

Audio Delivery and Territoriality in Collaborative Digital Musical Interaction

Robin Fencott
RobinFencott@eecs.qmul.ac.uk
Queen Mary University of London
Interaction, Media and Communication Group

Nick Bryan-Kinns
Nick.Bryan-Kinns@eecs.qmul.ac.uk
Queen Mary University of London
Interactional Sound and Music
Centre for Digital Music

This paper explores the design of collaborative musical software through an evaluation of the effects different audio delivery mechanisms have on the way groups of co-located musicians work together in real time via a software environment. Ten groups of three musically proficient users created music using three experimental interfaces. Logs of interaction provide evidence that changing the means of audio delivery had a statistically significant effect on the way users worked together and shared musical contributions. In addition, interview transcripts indicate a number of experiential differences between the audio delivery configurations. The findings and design guidelines presented in this paper are intended to inform future systems for musical collaboration, and also have implications more broadly for the design of multi-user interfaces for which sound is a fundamental component.

Collaboration, Design, Human Interaction, Evaluation, Music, Usability, CSCW, Audio

1. INTRODUCTION

Music is a fundamental part of human expression (Makelberge 2010) and although not always a collaborative activity (Makelberge 2010) the creation, performance and enjoyment of music is highly social (Healey et al. 2005; Jordà 2005). Musical interaction is frequently identified as creative, open-ended, process oriented and problem-seeking (Makelberge 2010; Sawyer 2003). Within the computer music community there is a history of developing collaborative musical interfaces (Weinberg 2005; Jordà 2005), whilst more recently laptop orchestras (Wang et al. 2009) and multi-touch interfaces (Xambó et al. 2011) have been the focus of much development. However there still is a paucity of research concerning the evaluation of such systems, and there is limited research into human interaction during computer supported collaborative music making.

This paper contributes a study to investigate how different audio delivery configurations (speakers, headphones) can afford understanding of the location, authorship and origin of musical contributions during real-time musical collaboration. Understanding the implications of how to present audio is essential for any form of sound based human computer interaction, however to date there have been no detailed or controlled user studies investigating the effect

different audio delivery mechanisms on the process of collaborative digital musical interaction. The study focuses on groups of co-located musicians using a networked graphical interface distributed across multiple computers. The findings are contextualised with reference to the concepts of awareness (Gutwin and Greenberg 2002), territory (Tse et al. 2004) and privacy (Dourish and Bellotti 1992). Quantitative analysis of interaction logs is used to study the effects different audio delivery configurations have on the way participants interact with the software, the degree to which they share their contributions, and their tendency to edit contributions made by other users. A multiple-choice questionnaire is used to gauge general preferences, and extracts from interviews with the participants are used to elaborate on relevant discussion points.

1.1. Privacy and Awareness

The Workspace Awareness (WA) framework (Gutwin and Greenberg 2002) describes the means by which people working together in shared physical workspaces gather awareness information about each others' activities. For instance the sounds produced during the execution of a task may indicate to other people that certain individuals are currently occupied or that certain artefacts are currently in use.

Musicians playing acoustic instruments provide a rich source of awareness through the gestures and movements associated with using their instruments, the sounds they produce and the way they orient around each other (Healey et al. 2005). However where musical interaction is mediated via software, generic input devices such as mice and keyboards may reduce opportunities for gathering and displaying awareness information. For instance Merritt et al. (2010) observe ensembles of skilled electronic music performers relying on crude visual indicators such as level meters to glean an understanding of who is responsible for which sounds in an unfolding musical improvisation. Collaborative interfaces for musical interaction should therefore provide additional awareness mechanisms to support users during real-time interaction (Fencott and Bryan-Kinns 2010; Bryan-Kinns and Hamilton 2009), although there are to date few studies investigating the design of such features.

1.2. Space and Territory

Many collaborative activities feature territorial behaviour. Territory has been identified as a signal of ownership or responsibility for objects, artefacts or spaces in shared document writing and other forms of collaborative activities (Thom-Santelli et al. 2009). Within HCI, territorial interaction can be reinforced through the re-appropriation of existing software functionality (Thom-Santelli et al. 2009), or may be the result of natural spatial partitions within an activity. For instance (Tse et al. 2004) showed that users of Single Display Groupware partitioned their workspace to avoid interfering with each another's work. The study presented in this paper investigates the role that spatial audio can play in providing information about the authorship of musical contributions within a shared interface. Our analysis also identifies and characterises territorial behaviour within a shared graphical interface.

1.3. Audio Delivery Devices for HCI

There is a paucity of research into auditory delivery in HCI. Kallinen and Ravaja (2007) investigate the effects of presenting news reports on headphones or speakers, reporting that headphones cause people to become more immersed in the task at hand and less conscious of their surroundings. Nelson and Nilsson (1990) report similar results in a single user simulated driving activity. Morris et al. (2004) compares shared speakers and shared speakers plus individual headphones to deliver audio in a collaborative multi-touch system, showing that parallel working styles were adopted when users were given a personal audio channel via an in-ear headphone bud.

