# “I Hear You”: Understanding Awareness Information Exchange in an Audio-only Workspace

**Oussama Metatla**
University of Bristol
Bristol, UK
o.metatla@bristol.ac.uk

**Nick Bryan-Kinns, Tony Stockman**
Queen Mary University of London
London, UK
{n.bryan-kinns, t.stockman}@qmul.ac.uk

## ABSTRACT
Graphical displays are a typical means for conveying awareness information in groupware systems to help users track joint activities, but are not ideal when vision is constrained. Understanding how people maintain awareness through non-visual means is crucial for designing effective alternatives for supporting awareness in such situations. We present a lab study simulating an extreme scenario where 32 pairs of participants use an audio-only tool to edit shared audio menus. Our aim is to characterise collaboration in this audio-only space in order to identify whether and how, by itself, audio can mediate collaboration. Our findings show that the means for audio delivery and choice of working styles in this space influence types and patterns of awareness information exchange. We thus highlight the need to accommodate different working styles when designing audio support for awareness, and extend previous research by identifying types of awareness information to convey in response to group work dynamics.

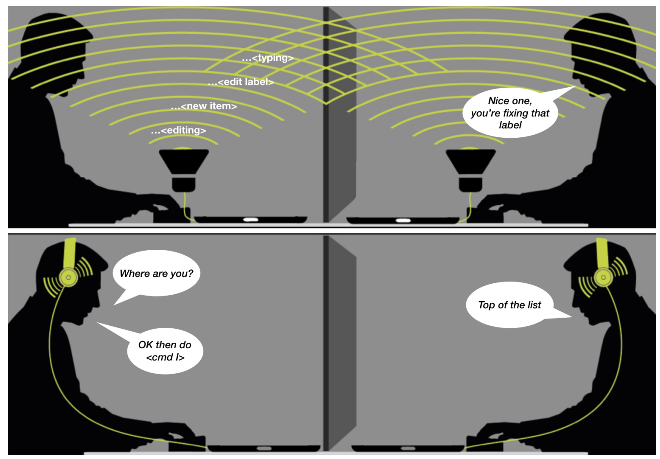


**Figure 1. Experimental setup in speakers and headphones conditions.**

### ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces—*Auditory (non-speech) feedback*; H.5.3 Information Interfaces and Presentation: Group and Organisation Interfaces—*CSCW, Synchronous interaction*

### Author Keywords
Audio-only, Interaction; Workspace Awareness; Collaboration

## INTRODUCTION
The rapid change in the nature of computing technology over the past few decades continue to challenge traditional HCI paradigms. It is now widely acknowledged that new and radical alternative interaction paradigms are needed for the varied interactional contexts that everyday technologies afford today. Most recently, Følstad and Brandtzæg [18] suggested that the latest wave in the field of HCI is the move toward Chatbots and AI-powered natural language interfaces. In effect, it seems that support for natural interaction is at the core of the tensions that arise from the traditional Graphical User Interface and its associated Windows Icons Menus and Pointers (WIMP) paradigm; when computers broke free from the desktop [1] they did so while constraining interactions to one predominant visual modality, which stood in direct contrast to the very nature of the everyday spaces they invaded. Auditory display took a back seat during these developments but then become a ubiquitous element in human-machine systems as a result in part of engineering improvements in audio delivery capability [46]. Audio forms a natural part of human interaction [6, 54], it has been shown to help overcome the lack of screen space [55, 9], and to even outperform visual displays in certain contexts [52, 57]. Yet, our knowledge of its potential as a medium of interaction remains largely confined to supplementing visual displays with alerts and notifications, and to supporting musical interactions or accessibility for people living with visual impairments [34, 36, 42]. But neglecting other modalities in interface design is effectively neglecting other ways in which humans convey meaning, and there is a need for characterising how non-visual modalities can more meaningfully contribute to future interaction paradigms.

In this paper, we thus sought to question the extent to which audio, by itself, can mediate collaborative interaction. We ask, how do people engage with one another in a collaborative task when we strip away - to a certain extent - all other modalities? Collaboration is a good case study for investigating the effectiveness of interactional modalities because the graphical user interface is prevalent in groupware systems. For instance, most support for awareness, which is a crucial aspect of col-



*CHI 2018,* April 21–26, 2018, Montreal, QC, Canada



ISBN 978-1-4503-5620-6/18/04. . . $15.00
DOI: https://doi.org/10.1145/3173574.3174120

laboration [26], relies on visual display techniques (e.g. [3, 43, 35, 27]). Audio in this case has actually been shown to support awareness in groupware systems when combined with graphics and to overcome limitations associated with visual displays [27]. But little is known about the practicality of audio as a sole medium for collaboration and a means for collecting and using awareness information to support joint activities.

Studying an audio-only workspace is an extreme example. But there are emerging technologies that are starting to rely on audio as the sole medium of interaction, e.g. Voice User Interfaces such as Amazon's Echo, and where audio would be advantageous, e.g. eyes-free interaction in-vehicle technologies. An audio-only interaction space is also attractive because of its lower complexity and cost [30]. We envision a future were these types of technologies will be extended to include collaborative interaction, not just across people and technology, but across people and people through technology; e.g collaboration through an Amazon Echo across rooms in a house or cars on the road. Thus, the audio-only context matters to HCI, and if audio by itself was indeed found to be usable and useful as a sole medium of interaction, then it would also be important to characterise the space it created.

We present a study that contributes to the characterisation of such a space. We examine awareness in an audio-only workspace to answer the following questions: first, how do collaborators maintain awareness when working in an audio-only workspace; second, how do the means for delivering audio to such a workspace influence patterns of awareness information exchange; third how do collaborators use audio as a resource to support their interactions. We make the following contributions. First, we provide empirical evidence that audio can be used as a sole means for supporting non-visual collaboration and for maintaining workspace awareness during such collaborations. Second, we identify which type of workspace awareness information is exchanged as a function of the type of audio channel used in the collaboration and the working style employed by collaborators. Finally, we present design recommendations for supporting awareness in non-visual interaction drawing on collaborative patterns mediated by audio.

## BACKGROUND

### Awareness in Collaboration

The study of awareness in collaboration has been diverse, ranging from general awareness about the physical and social environment [8], to co-presence and locus of attention [21, 32], and specific information about real-time joint activities [15, 26]. Dourish and Bellotti [15] described awareness as an *"understanding of the activities of others, which provides a context for your own activity"*[15, p.107]. According to this view, gaining an awareness *of something* in a collaboration involves acts of monitoring both the shared space and co-present individuals, as well as displaying one's own activities to others. Early ethnographic studies of joint activities by Heath and Luff [29] showed that competent collaborators typically adjust the levels of obtrusiveness in their monitoring of co-workers and explicitness in displaying their own actions to match the demands of the current task or state of the collaboration.

In general, research into awareness, whether through technological explorations or ethnographic studies have identified profound design tensions that remain a challenge for both users and designers [25]. A number of frameworks have been suggested to address some of these challenges. For instance, the Big Watch [33] and the Atmosphere [51] frameworks were proposed to support developers when implementing contextual awareness in asynchronous collaboration. Gutwin and Greenberg's framework [26] focused on workspace awareness and described what knowledge constitutes it, which perceptual mechanisms are used to extract such knowledge from a shared workspace and how it benefits collaboration. Their framework bounds the kind of knowledge that a person constructs and makes use of during shared activities to information about collaborators' interactions inside an immediate and synchronously shared workspace. However, the potential of audio as a sole means for collecting and using workspace awareness information has remains largely unexplored.

