Shaping Sounds: The Role of Gesture in Collaborative Spatial Music Composition

Thomas Deacon,
Nick Bryan-Kinns
Media and Arts Technology Centre
Queen Mary Univesity of London, UK
t.e.deacon@qmul.ac.uk

Patrick G.T. Healey
Cognitive Science Research Group
Queen Mary Univesity of London, UK
p.healey@qmul.ac.uk

Mathieu Barthet
Centre for Digtial Music
Queen Mary Univesity of London, UK
m.barthet@qmul.ac.uk

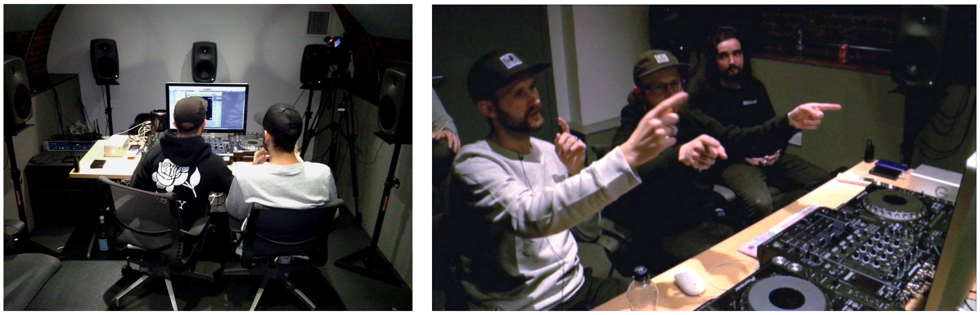


Figure 1. ***Commands*** **working in a Dolby Atmos music studio**

ABSTRACT

This paper presents an observational study of collaborative spatial music composition. We uncover the practical methods two experienced music producers use to coordinate their understanding of multi-modal and spatial representations of music as part of their workflow. We show embodied spatial referencing as a significant feature of the music producers' interactions. Our analysis suggests that gesture is used to understand, communicate and form action through a process of shaping sounds in space. This metaphor highlights how aesthetic assessments are collaboratively produced and developed through coordinated spatial activity. Our implications establish sensitivity to embodied action in the development of collaborative workspaces for creative, spatial-media production of music.

CCS Concepts

•**Human-centered computing** → **Computer supported cooperative work;** *Collaborative content creation;*

CC '19, June 23–26, 2019, San Diego, CA, USA

ACM ISBN 978-1-4503-5917-7/19/06.

DOI: https://doi.org/10.1145/3325480.3325493

Author Keywords

Collaborative Creativity; Music Production; Spatial Audio; Ethnomethodology; Gesture;

INTRODUCTION

New consumer technologies often prompt music practitioners to adapt and develop new ways of working [11, 43]. One growing area is the production of spatial music for immersive content [5]. Tools and practices have existed in this area for decades, but often the design of systems does not acknowledge that modern song-writing practice is collaborative [7]. Also, modern music production is complex and depends on computer systems to support creativity [9, 36, 35]. As researchers, we need to understand how to *embed* collaboration into new systems supporting spatial music making. Design that prioritises co-creativity enables musicians to develop new skills, maintaining professional standards in an emerging field.

In this paper, we investigate current music practice to improve the design of new collaborative workspaces. The practice is modern electronic music production using Digital Audio Workstations (DAWs). Modern electronic music practice encompasses a spectrum of genres, spanning avant-garde electronics and commercial pop. This paper focuses on how two experienced composers work with new spatial audio production tools. We use ethnomethodologically-informed design ethnography [13], alongside detailed video analysis [31], to develop an understanding of co-creative work in action. Our main concern in this paper is to establish how one aspect of collaboration, gesture, is of importance to work in a spatial music studio. Focussing on the collaborative decisions made about spatial audio in music composition, we develop our analysis on how spatial referencing gestures are used to build a shared space for creative decision making. We describe how collaborators "Shape Sounds", integrating gesture, speech and computer interfaces. Using this metaphor, we discuss the relationship of spatial sound to musical content and collaborative process. This offers insight into how people collaborate to create complex media content. Such an analysis builds a set of sensitising concepts that can support the design of the interfaces. We aim our design implications to support new collaborative workspaces for spatial audio using Extended Reality (XR) technologies[1].