Alexandraki and Kalantzis (2007) use a questionnaire to ascertain musicians' preferences for audio delivery, noting a slight preference for multi-channel audio. Blaine and Perkis (2000) compare headphones and speakers through informal user testing and suggest that headphones caused participants to be less communicative and more isolated from the group. They also argue that spatialised audio might help users attribute ownership to musical contributions, although their results show that non-musicians encountered difficulties identifying the effects of their actions. Conversely, Merritt et al. (2010) states that their groups of laptop musicians rejected the idea of personal speaker channels in favour of a combined mix from a single set of speakers. Finally, the importance of coupling the performer with a localisable sound source has been advocated by laptop orchestras (Wang et al. 2009), although this research lacks evaluation.

2. STUDY

We conducted a study to investigate how different audio delivery configurations contribute to the way groups of musicians engage in musical collaboration. This section describes the collaborative music software developed to run the study, and presents the hypotheses and experimental design.

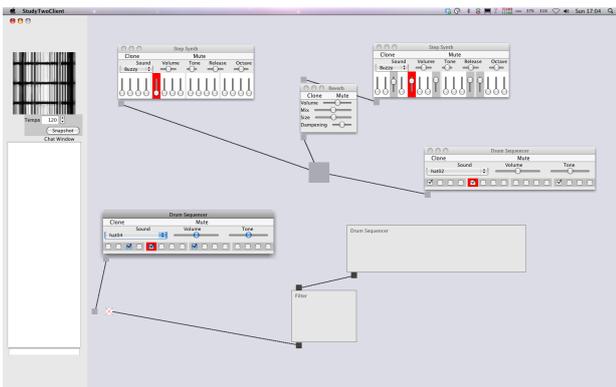
2.1. Collaborative Music Software

In order to conduct the study we developed a music environment which allows users working on separate computers to create music via a shared workspace. Our justification for developing a piece of bespoke software is presented in Fencott and Bryan-Kinns (2012). The interface was written in Java and SuperCollider. To make music, virtual musical instruments (drum machines and step-sequencer based synthesisers) are created in the on-screen workspace which is duplicated on the screens of all connected computers (see Figure 1). The 'instruments' (also referred to as 'Modules') provide control over the contents of a looping musical bar, and offer various sound synthesis controls. The looping nature of the sequencer and synthesised sounds makes the software especially suited to the creation of 'electronica' style music. Multiple instantiations of each instrument can be created to build up complex layered and harmonised parts, and instruments can be patched through audio effects to create rich sonic textures, and provide opportunities for musical contributions from different users to be interconnected and inter-associated with each other. Instruments and effects appear in the same location on all users' screens, and changes to the parameters (e.g. slider movements) are immediately updated for all users. The software also features a tempo control

which operates globally for all users. Incorporating fine grained control over tone, metric placement of beats and chromatic pitch of notes conforms to the notion that 'good instruments should be able to make bad sounds'(Overholt 2009), and is in contrast to previous studies of collaborative musical interaction which have typically focused on non-musician users (Weinberg 2005; Bryan-Kinns and Hamilton 2009).

The software features Public and Personal audio outputs for each user to facilitate individual and group level working. Patching a module to the Public output causes its audio to be routed to all participants, whereas if a user patches a module to their Personal output it is routed exclusively to their headphones or speaker. The Public output is represented as a grey rectangle in the centre of the workspace, while Personal outputs appear in the bottom left of the screen. Patching modules to the Public and Private outputs also affects their availability for editing. Modules patched to the Public output are freely editable by all users, while modules patched to a user's Personal output are rendered non-editable by other users, and appear as grey boxes on all other participants' computer screens.

Figure 1: Screenshot of the collaborative interface



2.2. Hypotheses

The study investigated the following hypotheses:

H1 Presenting audio exclusively through headphones will encourage more individual work. Indicators of individual work are: more use of Personal output; less co-editing of contributions; less verbal acknowledgements.

H2 An interface with the most auditory distinction between public and private audio channels will be preferred by participants. This hypothesis is based on the assumption that privacy is a key factor in the user's preference (Fencott and Bryan-Kinns 2010).

H3 An interface which presents personal and public audio entirely through speakers will cause

participants to work more collectively, as evidenced by: more audio played in the Public channel; less use of privacy features; increased discussion of audio sources.

2.3. Methodology

The study employed a three condition within-subjects design in which all participants were exposed to each of the conditions. Aside from the three configurations of headphones and speakers for audio delivery described below, the musical software remained identical throughout all conditions. To counteract ordering effects the sequence of conditions was permuted between groups.