### Audio and Awareness in Collaboration

Early research on media spaces provided initial evidence of how people use incidental sounds to construct a social space where audio helps in communicating information about ongoing activities across locations [7, 22, 17, 53]. Capturing and integrating auditory cues back in remotely shared spaces was also often found to support awareness and to enrich collaborative experiences [13, 21, 45]. In the Audio Aura system [45] physical actions in the environment were captured and translated into background auditory cues for monitoring purposes, and in ShareMon [13], file-sharing was augmented with auditory icons to notify users of ongoing group activities in an unobtrusive manner. In the Thunderwire system [30], auditory display was shown to provide a usable and sociable space for interaction even in the complete absence of other modalities.

#### *Grounding*

Audio is an integral part of building common ground, which is the set of mutual knowledge, beliefs and mutual assumptions shared by two or more parties [11]. For instance, speakers monitor addressees for understanding and, when necessary, alter their utterances in progress [12] and infer shared comprehension by observing changes in behaviour or to shared objects [24]. Studies have also showed that when people engage in collaborative problem-solving over video communication, they structure their tasks differently depending on whether they could see and/or hear each other [14]. This is significant given that studies have also shown that people tend to switch off video channels and maintain audio channels whenever the quality of online communication deteriorates [48, 37].

#### *Impact of Audio Channels*

A natural question that arises with regards to using audio in collaboration is whether the means for delivering audio has any impact on collaborative work. Hancock *et al* [28] manipulated a speaker setup around a tabletop display, assigning a different timbre to each user. Their results showed that adding audio feedback increases group awareness but at the cost of decreasing individual performances. In evaluating an AudioCave system, Ramloll and Brewster [50] setup audio output such that collaborators wore headphones and heard each other's speech as a localised output in one condition, and as coming from a fixed point at the centre of the workspace in another. Their results showed that participants used the spatial position of their peers' speech to maintain awareness of their interactions with the shared space. However, these studies report little details about which awareness information was extracted from the localised speech and how this information was then used in the collaborations.

Morris *et al* [44] compared the use of headphones and shared speakers to deliver auditory feedback to a group of tabletop users. Their system allowed users to browse and manipulate movie scenes and music files to create soundtracks. They found that group task strategies changed when users wore headphones, and that using headphones does not impede group communication. These results contrast with those reported by Blaine and Perkis [4], who compared three different audio setups for delivering audio to a group of users collaborating to create music on a shared tabletop. Blaine and Perkis reported that using a shared speaker to display a mix of all collaborators' audio made it difficult for users to identify their own sounds, while using a headphones-mix showed improvements, but only for users with musical experience and at the cost of impeding group communication. Thus, both studies acknowledge the potential impact of using speakers and headphones on public and individual awareness of actions, and consequently on the dynamics of the collaboration, but neither explore the details of such impact on workspace awareness information exchange.

### Working Styles & Audio in Collaboration

In describing joint work, Gaver [21] distinguished between focused collaboration and divided labour and asserted the potential of using auditory cues to communicate serendipitous information that increases awareness of activities and events. Similarly, Dourish and Bellotti [15] pointed out that designing awareness mechanisms in collaborative systems should allow users to move smoothly between close and loose collaboration. This potential was demonstrated in the ARKola simulation [23], in which pairs of users manipulated a shared workstation displaying a number of machines in a virtual factory plant for handling the production of a virtual soft drink. The results showed that confirmation sounds provided an awareness of partners' actions that was absent from the graphics-only condition.

Huang *et al.* [31] examined the combined effect of auditory and haptic feedback on collaboration and found that they impact grounding strategies and processes. McGookin and Brewster [38] examined users collaborating to create graphs using an audio-haptic tool. They reported that their participants used different strategies for monitoring their partners' activities depending on the working style they employed. To reconcile the reported advantages and disadvantages of sharing audio, McGookin and Brewster suggested that *"the amount and type of shared audio should be altered dependent on the strategy that participants adopt"* (p.2577). Exactly which type of awareness information and how much of it should be altered remains an open question, however. Their study was also of a small scale (4 pairs) and more thorough studies are needed to confirm these initial findings.

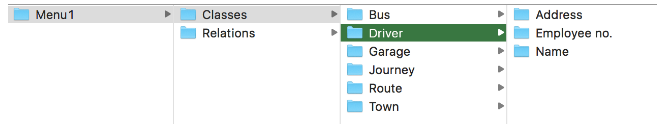


**Figure 2. Example of a menu used in the study. Participants could hear but not see the menu as they collaboratively accessed and edited it.**

## STUDY

To contribute to addressing the gaps highlighted above, we ran a study to examine workspace awareness in an audio-only collaboration, focusing on the impact of using private (headphones) and public (speakers) audio channels and varying working styles on workspace awareness information exchange. We chose the basic task of accessing and editing shared menus. Hierarchical menus continue to be a common form of organising content across a variety of application areas, including familiar interfaces such as Interactive Voice Response systems. We asked pairs of participants to use an audio-only tool to access and edit random menu structures in a co-located setting where they could hear but not see each other, and we hypothesised that: H1: Concealing audio output through headphones in the audio-only workspace will increase verbal exchange of workspace awareness information; and that H2: The choice of working style employed by participants will influence the type of workspace awareness information exchanged. Participants collaborated under two conditions in a within-groups experimental design; Condition 1: the audio output of each participant's interactions with the tool was delivered through speakers, so their partner could hear what they were doing; Condition 2: audio was delivered through headphones, concealing it from the workspace and making it only accessible to the participant who produced it.

### Overview of the Audio-only Tool and Workspace

Our experimental tool allows a user to navigate a menu structure using the keyboard cursor keys and to add or remove menu items, or edit their labels using keyboard commands. The design of the tool was based on Metatla *et al*'s approach to the sonification of hierarchical menu structures [39, 40]. More specifically, the tool functioned in two modes; inspecting and editing a menu. We used speech to read the labels of menu items, and a mixture of earcons [5] and auditory icons [20] to display user's inspection and editing actions. We further used two different continuous sounds (wind sound and water bubbling sounds) and a male and a female voice to differentiate between each mode of interaction. Table 1 shows examples of the audio display of our tool. To support collaboration, the tool provides two users with independent access to a shared audio menu. Each user can access or edit any item on the menu so long as it is not currently being edited by their partner; the tool displays an audio alert when a user attempts to edit an item that is currently being edited by their partner. Each user thus triggers a unique set of audio output corresponding to their inspection and editing actions on the menu. We used Java FreeTTS to implement speech output and JSyn audio synthesis API to implement the earcons, auditory icons.

As shown in Figure 1, participants sat facing each other, they had a set of computer speakers or single-ear headphones to use

**Table 1. Example actions on the audio-only menu tool**

| User action | Non-speech and Speech audio | Mode |
|---|---|---|
| Navigation: <left cursor> | (Navigation earcon) + "[reads item label]" Female voice | (Wind) (Continuous sound) |
| Editing: <cmd S> | (Selection earcon) + "[item] Selected" Male voice | (Bubble begins) (Continuous sound) |
| Deleting: <cmd D> | (Deletion auditory icon) + "[item] Deleted" Male voice | (Bubble stops) (Continuous sound) |

in each experimental condition. There was no visual display of the menus in both conditions. The keyboard, speakers and headphones were connected to two computers, one for each participant with one of those acting as a server and linking the pair to a shared menu. The speakers were positioned in front of each participant in the first condition and displayed the audio output associated with their actions with the volume adjusted so that their partner could also clearly hear it. An opaque board was placed between the two participants to eliminate any form of visual communication (body language, facial expressions, etc.). Participants could hear each other's audio output (when using speakers) as well as converse comfortably in both conditions.