LITERATURE

Spatial Audio Reproduction and Composition

Spatial audio approaches attempt to create the impression of spatially displaced sound sources by using speakers or headphones. Spatial audio supplies some of the localisation cues that we use to decode source direction in environmental sound; allowing a user to pinpoint where sound is coming from [42]. Our study deals only with one method of spatial audio rendering, Dolby Atmos [1]. Though originally a cinema sound technology, recent developments have seen Dolby Atmos expanded into game sound, virtual reality (VR) sound design, and live spatial music nightclub events. This has meant Dolby have been expanding Atmos music mixing support across a variety of platforms and levels of engagement. Dolby Atmos is a mixture of channel-based and object-based audio rendering. Channel-based audio is discrete streams of audio data, each associated with a specific loud-speaker position [1]. An object-based approach represents the sound scene as a set of independent sounds [37], where sound sources are accompanied by metadata that contains features such as level and position. An overview of the Atmos setup can be seen in figure 2.

Using spatial audio, what composers attempt to recreate are "acoustically complex scenes" [14]. To do this, composers must create virtual acoustic spaces and arrange sources over time. For composers, spatialisation requires the specification of many features and parameters, such as a sound source's: (i) spatialisation method (channels or objects); (ii) location; (iii) orientation; (iv) directivity; (v) reverberant relationship to virtual space. In order to produce motion or other dynamic effects, the temporal evolution of these parameters must also be specified. Given the level of control, working at a musical level, the construction of spatial music still remains challenging [15]. This has implications for how composers work with sonic materials and their ideas. For example, to momentarily focus in on a single audio object, or zoom out to understand the artwork as a whole, composers need to balance perceptual, technical and aesthetic decisions.

Social Interaction at Work in the Studio

Collaborative spatial music composition requires communication, imagination and action in a shared space, making it an interesting site for the exploration of creative sense-making. Previous ethnographic studies of collaborative interaction in music production have highlighted the importance of mobile devices and social media in distributed workflows [36], and how contextual metadata are created that relate analog and digital materials [35]. Alongside these studies, research needs to address design requirements for collaborative workspaces that mediate levels of human interaction, in remote or co-located situations. For instance, co-located teamwork relies not only on verbal communication but also spatially oriented interactions around shared physical artefacts [26, 25, 51]. So to design for *spatial* collaborative systems, analysis needs to understand user's interactions within social space and heterogeneous device ecologies.

In professional song-writing for modern pop music, six non-linear and interacting processes feature - stimulus, approval, adaptation, negotiation, veto and consensus [6]. The skills to generate a suitable *Stimulus*, in relevant media, is the traditionally "musical" part of creative composition. But it is not only audible stimulus that are a resource for joint musical creativity [40]: computer interface feedback [10], drawings [46], posture [21] and gesture [41] can each provide relevant cues. All these externally represented stimuli can be used to structure co-writing. The processes of social evaluation in the stimulation evaluation model of musical co-writing (Veto, Consensus, Approval, Adaptation, Negotiation) retain common language meanings but are used with respect to stimuli.

Ethnomethodological [13] and Distributed Cognition [29][2] analysis can provide granular understanding of collaborative action and process in music making [40, 8, 9]. Such studies focus on the social and material structure of collaborative interaction, paying particular attention to the sequential character of work in creative collaboration. For our purposes, the general process of work is co-authoring spatial music using digital interfaces, but interactionally this resolves to member's methods of negotiating an understanding and acting on it. As analysis of everyday social action [18], the interactional process of arriving at a decision suitable to collaborators is iterative, and based on jointly available resources for communication [28]. The moment-by-moment process of setting problems, resolving them and acting, highlights the *indexicality* of musical activity [19, 8]. Simply, musicians can seize opportunities as they emerge interactively, and build on each others contributions. It is gesture's linguistic, spatial, and musical indexicality that this paper explores, demonstrating it as a relevant form of action to understand collaborative spatial music composition.