Condition C1 - Speakers only: Each participant has their own speaker, which is used to present both Personal audio and Public audio. This is similar to conventional instrumental playing where each person's instrument comes from a distinct spatial location. Using individual speakers creates a situation in which participants each have a personalised version of the music which can be overheard by the other participants.

Condition C2 - Headphones and Public Speakers: Each participant has their own speaker for Public audio. Personal audio for each participant is played through headphones. This arrangement is similar to the DJ practice of using headphones to cue new records in private before crossfading to speakers for the audience to hear (Pfadenhauer 2009).

Condition C3 - Headphones Only: Public and Personal audio are routed through headphones. Musical contributions routed to the Public Channel go to all headphones, while contributions routed to a user's Personal channel are only played through that user's headphones.

2.4. Participants and Recruitment

Thirty individuals were recruited via e-mail lists. The recruitment e-mail asked for 'people with an interest in creating music, for instance composers, musicians, DJs, and students of Music, Music Technology or related fields'. Each participant received financial compensation for taking part. The decision to study users with knowledge of music and music technology represents a departure from studies of non-musician users engaging with simplified musical interfaces (e.g. (Weinberg 2005)).

Participants were organised into groups of three. 68% of participants were male (based on 29 responses). The average age was 33 (based on 25 responses). 69% of participants could play a musical instrument. 22% classified themselves as of beginner level musical proficiency, 40%

classifying themselves as of 'intermediate' level, and approximately 18% for both 'semi-professional' and 'professional' levels. 86% of participants had composed songs individually, while 75% had composed songs with others. 58% of participants identified their level of computer literacy as 'intermediate', and 37% identified as 'expert'. One participant identified as a 'beginner'. 27.59% of participants had not previously used collaborative software. 25% had played online multi-player computer games. 14.29% had used collaborative document editors and 7.14% had used collaborative writing software. 50% had used collaborative music software.¹

2.5. Experiment Task

To spark discussion and provide a common ground for the participants, a challenge was set to compose music to compliment a short video animation displayed in the top left of the interface (see top left in fig 1). A different video animation was used for each condition, and the sequence of video animations was ordered independently of the condition ordering.²

2.6. Measures

2.6.1. Questionnaire data

A post-test questionnaire based on the Mutual Engagement Questionnaire (Bryan-Kinns and Hamilton 2009) gathered information about the participants experiences with the experimental conditions by asking participants to order the experimental conditions in terms of how they related to a list of statements.

2.6.2. Interaction Log Analysis

The following interaction features were logged by the software: Creating a module, Deleting a module, Modifying a module control (e.g. moving a slider, pressing a button), Connecting a module to the Public output, Connecting a module to a Personal output, Patching to and from effects, Movement and spatial position of modules, Tempo changes.

2.6.3. Group Discussions

Video recorded discussions were held at the end of each session. The discussions focused on preferences, perceived differences between the speakers and headphones, use of the Personal and Public channels, awareness of each other's activities, roles, working strategies and spatial use of the shared on-screen workspace.

2.7. Procedure and Apparatus

Sessions started with a verbal introduction by the researcher. Participants were then presented with

a pre-test questionnaire to collect demographic information. A 10 minute training period with the software followed. Participants were then given fifteen minutes with each experiment condition. The post-test questionnaire was presented after the conditions, and finally a group discussion was held. The researcher sat in a visually occluded control room while participants completed the questionnaires and engaged in the experimental conditions.

The software ran on three Apple Mac-Mini computers with 21" widescreen displays, placed on a large round table (see fig 2). The displays were lowered to allow participants to see over them. Studio quality Yamaha MSP5 monitors were used for conditions requiring speakers. These were positioned to the right of each display. Sony MDR-7509HD headphones were used for conditions C2 and C3. These headphones could be worn over the head or held to the ear in the style of a DJ.

Figure 2: Equipment used for user study



3. RESULTS

3.1. Post-Test Questionnaires

Twenty-eight participants completed the post-test questionnaire in full, one participant provided responses to almost all statements and one participant provided no responses. The Friedman test was used to identify statements which elicited a statistically significant trend. A significant number of participants identified condition C2 as the one in which they 'lost track of time' ($p=0.0406$, $df=2$, $csq_r=6.41$). A significant number of participants ($p=0.0211$, $df=2$, $csq_r=7.72$) rated condition C2 as the one in which they 'had the most privacy'. No other statements provoked a statistically significant effect. Table 1 presents the post-test questionnaire results.

¹Percentages do not add up to 100 for multiple choice questions

²Videos created using 'Mother' <http://www.onar3d.com/mother/>

Table 1: Post-Test questionnaire results summarised to rank averages. Significance of $p < 0.05$ highlighted in bold.

