Individual Audio Channels with Single Display Groupware: Effects on Communication and Task Strategy

Meredith Ringel Morris, Dan Morris, Terry Winograd
Stanford University
353 Serra Mall, Stanford, CA, USA
{merrie, dmorris, winograd}@cs.stanford.edu

ABSTRACT
We introduce a system that allows four users to each receive sound from a private audio channel while using a shared tabletop display. In order to explore how private audio channels affect a collaborative work environment, we conducted a user study with this system. The results reveal differences in work strategies when groups are presented with individual versus public audio, and suggest that the use of private audio does not impede group communication and may positively impact group dynamics. We discuss the findings, as well as their implications for the design of future audio-based "single display privacyware" systems.

Categories and Subject Descriptors
H.5.3 [**Information Interfaces and Presentation (e.g., HCI)**]: Group and Organization Interfaces – *computer-supported cooperative work.*

General Terms
Design, Experimentation, Human Factors.

Keywords
Single display groupware, audio interfaces, single display privacyware, multimodal interfaces, tabletop interfaces.

1. INTRODUCTION
Single Display Groupware (SDG) [23] allows several co-located people to work together using a single, shared display. In comparison to a working environment based around individual workstations, this shared context can increase productivity and facilitate group communication. However, SDG systems also present numerous challenges, such as clutter caused by limited display real estate and the inability to convey private or personalized information to members of the group. Since the group shares a single surface, all information is visible to all group members. Channels for conveying private information have several practical applications, including transmission of private or secure data and reduction of problematic clutter.

Single Display Privacyware [21] (SDP) extends the notion of Single Display Groupware to incorporate auxiliary mechanisms for conveying private or customized content to individual users of a shared display. Several examples of privacyware have been explored, including systems using specialized shutter glasses [2, 21], systems using auxiliary displays such as PDAs and laptops [6, 11, 13, 19], and systems using physical partitioning of the shared surface [18, 25]. The use of multimodal interfaces as a solution to the single display privacyware problem is a relatively unexplored area. A few systems using audio for entertainment purposes [3, 9] have been developed, but their utility has not been formally evaluated, nor have systems using private audio channels to support group productivity tasks been explored. We discuss these previous systems in more detail in Section 6.

This paper describes a multimodal approach to SDP. Our system uses individual sound channels to provide private information to specific users. We discuss the implementation of our system, and we present the results of an initial study that demonstrates the applicability of this approach to a collaborative task. The quantitative and qualitative results suggest that private audio has potential as a means of supplementing shared displays. We conclude with a discussion of related work.

2. SYSTEM DETAILS
2.1 Hardware
Our system currently employs a DiamondTouch [4] table, although the concepts we describe are applicable to other forms of Single Display Groupware. The DiamondTouch is a top-projected display that also serves as an input device. The touch-sensitive surface accepts simultaneous input from up to four people. Users sit on custom chairs, through which they are capacitively coupled to the table. This configuration allows identification information to be associated with each touch event.

This user-identification feature is important to our system as a mechanism for determining which user should receive certain audio content, based on his actions on the table surface. Other existing hardware for SDG would need to be augmented in some manner (e.g., by adding cameras or by adding support for multiple mice) in order to be able to associate user identification information with each input event.

Four chairs (color-coded with conductive seat covers) surround our DiamondTouch, one along each side of the table. The color of a user's chair is associated with him in our system's user interface. Figure 1 depicts our system configuration.

Our system runs on a consumer-grade PC (3.0 GHz Pentium 4 with 1 GB RAM), to which we have added five off-the-shelf soundcards. One of the soundcards is connected to a set of standard PC speakers, while each of the other four is connected to

Figure 1. Four users sitting around a tabletop display can receive private information over their individual earbuds.

an earbud-style headset. We chose to use earbuds (small knobs that fit inside the ear) rather than standard headphones (which cover the entire ear) in order to facilitate collaboration. Users of our system wear a single earbud in one ear, so they can still converse at normal volumes with their co-workers. Our decision to use single-ear audio is reinforced by a study of Sotto Voce [1] (a PDA-based museum guide system), which found that using one-eared headsets allowed users to comfortably converse with each other. Additional literature [5] suggests that listeners are better able to differentiate multiple audio sources if they are directed to different ears; our single-earbud approach leverages this fact by presenting system-generated audio to one ear and allowing conversation to be perceived contralaterally.

2.2 Software

2.2.1 Sound API

We have implemented a Java library that allows sound clips (wav, mp3, MIDI, etc.) and text-to-speech requests to be sent to one or more sound channels. To play a sound, the programmer specifies either the text to be spoken or the sound file to be played, along with a bit mask indicating which subset of the soundcards should output the sound. In this manner, it is possible to specify that sound X should, for example, be played only over soundcard 1 (connected to the first user's earbud), while sound Y should be played over soundcards 3 and 4 (user 3 and 4's earbuds). Multiple sounds can be simultaneously mixed over each of these channels. Our library provides several ways to control sounds – in addition to the ability to play, pause, and stop the audio, it is also possible to seek to an absolute or relative offset within each audio stream.

We created a Java library so that our private sound API would be compatible with DiamondSpin [20], a Java toolkit for developing multi-user tabletop interfaces that we used to develop our prototype tabletop application. Because the current implementation of the Java Sound API does not provide access to individual audio devices, our library uses the Java Native Interface to pass requests to a C++ back-end. The C++ module uses Direct Show and the Microsoft Speech API to route audio clips and text-to-speech requests to individual sound devices. Each sound is loaded or rendered to a shared data buffer, and asynchronous playback requests are submitted for each requested output device.