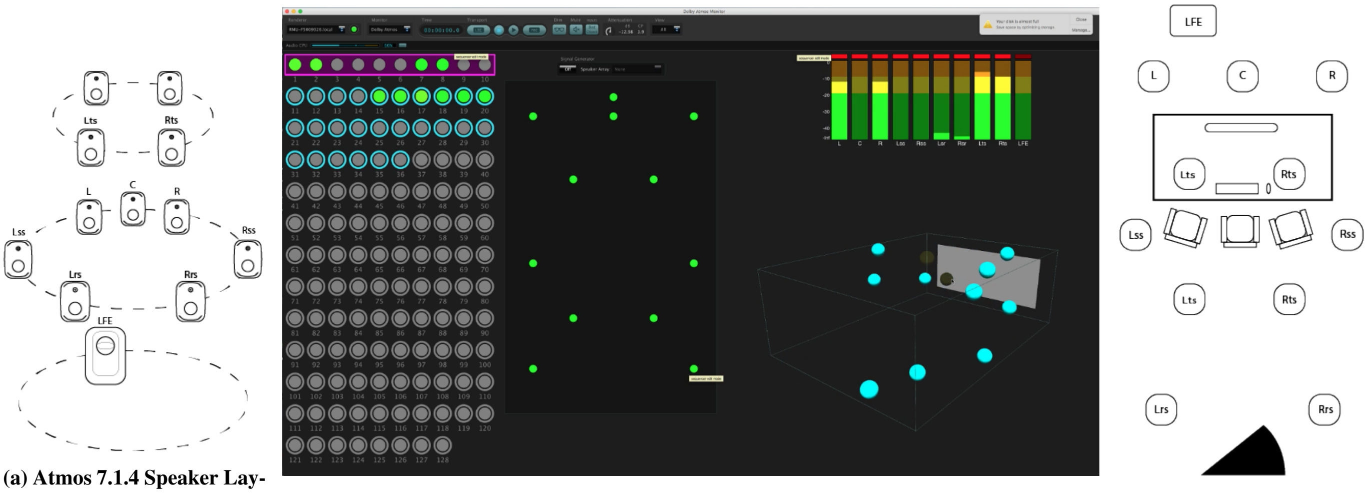
Sketching sound in space through gesture

Sketching is a well documented creative process, where there is dynamic interplay between design ideas and sketches [16]. Music practice has explored sketches and visual notation for a long time [47]. But also non-musical gesture[3] has its place

[1] Extended reality (XR) is an emerging term to encapsulate augmented, virtual, and mixed reality technologies [30]. These are typically immersive 3D interaction environments that utilise spatial computing with forms of audio-visual representation.

[2] Distributed Cognition is system-level analysis of cognition focussed on propagation and transformation of knowledge through representations, both inside the heads of individuals and in the world as embodied agents.

[3] Actions that do not produce sound



(a) Atmos 7.1.4 Speaker Layout **(b) Atmos Monitor with Speaker Layout, Monitoring and Room Visualisation** **(c) Room layout**

Figure 2. Dolby Atmos Setup. Speaker position abbreviations: Left (L); Right (R); Center (C); Low-Frequency Effects (LFE); Left side surround (Lss); Right side surround (Rss); Left Rear Surround (Lrs); Right Rear Surround (Rrs); Left Top Surround (Lts); Right Top Surround (Rts).

in certain practices. For instance in Indian raga music, improvised melodic action gestures form an embedded part of musical pedagogy [41]. By tracing curves in space, stretching virtual materials, sculpting virtual objects, gestures serve as three-dimensional, kinetic representations of melody. But why is gesture important to co-writing spatial music?

Gesture plays an important role in spatial reasoning and communication [4]. A series of concepts have been proposed to describe gesture's properties for supporting spatial activities, some include: "virtual diagram" [32], "tracking gestures" [39], and "virtual maquettes"[22]. Kang et. al. [32] coined the phrase "virtual diagrams created in the air" for gestures that relate information about systems. Like paper or computer diagrams, gestures can display elements and relationships using a publicly available sub-strate, air [32]. At a simple level, points and lines provide basic gestural primitives to construct virtual diagrams [49], for instance telling someone directions by defining landmarks with points and lines that interconnect them. "Tracking" gestures are a behaviour architects' used to develop a three dimensional understanding from two dimensional schematics [39]. Participants gesture, talk and draw, to establish a collective imagination of the domain problem, a room. A "tracking" gesture, tracing along features in the air above a paper diagram, alludes to information not present in the physical materials. In another study of architecture [22, 23], the progressive interactional sequences of talk, gesture and drawing resolved 'sub-spaces' of the work [23], where architects' create malleable 3D "virtual maquettes" through temporary combinations of schematics, paper sketches, and gestures [22].