Statement	C1	C2	C3	P
The best music	2	1.9	2.2	0.56
I felt most involved with the group	1.9	1.9	2.1	0.52
I enjoyed myself the most	2.1	1.9	2.1	0.66
I felt out of control	2.2	1.8	2	0.34
I understood what was going on	1.9	2	2.1	0.8
I worked mostly on my own	2	1.9	2.1	0.64
I lost track of time	2.2	1.6	2.1	0.04
Other people ignored my contributions	1.9	2	2.1	0.58
We worked most effectively	2.1	2	2	0.9
The interface was most complex	1.9	2.1	2	0.79
I had the most privacy	2.4	1.7	1.9	0.02
I knew what other people were doing	2	2	2	0.97
We edited the music together	2	2.1	1.9	0.57
I made my best contributions	2	2	2	0.07
I was influenced by the other people	2	1.9	2	0.9
The condition I preferred the most	2.1	1.9	2	0.73

3.2. Interaction Log Analysis

Interaction log analysis using the Friedman Test showed that where participants were given Speakers Only (Condition 1) they made significantly less use of the Personal channel to listen to musical contributions ($p=0.0253$, $df=2$, $csq_r=7.35$). There was no significant difference in the amount of modules created ($p=0.7225$, $df=2$, $csq_r=0.65$), the amount of modules deleted ($p=0.5169$, $df=2$, $csq_r=1.32$), between the amount of editing which individuals performed on their own modules ($p=0.8395$, $df=2$, $csq_r=0.35$), or the instances of co-editing which took place ($p=0.3413$, $df=2$, $csq_r=2.15$). There was no significant effect on the number of Module coordinate position movements ($p=0.4677$, $df=2$, $csq_r=1.52$). There was no significant difference in the use of the tempo control between conditions ($p=0.5916$, $df=2$, $csq_r=1.05$). Finally there was no significant effect on the amount of times participants patched modules to the Public channel ($p=0.1496$, $df=2$, $csq_r=3.8$). Table 2 summarises these results.

3.3. Spatial Workspace Organisation

Visualisations were produced to plot the co-ordinate position of each music module over the course of the interaction. Colour (red, green, blue) was used to signify which of the three participants created each module (See Figure 3). Due to the low frequency of

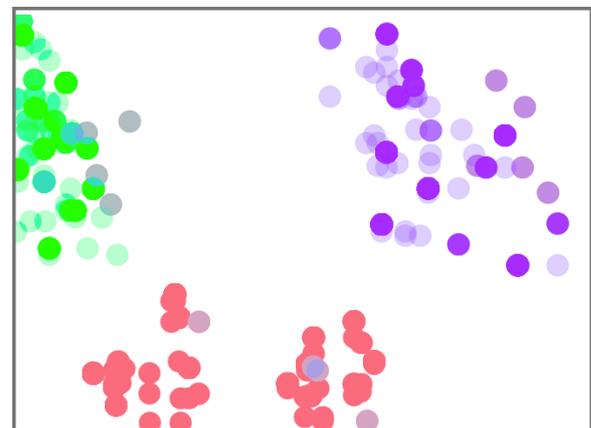
Table 2: Interaction log results summarised to rank averages. $df=2$ in all cases. Significance of $p < 0.05$ in bold.

Feature	C1	C2	C3	csq _r	p
Creations	1.9	2	2.1	0.65	0.72
Deletions	2.1	1.8	2.1	1.32	0.51
Individual editing	1.9	2	2.1	0.35	0.83
Co-editing	2.2	1.9	1.9	2.15	0.34
Movements	2	1.8	2.2	1.52	0.46
Patch to Public	2	2.3	1.8	3.8	0.14
Patch to Personal	1.6	2.3	2.2	7.35	0.02

co-editing instances (people editing modules created by others), the initial module creator was assumed to provide a good indicator of authorship.

A variety of spatial patterns were informally identified, the most common being contributions by individuals arranged in corners of the screen, in horizontal or vertical stripes, or randomly spaced. The visualisations were manually coded as 'grouped' or 'intermingled' sets based on the degree to which the areas of coloured dots appeared to be grouped together. The categorisations were performed independently by a third party rater, and a Cohen's Kappa test produced an inter-rater reliability of 0.6667 (0 indicates total disagreement, 1 indicates complete agreement).

Figure 3: Visualisation of workspace territory. Circles represent position of modules, colours indicate individuals.



The Mann-Whitney U test was then used to compare the interaction logs for grouped and intermingled data sets. It was found that groups who intermingled their modules performed significantly more co-editing than groups who spatially separated their modules ($U_a = 1354.5$, $z=-2.76$, $p_1=0.0029$, $p_2=0.0058$). Groups with more pronounced spatial partitioning also created more modules ($U_a = 696$, $z=2.55$, $p_1=0.0054$, $p_2=0.0108$) and made more use of the public channel ($U_a = 697.5$, $z=2.54$, $p_1=0.0054$, $p_2=0.0111$).