2.2.2 SoundTracker

We developed a prototype application in order to explore the feasibility of using private audio to supplement a shared display. Our application, SoundTracker, allows up to four users to browse a collection of photographic stills from a movie (each representing a particular scene) and a collection of mp3 songs (represented on-screen by their titles). Songs can be bound to scenes by dragging song titles onto images, allowing users to choose a "soundtrack" for the film.

This application is representative of a broader class of groupware that supports tasks where multiple users are involved in collaborative decision-making involving a large number of documents or other objects. Although the SoundTracker application's use of an audio-centric task limits its generalizability, it was designed primarily to focus on low-level interface issues and to explore the impact of private audio on group behavior in a controlled setting. With these larger goals in mind, we conducted a user study using our private audio system and the SoundTracker application in order to ascertain whether multiple private audio channels had potential as a useful way to augment single display groupware. This study addressed several questions:

- What effect does the use of private audio channels have on work strategies as compared to the use of a system with no private source of information?
- How does wearing earbuds and listening to different audio sources affect communication among group members?
- How does the use of private audio channels affect group productivity as compared to a system with no private source of information?
- How does the use of private audio affect the overall usability of the system?

The remainder of this section describes additional features of SoundTracker; in Section 3, we describe details of the user study conducted with this application.

2.2.2.1 Song Objects

A song is represented by a label containing the song's title and a musical note icon (Figure 2a). To move a song around the table, a user can touch the title area with his finger and drag the song object to its new position. In order to play the song, a user can touch the musical note icon, and the song will be played through that user's earbud. Touching the icon a second time will stop the song. When a song is played, its note icon changes color, and a slider appears, which can be used to navigate forward and backward within the song (Figure 2b).

Multiple users can play the same song simultaneously by each touching the musical note icon. Each user will have his own seek slider (color-coded according the color of each user's chair) for that song (Figure 2c). If one user touches the note a second time to turn off the song, it turns off only for him, and continues playing for any other users who had previously activated it.

2.2.2.2 Scene Objects

Scene objects represent scenes from a movie. Each scene object consists of a "tray" (an initially blank area where song icons can be placed), a photographic still, and a "speech bubble" icon (Figure 3a). Touching the photo or its tray with a finger allows

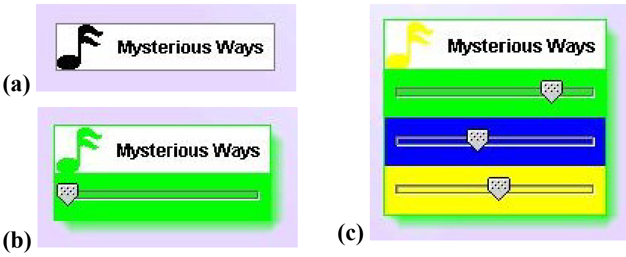

Figure 2. (a) A song object, in the off state. Touching the title area allows a user to move the song object around the table, and touching the note icon plays the song over that user's earbud. (b) A song object, being played by the user in the green chair. When a song is played, a slider appears that allows the user to seek forward and backward within the song. (c) More than one user can simultaneously play the same song. By using their individual slider bars, users can each listen to different sections of the same song.

users to move it about the table in the same manner as the song objects. Touching the speech bubble plays a brief caption (ranging from 2 to 7 seconds), which summarizes the plot of that scene, over the earbud of the user who touched it. If the user does not want to play the entire caption, he can touch the speech bubble icon a second time in order to turn it off. As with the song objects, more than one user can simultaneously play the same caption.

Users can associate songs with scenes by dragging a song object into the tray. The song will then "snap" into the bottom of the tray (Figure 3b) and will remain attached if the scene object is moved around the table. A song can be disassociated manually by dragging it outside the borders of the tray, or by replacing it with a new song.

It is possible for a single user to play both a song and a scene caption over his earbud at the same time. However, we imposed a restriction that each user can play at most one song and one caption. We imposed this limit because we found during our pilot testing that it was possible to attend to both a song and a caption simultaneously, but multiple songs or multiple captions became muddled and difficult to comprehend.

Because users are sitting at four different angles around the table, there is no single "correct" orientation for scene or song objects. To address this difficulty, we used the user-identification capability of the DiamondTouch combined with our prior knowledge of the fixed locations of the chairs. Whenever a user touches a scene or song object, that object is re-oriented to face that user, using the transformations provided by the DiamondSpin toolkit.

3. USER STUDY

3.1 Participants

We recruited sixteen paid subjects (thirteen men and three women), ranging in age from eighteen to twenty-four years. None of the subjects had prior experience using a DiamondTouch table, but all were experienced computer users. All had normal hearing and normal color vision. The sixteen subjects were divided into four groups of four users each. Twelve additional users (three groups of four) served as pilot subjects.

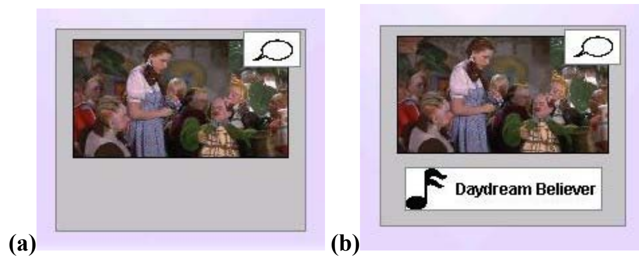

Figure 3. (a) A scene object consists of a picture, a speech bubble icon, and a tray. Touching the speech bubble plays a brief caption summarizing the scene's plot. (b) Dragging a song object onto the scene and releasing it adds it to the scene's tray.

3.2 Measures

Several types of quantitative and qualitative data were gathered. The SoundTracker software was instrumented to log time-stamped records of all interactions, including events related to moving songs and scenes about the table, events related to playing songs and captions, and associations and disassociations of songs and scenes. All groups were videotaped and observed by the experimenter, who took notes throughout. Finally, after using the system, all participants completed a questionnaire containing both Likert-scale and free-form questions.