So what conceptualisation of gesture is useful to understand spatial music production? The topology of a virtual diagram is rooted in static visual communication, making it too restrictive and possibly misleading for the type of work we observe. In the architecture studies (tracking and maquettes), the volumetric quality and relationship to socio-material resources is relevant. But again, these conceptualisations were forged for the purpose of describing a spatio-visual object of work (a building). The requirement to relate action to the fluid/plastic medium of sound to space and socio-material interaction forces a disambiguation from previous terminology. For our work, we need to situate the nature of sound into a conceptualisation. Sound is ephemeral and spatial, but production processes rely on visual "drawing" tools. We propose the notion of "Shaping Sounds" to situate sound and gesture in imaginary constructions that are collaboratively manipulated. "Shaping" consists of gestural sketches and sculptural actions that occupy space. We imply a dynamic interplay of ideas through action, but rather than just static schematics, these actions are temporal, fluid and possibly volumetric. Using "Shaping Sounds" as a term, we imply that perceptual, spatial, temporal and aesthetic entities can be worked with. As part of a co-creative activity we draw attention to how gestural shapes are available for collaborative manipulation. They retain spatial information under transformations, linking material and imaginary objects with respect to music. As a conceptualisation, "Shaping Sounds" inhabits aspects of all the previous types presented. Using this metaphor we can analyse what abstractions are alluded to through specific actions. These could include musical note and rhythm relationships, spatial positions of objects, or a mixture of both.

RESEARCH APPROACH

Design and Methodology

Research fieldwork involved shadowing, recording videos, some discussion during process, and conducting of interviews in-situ. This paper focuses on one eight hour session related to spatial audio, though other fieldwork sessions were conducted with participants that covered the more normal practices of professional music production (about 50 hours). As ethnomethodologically-informed design ethnography [13], our work explores the situated interactional and material resources of creative collaboration, analysing how they are used, and how they characterise the work of professional music making. Ethnomethodological approaches are used widely in the study of collaborative work [26, 27]. As our primary analysis method, we use a subset of this field video-based Interaction

Analysis [31]. These methods have previously been used for multi-modal analysis of collaborative musical interaction [34, 52] and in-the-wild creativity research [44]. The interaction analysis involved the following phases:

1. Collect data - Videotape naturally occurring encounters as part of a broader ethnographic study, using participant observation with informed consent. This session used three camera angles of the room and a screen recording of DAW interaction. Cameras where positioned approximately at the Left, Centre, and Right speaker positions indicated in 2c. Also, room sound and participant voices were recorded as audio.
2. Make a content log - watch through the videos and describe each distinct feature with a summary of events. In design ethnography terms, this would be horizontal slicing [13].
3. Identify patterns - sequences of interaction that occur repeatedly and that provide insight into the nature of distributed creativity, in our case use of gesture.
4. Transcribe - Select data for transcription and annotation.
5. Collaborative review - Discussion of video segments in collaborative data sessions with other researchers [31, 24, 3]. Sessions focus on short segments of video data, with removal of any inferences collected during our analysis, allowing the groups analysis to surface findings in the data or errors in transcription.
6. Follow-up - Discussion of findings with participants, using videotaped segments. Goal is to elicit perspectives from the participants about our analysis and gather their reflections on the music making process.

Unit of analysis

The goal of the present study is to show how participants orchestrate communicative channels to create temporal explanations of their spatial music work. The findings presented are not a thorough description of audio production processes and techniques. Instead, we provide vignettes that include turns of speech, gesture, gaze, posture, and interface interaction. The combined transcription and analysis aims to address how sense-making is socially distributed and reactive to gesture and interface use. As a unit of analysis, we highlight gestural spatial referencing in the process of spatial music creation.We utilise a detailed form of gestural annotation that layers discrete points in time of gestures on top of each other, and we draw the gestural trajectories across time. The approach is influenced by previous work in cognitive science that evaluates gesture use in-the-wild [17, 38, 23].Our interpretation acknowledges the "contingency" of action in the moment [44], meaning that each participant did not know what the other was about to do at any moment in time. Simply, we analyse gestures occurrence and ask "why this?" and "why now?".

Participants

The participants were a professional production duo based in London called *Commands* (Kyle & Keir). They work regularly with international pop artists and record companies. Both are trained musicians with producers with live performance experience. The duo has extensive production experience in electronic dance and pop music. The pair have worked together on and off for over 10 years, with intensive professional collaborative work being conducted in the last 5 years. Before this session, *Commands* had never used the Dolby Atmos system, nor worked professionally with spatialised audio. At the beginning of the session, a Dolby Atmos music engineer introduced the technical setup and some initial workflow suggestions.