3.4. Video Observation and Group Interviews

Video of the interaction and group discussion was manually transcribed and coded using Grounded Theory methods (Muller and Kogan 2010), although space limitations prevent a full report of this data. Instead, extracts from the interviews and group interaction are presented throughout the following section to elaborate on specific findings.

4. DISCUSSION

The results suggest that manipulating the way audio was delivered changed the way groups collaborated, and that it also influenced the perceived quality of the interaction. This section begins by using questionnaire, interaction log and transcription extracts to assess the study hypotheses, before expanding on more general issues surrounding the findings of the study.

4.1. Hypothesis H1

Hypothesis H1 stated that 'Participants will work more individually when audio is presented exclusively through headphones'. Analysis of the interaction logs indicate that the experimental conditions did not influence the amount of editing or co-editing which took place, and there was also no effect on the use of Personal audio channels when audio was routed through headphones. Participants did however report having least privacy in the speakers only condition, and during discussion the participants identified a number of distinctions between working in headphones and working in speakers.

Some participants noted that the headphones encouraged more concentrated listening than with speakers, causing them to 'hear close things', or focus on 'texture' and 'tiny details'. The property of headphones to facilitate a more intimate, close and immersive listening experience was also identified in (Kallinen and Ravaja 2007). The tendency for headphones to promote more focused listening had contradictory effects on reported experiences of involvement in the group. On the one hand some participants reported being less involved with the group as they became more focused on the details of the sounds they were creating, at the expense of engaging with the publicly shared music. For other participants, more concentrated listening resulted in their becoming more attentive to the changes made by others, and consequently they noted feeling as though they were working more as a group when using headphones. These two effects are illustrated by the following transcript extracts:

B: yeah, and I think you're much more aware of, of other people's changes, in the last one, having the

headphones on [...] and I think, my feeling was that that encourages, encouraged us to change more in each others'. [...] that was my feeling anyway, that there was a bit more collaboration with each others' sounds

C: well, erm, I see, the headphones definitely, shut me off from, erm, I mean I was able to concentrate solely on what I was doing, but I wasn't as involved as a group

B describes a situation in which headphones caused him/her to engage more with the group, as the close listening drew attention to the changes others were making to the sounds. Conversely, C states that the headphones allowed him/her to concentrate on his/her own ideas, but at the expense of less group involvement. The contradictory statements from participants do provide reassurance that the interview questions were not leading participants towards particular answers (Furniss et al. 2011), however they also suggest that there was a wide range of different experiences of the interaction.

4.2. Hypothesis H2

Hypothesis H2 posits that an interface which provides the greatest auditory separation between personal and public work would be preferred by participants. The post-test questionnaire data does not indicate that any of the conditions affected preference based on results for the statements 'The condition I preferred the most' or 'I enjoyed myself the most'. Analysis of post-test responses indicate a significant proportion of participants rated C2 (Speakers + Headphones) as the one they most lost track of time in, which has previously been interpreted as an indicator of engagement (Bryan-Kinns and Hamilton 2009).

Participants expressed mixed responses to the C2 condition. Some participants expressed difficulty switching between headphones and speakers, noting that it caused a disruption in their ability to focus on the shared aspects of the music. One participant described using headphones initially to experiment with the software, before switching back to speakers to work with the group.

C: yeah yeah, because I, I was just trying to get, see how it all worked, and, which I preferred, and then I found it was much better to just be on the same wavelength as everybody else

Another participant noted that the headphones made them concentrate more on what they were doing individually, at the expense of formulating a musical contribution which was coherent with the music playing through the public channel on speakers:

D: I was more concentrating on what I was doing, and when I tried to erm, add it to the, you know, public thing, it was, it just didn't sound right.

It seems more accurate to state that participants noted feeling more involved with the group when they were most aware of and attentive to the changes being made by others. This appears to have occurred more often when audio was presented via speakers or headphones, rather than when public and personal audio were split into separate devices.

4.3. Hypothesis H3

Hypothesis H3 proposed that 'An interface which presents personal and public audio entirely through speakers will cause participants to work more collectively'. When Public and Personal audio was presented entirely through speakers (C1), interaction log analysis showed that participants made significantly less use of the Personal channel to listen to musical contributions than they did in either of the other conditions. This suggests that the personal channel was less useful when delivered via speaker, or perhaps that working with speakers discouraged personal working. This supports the hypothesis that participants would make less use of the privacy functions when audio was delivered this way. Furthermore, response to the questionnaire statement 'I had the most privacy' shows participants reported experiencing least privacy in Condition 1. During discussion, participants noted difficulty determining the spatial location of sounds from individual speakers.