3.3 Procedure

When a group arrived for the study, the four group members were seated one on each side of the DiamondTouch table. Participants were told they would be working together as a group during the study, and were asked to introduce themselves to the rest of the group. The group then completed a tutorial in which they were introduced to the basic functionality of the SoundTracker application.

After the tutorial, each group was presented with seventeen images captured from a popular movie, each representing a particular scene from the film, and thirty-four icons representing songs selected from a popular music collection. The group was instructed to construct a "soundtrack" for the film by assigning songs to images. The criteria for a good soundtrack were subjective, but groups were instructed to consider elements such as the song's tempo and emotional content and how they might fit with the mood of a scene. To further motivate subjects to make careful selections, they were told that after all groups had finished the experiment, a panel of judges would review each group's final soundtrack selection and would vote on which one was the "best," awarding a prize (gift certificates) to the winning group. Participants were instructed to notify the experimenter when they felt they had reached a consensus on the final soundtrack selection. A twenty-minute time limit was enforced.

Each group was asked to perform this task twice: In one pass through the task, the "private sound" condition, captions and songs were played over individual earbuds. In the "public sound" condition, captions and songs were played over a single, shared speaker. The restrictions on per-user songs and captions were the same in both conditions. The ordering of the two conditions was balanced among groups.

Scenes were selected from a different movie in each condition – "The Princess Bride" (MGM, 1987) and "Titanic" (Paramount Pictures, 1997) – and two disjoint sets of songs were used. The

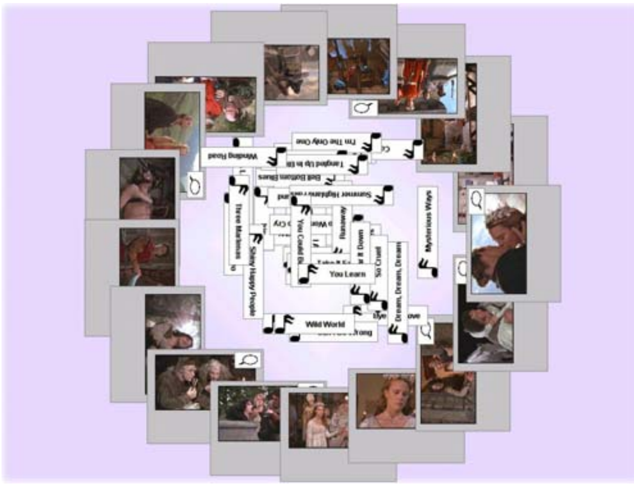

Figure 4. The initial layout of the table in each condition has the scene objects arranged in a circle with the song objects piled in the center.

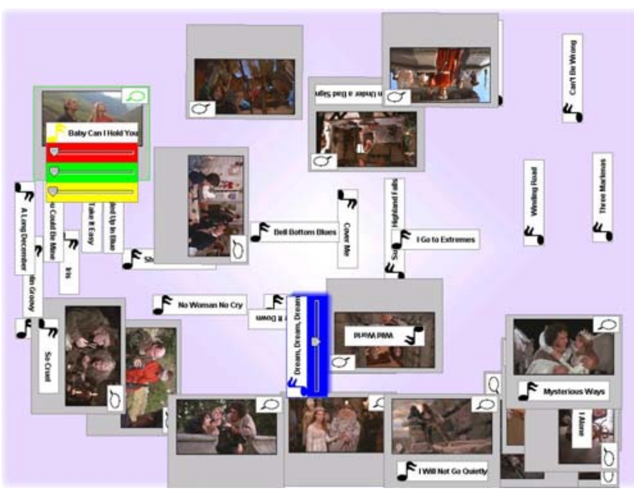

Figure 5. A typical example of the table's layout partway through the study. Some songs and scenes have been paired, and songs and captions are being played.

association between movies and conditions was also balanced among groups. Each movie always appeared with the same set of songs, which were not selected from the film's original soundtrack. Three of the sixteen participants had never watched "The Princess Bride," and a different three subjects had never seen "Titanic" – however, these individuals were distributed such that in each group at least two group members (and usually three or four) had seen each of the films. Also, before beginning the application, the experimenter read a summary of the movie's plot to the group.

Figure 4 shows the configuration of the table at the beginning of an experiment. The seventeen scene objects are arranged in a circle, facing the outer edges of the table, ordered chronologically according to their order in the film. The thirty-four song objects are piled randomly in the middle of the circle. Figure 5 shows a screenshot of a table configuration captured several minutes into an experiment – some song-scene pairings have been made, and users are playing some songs and captions. Figure 6 shows a typical end-of-condition scenario, where all scenes have been

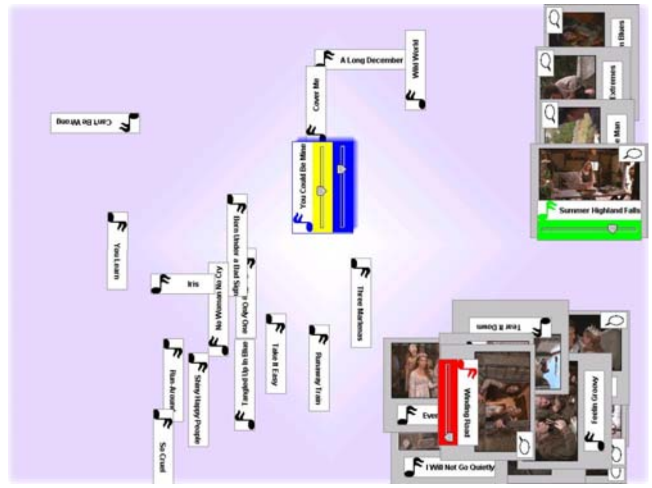

Figure 6. A typical table configuration near the end of the study. Scenes have been assigned songs and are piled in one area of the table, in order to reduce clutter. Users are playing some songs to verify their agreement with the group's final selections.

associated with songs, and have been piled in one area of the table to reduce clutter.