### Procedure

Participants were first trained on how to use the tool to access and edit menus until they were familiar with the various commands and corresponding audio output. They were then presented with a simple menu and a written description – similar to those used in the testing part – and were given time to use the tool while being closely assisted by the experimenter until they felt comfortable with both the tool and the task. The training part lasted for up to thirty minutes. They were then asked to perform the experimental task. The pairs collaboratively accessed and edited two menus, one at a time, under each of the two experimental conditions, the order of the conditions was randomised. In the testing part, an initial menu was loaded on the tool (e.g. a portion of the menu shown in Figure 2) at the start of the task and gave participants two written descriptions explaining how the menu should be populated. The descriptions contained complementary information; for instance, one participant might have had information about the name of a menu item at one level of the menu hierarchy, while their partner had information about its content at lower levels. This was done to ensure participants talk to each other during the task. Participants were instructed to consult the descriptions and to complete the menus using the audio-only tool. They were informed that they had complementary information and therefore needed to consult with one another. They were given no time limit. Initial menus contained eight items, final menus contained 24 items across four hierarchical levels.

### Participants

We recruited 32 participants (24 males) to make up 16 pairs. Participants were students at the authors' institution and received a cash incentive for their participation. All participants indicated prior experience with audio menu, mainly from interacting with Interactive Voice Response systems.

### Data Gathering, Measures & Coding Scheme

We video recorded the tests, logged and timestamped all interactions with the tool, and conducted informal interviews with the participants at the end of each session to gather personal reflections on the collaborative experience. We transcribed and coded the videos to capture the content and patterns of workspace awareness information exchange between participants during the collaborations. We developed a coding scheme based on Gutwin and Greenberg's framework for workspace awareness (WA) [26] as outlined in Table 2.

We focused the coding scheme on *Actions*, *Intentions* and *Locations*. In relation to *Actions* and *Intentions*, the coding scheme captured indicators of past, current and future activities as well as indicators of action completion statuses. That is, all explicit references made by participants where they supplied or requested information from their partner about what actions they had undertaken, what actions they are currently performing, what actions they plan to perform, and when they completed an action. In relation to *Location*, the coding scheme captured supplied and requested information about position on the menus. We thus further divided the coding of WA information exchange into *Supplied* and *Requested* types. The supplied type refers to information provided by a participant to their partner without the latter having asked for it, and the requested type refers to instances where a participant explicitly asks their partner for information regarding their actions, intentions or location. We asked two independent coders to use the developed scheme to identify and label WA information exchange on video transcripts from two pairs and performed an inter-rater reliability analysis using the Kappa statistic to determine reliability among coders. This revealed high levels of reliability, Kappa = 0.93 ($p<0.01$).

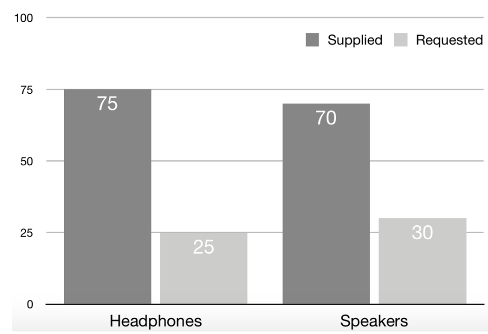


**Figure 3. Proportions (%) of supplied vs. requested types of WA exchange when using speakers and headphones.**

## RESULTS

All pairs successfully completed the experimental task under each condition. Data from one pair was partly lost due to a system failure and was therefore excluded from the reported results. We first report on the overall rates of WA information exchange before examining how this was affected by working styles. We used Bonferroni corrections to account for family-wise error rates for p-testing conducted between-and-across combinations, based on the number of comparisons we had, we consider significance levels at $p<0.0127$ in the following. A related-sample Wilcoxon Sign Ranks test confirmed that overall participants exchanged significantly more WA information when using headphones ($W=25.5$, $p=0.005$). This result supports hypothesis H1. As shown in Figure 3, participants supplied significantly more WA information to each other

Table 2. **The refined version of Gutwin and Greenberg's workspace awareness elements and indicators used in the developed coding scheme.**

| Type | Element | Description |
|---|---|---|
| **Supplied** | What I Did | A participant describes the editing actions that they have just completed without explicit request from their partner; this information is supplied after the action is completed. *E.g."ok, I added three items, gene tag and article"*. |
| | What I'm Doing | A participant describes the actions that they are currently performing without explicit request from their partner; information is supplied after the action is initiated and while it is still being performed. *E.g."I'm selecting that last item to delete it."* |
| | What I will Do | A participant describes what action they are about to perform without explicit request from their partner; this information is supplied before the action is initiated. *E.g."ok I'm gonna go and select Bus"*. |
| | Where I Am | A participant describes their current position on the menu without explicit request from their partner. *E.g."So I'm on the Therapy item.* |
| | Completion Status | A participant informs their partner that an action has been completed without explicit request from their partner and without explicit reference to which action has been completed. *E.g."OK, I'm done."* |
| **Requested** | What Did You Do? | A participant requests information from their partner about what actions they have completed. *E.g."Which item did you delete?"* |
| | What Are You Doing? | A participant requests information from their partner about what action they are currently performing. *E.g."are you adding the child items?"* |
| | What Will You Do? | A participant requests information from their partner about what action they intend to perform. *E.g."which bit are you going to do next?"* |
| | Where Are You? | A participant requests information about where their partner is on the menu. *E.g."which item are you on?"*. |
| | Completion Status | A participant requests information from their partner about whether they have finished performing an action but without explicit reference to a particular editing action. *E.g."are you done?"* |

than they requested from one another in both conditions (70% supplied vs. 30% requested when using speakerss; W=79.5, p=0.005, and 75% supplied vs. 25% requested when using headphones; W=5, p=0.005).

A separate comparison of the supplied and requested types of WA information across the two conditions revealed that the supplied type of exchanges were significantly higher when participants used headphones (W=23, p=0.005). In particular, two types of information were exchanged at significantly higher rates when audio was delivered through headphones; *"What I Did"* (W=25.5, p=0.005) and *"Supplied Completion Status"* (W=15, p=0.005); and one type at marginal significance *"What I Am Doing"* (W=81, p=0.025) (Figure 4).

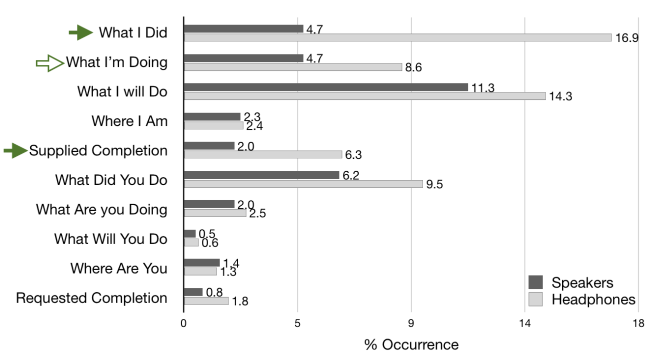


**Figure 4. WA information exchange when using speakers and headphones. Arrows = statistical significance at p=0.005 and p=0.025.**

## Patterns of Interaction & Working Styles

We analysed participants' interaction patterns in order to extract and examine the impact of working styles on WA information exchange. We used the timestamped interaction logs to plot participants activities, an example of which is shown in Figure 5. The plots show different working styles for pairs 11 and 15, particularly noticeable in the way pairs organised turn taking when audio was delivered through speakers; participants in pair 11 interacted with the tool in a sequential manner, where only one participant interacted with the shared menu at any given moment; On the other hand, pair 15's collaborative interaction shows a different pattern. Even though participant P15A issued the first three editing commands between minute 2:00 and 2:30, their partner P15B joined in at minute 2:45 of the collaboration and both participants simultaneously executed editing actions 5, 6, 7, 8, 9 and 10. The pair continued working in parallel throughout the extract, both when inspecting and editing the menu. The interactions patterns of pairs 11 and 15's when they used headphones do not manifest the same evident difference in the way collaborative work was organised. When audio was delivered through headphones, some of the pairs' editing actions occurred sequentially while others were executed in parallel, but almost all editing actions occurred when one participant was inspecting the menu while their partner was issuing editing commands.