4.4. Workspace Organisation

Participants were free to organise the on-screen music modules arbitrarily, and it is important to emphasise that the spatial position of modules did not influence the musical outcome of the software. The categorised visualisations provide evidence that participants in half the groups employed strong spatial organisation within the shared workspace to separate contributions based on ownership. The visualisations also show that groups used a similar spatial arrangement in every condition, and the spatial arrangements appear not to have been influenced by the audio presentation modes. During interviews, individuals often seemed aware that they were working in a particular area of the screen, although they did not always know where other group members were working. Studying the video recordings taken during engagement with software, it appears that in many cases the spatial organisation was the result of unspoken, tacit or implicit agreement, rather than the verbal negotiation.

Unlike in the case of people sitting or standing around a shared screen or table, the circular seating arrangement (see Figure 2), combined with the consistent spatial position of modules within the shared and distributed workspace means that the physical position of participants around the table could not have contributed to the way participants organised the spatial layout of elements on the screen, while the primarily auditory activity of music-making presents no inherent cues or suggestions towards specific spatial arrangements. It is possible that the central position of the Public output patching block may have directed participants towards a natural spatial partitioning strategy around the centre of the screen, however the plurality of layout approaches (corners, horizontal and vertical stripes, non-uniform) suggests that this was not a major contributing factor.

Participants may have spatially partitioned the interface to reduce interference with one-another's work. This statement is supported by interaction log evidence; using the Mann-Whitney test to compare instances of co-editing between the sets of partitioned and intermingled groups reveals that participants in groups with weaker spatial organisation strategies were more inclined to edit modules created by others than were participants in groups with more strict spatial organisation. However, conversation during the sessions indicates that the role of spatial partitioning stretched beyond the minimisation of interference, and was used to signify and help manage awareness of authorship of musical contributions. This follows Thom-Santelli et al. (2009), who argues that during collaboration, territoriality serves the communicative function of indicating ownership over a particular object or space. The following extract demonstrates how participants adopted a spatial ordering strategy to counter difficulties maintaining awareness of each other's actions.

E: it's just hard to keep up with so much, what's going on. You don't know who's, who's is doing what, you know?
F: (laughs)
G: ok, ok well how about this, how about this.
E: uh uh
G: Why doesn't everybody, like, lets say, you go to one side with yours, I go to one side with mine, and you go to one side with yours. Like, it's to move it to one side so we know what everyone is doing.
E: yeah, yeah
G: that make sense?
F: yeah

This extract demonstrates the way in which participants negotiated an informal mechanism or agreement to scaffold authorship awareness through partitioning of the workspace. The participants used

the spatial division of the shared interface to create or claim personal workspaces for themselves within the shared interface, even though such workspaces were not provided explicitly via the interface. This extract also highlights a common vocabulary used by participants to discuss the screen layout. During interviews, participants often described themselves in terms of working 'in the bottom left', 'top right', and so on, although some participants appeared not to have any sense of territory within the interface and talked about working 'all over the place', or 'putting stuff anywhere there was space'. These were points at which the spatial organisation broke down, for instance one participant noted during the interview

K: you watch around for space, and start just putting stuff wherever

Participants occasionally used the spatially congruent layout of the interface as a resource to discuss aspects of the arrangement, as demonstrated in the following excerpt. Here A uses the spatially consistent workspace as a resource to draw B's attention to a particular music module:

H:*pointing with both fingers all over his screen*
I can see what everyone else has got on their parameters

J: yeah, that's right

H:*pointing at right hand side of screen*
because I've put, I've got, the second step sequencer, down on the right hand side, I've put the notes where your kicks are,
.....*pointing to left of screen*

H: so it's sort of

J: ah, right, paralleled

Due to the physical arrangement of screens on the circular table, J is unable to ground H's deictic reference to the information on his screen. H therefore verbally refers to the music module's spatial position within the shared on-screen workspace, by stating 'the second step sequencer, down on the right hand side' of the workspace. In this way H and J use their intersubjective knowledge of the spatially consistent layout of the workspace to discuss an aspect of the shared interface. J then notes that H's sequencer is 'paralleled', to acknowledge the observation that H's step sequencer is playing notes at the same time as the kick drums from J's drum module. This extract also demonstrates that participant H was using visual access to J's music modules as a resource for creating musically coherent contributions.

4.5. Identification of Contributions

In this study, musical material was created by editing on-screen sequencers. This poses two problems related to the gathering of workspace awareness:

a lack of feedthrough awareness at the time of creation, and following this the subsequent existence of an autonomous agent (the sequence) which proceeds to generate music independently of its creator and provides limited information about this autonomous process. Even though the interface was consistently distributed across computer screens, participants reported having difficulty identifying specific music modules within the interface. Some participants reported using their personal channel to discover which modules were responsible for which sounds, and during group interviews participants talked more about the importance of knowing what was making a specific sound than they talked about knowing who was responsible for a contribution.