After completing both experimental conditions, all sixteen subjects individually completed a questionnaire that contained both Likert-scale questions and free-form response questions about subjects' experiences. Table 1 summarizes the results of the Likert-scale questions.

4. RESULTS

The questionnaire, log, and observation results paint an interesting picture of the effects of private versus public audio regarding task strategies, communication, productivity, and usability.

4.1 Task Strategies

Subjects were asked several free-form questions on the post-study questionnaire. One such question asked, "Please describe how your strategy for assigning the soundtrack differed between the headphones and public speakers conditions." Responses followed a consistent pattern, indicating that in the private audio condition groups tended to use a divide-and-conquer strategy while following a more serial strategy in the public audio condition. For example, one subject wrote, "With headphones, we worked individually until we found something we liked and then shared with the group. With speakers we went through each song and picture together." In the public condition, whenever someone accidentally began playing a song while another song was already playing, he immediately turned it off and apologized to the group. No groups ever intentionally played multiple songs simultaneously over the speakers, but they did sometimes play one song and one caption simultaneously without apparent comprehension difficulty.

Another strategic difference we observed was that the private audio created a more "democratic" setting, where all users participated in selecting interesting songs and scenes and suggesting pairings. Shy users who were not as willing to speak up in the public condition participated more in the private condition, non-verbally making suggestions by creating scene-song pairings that were later discussed by other group members.

Table 1. This table shows the mean responses to the Likert-scale questions completed by each of the sixteen participants from 1 = strongly disagree to 5 = strongly agree.

	Mean
I found it difficult to communicate with my group when wearing headphones	**1.88**
I found it uncomfortable to wear headphones	**2.25**
I enjoyed using headphones to complete the task	**4.0**
I enjoyed using public speakers to complete the task	**3.31**
I found it easy to complete the task using headphones	**3.88**
I found it easy to complete the task using public speakers	**3.19**
I felt satisfied with the group's soundtrack selection in the headphones condition	**4.19**
I felt satisfied with the group's soundtrack selection in the public speakers condition	**4.06**

By contrast, in the public condition, one or two group members often took on a leadership role, suggesting pairings and controlling the interface. Rogers and Rodden [15] note that use of traditional, shoulder-to-shoulder single display groupware (such as electronic whiteboards) typically results in situations where the more dominant group member controls most of the interaction while others play a supporting role. Informal observations by Rogers [14] suggest that tabletop groupware promotes more participation by all group members than shoulder-to-shoulder displays. Our observations further this line of inquiry by suggesting that participation by all group members can be further increased by the addition of private audio channels.

The software logs of user actions support our observations that groups had more "democratic" behaviour when using headphones as compared to the public condition. To assess the degree to which all four users were participating in the task, we measured the number of song play events and song/scene pairing events that each user was responsible for in each condition. We propose that a uniform distribution of these events among users suggests more uniform participation in the task. For each group, we counted the percentage of table events that were associated with each user. We then computed the standard deviation of those percentages within each group. We call this value a "table-dominance" score. A higher "table dominance" value indicates a less uniform distribution of events among users, while a lower score reflects more equal contributions among group members. The results are summarized in Figure 7, averaged over the four groups. For each type of event (playing songs and creating song/scene pairs), the "table dominance" value was significantly higher for the public sound condition ($p<.05$), indicating that a subset of users tended to dominate the manipulation of items on the table. This may reflect the fact that shy or unassertive users felt more empowered to contribute in the private case.

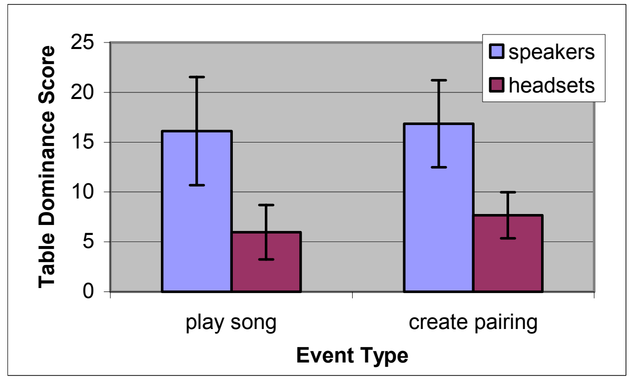

Figure 7. The lower table dominance score (standard deviation among the percent of song plays and song/scene pairings initiated by each group member) indicates more equitable participation in the private condition as compared to the public condition. A higher score indicates less equal participation among group members.

Groups replaced songs assigned to scenes (e.g., after a song had been associated with a scene, it was removed and a new song was added instead) more frequently ($p<.05$) in the private condition (an average of 69.5 replacements per group) than in the public condition (an average of 29.25 replacements per group) (Figure 8). This could be indicative of several factors – perhaps the groups just got it "right" the first time when they all focused on one task together in the public condition. Or, it could reflect the fact that, in the private condition, groups collaborated on the task by actively reviewing – and often replacing – choices made by other group members. Of the replacements made in the public condition, 68.4% were self-replacements (a user replacing a song-scene pairing that he had created) and 31.4% were other-replacements (replacing a pairing that had been established by another user). In the private condition, 58.3% were self-replacements and 41.7% were other-replacements.