We applied a similar analysis of interaction to the remaining pairs to determine their dominant working style. We found that the majority of pairs worked either sequentially or in parallel under each condition, though some pairs did not exclusively employ a single style of interaction per condition. For pairs who employed both styles, one style was used more dominantly than the other, where in most cases the less dominant style was used either when editing particular items of the menus or when inspecting rather than editing the menus. Most pairs used a parallel working style as a dominant style when they used headphones, but worked sequentially and/or in parallel when using speakers. Indeed, the proportion of overlapping interaction times – as extracted from the interaction logs – were significantly higher in the headphone condition (42.56% vs. 26.22%; t=2.841, p=0.013), but this result changed when pairs were grouped on the basis of their dominant working style in the speakers condition. For pairs classified as *Sequential*, the proportion of overlapping interaction was significantly higher in when they used headphones (43.29% vs. 8.03%; t=8.219 at p<0.001). For pairs classified as *Parallel*, the difference of overlapping interaction times between the two conditions was not statistically significant

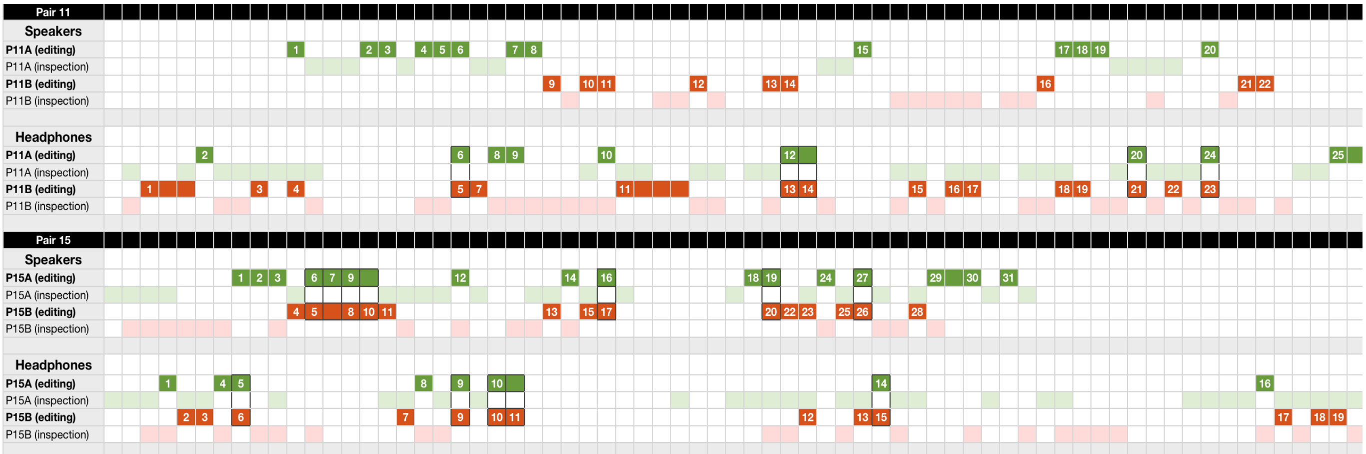


**Figure 5. Example plots of interaction patterns for the first 17 minutes of pairs 11 and 15. The horizontal axis represents time with 15 seconds intervals. Lighter coloured boxes represent menu inspection actions, darker ones represent editing actions. Unnumbered dark boxes are continuations of the editing action that preceded them (i.e. editing actions that took longer than 15 seconds to perform). The numbers give the order in which editing actions occurred. For each pair, the top part of the plot depicts patterns when using speakers and the bottom part when using headphones.**

(42.22% vs. 42.11%; t=0.2, p=0.98). We took these significant differences between overlapping interaction times as an indication that our grouping of sequential v. parallel pairs based on dominant working styles when using speakers was adequate. The observed difference in working styles and the emergence of two distinct groups based on interaction patterns then allowed to address hypothesis H2 through a detailed examination of how patterns of WA information exchange was affected by working styles.

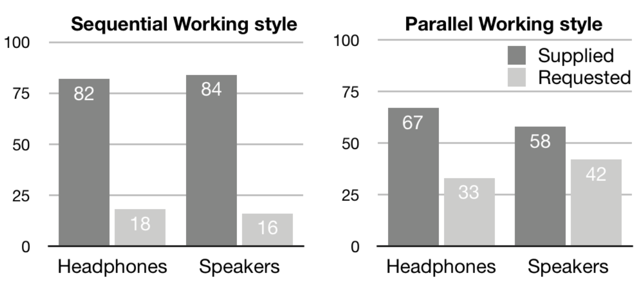


**Figure 6. Proportions (%) of Parallel and Sequential pairs' supplied vs. requested types of WA information exchanges when using speakers and headphones.**

## Results for Parallel Pairs

A total of eight pairs were classified as *Parallel*. They exchanged more WA information when they used headphones at marginal significance (W=27.5 for N=16, p=0.025). As shown in Figure 6, they supplied more WA information to each other than they requested from one another in both conditions, but the difference was only statistically significant when they used headphones (58% supplied vs. 42% requested with speakers, 67% supplied vs. 33% requested with headphones; W=5 for N=15, p=0.005). In particular, two out of five elements were exchanged at significantly higher rate: *"What I Did"*(W=12.5 for N=14, p=0.005), and *"Supplied Completion Status"* (W=7 for N=10, p=0.005).

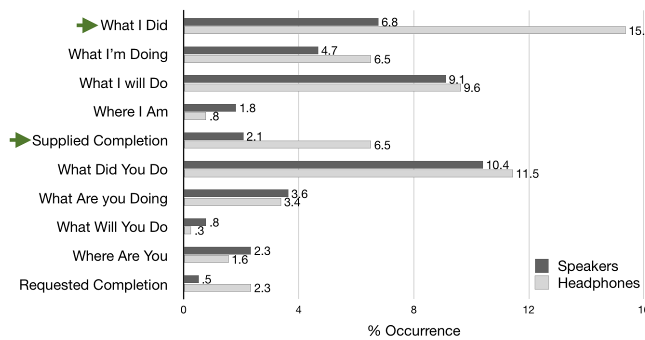


**Figure 7. Parallel pairs' WA information exchanges when using speakers and headphones. Arrows = statistical significance at p=0.005**

## Results for Sequential Pairs

Seven pairs were classified as *Sequential*. They also exchanged significantly more WA information when they used headphones (W=3 for N=14, p=0.005). As shown in Figure 6, they supplied more WA information to each other than they requested from one another in both conditions (84% vs 16% with speakers, and 82% vs. 18% with headphones). In particular, sequential pairs supplied significantly more WA information of type *"What I Did"* (W=4 for N=14, p=0.005) and *"Supplied Completion Status"* (W=2 for N=10, p=0.005) and of type *"What I Am Doing"* (W=17 for N=13, p=0.025) at marginal significance. They also requested significantly more WA information of type *"What Did You Do"* (W=2.5 for N=12, p=0.005) when they used headphones (Figure 8).