Using a mouse to create a sequence does not indicate to other co-located musicians what activities are being undertaken, and indeed depending on the physical layout of workstations the action of moving the mouse may itself be non-visible. One participant in this study noted correlating on-screen activities such as fader movements with feedthrough awareness provided by the sounds of other participants' mouse-clicks to attribute authorship to certain music modules.

4.6. Key Findings

- The form of audio delivery influenced the degree to which participants used their personal audio channel. When Personal audio was routed to individual speakers next to each participant and public audio was routed to all speakers, participants made significantly less use of the personal channel.
- The spatially and visually consistent layout of the interface was exploited in several ways to support collaboration; primarily as an aid to joint attention, and as a means of indicating ownership over specific music modules.
- Most groups adopted similar spatial arrangements in each interface condition, although there was no evidence that the layout was influenced by the way audio was delivered.
- Strong territorial behaviour was identified in half the groups. These groups performed less co-editing, created more modules and made more use of the personal audio channel.
- Identification of contributions and identification of ownership appear to be two distinct issues.

5. IMPLICATIONS FOR DESIGN

5.1. Layout Features

Given the importance of module layout during collaboration, a redesigned interface could incorporate additional layout and organisation features to support

additional scaffolding for collaboration, awareness and joint attention. These features would not have a direct influence on the sonic output of the software, but could aid groups of people in structuring a collection of interface elements. Layout features could include user configurable dividers, workspaces, partitions, annotations and colour coded areas. The ability to group or bundle associated modules together (e.g. a collection of drum sequencers forming a rhythm) could be another useful feature. A new research direction might be to investigate the extent to which these organisational features need to be consistently duplicated across all connected computers, and how groups or individuals might exploit the affordances of these features.

5.2. Multiple Devices for Audio Delivery

In single user performance contexts such as DJing, the separation of audio into different devices has been identified as a central aspect of the practice. However in a real-time, co-located collaborative context, where multiple people are listening to a variety of sound sources simultaneously, splitting audio across multiple devices appears to be problematic from a design and usability perspective. Although results from previous studies implied that this separation might be beneficial (e.g. (Fencott and Bryan-Kinns 2010)), our study suggests that the separation of audio into different devices is detrimental to a groups' ability to coordinate and manage their collaboration, and may contribute to feelings of less involvement and less awareness.

Secondly, problems arose due to switching between headphones and speakers, balancing the level between speakers and headphones, the auditory disruption of headphone wearing on conversation and monitoring of audio played through speakers. These issues could be counteracted with less acoustically isolated headphones, providing individuals with control over the level of their audio outputs, and using wireless headphones to make switching between headphones and speakers less awkward.

Finally, a clear drawback of our findings of this study is that although the participants were musically inclined, they had limited experience of collaborative software, and limited experience using the software developed for this study. Had the participants become accustomed to the split audio design of Condition 2 they may have developed ways to deal with the problems they encountered. Consequentially, a strong implication is that for first time users, audio should be presented via a single device where possible (either headphones or speakers) as it appears to encourage stronger feelings of group involvement, and a greater sense of awareness.

5.3. Single Device for Audio Delivery

Interaction log analysis of the data suggests that using speakers for shared and individual audio presentation discouraged people from using personal audio channels, although previous research suggests that incorporating the ability to work in auditory isolation allowed participants to formulate more complete contributions before sharing them (Fencott and Bryan-Kinns 2010). Designers must therefore balance the choice to discourage individual work against the benefits of allowing users to control how and when their ideas are shared with the rest of the group. If the system is intended to promote open and collective group interaction then using speakers might be preferable, while a system which is designed to promote more focused individual work might benefit from headphone presentation, as this allows users to concentrate on their own contributions and take advantage of the detailed sound provided by headphones. A collaborative system could also incorporate a switching mechanism which allows a group to jointly transfer audio from their individual headphones to a speaker system at a point where they can combine their contributions.

5.4. Ownership and Identification

Participants seemed less concerned with who was contributing what, although they commented on using the personal audio channel to discover which interface elements were responsible for which sounds. This suggests that interfaces should provide separate mechanisms for identifying 'who' is doing what, and 'what' is doing what within the interface. Our subsequent research has pursued this issue.

6. CONCLUSION

Designing to support group musical interaction necessitates a careful consideration of how audio should be presented. Using an experimental design, this study has identified a number of ways in which different audio delivery mechanisms influence group musical interaction among ten groups of musically inclined users. This informed the synthesis of design implications for the way sound should be presented to support collaboration. In addition, analysis of the way groups configured, managed and discussed the shared interface points to a number of other design considerations for future collaborative systems.