Not surprisingly, in the public condition groups were unlikely to play more than one song and/or one caption at a time, while in the private condition several users simultaneously played sounds. As one would expect, in the private condition, users played songs and captions more frequently (an average of 221 songs and 78.25

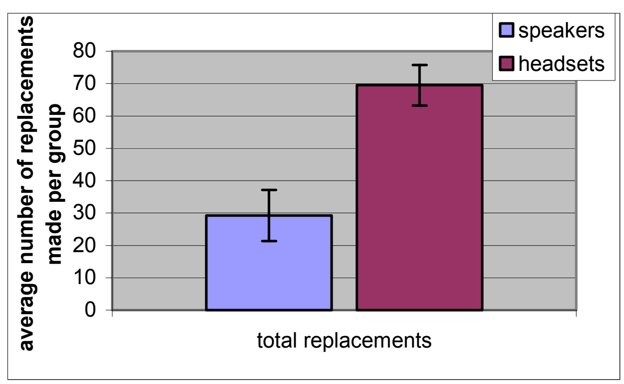

Figure 8. Groups in the private condition were more likely to replace previously established song/scene pairings than groups in the public condition.

captions per group) than in the public condition, in which they played an average of 93.5 songs and 36.5 captions per group (songs: $p<.02$, captions: $p<.01$). Longer clips of songs were played in the private audio condition, with an average duration of 11.56 seconds, as compared to 7.45 seconds in the public case ($p<.05$). While we anticipated that there would be more song play events in the private condition, we had also expected that there would be greater coverage of the songs in the private case, since we thought that the lower number of play events with public audio might also mean that some songs were not explored at all. However, we were surprised to see that this was not the case – nearly all of the thirty-four songs were played at least once in both conditions (an average of 33.75 per group in the private condition, and 33.25 in the public condition). It is possible that this is a ceiling effect – it would be interesting to see whether both conditions still result in equal coverage of songs if either the number of songs were increased or if the allotted task time were shortened.

4.2 Communication

Subjects did not seem to feel that private audio reduced group communication. They disagreed (mean = 1.88) with the statement "I found it difficult to communicate with my group when wearing headphones." Another indication that participants felt the group communicated and worked well together in the private audio condition is their agreement (mean = 4.19) with the statement "I felt satisfied with the group's soundtrack selection in the headphones condition." Respondents agreed with the corresponding statement about the public condition (mean = 4.06), indicating that participants felt the group reached a consensus about their final selections in both conditions.

On the post-study questionnaire, subjects' free-form responses to, "Please describe how your level of and quality of communication with the other participants differed between the headphones and public speakers conditions" varied. Three respondents indicated that one nice aspect of communication in the public condition was the fact that "everyone is focused on the same task," although another subject wrote that in the private condition "we could imitate a speaker-like effect by all listening to the same music clip or scene caption." Another person wrote, "With the speakers, it was hard to communicate because all of us had to listen to one song at a time and we all had to hear it. With the headphones, one person could listen to a song while the other three talked. There were more combinations of listening and communication possible." However, most participants indicated that communication levels were about the same in each condition.

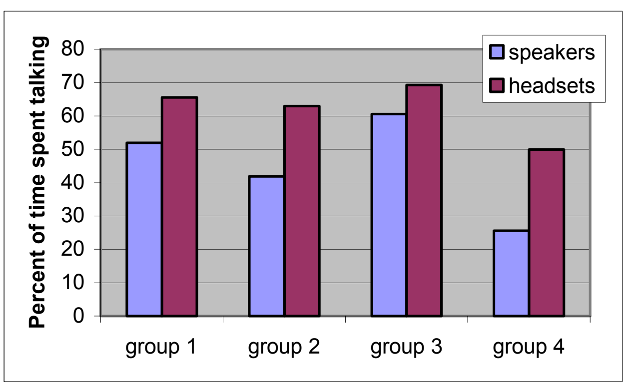

Figure 9. Each group spent more time talking in the private condition than in the public condition.

Our observations during the study and of the additional three pilot groups supported the self-reports that earbuds didn't impede communication. In fact, analysis of the videotapes shows that all groups spent more time talking with each other in the private condition than in the public condition. Figure 9 shows, for each group, the percentage of time that group talked during the private condition and the percentage of time that group talked during the public condition – all groups spoke more with the headsets than with the speakers. The fact that groups in the public condition tended not to speak while a song was playing over the speakers likely accounts for this difference, although it could also reflect the fact that groups in the private condition needed to talk more to accomplish the task because they lacked the shared context present in the public condition.

In addition to the differences in amount of conversation between conditions, the nature and distribution of the conversation also differed. In the public condition, groups typically played one song at a time for several seconds without speaking, and then would turn the song off and discuss its merits for pairing with a particular scene. This produced a pattern of music-talk-music-talk.

In the private condition, however, groups usually were quiet for the first few minutes of the session, as they spent this time exploring a subset of the songs and scenes in parallel. After that initial quiet period, they spoke frequently, with conversation falling into the following categories:

- *Advertising* – users often told the rest of the group about a certain genre of song they had discovered, to see if others might be able to match it to a scene.
- *Querying* – users often asked the rest of the group whether anyone else had found a song fitting certain criteria in order to match a certain scene.
- *Verifying* – users often asked the rest of the group to verify that a song-scene pairing they had created was appropriate. Other group members would then listen to the song and caption and discuss their suitability as a match.
- *Strategizing* – one or more group members would propose a strategy, such as creating piles of songs that matched certain criteria (e.g., a happy pile and a sad pile), or creating piles of verified pairings in order to reduce clutter and avoid repeating work.

4.3 Productivity

Subjects found it significantly easier ($p<.05$) to complete the task in the private condition – the statement "I found it easy to complete the task using headphones" received a mean score of 3.89, while "I found it easy to complete the task using public speakers" received only a 3.19. Subjects' free-form comments on the questionnaire indicated that they perceived the private condition as allowing them to complete the task more efficiently.