## Parallel vs. Sequential Pairs

A Mann-Whitney test showed no significant difference in the overall amount of WA information exchanged between parallel and between sequential pairs when they used speakers. However, when comparing the supplied and requested types of exchange separately, parallel pairs requested significantly more WA information of type *"What Did You Do"* than the sequential pairs (U=58.5, p=0.01). When using headphones,

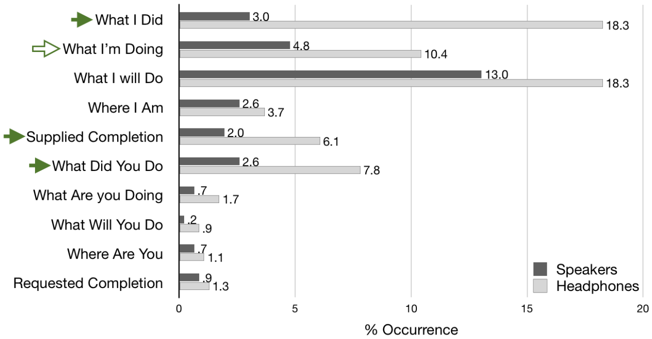


**Figure 8. Sequential pairs' WA information exchanges when using speakers and headphones. Arrows = statistical significance at p=0.005 and p=0.025.**

sequential pairs exchanged significantly more WA information than parallel pairs (U=50, p=0.01). This difference was significant for the supplied type (U=38, p=0.01) but not the for requested type (U=102.5). In particular, sequential pairs supplied significantly more WA information of type *"What I Will Do"* (U=40.5, p=0.01). These results support H2.

## QUALITATIVE ANALYSIS: HOW AUDIO MEDIATED COLLABORATIVE INTERACTIONS

In the following we highlight examples of how WA information was extracted and used during the collaborations that we observed across all pairs.

### Extracting information about actions

When a participant interacted with the collaborative tool, they were either inspecting or editing the menus. Two auditory display techniques were used to distinguish between the inspection and editing modes. The first was the timbre of an ambient sound that was continuously displayed while a user was in a given mode of interaction; the second was the gender of the speech display that alternates between a male voice for the inspection mode and a female voice for the editing mode. To know about current actions, a participant would listen out for either one or both of these audio clues and to the details of the spoken information. Different information is extracted from such output depending on how much attention a participant was paying to the audio output, or on how much an audio output grabs the attention of a participant. This

**Table 3. Extracting detailed information about partner's actions from their audio output source – attentive listening, listening-in-search. <user action>; *(non-speech sound)*; *"system speech output"*.**

| Time | P6A | P6A's Audio | P6B | P6B's Audio |
|---|---|---|---|---|
| 05:10 | | | <down> | *"capacapacity"* |
| 05:12 | | | <delete> | |
| 05:13 | | | | *"capacapacity removed"* |
| 05:18 | | | <new item.> | *(editing)"new item* |
| 05:19 | | | | *(editing) enter label"* |
| 05:20 | | | <typing> | |
| 05:24 | | | | *(success)"capacity* |
| 05:26 | **nice one, you** | | | *added to item bus"* |
| 05:27 | **fixed that label** | | | |

behaviour could be described by Truax's three types of listening attention to account for various listening experiences [54]. An active level of listening, *listening-in-search*, involves a conscious search in the environment for significant cues; an intermediate level of listening, *listening-in-readiness*, where *"the attention is in readiness to receive significant information, but where the focus of one's attention is probably directed elsewhere"*(p.22); and *background listening*, where the occurrence of a sound bears no special or immediate significant to the listener. In the extract shown in Table 3, pair 6 are working using speakers, participant P6B rectified a spelling mistake by removing the item "Capacapacity" and replacing it with a correctly spelled one. During this time, their partner P6A partner remained interactively idle while P6B completed this editing action. Here, P6A seems to have switched to a listening-in-search mode and, as soon as the action is completed, acknowledges it by saying: *"Nice one, you deleted Capacapacity"*. The presence of the audio output through speakers thus communicates partner's activity information at two levels: 1) information about whether a partner is active, and 2) information about what activity the partner is engaged in, with each level requiring a different level of attention, and thus a different type of listening. This information was more difficult to obtain when a participant had access only to their audio output through headphones and should therefore be considered as a potential WA information to convey under such conditions.

### Extracting Information about Locations

The tool provided two auditory display techniques that conveyed information about a user's position on a given menu. The first was the continuous sound that distinguishes between interaction modes, the second was the mixture of speech and non-speech sounds that display the current item under focus. Detecting and recognising the timbre of the continuous sounds (wind vs. water bubbles) could potentially help a user infer which part of the menu their partner was active on, whereas listening to the details of the displayed speech and non-speech sounds enabled them to establish details about their exact position by matching this to their own mental model of the menu structure.

**Table 4. Determining partner's location on the menu.**

| Time | P13A | P13A's Audio | P13B | P13B's Audio |
|---|---|---|---|---|
| 04:43 | | | <down> | *(down)"design"* |
| 04:44 | so a Tag can be used | | | |
| 04:45 | by many Doctors | | | |
| 04:46 | for many Therapies | | | |
| 04:47 | | | <open> | *(open)"three"* |
| 04:48 | | | <down> | ***(down)"therapy"*** |
| 04:50 | **Therapy** | | erm | |
| 04:51 | **OK stay there** | | yeah | |

Table 4 shows an example of using audio to determine partner's position on the menu. In this extract, pair 13 are working sequentially and using speakers. The pair are in the process of discussing the suitable details that should be added onto an item. Participant P13B inspects the menu while participant P13A is idle and reasons about the content to be added. When her partner browses the item in question reaching one of its child items "Therapy" at minute 4:48, P13A immediately asks her partner to remain in that position: *"Therapy, ok stay there"*. P13A in this case has correctly inferred that the position of her partner is the correct one to be at in order to execute the editing action in question. This example could be understood in terms of what has been elsewhere referred to as *Artefact Feedthroughs* [2, 26], where the manipulations of an artefact communicates information about what actions incurred changes to its state, and *Outlouds* [29], where verbal communication is intentionally and explicitly exchanged to support this process. We observed instances where, in order to compensate for the lack of shared audio in the headphones condition, participants "followed" or "tracked" their partners by repeatedly inspecting areas of the menu structure they knew were being edited in order to listen out for incurred changes as and when they appeared.

### Detecting Errors

Participants made a variety of errors when using the collaborative tool and the auditory display of such errors enabled their partners to detect them. Such awareness typically led to providing assistance and explanations, suggesting solutions, or overtaking the executing of the task that is posing difficulty to a partner. We observed these patterns in pairs 2, 4, 5, 6, 8, 9, 11, 12, and 16. In the extract on Table 5 for instance, participant P8B was able to identify that her partner is encountering a problem after hearing the error sound displayed twice in a row, this prompted her to enquire: *"What's the problem?"*. Her partner had encountered an *interaction error* having issued an invalid inspection command. Other types of errors that participants were able to extract from the audio output of their partners' interactions included *procedural errors*, in which a participant mixes the order of the steps required to execute a command, and *content errors* in which they issued the correct command in the correct procedural order, but failed to input the correct content on the menu. In all cases, participants in our study were able to provide timely assistance while relying on audio as the only resource for doing so.