7. ACKNOWLEDGEMENTS

Robin Fencott is funded by an EPSRC Doctoral Training Account Award.

8. REFERENCES

- Alexandraki, C. and Kalantzis, I. (2007) Requirements and application scenarios in the context of network based music collaboration. In *Proceedings of the AXMEDIS 2007 Conference. i-maestro workshop*, pp. 39-46. Barcelona, Spain.
- Blaine, T. and Perkis, T. (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*, DIS '00, pp. 165-173. ACM, New York.
- Bown, O., Eldridge, A., and McCormack, J. (2009) Understanding interaction in contemporary digital music: from instruments to behavioural objects. *Organised Sound*, 14(2), 188-196.
- Bryan-Kinns, N. and Hamilton, F. (2009) Identifying mutual engagement. *Behaviour & Information Technology*. 31(2),101-125.
- Dourish, P. and Bellotti, V. (1992) Awareness and coordination in shared workspaces. In *CSCW '92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, ACM, pp. 107-114. New York, NY, USA.
- Fencott, R. and Bryan-Kinns, N. (2013) HCI, CSCW and Collaborative Digital Musical Interaction. In Simon Holland, Katie Wilkie, Paul Mulholland, Allan Seago (eds), *Music and Human-Computer Interaction*. Springer Verlag, London.
- Fencott, R. and Bryan-Kinns, N. (2010) Hey man, you're invading my personal space! privacy and awareness in collaborative music. In *Proceedings of the 10th International Conference on New Interfaces for Musical Expression (NIME 2010)*, pp. 198-203. Sydney, Australia.
- Furniss, D., Blandford, A. and Curzon, P. (2011) Confessions from a grounded theory phd: Experiences and lessons learnt. *CHI*, pp. 113-122. Vancouver, BC, Canada.
- Gutwin, C and Greenberg, S. (2002) A descriptive framework of workspace awareness for real-time groupware. *Comput. Supported Coop. Work*, 11(3): 411-446.
- Healey, P., Leach J. and Bryan-Kinns, N. (2005) Inter-play: Understanding group music improvisation as a form of everyday interaction. In *Proceedings of Less is More — Simple Computing in an Age of Complexity*. Microsoft Research Cambridge.
- Jordà, S. (2005) Multi-user instruments: Models, examples and promises. In *Proceedings of the 2005 conference on New Interfaces for Musical Expression (NIME'05)*, pp. 23-26. ACM, Singapore.
- Kallinen, K. and Ravaja, N. (2007) Comparing speakers versus headphones in listening to news from a computer - individual differences and psychophysiological responses. *Computers in Human Behaviour*, 23(1), 303-317.
- Makelberge, N. (2010) 'le concert, c'est moi'. In *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*, DESIRE '10, pp. 150-153. Lancaster, UK.
- Merritt, T., Kow, W., Ng, C., McGee, K. and Wyse, L. (2010) Who makes what sound?: supporting real-time musical improvisations of electroacoustic ensembles. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, pp. 112-119. ACM, New York.
- Morris, M. R., Morris, D. and Winograd, T. (2004) Individual audio channels with single display groupware: effects on communication and task strategy. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp 242–251, ACM, New York.
- Muller, M. J. and Kogan, S. (2010) Grounded theory method in hci and cscw. Technical Report 10-09, IBM Watson Research Center.
- Nelson, T. M. and Nilsson, T. H. (1990) Comparing headphone and speaker effects on simulated driving. *Accident Analysis & Prevention*, 22(6), 523-529.
- Overholt, D. (2009) The musical interface technology design space. *Organised Sound*, 14(2), 217-226.
- Pfadenhauer, M. (2009) The lord of the loops. observations at the club culture dj-desk. *Forum: Qualitative Social Research*, 10(3).
- Sawyer, R. K. (2003) *Group Creativity: Music, Theater, Collaboration*. Psychology Press.
- Thom-Santelli, J., Cosley, D. R. and Gay, G. (2009) What's mine is mine: territoriality in collaborative authoring. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pp. 148-1484. ACM, New York, USA.
- Tse, E, Histon, J., Scott, S. D., and Greenberg, S. (2004) Avoiding interference: how people use spatial separation and partitioning in SDG workspaces. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 252-261. ACM, New York, NY, USA.
- Wang, Ge., Bryan, N., Oh, Jieun. and Hamilton, R. (2009) Stanford laptop orchestra (slork). In *Proceedings of the International Computer Music Conference*. Montreal.
- Weinberg, G. (2005) Interconnected musical networks: Towards a theoretical framework. *Computer Music Journal*, 29(2), 23-39.
- Xambó, A., Laney, R., Dobbyn, C. and Jordà, S. (2011) Collaborative music interaction on tablets: an HCI approach. In *25th BCS Conference on Human-Computer Interaction*.