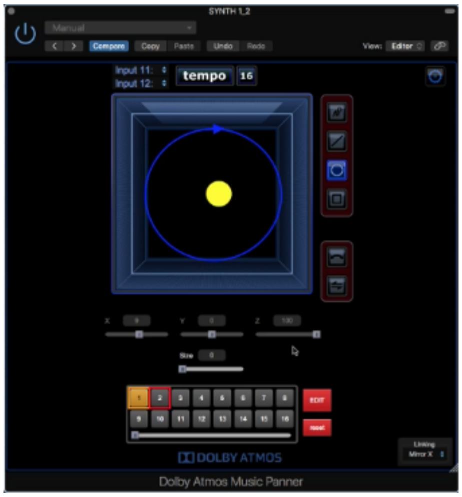


Figure 3. Atmos panner open as pop-up window. Atmos Panner has circular trajectory.

Setting

This fieldwork was conducted at the Dolby UK office in London, in their Dolby Atmos music mixing room; views of the studio can be seen in figure 1. The room contains a Dolby Atmos speaker array, figure 2a, and a computer workstation with a single monitor.

The work discussed covers a single eight hour session at the site where pop music producers *Commands* were invited to mix a track of previously produced music in Dolby Atmos. All the musical content was originally made without the intention of it being a spatial audio mix. While this work is not a standard practice for *Commands*, Dolby's expansion of the Atmos architecture has established new roles and a variety of workflows to support spatial music composition and mixing[4].

Tools

All work done by *Commands* in the data used the following tools within the computer, using a mouse and keyboard with no external physical tools or mixing equipment being utilised.

Logic Pro X Digital Audio Workstation (DAW) During the session only Logic Pro X (LPX) was used as the audio application to play, edit and mix audio files.

Dolby Atmos Panner A DAW plug-in that lets users spatially position audio objects in a Dolby Atmos mix. The plug-in provides a virtual room in its UI that is used when inputting

[4] Interested readers can view a Dolby Atmos music production session with Deadmau5 at https://www.youtube.com/watch?v=pp8RPrBWYEo

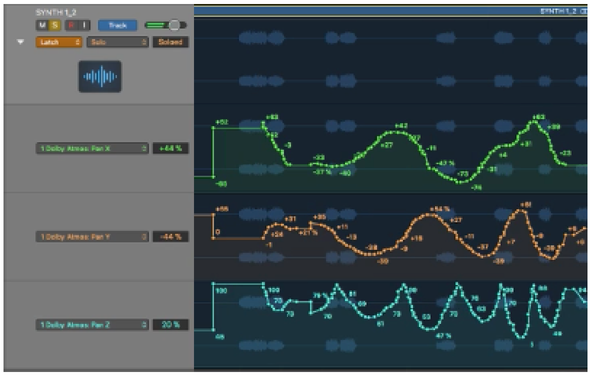


Figure 4. Example of automation data recorded by panner input.

panning position for an object, or monitoring the object position during automation playback [2]. The plug-in can be viewed in figure 3. An example of automation data recorded in from the panner to LPX can be seen in figure 4.

Dolby Atmos Monitor An application that lets users visually monitor an Atmos mix as it renders audio. A view of the monitor application can be seen in figure 2b, and the relationship of speakers to room space can be seen across figures 2a to 2b.

FINDINGS

Gestural Exchange

Figure 5 is a representation of *Commands* shaping sound phenomena through talk, gesture and tool use, as they work to inscribe spatialisation events in the composition. In this segment, *Commands* are working on the first sound source of the project, a synthesizer sound, after being introduced to the software by the engineer. This is still an exploratory phase, both of, the newly introduced features, and the content they wish to use them on. Outline of key hand gestures, related to lines (L) in figure 5:

L2-3 Drag out - Keir suggests how a panner drawing function works in the GUI, supported by hand gesture, rotating an open hand from left to right.

L6 Forward circle - Keir draws a circle in the screen space, his statement maybe a query of how it will work, gesture is on wrong plane for the actual sound movement.

L9-10 Arc - Kyle draws in the speaker space, this perhaps acts as a spatial paraphrase of Keir's proposal in L6.

L11-14 Left right points - Kyle uses series of left and right pointing gestures. Gesture process seems to allow Kyle to access words relating to intention. Phase may be rejection of the circle idea and proposal of an alternative. In follow-up, Kyle mentioned the shape of the hand and difference in scale of this movement in gesture 10 & 11 indicated a larger panning change, compared to smaller pointing gestures.