The number of times groups "changed their minds" about a particular song-scene pairing by replacing one song with another could be taken as a measure of the quality of their final soundtracks. As mentioned in Section 4.1 (Figure 8), groups replaced assignments significantly more frequently in the private condition than in the public condition, which may indicate that groups were able to put more thought and effort into the soundtracks produced in the private condition.

Pilot studies revealed that groups given unbounded time to complete the task took much longer using public speakers than private earbuds. This is probably a result of the increased efficiency of the divide-and-conquer strategies enabled by private audio, although it could also indicate that groups spent more time discussing and debating each choice in the public condition, perhaps because they were all always focused on the same task. However, because the pilot groups took such a long time with the task, we imposed the twenty-minute time limit, reminding people when five minutes and one minute remained. In the public condition, groups had to hurry a great deal when they got these reminders, whereas in the private condition groups were nearly finished by this time anyway, and usually used the remaining time to thoroughly review their chosen pairings

4.4 Overall Usability

In addition to reporting that it was easier to complete the task in the private condition (as discussed in Section 4.3), subjects also found the private condition slightly more enjoyable than the public condition (p=.085), giving a mean score of 4.0 to the statement "I enjoyed using headphones to complete the task," but only giving a mean of 3.31 to "I enjoyed using public speakers to complete the task." Overall, subjects felt that wearing the earbud was not particularly uncomfortable, as suggested by their disagreement (mean = 2.25) with the statement "I found it uncomfortable to wear headphones."

The questionnaire also asked subjects, "Which session did you prefer? Please comment on the reasoning behind your choice." Ten of the sixteen participants said they preferred the private condition, while six said they preferred the public condition. For those who preferred the private condition, a common justification was greater efficiency due to the ability to work in parallel on parts of the task, and feeling more comfortable exploring songs without worrying about bothering other users. People who preferred the public condition felt it helped the group focus more on common tasks.

4.5 Tabletop Use

Although observing the impact of private audio channels was our primary interest, we also observed interesting patterns in the use of space on the shared tabletop. Clutter on the table was a significant issue – it was not possible to spread all seventeen scene objects and thirty-four song objects across the table without overlapping them. Groups consistently came up with piling strategies to help reduce clutter. Three of the four groups created a "finished" pile, where they put scene-song pairings that they had all agreed were final, both to save space and to prevent wasting time by unnecessarily revisiting those decisions. Two of the groups also created a "rejection" pile of songs that everyone agreed would not be appropriate matches for any of the scenes.

We also analyzed the spatial distribution of object interactions. The table was divided into five equally-sized zones, illustrated in Figure 10. Four of these zones correspond to the "local space" of each user, and a fifth zone represents a center "neutral" area. These zones are purely an analytical construct, and were not reflected in the software's UI. Each time a user manipulated an object on the table, the logged event was tagged as an "own-area" event (if it took place in the user's local area), an "other-area" event (if it took place in another user's local area), or a "neutral-

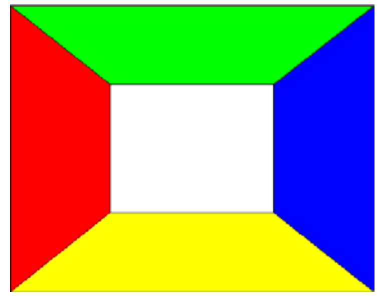

Figure 10. An illustration of the analytical division of the table into "local areas" for each user.

area" event. We found a disproportionately small number of "other-area" events. Only 27.9 percent of song play events, for example, were tagged as "other-area", despite the fact that other users' local areas constitute 60 percent of the tabletop. We do not expect that this was the result of reach length limitations; the DiamondTouch has an 88 cm diagonal, so it is small enough for even petite adults to comfortably reach across the entire table. This tendency to avoid other users' local areas is in keeping with observations of physical table use that suggest that people establish personal territories [17, 24]. There was no significant difference in the spatial distribution of events between the public and private conditions.

5. DISCUSSION

Although further experimentation will be required to draw broad conclusions, our results indicate that individual audio channels can be a useful addition to single display groupware. With private audio, group members participated more equitably in the task, spoke to each other more frequently, and managed the available time more effectively than when individual audio channels were not available. Users also indicated that they found the system enjoyable and easy to use.

A next step is to evaluate the use of individual audio in situations where the audio conveys text, rather than music. It is possible that people are less able to focus on their conversations with other group members when they are listening to speech; however, based on the use of the scene captions in our application, we suspect that occasionally listening to brief text clips that supplement the information on the shared display may not be overly distracting.

5.1 System Improvements

In our original design, the "seek" sliders were not available for song objects – songs always played from their beginning until they completed or were turned off. However, pilot testing revealed that this was frustrating to users who wanted to briefly browse songs. Since the early seconds of a song are often not representative of the overall tempo or mood, the ability to seek is important. This brings up a point applicable for the design of more general interfaces for augmenting shared displays with private audio – because audio information must be reviewed serially and cannot be quickly scanned like visually presented material, providing interfaces that allow users to navigate within (and perhaps even change the playback rate of) their audio stream, whether it is music or speech, is critical.

Although users indicated in the questionnaires that they did not find the earbuds uncomfortable, improvements could still be made that would make the system more appealing for everyday or long-term use. Using wireless headsets would be a large improvement, since it would allow users greater mobility as well as reducing the likelihood of tripping over long wires. There are also less invasive, though more expensive, alternatives to using headsets as a means of delivering the private audio – for example, the Audio Spotlight [12] can "shine" a sound beam at a specific person.

One common suggestion provided by the free-form questionnaire comments was to provide a mechanism for users to "push" the sounds they were hearing to other users' headsets. In our study, if a user wanted others to hear the same thing she was listening to, she would have to ask out loud for others to touch the same song, and would sometimes even give instructions about how far to seek ("Everyone go to ¾ of the way from the beginning.").

