, scrolling, and gesturing can all be richly represented with continuous synthesized sound. This broad applicability has long been stated by audio researchers; with techniques such as granular synthesis, however, it becomes considerably easier for groupware designers to take advantage of these awareness benefits. Dynamic audio awareness also presents particular opportunities for improving interactive richness in game environments.

The audio-awareness approach is made more widely applicable because of the simplicity of the granular synthesis technique. We are not sound designers, but were easily able to simulate many types of chalk sounds with consumer recording equipment and a few hours' work in obtaining real-world samples and testing the resulting synthesized sounds. Audio awareness can be extended to new actions simply by recording new real-world sources, and identifying the input parameters from the digital environment that should influence the synthesis engine.

Potential limits to audio awareness

Although dynamic audio information can be a valuable addition to the awareness support in groupware, there are several situations in which the technique has limits.

- *Actions with no obvious sound analogue.* There are many actions in computational environments that do not have real-world sound analogues – for example, there is no obvious sound for converting a picture to grayscale. Some of these actions are symbolic commands that occur instantaneously (and so have no timescale for continuous sound), and others do not have a clear real-world sound source. In these situations, we believe that it is still possible to provide reasonable auditory awareness, but not through granular synthesis alone. Although we focus on continuous synthesized audio, it is clear that other techniques (such as parameterized auditory icons [11], symbolic sounds [6], representative soundscapes [9], or genre sounds [7]) can be used in conjunction with dynamic synthesis. By using the range of audio techniques that have been explored in previous literature, it should be possible to provide awareness of a large majority of actions in groupware systems.
- *Potential for distraction*. The other side of all awareness support is the possibility that the awareness information will distract users from their individual work. This is certainly also true of audio information, and the fact that audio is not tied to visual attention (which was one of its main strengths in our studies) could also be a main drawback – that is, people might be unable to simply ignore the information by focusing elsewhere. However, the advantage of using audio awareness in a computational environment is that it can be adjusted according to the user's preferences – for example, users could simply turn off the audio, or could set volume levels for different people or different types of activity.
- *Auditory clutter*. Our study looked at the effects of visual workspace clutter, but we did not examine situations where there are several sounds happening at once – which could make it difficult for people to determine specifics of any one activity. It seems clear that this kind of clutter can cause problems, and that more clutter will make interpretation more difficult. However, there are also some mitigating factors suggesting that clutter will not be a major problem for the approach: first, people are able to deal with multiple sound sources in the real world, particularly when the sounds have different qualities (e.g., chalk sounds are easy to distinguish from chalk erasers), and second, groupware environments will allow people to set preferences on audio feedback, as mentioned above. We note that in our second study, the participant's own individual drawing sounds were always audible, so our results are already based on the presence of one set of distracter sounds in the workspace.
- *Technical limitations of granular synthesis.* The granular synthesis system has limits in terms of the number of different sounds that it can play at once. Since the engine plays multiple grains simultaneously, the number of different types of sound that can be generated is limited by the number of simultaneous voices in the sound hardware. This limits the number of different activities that can be represented with synthetic sound. If people are generating the same type of sound, then grains can be combined into a shared synthesis engine – using this approach, we estimate that current hardware could support eight simultaneous users if all were using chalk. We note that it is also possible to combine granular synthesis with abstract sounds (and in fact we do this with the initial contact sound for a chalk stroke).

Lessons for designers

Our main lesson for groupware designers is to reiterate and add to the findings of earlier audio investigations – that auditory information can significantly improve group awareness in situations where it is difficult to see or attend to visual displays. The relative simplicity of granular synthesis means that designers can easily consider including this kind of information in their groupware systems, whenever they foresee the co-occurrence of awareness need and visual difficulty.

Designers should also consider audio awareness information in situations where visual awareness cannot be provided at all. For example, in small-screen devices such as a smartphone, there may be no room for a visual display; in these cases, synthetic audio could still provide participants with a reasonable understanding of what others are doing in the workspace.

Last, designers should make use of the granular synthesis technique for sound production, which has proven to be a simple mechanism for developing and synthesizing realistic and rich action sounds. This technique can feasibly be added to most groupware environments, and should be easy to add to many groupware toolkits. (Our synthesis engine is available at hci.usask.ca/audio-awareness/).

CONCLUSIONS AND FUTURE WORK

Workspace awareness is difficult to maintain in distributed settings where collaborators may be working in other parts of the workspace. Although awareness displays exist to reduce the problem, visual presentations are problematic when users cannot see or attend to the information. We investigated the use of audio information as a way to improve awareness support in these situations. We tested audio awareness in a realistic groupware setting, and showed that adding audio to a visual view can significantly improve awareness of off-screen actions.

Our research suggests several directions for further study. First, we plan to test our granular synthesis system with other types of sound, extend the system to make it easier for designers to create and use new sounds, and incorporate the engine in a groupware toolkit. Second, we will expand our evaluation to other situations, including multiple collaborators active in the workspace, multiple different types of activity, and different types of groupware application such as diagram editors and board games. Third, we will study the effectiveness of audio awareness in real-world scenarios, and will make use of other mechanisms for assessing awareness tools such as using an eye tracker to explore in more detail the way that audio and visual awareness presentations work together.

ACKNOWLEDGMENTS

Our thanks to John Willamson for granular synthesis code, David McDine for assistance with the studies, and the anonymous reviewers for many valuable suggestions. This work was supported by NSERC, the GRAND NCE, and the SurfNet research network.