**Table 5. Detecting an interaction error through the error sound.**

| Time | P8A | P8A's Audio | P8B | P8B's Audio |
|---|---|---|---|---|
| 17:21 | <down> | *(down)"Associate"* | | |
| 17:22 | <open> | *(open)"two"* | | |
| 17:23 | <open> | ***(error)*** | | |
| 17:24 | <open> | ***(error)*** | | |
| 17:25 | | | **what's the problem?** | |

### Guiding Partners

Participants also used the WA information extracted from the auditory display to provide their partners with detailed guidance. This included guiding their movements around the menus, as well as through the steps required for executing editing actions. We observed these patterns in pairs 1, 2, 6, 7, 8, 9, 11, 12, and 13. The extract in Table 6 illustrates an example of this. Participant P8B was assigned the task of inserting the item "Number" under the parent item "Article" and is guided by P8A to complete this task. At every step of the interaction, P8A listens attentively to his partner's audio output to obtain information about her current location, and instructs her movements accordingly *"go inside", "come down"*. When he realises that she had reached the appropriate position on the menu, he instructs her to issue the corresponding editing command for setting the item question *"erm cmd I"*. Notice how the participant executing the editing action provided her partner with little to no explicit verbal clues about her actions or locations. This is is an indication that she was expecting him to be attentively listening to her interaction, and so did not feel the need to supply him with any WA information. Audio in this case was an integral part of building common ground.

**Table 6. Guiding partner's movements when using speakers.**

| Time | P8A | P8A's Audio | P8B | P8B's Audio |
|---|---|---|---|---|
| 34:55 | | | | *"article open"* |
| 34:56 | **go inside** | | ok | |
| 34:57 | | | <down> | *(down)"attributes"* |
| 34:59 | **come down** | | | |
| 35:00 | | | <open> | *(open)"one"* |
| 35:01 | | | <down> | *(down)"number"* |
| 35:02 | **yeah control erm** | | | |
| 35:03 | **cmd I** | | | |
| 35:04 | | | hmm <Insert Item> | *(success)"attribute number inserted"* |
| 35:05 | | | | |

There were also instances where guidance occurred between pairs working in parallel when using speakers, and between pairs working when using headphones, but these had noticeable differences. Parallel pairs working with speakers typically switched to the sequential working style during guidance interaction, and expectedly, guiding one's partner when using headphones involved extensive exchange of WA information (both supplied and requested), particularly of the types *"Where I Am/ Where Are You"* and *"What I Have Done/ What Have you Done"*. Often, information that was provided in the form of feedthroughs when using speakers was replaced by explicit requests and outlouds when using headphones. Table 7 shows an extract where a participant guides their partner when using headphones.

**Table 7. Guiding partner when using headphones.**

| Time | P11A | P11A's Audio | P11B | P11B's Audio |
|---|---|---|---|---|
| 21:16 | | | let's start again | |
| 21:17 | go into Entities | | <switch> | *(switch)"entities"* |
| 21:18 | | | **hmm** | |
| 21:19 | Therapy | | **yeah i'm** | |
| 21:21 | and do.. | | | |
| 21:22 | ..cmd.. | | <down> | *(down)"therapy"* |
| 21:23 | ..S | | | |
| 21:24 | | | **ok i'm in therapy** | |
| 21:25 | do cmd S | | | |
| 21:26 | to select it | | | |
| 21:27 | | | **cmd and S** <select> | *(success)"therapy selected"* |
| 21:28 | **does it say** | | | |
| 21:29 | **it's selected?** | | it says | |
| 21:30 | | | **therapy selected yes** | |
| 21:31 | then go to Gene | | | |
| ... | | | | |
| 21:37 | | | <down> | *(end of list)"gene"* |
| 21:38 | | | **now i'm in gene** | |

### Coordinating Collaborative Actions

Extracting information about partners' actions and locations from the auditory display allowed pairs to coordinate their collaborations in a variety of ways; information about the content to be edited could be exchanged swiftly, when needed, and unprompted; editing episodes could be organised fluidly; and interdependent editing actions could be coordinated efficiently. We observed these patterns in pairs 1, 2, 4, 5 (both when using speakers and headphones based on keyboard typing sounds), 6, 7, 8, 9, 10, 11, 12, 13, and 16. Table 8 shows an extract where one participant was editing content based on the information described on their partner's text. The partner, unprompted, supplies the content information “Tissue Type” at the exact moment when it is needed.

**Table 8. Supplying content during an editing action.**

| Time | P8A | P8A’s Audio | P8B | P8B’s Audio |
|---|---|---|---|---|
| 05:54 | I add one more | | | |
| 05:55 | attribute right? | | | |
| 05:56 | | | add one more attribute | |
| 05:57 | <new item> | *(editing)“new item”* | | |
| 05:58 | | *(editing)“enter label”* | | |
| 05:59 | | | **Tissue Type** | |
| 06:00 | <typing> | | | |

The availability of the audio in the workspace allowed the two participants to have joint attention over the progress of an editing task, which would otherwise have been exclusively accessible to one participant. WA information about actions and locations in this case are not explicitly requested nor supplied, but extracted and used efficiently in the collaboration. This is a manifestation of an ability to move between focussed individual work and joint work [21, 15] and another instance in which audio was instrumental in helping partner build common ground. Furthermore, discussions about new content and sometimes labour typically marked the transition from one editing episode to the next. Effectively extracting information about partners’ actions and locations facilitated such transitions. An example of this is illustrated in Tables 9. Participant P9B detects that the process of deleting the item “Employee ID” from the parent “Driver” has been completed at minute 46:47, and that it was time to move on to the next piece of content; the “Journey” item. In this example the pair moved fluidly between editing episodes without explicitly requesting or supplying information about progress or completion status; they were able to detect the completion of the editing actions through a feedthrough that otherwise would have been supplied or requested as a *“Completion Status”* update.

**Table 9. Organising editing episodes.**

| Time | P9A | P9A’s Audio | P9B | P9B’s Audio |
|---|---|---|---|---|
| 46:42 | <down> | *(down)“employee ID”* | | |
| 46:45 | <cmd D> | *(success)“employee ID* | | |
| 46:46 | | *deleted”* | | |
| 46:47 | | | **ok cool, I hear you, erm** | |
| 46:48 | | | **so that’s Driver** | |
| 46:49 | | | should we move go to | |
| 46:50 | yup, Journey | | the | |

## DISCUSSION

Previous research highlighted the important role audio plays in supporting awareness (e.g. [21, 27, 28]). However, exactly how audio contributed to supporting collaborations in such instances is not always clear. In particular, it is not clear which type of workspace awareness information is relevant in a non-visual collaboration, how collaborators maintain awareness in a non-visual workspace, and how this is influenced by working strategies. Insights about these questions are important for the effective design of awareness displays in collaborative systems [26, 27]. Our study attempted to address these questions by examining the extent to which audio can be a practical modality when it is used as the sole means for mediating collaboration.

Previous research also showed that audio channel setups have an impact on collaboration, e.g. in terms of performance [28] and awareness [50, 4]. The way audio was delivered to the workspace in our study allowed us to go a step further by examining the impact of concealing or exposing partners’ activities through headphone and speaker display in rich quantitative and qualitative descriptions. As we anticipated, the proportions of WA information verbally supplied was significantly higher when participants used headphones. But not all types of WA information were supplied at an equal rate. Three elements (*“What I Did”*, *“What I Am Doing”* and *“Supplied Completion Status”*) were exchanged at a significantly higher rate when audio was delivered through headphones. We could thus argue that exposing audio in the workspace through speakers afforded the deliverance of these three types of information. That is, sharing audio through speakers provided participants with the ability to obtain information about their partner’s past and current actions, and about the progress of those actions without them having to explicitly ask for it. However, it is important to highlight that how this WA information is used varied depending on working strategies and was not merely a function of the audio setup (e.g. [44]). For instance, similar behaviour was observed when audio was exposed but collaborators employed a parallel working strategy. The loose character of parallel pairs’ collaborations often meant that participants felt a greater need to find out about each other’s past actions and frequently supplied each other with information in the form of updates about what has happened. On the other hand, sequential pairs’ collaborations were focused, and information was often supplied in the form of descriptions about what was currently happening or what was about to happen in the immediate future.