6. RELATED WORK

6.1 Visual Privacyware

There are several systems that provide a private source of visual information to users of Single Display Groupware. The three main approaches for adding private visual data are the use of shutter glasses or head-mounted displays, the use of several smaller, auxiliary displays, and physically partitioning the shared space.

Shoemaker and Inkpen [21] use alternating-frame shutter glasses to present disparate views of a shared display to two users – each user sees the same basic information on the display, but each sees only his own cursor, contextual menus, and user-specific task instructions. Agrawala et al. [2] use a similar technique to present two users with stereo views of the same 3-D model from different perspectives based on where each user is standing.

Auxiliary displays can also be used to provide privacy, and are analogous to the personal pads of notepaper that people bring with them to traditional meetings. Greenberg et al.'s SharedNotes [6] lets users make personal notes on PDAs that they can selectively transfer to a large, shared display. Myers et al.'s PebblesDraw [11] allows users to simultaneously operate a shared drawing program from individual PDAs. Rekimoto's Pick-and-Drop technique [13] allows users to "pick up" and "drop" information using a stylus in order to transfer data between a PDA and a large display. The UbiTable [19] uses laptops as auxiliary devices – users keep private information on their laptops, but can wirelessly transfer items they wish to collaboratively discuss onto a shared tabletop display. iROS [7] allows people to use their laptops to post content to large shared displays using its "multibrowse" mechanism.

A third option for visually presenting private information is to physically partition the shared display. The Personal Digital Historian [18] has a central area where commonly referenced digital photos can be displayed and manipulated. The corners of the display, however, are semi-private spaces where an individual user can keep another collection of photos. This is an affordance of standard physical tables as well – papers situated on far sides of the table and oriented toward other users effectively become private [8]. Wu and Balakrishnan [25] take a different approach to physical partitioning; when a user places his hand vertically and slightly tilted on top of a top-projected tabletop display, the system detects this gesture and takes advantage of the top-projection to project "secret" information onto the user's tilted palm.

There are several drawbacks associated with visual privacyware solutions. The use of alternating-frame shutter glasses does not generalize well to more than two users, because presenting private data to n users reduces the effective maximum refresh rate by a factor of $(1/n)$, causing perceptible flickering. Also, this requires users to wear specialized goggles or headmounts, which many people find invasive. Requiring specialized glasses may also reduce eye contact among users and thus reduce group collaboration. The use of auxiliary displays such as PDAs and laptops has drawbacks as well – as Shoemaker and Inkpen point out in [21], these devices do not support the ability to provide information within the context of the shared display, and thus may not be appropriate for certain types of user-specific information, such as cursors or contextual menus, which are only relevant in relation to other items on the main display. Collaboration may also be inhibited by the distraction of each person looking at his individual device. Furthermore, the need to look back and forth between the laptop/PDA and the main screen may create extra cognitive load, reducing overall productivity. Using PDAs to convey private information also requires users to look away from the main display to examine the PDA, thus revealing to other users that they are examining private data.

6.2 Audio Privacyware

Multimodal SDG interfaces that use private audio channels to convey personalized information to group members are relatively unexplored. Magerkurth et al [9] mention that users playing their competitive tabletop computer board game wear headphones to receive secret game-related data, and that informal observations of game play suggested this was well-received by players.

The Jam-O-Drum system [3], an environment for collaborative percussion composition, allows sound to be distributed to individual users via user-directed speakers. The Jam-O-Drum creators mention that they tried having each drummer wear a headset that would play their own drum music more loudly than the drum music of the rest of the group, in order to help them better monitor their performance. Their informal observations suggest that this seemed to reduce communication among group members. Although the Jam-O-Drum's negative experience with using headsets in a groupware environment has discouraged others from pursuing this avenue, we have found from our user experiment that the use of headsets did not impede communication, suggesting that this idea deserves reexamination as a potential interface. There are several differences between our system and Jam-O-Drum that could have lead to this difference in communication levels – two of the most salient differences are that (1) we used an earbud in a single ear to convey audio, while they used headphones that covered both ears, which may have made it more difficult to communicate with other group members, and (2) Jam-O-Drum continuously played audio over all of the headphones, since audio was the focus of their application. In contrast, users in our experiment received audio only on-demand, as a means to help them complete a multimodal task.

The Sotto Voce system [1], a museum exploration system, presents the converse of the system described in this paper; users with private visual displays (PDAs) can passively share audio data.

All of these different types of privacyware – shutter glasses, auxiliary displays, physical space partitioning, and individual audio channels – have unique advantages and disadvantages. Exploring which types of privacyware would be most applicable to different groupware scenarios is an area that warrants further study.

6.3 Audio and Ambient Awareness

One common use of audio in groupware has been to provide people with ambient awareness of other group members' activities. Examples include explorations of using sound to provide ambient awareness in *media spaces* (systems that use media such as video and audio to create a shared "space" for distributed work groups) [22], and using spatialized, non-speech audio to provide awareness of the activities of users working on different segments of a very large display [10]. An avenue for future work would be to compare our current implementation of private audio with an implementation that included mechanisms for ambient awareness of other group members' activities, perhaps by mixing in portions of other users' private sound at a lower volume. Sotto Voce [1] and the Jam-O-Drum [3] used differential volume to distinguish between a user's own sounds and sounds generated by other users, but did not explore how this awareness information altered behavior as compared to only providing a user's own sounds.

7. CONCLUSION

We have introduced a system for augmenting a shared tabletop display with individual audio channels as a means of conveying private or personalized information to individual members of a group. We conducted a user study with a prototype application – SoundTracker – that utilizes private audio channels. Quantitative and qualitative results showed that private audio, as compared to using a single set of standard PC speakers, resulted in changes in groups' task strategies, and did not impede group communication. We are encouraged by these results, and plan to further explore the potential for supplementing single display groupware applications with private audio channels.