L16 Screen point - Kyle points at the screen and makes a curved left to right sweeping gesture paired with speech. This acts as a continuation of the previous proposal for motion of the sound object's sequence.

L19 Circular - Kyle draws in the screen space with his hand while Keir changes posture. This phase maybe an integration of Keir's forward circle suggestion (L6) but at a later phase of the proposed sound object sequence.

L19-20 Vocal beat points - Kyle indexes musical rhythm features using vocalisation (da da da) precisely timed with room space pointing gestures.

This gesture sequence describes a process of negotiation on how to position and locate-in-time the synthesizer sound. Each gesture highlighted offers spatial stimuli to apply to the sound object sequence. These gesture stimuli require knowledge of the spatial context they refer to, in order to be approved, adapted or vetoed. Keir demonstrates his perspective in an allocentric way, using the screen space to bracket how the sound should behave. This use of gesture and GUI representation allows Keir to "draw out" from the screen, constructing a relationship of spatial movement to the current object in the DAW. Contrasting this, Kyle's first gesture (Arc) starts using an egocentric mode, in the body/room/sound space. Additionally, Kyle's Arc gesture is interesting as it is a form of spatial paraphrasing of Keir's previous contribution. But Kyle's use of room space is not constant throughout all subsequent gestures, the Screen point (L16) highlights a transition back to an allocentric demonstration. This may be a function of it being physically easier to represent circular motions using the screen space. This example highlights the space of action as a site where multiple spatial representations must be integrated in relation to possible tool actions. Kyle and Keir construct differing maps of space and sound action that are mutually recognisable and available for collaborative manipulation. The key issue at this phase is translating mental representations into the sound space of the audio renderer by sequencing object behaviours in the GUI.

Vocal and Point

In this segment, *Commands* continue the previous phase of work on the synthesizer sound, implementing the previously described spatial pattern. Figure 6 represents a phase of action where they enter the data into the panner sequencer. Outline of gesture events related to lines in figure 6:

L10-L13 Points - Kyle, with his left hand, times finger pointing in room/body space while vocalising musical notes. The speed across gestures relates to musical and speech timing. Gestures also provide placeholders for past GUI actions.

L12 Arc - Kyle arcs his left hand to the left, with a pointed finger, in time with a staggered vocalisation. Perhaps indicating the intended level of movement within the trajectory.

L13 Stab - Kyle points at the right speaker, concurrent with speech.

L13-14 Sweep - Kyle slowly sweeps pointed finger left to point at L speaker, gesture reaches stop before clicks and speech. Gesture provides placeholder for future GUI actions.

What Kyle does is draw out an audio object's temporal, rhythmic, tonal, and spatial trajectories. It is an interwoven activity