In general, the coding scheme that we used captured instances in the collaborations where participants explicitly exchanged information pertaining to workspace awareness, and this provided a means for establishing which elements of WA information were used during the collaborations. Addressing the awareness problem in groupware design is an important and difficult problem and part of the solution is to provide collaborators with more information about what is going on in the shared space [26]. But designers must carefully determine which information should be conveyed. Being able to tease out which WA information is relevant is thus crucial when designing awareness support in non-visual environments since overloading users with too much information can be detrimental when users engage in activities that compete for attention and memory resources [34, 49]. This is also particularly crucial when audio is used to support collaboration because sound can be annoying when it does not convey useful or relevant information [10]. Indeed, annoyance and interference are two issues that often stand in the way of exploiting the auditory modality in HCI [19], and “relevance” is often a determining factor between a good and a bad representation of information [56, 47]. Our results have thus confirmed previous findings and extended them in a number of ways; first, we provided empirical evidence that audio, on its own, can efficiently mediate collaborative interaction with audio menus and supported pairs or participants in building and maintaining common ground. Second, we showed that the reported impact of audio channels occurs in an audio-only environment and not only in environments where audio is a supplementary modality. Finally, we established which WA information was relevant and how it is influenced not only by the audio channel setup but also by collaborative strategies.

### Design Recommendations

We compiled the following design recommendations on the basis of the analysis and discussion of our findings:

*Types of WA information when concealing audio:* If headphones conceal partners' audio during non-visual collaboration, the collaborative system should be designed to convey and/or allow users to request WA information of types *"What I Did"/"What Did You Do"*, *"What I Am Doing"/"What Are You Doing"* and actions' *"Completion Status"*.

*Adapt both the type and amount of WA information to match working styles:* When a non-visual collaborative interaction occurs through a parallel working style, the system should provide users with WA information of types *"What I Did"/"What Did You Do"* and actions' *"Completion Status"*. The collaborative system should also be designed to detect sequential interaction and reduce the amount of WA it conveys to users, since this information can be redundant in such a case. Alternatively, the system should provide users with a means to control the amount of WA information that is conveyed to them to match their needs.

*Detect guidance and supply WA information accordingly:* A non-visual collaborative tool should detect guidance modes of interaction or allow users to manually switch to such a mode. If audio is concealed, the guidance mode should explicitly convey and/or allow users to explicitly track WA information of types *"What Did You Do"*, *"What Are You Doing"* and actions' *"Completion Status"*.

*Provide private and public workspace areas and a means to switch between them:* Collaborators should be able to work privately, such that they control whether their partners can hear their output, but also be able to expose their audio to their partner, for example, to support guidance interaction. If the collaboration involves more than two people, then collaborators should be provided with a means to select a workspace to switch to from available users, or invite specific users to access their private workspace.

*Provide a log of contributions to shared tasks:* Both working styles and the means for delivering audio to the workspace had an impact on such an ability to keep track of self and partner's contributions to the task. In situations where this ability is crucial to the task at hand, the non-visual collaborative tool should be designed to convey and/or allow users to request WA information of types *"What Did You Do"/ "What I Did"* or allow users to requested such information.

*Provide users with a means to display WA information of their choice to their partners':* In order to support coordination, particularly when collaborators work in parallel or audio is concealed, users should be provided with a means for choosing which WA information to expose and when such information is displayed to their partners.

*Provide a means for following or tracking partners' interactions in a workspace:* Non-visual collaborative systems should provide users with the ability to allow their partners to control the audio output that they receive, particularly when audio is concealed, i.e. handing the control of their display to their partners, which should allow users to monitor shared activities.

### Limitations and Future Work

Our findings are limited to instances where pairs of users collaborate to access and edit hierarchical menus. It remains unclear if such findings would extend to collaborations between a large group and other coordination and negotiation tasks. Certain aspects of the observed findings are likely to reoccur in such scenarios, such as choices of working styles and their influence on the collaborative process. For instance, collaboration between a large group of users could break up into smaller subgroups who focus on different subtasks. An avenue for future research is therefore to explore the practicality of the auditory modality for supporting collaborations between larger groups of users. Our findings are also limited to two settings only in which audio is either concealed or shared within the workspace through headphones or speakers. Future work should explore other audio settings. Also, while focusing only on audio provides insights about the practicality of this modality, this is an extreme scenario that serve to tease out the capabilities of audio as a medium of collaboration and to characterise the collaborative space it affords. These insights do not automatically translate to other situations where audio is combined with other modalities, for example it does not account for potential crossmodal effects [41, 16]. Future work should therefore examine how our findings generalise to multisensory and crossmodal systems.

## CONCLUSION

We presented a study that examined workspace awareness information exchange during an audio-only collaborative task. The results showed that concealing or exposing audio output in such a workspace had an impact on which workspace awareness information was exchanged between partners, and that this observed impact was also dependent on the working style partners chose to employ. These results highlighted that the relevance of workspace awareness information is not static but dynamic, changing according to how collaborators choose to work with sounds. Our findings provide empirical evidence that audio can be used as a sole means for supporting non-visual collaboration with shared menus and for maintaining workspace awareness and common ground during such collaborations. Further, we identified which type of workspace awareness information was exchanged under different conditions, and thus further extended previous research by identifying what information is relevant and when, and should therefore be captured about collaborators' interactions and conveyed to partners when designing workspace awareness support for non-visual collaboration.

## ACKNOLWEDGEMENT

We would like to acknowledge the support of EPSRC Fellowship Grant EP/N00616X/2 "CRITICAL" project, and the Algerian Ministry of Higher Education and Scientific Research (MESRS).

## REFERENCES

1. Dix Alan. 2010. Human-computer interaction: A stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers* 22 (2010), 13–27.
2. Dix Alan, Finlay Janet, Abowd Gregory, and Beale Russell. 2004. Human-computer interaction. *England: Pearson Education Limited* 5 (2004).
3. Ronald M Baecker, Dimitrios Nastos, Ilona R Posner, and Kelly L Mawby. 1995. The user-centred iterative design of collaborative writing software. In *Human-computer interaction*. Morgan Kaufmann Publishers Inc., 775–782.
4. Tina Blaine and Tim Perkis. 2000. The Jam-O-Drum interactive music system: a study in interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 165–173.
5. Meera M Blattner, Denise A Sumikawa, and Robert M Greenberg. 1989. Earcons and icons: Their structure and common design principles. *Human–Computer Interaction* 4, 1 (1989), 11–44.
6. Sara A Bly, Steven P Frysinger, David Lunney, Douglass L Mansur, Joseph J Mezrich, Robert C Morrison, and William Buxton. 1985. Communicating with sound (panel session. In *ACM SIGCHI Bulletin*, Vol. 16. ACM, 115–119.
7. Sara A Bly, Steve R Harrison, and Susan Irwin. 1993. Media spaces: bringing people together in a video, audio, and computing environment. *Commun. ACM* 36, 1 (1993), 28–46.
8. A Boming and M Travers. 1991. Two approaches to casual interaction over computer and video networks. In *International Conference on Human Factors in Computing Systems (CHI91)*. 13–19.
9. Stephen Brewster. 2002. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing* 6, 3 (2002), 188–205.
10. William Buxton. 1989. Introduction to this special issue on nonspeech audio. *Human Computer Interaction* 4, 1 (1989), 1–9.
11. Herbert H Clark, Susan E Brennan, and others. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991 (1991), 127–149.
12. Herbert H Clark and Meredyth A Krych. 2004. Speaking while monitoring addressees for understanding. *Journal of memory and language* 50, 1 (2004), 62–81.
13. Jonathan Cohen. 1993. "Kirk here": using genre sounds to monitor background activity. In *INTERACT'93 and CHI'93 Conference Companion on Human Factors in Computing Systems*. ACM, 63–64.
14. Gwyneth Doherty-Sneddon, Anne Anderson, Claire O'malley, Steve Langton, Simon Garrod, and Vicki Bruce. 1997. Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of experimental psychology applied* 3 (1997), 105–125.
15. Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. ACM, 107–114.
16. Daniel J Finnegan, Eamonn O'Neill, and Michael J Proulx. 2016. Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 200–212.
17. Robert S Fish, Robert E Kraut, and Barbara L Chalfonte. 1990. The VideoWindow system in informal communication. In *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*. ACM, 1–11.
18. Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the new world of HCI. *interactions* 24, 4 (2017), 38–42.
19. Christopher Frauenberger, Tony Stockman, and Marie-Luce Bourguet. 2007. A survey on common practice in designing audio in the user interface. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1*. British Computer Society, 187–194.
20. William W Gaver. 1986. Auditory icons: Using sound in computer interfaces. *Human-computer interaction* 2, 2 (1986), 167–177.
21. William W Gaver. 1991. Sound support for collaboration. In *Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW?91*. Springer, 293–308.
22. William W Gaver. 1992. The affordances of media spaces for collaboration. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. ACM, 17–24.
23. William W Gaver, Randall B Smith, and Tim O'Shea. 1991. Effective sounds in complex systems: The ARKola simulation. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. ACM, 85–90.
24. Darren Gergle, Robert E Kraut, and Susan R Fussell. 2004. Action as language in a shared visual space. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 487–496.
25. Tom Gross. 2013. Supporting effortless coordination: 25 years of awareness research. *Computer Supported Cooperative Work (CSCW)* 22, 4-6 (2013), 425–474.
26. Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)* 11, 3-4 (2002), 411–446.