REFERENCES

1. Baecker, R., Nastos, D., Posner, I., and Mawby, K., The User-Centred Iterative Design of Collaborative Writing Software, *Proc. CHI 1993*, 399-405.
2. Barrass, S. and Adcock, M., Interactive Granular Synthesis of Haptic Contact Sounds, *Proc. AES Virtual, Synthetic and Entertainment Audio 2002*, XX-YY.
3. Beaudouin-Lafon, M. & Gaver, W., ENO: Synthesizing Structured Sound Spaces, *Proc. UIST 1994*, XX-YY.
4. Bencina, R., Implementing Real-Time Granular Synthesis, in *Audio Anecdotes III*, K. Greenebaum and R. Barzel, eds., AK Peters, 2007.
5. Beaudouin-Lafon, M., and Karsenty, A., Transparency and Awareness in a Real-Time Groupware System, *Proc. UIST 1992*, 171-180.
6. Blattner, M., Sumikawa, D., and Greenberg, R., Earcons and Icons: Their Structure and Common Design Principles, *HCI*, 4(1), 1989, 11-44.
7. Cohen, J., 'Kirk Here:' Using Genre Sounds to Monitor Background Activity. *Proc. CHI 1993*, 63-64.
8. Cohen, J. Monitoring Background Activities. In *Auditory Display*, G. Kramer, ed., Addison-Wesley, 1994, 499-531.
9. Cohen, J. Out to Lunch: Further Adventures Monitoring Background Activity, *Proc., ICAD 1994*, XX-YY.
10. Dourish, P., Bellotti, V., Awareness and Coordination in Shared Workspaces, *Proc. CSCW 1992*, 107-114.
11. Gaver, W., Sound Support for Collaboration, *Proc. ECSW 1991*, 293-308.
12. Gaver, W., The SonicFinder: An Interface That Uses Auditory Icons, *HCI*, 4, 1, 1989, 67-94.
13. Gaver, W., Smith, R., and O'Shea, T., Effective Sounds in Complex Systems: The ARKola Simulation, *Proc. CHI 1991*, 85-90.
14. Gaver, W., Synthesizing Auditory Icons, *Proc. CHI 1993*, XX-YY.
15. Gaver, W., How Do We Hear In The World? Explorations of Ecological Acoustics. *J. Ecological Psychology*, 5, 4, 1993, 285-313.
16. Gutwin, C., Greenberg, S., and Roseman, M. Workspace Awareness in Real-Time Distributed Groupware, *Proc. People & Computers 1996*, 281–298.
17. Gutwin, C., and Greenberg, S., A Framework of Awareness for Small Groups in Shared-Workspace Groupware, *CSCW*, 3-4, 2002, 411-446.
18. Gutwin, C., and Greenberg, S., Effects of Awareness Support on Groupware Usability, *ToCHI*, 6, 2, 1999, 243-281.
19. Gutwin, C., Roseman, M., Greenberg, S., A Usability Study of Awareness Widgets in a Shared Workspace Groupware System, *Proc. CSCW 1996*, 258-267.
20. Hindus, D., Ackerman, M., Mainwaring, S., and Starr, B., Thunderwire: a Field Study of an Audio-only Media Space. *Proc. CSCW 1996*, 238-247.
21. Keller, D. and Truax, B., Ecologically-Based Granular Synthesis, *Proc. ICMC 1998*, XX-YY.
22. Oakley, I., Brewster S.A. and Gray, P., Communicating with feeling. *Haptic Human-Computer Interaction*. Brewster, S.A. and Murray-Smith, R. (Eds.), LNCS 2058, 2001, 61-68.
23. McGookin, D. and Brewster, S.A., An Initial Investigation into Non-Visual Computer Supported Collaboration. *Proc. CHI 2007*, 2573 - 2578.
24. Morris, M., Morris, D., and Winograd, T., Individual audio channels with single display groupware: effects on communication and task strategy. *Proc. CSCW 2004*, 242-251.
25. Ramloll, R., Mariani, J., Do Localised Auditory Cues in Group Drawing Environments Matter? *Proc. ICAD 1998*, icad.org/websiteV2.0/Conferences/ICAD98/.
26. Roads, C. *The computer music tutorial*. MIT Press, 1966.
27. Roads, C., *Microsound*, MIT Press, Cambridge, 2002.
28. Schlienger, C., Conversy, S., Chatty, S., Anquetil, M., and Mertz, C., Improving Users Comprehension of Changes with Animation and Sound: an Empirical Assessment. *Proc. Interact 2007*, 207-220.
29. Smith, R., O'Shea, T., O'Malley, C., Scanlon, E., Taylor, J., Preliminary experiences with a distributed, multi-media, problem environment, *Proc. ECSCW 1989*.
30. Stefik, M., Bobrow, D., Foster, G., Lanning, S., and Tatar, D., WYSIWIS Revised: Early Experiences with Multiuser Interfaces, *ToIS*, 5(2), 1987, 147-167.
31. Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., and O'Modhrain, S. 2005. GpsTunes: controlling navigation via audio feedback. *Proc. MobileHCI 2005*, 275-278.
32. Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., and Baudisch, P., Earpod: eyes-free menu selection using touch input and reactive audio feedback. *Proc. CHI 2007*, 1395-1404.