1 Kyle Right, object (1.0)

2 Keir Do a circle (0.7) that will {1} be **drag out**
3 {2} (1.0)

4 Kyle [A: draws circle from left to right (1.9)]
5 **Left** (1.7) that's guna be quite tight (0.5)

6 Keir That's guna go {3} **like that yeah** {4}

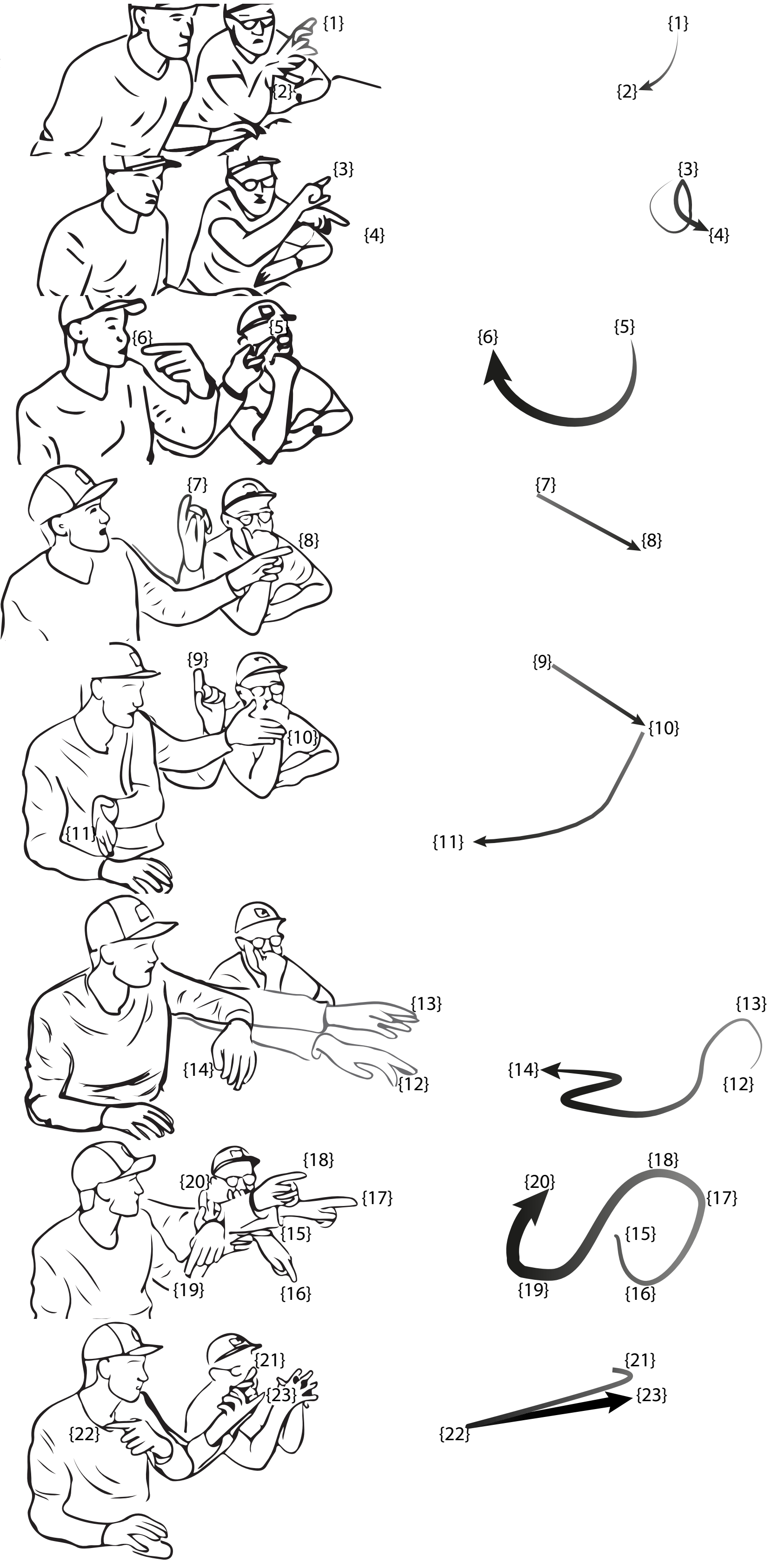
7 Kyle Yeah, well I think that's (0.5)
8 [[B: drags pointer]
9 [**well a circle is guna be** just be like {5}
10 **once around** {6} but what I'm thinking is
11 if we have just the {7} **object** {8} **in the**
12 **left**, any-if we have just some, you know,
13 **it** {9} **playing** from {10} **the left hand**
14 **side** {11} **then the right**

15 Keir Mm::hmm

16 Kyle {12} **Then** we can do {13} **a left to** {14}
17 **right one**, [like a {15},
18 Keir [Mm::hm

19 Kyle {16} **sp** {17} **spin** {18} one for {19} the
20 {20} **da**{21} **da**{22} **da**{23} I [guess
21 Keir [Yeah

22 Kyle And if we put the lock on [Starts audio
23 playback] (1.2) [Stops audio playback]
24 thats way slower than I thought [Changes
25 sequence timing division]



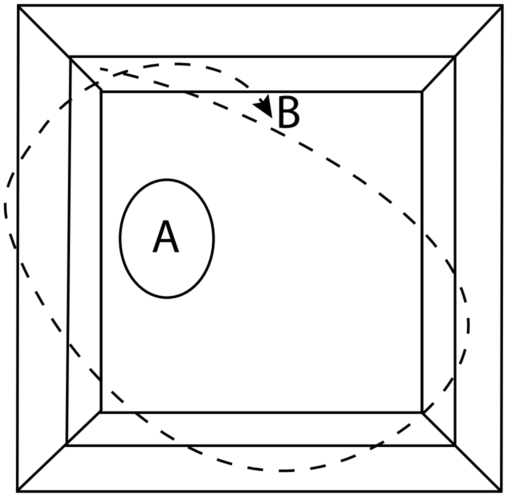


Figure 5. Gestural Exchange. Line numbers (L) listed on left. The simplified panner space is referred to as the virtual environment (VE). Gestures annotated with numbers in braces, GUI actions annotated with letters in square brackets. All gesture trajectory lines drawn in perspective of video, unless within the VE. Virtual diagrams are aligned in space. Gaps in seconds indicated in parethesis. Bold set speech transcriptions indicates what speech adjacent gestures emphasise.