27. Carl Gutwin, Oliver Schneider, Robert Xiao, and Stephen Brewster. 2011. Chalk sounds: the effects of dynamic synthesized audio on workspace awareness in distributed groupware. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 85–94.

28. Mark S Hancock, Chia Shen, Clifton Forlines, and Kathy Ryall. 2005. Exploring non-speech auditory feedback at an interactive multi-user tabletop. In *Proceedings of Graphics Interface 2005*. Canadian Human-Computer Communications Society, 41–50.

29. Christian Heath and Paul Luff. 1992. Collaboration and control Crisis management and multimedia technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW)* 1, 1 (1992), 69–94.

30. Debby Hindus, Mark S Ackerman, Scott Mainwaring, and Brian Starr. 1996. Thunderwire: a field study of an audio-only media space. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*. ACM, 238–247.

31. Ying Ying Huang, Jonas Moll, Eva-Lotta Sallnäs, and Yngve Sundblad. 2012. Auditory feedback in haptic collaborative interfaces. *International journal of human-computer studies* 70, 4 (2012), 257–270.

32. Hiroshi Ishii, Minoru Kobayashi, and Jonathan Grudin. 1993. Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Transactions on Information Systems (TOIS)* 11, 4 (1993), 349–375.

33. Manuele Kirsch-Pinheiro, José Valdeni De Lima, and Marcos RS Borges. 2003. A framework for awareness support in groupware systems. *Computers in Industry* 52, 1 (2003), 47–57.

34. Gregory Kramer. 1994. An introduction to auditory display. In *Auditory Display: Sonification, Audification and Auditory Interfaces, SFI Studies in the Sciences of Complexity, Proceedings*. Addison-Wesley, 1–77.

35. Jiannan Li, Saul Greenberg, and Ehud Sharlin. 2014. Enhancing Workspace Awareness on Collaborative Transparent Displays. (2014), Research report 2014–1065–16.

36. Jack M Loomis, Reginald G Golledge, and Roberta L Klatzky. 2006. Navigation system for the blind: Auditory display modes and guidance. *Navigation* 7, 2 (2006).

37. Paul Luff, Hideaki Kuzuoka, Christian Heath, Keiichi Yamazaki, and Jun Yamashita. 2009. Creating assemblies in media space: Recent developments in enhancing access to workspaces. In *Media Space 20+ Years of Mediated Life*. Springer, 27–55.

38. David McGookin and Stephen Brewster. 2007. An initial investigation into non-visual computer supported collaboration. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2573–2578.

39. Oussama Metatla, Nick Bryan-Kinns, and Tony Stockman. 2008. Constructing relational diagrams in audio: the multiple perspective hierarchical approach. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 97–104.

40. Oussama Metatla, Nick Bryan-Kinns, and Tony Stockman. 2012. Interactive hierarchy-based auditory displays for accessing and manipulating relational diagrams. *Journal on Multimodal User Interfaces* 5, 3-4 (2012), 111–122.

41. Oussama Metatla, Nuno N Correia, Fiore Martin, Nick Bryan-Kinns, and Tony Stockman. 2016a. Tap the ShapeTones: Exploring the Effects of Crossmodal Congruence in an Audio-Visual Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1055–1066.

42. Oussama Metatla, Fiore Martin, Adam Parkinson, Nick Bryan-Kinns, Tony Stockman, and Atau Tanaka. 2016b. Audio-haptic interfaces for digital audio workstations. *Journal on Multimodal User Interfaces* 10, 3 (2016), 247–258.

43. Pascal Molli, Hala Skaf-Molli, and Christophe Bouthier. 2001. State treemap: an awareness widget for multi-synchronous groupware. In *Groupware, 2001. Proceedings. Seventh International Workshop on*. IEEE, 106–114.

44. Meredith Ringel Morris, Dan Morris, and Terry Winograd. 2004. Individual audio channels with single display groupware: effects on communication and task strategy. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 242–251.

45. Elizabeth D Mynatt. 1994. Designing with auditory icons: how well do we identify auditory cues?. In *Conference companion on Human factors in computing systems*. ACM, 269–270.

46. Michael A Nees and Bruce N Walker. 2011. Auditory displays for in-vehicle technologies. *Reviews of human factors and ergonomics* 7, 1 (2011), 58–99.

47. Donald A Norman. 1993. *Things that make us smart: Defending human attributes in the age of the machine*. Basic Books.

48. Daniele Pagani and Wendy E Mackay. 1993. Bringing Media Spaces into the Real World.. In *ECSCW*. 347.

49. Michael J Proulx. 2007. Bottom-up guidance in visual search for conjunctions. *Journal of Experimental Psychology: Human Perception and Performance* 33, 1 (2007), 48.

50. Rameshsharma Ramloll and Stephen Brewster. 2002. An environment for studying the impact of spatialising sonified graphs on data comprehension. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*. IEEE, 167–174.

51. Markus Rittenbruch. 2002. Atmosphere: a framework for contextual awareness. *International Journal of Human-Computer Interaction* 14, 2 (2002), 159–180.

52. Charles Spence and Jon Driver. 1997. Audiovisual links in exogenous covert spatial orienting. *Attention, Perception, & Psychophysics* 59, 1 (1997), 1–22.

53. R Stults. 1986. Media space (Technical Report). *Palo Alto, CA: Xerox Palo Alto Research Center* (1986).

54. Barry Truax. 2001. *Acoustic communication*. Vol. 1. Greenwood Publishing Group.

55. David D Woods. 1995. The alarm problem and directed attention in dynamic fault management. *Ergonomics* 38, 11 (1995), 2371–2393.

56. Jiajie Zhang. 1991. The interaction of internal and external representations in a problem solving task. In *Proceedings of the thirteenth annual conference of cognitive science society*, Vol. 88. Erlbaum Hillsdale, NJ, 91.

57. Shengdong Zhao, Pierre Dragicevic, Mark Chignell, Ravin Balakrishnan, and Patrick Baudisch. 2007. Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1395–1404.