8. ACKNOWLEDGEMENTS

We would like to thank Mitsubishi Electric Research Labs for donating a DiamondTouch table and other computer equipment used in this study, and for providing a license to use the DiamondSpin tabletop interface toolkit. We would also like to thank Rachel Weinstein for sewing our colorful seat covers, and Dan Maynes-Aminzade for his projector-mounting expertise. We appreciate the efforts of all the people who participated in our user study and who offered comments on drafts of this paper – the reviewer comments were particularly helpful. We acknowledge the financial support of the NDSEG, AT&T, and NSF fellowship programs.

9. REFERENCES

[1] Aoki, P., Grinter, R., Hurst, A., Szymanski, M., Thornton, J., and Woodruff, A. *Sotto Voce*: Exploring the Interplay of Conversation and Mobile Audio Spaces. *Proceedings of CHI 2002*, 431-438.

[2] Agrawala, M., Beers, A., McDowall, I., Frohlich, B., Bolas, M., and Hanrahan, P. The Two-User Responsive Workbench: Support for Collaboration Through Individual Views of a Shared Space. *Proceedings of SIGGRAPH 1997*, 327-332.

[3] Blaine, T. and Perkis, T. The Jam-O-Drum Interactive Music System: A Study in Interaction Design. *Proceedings of DIS 2000*, 165-173.

[4] Dietz, P. and Leigh, D. DiamondTouch: A Multi-User Touch Technology. *Proceedings of UIST 2001*, 219-226.

[5] Ericson, M. and McKinley, R. The Intelligibility of Multiple Talkers Separated Spatially in Noise. In *Binaural and Spatial Hearing in Real and Virtual Environments*, R.H. Gilkey and T.A. Anderson (eds.), L. Erlbaum Associates, Mahwah, NJ, 1997, 701-724.

[6] Greenberg, S., Boyle, M., and Laberge, J. PDAs and Shared Public Displays: Making Personal Information Public, and Public Information Personal. *Personal Technologies*, 3(1), 1999, 54-64.

[7] Johanson, B., Fox, A., and Winograd, T. The Interactive Workspaces Project: Experience with Ubiquitous Computing Rooms. *IEEE Pervasive Computing* 1:2 (April – June 2002), 67-75.

[8] Kruger, R. and Carpendale, M.S. Orientation and Gesture on Horizontal Displays. *UbiComp 2002 Workshop on Collaboration with Interactive Walls and Tables.*

[9] Magerkurth, C., Memisoglu, M., Stenzel, R., and Streitz, N. Towards the Next Generation of Tabletop Gaming Experiences. *Proc. Graphics Interface 2004*, in press.

[10] Müller-Tomfelde, C. and Steiner, S. Audio-Enhanced Collaboration at an Interactive Electronic Whiteboard. *Proceedings of the 2001 International Conference on Auditory Display*, 267-271.

[11] Myers, B., Stiel, H., Gargiulo, R. Collaboration Using Multiple PDAs Connected to a PC. *Proceedings of CSCW 1998*, 285-294.

[12] Pompei, F.J. The Use of Airborne Ultrasonics for Generating Audible Sound Beams. *Proceedings of the 105th Audio Engineering Society Convention*, San Francisco, CA, 1998.

[13] Rekimoto, J. A Multiple Device Approach for Supporting Whiteboard-Based Interactions. *Proceedings of CHI 1998*, 344-351.

[14] Rogers, Y., Hazlewood, W., Blevis, E., and Lim, Y. Finger Talk: Collaborative Decision-Making Using Talk and Fingertip Interaction Around a Tabletop Display. *Proceedings of CHI 2004 Extended Abstracts*, in press.

[15] Rogers, Y. and Rodden, T. Configuring Spaces and Surfaces to Support Collaborative Interactions. In O'Hara, K., Perry, M., Churchill, E., and Russell, D. (eds.) *Public and Situated Displays*. Kluwer Publishers (2004), 45-79.

[16] Scott, S., Grant, K., and Mandryk, R. System Guidelines for Co-located Collaborative Work on a Tabletop Display. *Proceedings of ECSCW 2003*, 159-178.

[17] Scott, S. Territory-Based Interaction Techniques for Tabletop Collaboration. *UIST 2003 Conference Companion*, 17-20.

[18] Shen, C., Lesh, N., Vernier, F., Forlines, C., and Frost, J. Sharing and Building Digital Group Histories. *Proceedings of CSCW 2002*, 324-333.

[19] Shen, C., Everitt, K., and Ryall, K. UbiTable: Impromptu Face-to-Face Collaboration on Horizontal Interactive Surfaces. *Proceedings of UbiComp 2003*, 281-288.

[20] Shen, C., Vernier, F., Forlines, C., and Ringel, M. DiamondSpin: An Extensible Toolkit for Around-the-Table Interaction. *Proceedings of CHI 2004*, 167-174.

[21] Shoemaker, G. and Inkpen, K. Single Display Privacyware: Augmenting Public Displays with Private Information. *Proceedings of CHI 2001*, 522-529.

[22] Smith, I. and Hudson, S. Low Disturbance Audio for Awareness and Privacy in Media Space Applications. *Proceedings of ACM Multimedia 1995*, 91-97.

[23] Stewart, J., Bederson, B, and Druin, A. Single Display Groupware: A Model for Co-present Collaboration. *Proceedings of CHI 1999*, 286-293.

[24] Tang, J. (1991) Findings from Observational Studies of Collaborative Work. *International Journal of Man-Machine Studies*, 34, 143-160.

[25] Wu, M. and Balakrishnan, R. Multi-Finger and Whole Hand Gestural Interaction Techniques for Multi-User Tabletop Displays. *Proceedings of UIST 2003*, 193-202.