1 Kyle so lets just try this, wuh, uh [A: clicks left
2 of VE] **one**

3 Keir mmhmm (1.6)

4 Kyle [[B: clicks right of VE]
5 **two** (2.9) **a line across** [drags mouse
6 pointer right to left]

7 Keir mmhmm (0.5) [**dun** {nods head to left}
8 Kyle [C: draws left arrow]
9 duhn dun (1.2)
10 {1} **duhn**
11 {2} **dahn** (0.2)
12 **duh duh** {3} {4} (0.5)
13 {5} **right then**

14 Keir mmhmm (0.2)

15 Kyle [D: click right] **duhn** (1.5)
16 {6} (1.1) [E: click left] (0.3) **duuhn**, and
17 then (1.6)

18 Keir thats you up to [five again
19 Kyle just **put** [F: clicks centre three times] so if i do eight
20

21 Keir yeah thats it {leans backwards}

22 Kyle [Starts audio playback]

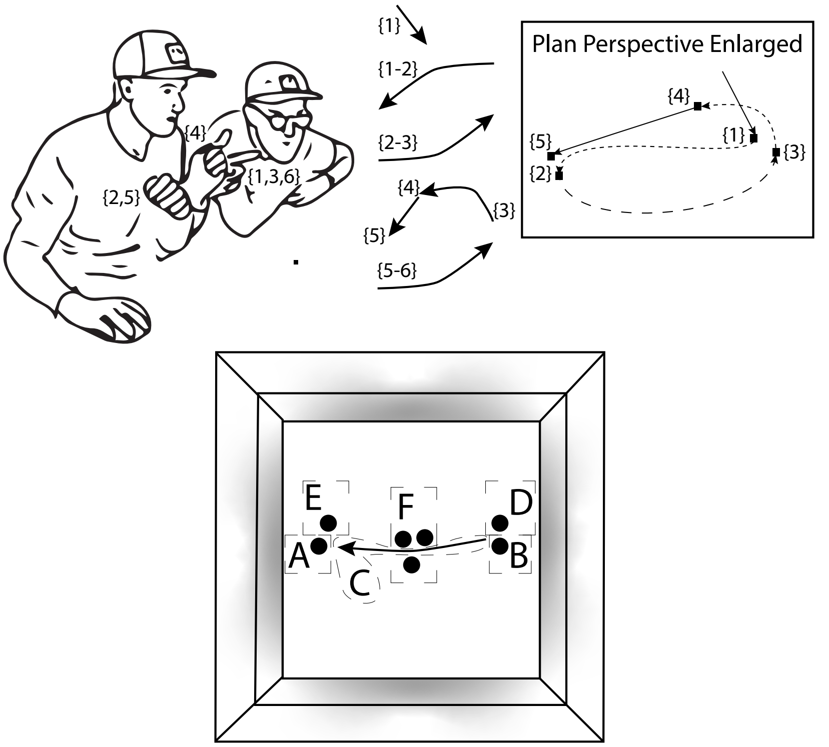


Figure 6. Vocal and Point. Gestures annotated in braces with numbers, GUI actions annotated with letters in square brackets. All gesture trajectory lines drawn in perspective of video, unless within the VE or inset. The dashed lines in the inset trajectories indicate crude speed differences: solid lines being fastest, tightly packed dashes being slower than lines, and loosely packed dashes slower still.

using tools, gesture, and internal representations that provides collaborative access to the work at hand. But analysis cannot focus on the gesture alone, or some possible mental state the speaker is externalising. It requires simultaneous attention to the activity participants are working on and the framework of action used. The activity is spatialising the synthesizer sound after consensus on a course of action. The framework of action includes the following components:

1. semantic descriptions, e.g. talk that sets its addressee a placement problem "on the right then";
2. vocalised musical descriptions e.g. musical mimicking L9-12 ("duhn dun");
3. a complex perceptual field where further action is to be located (gestures, tool interaction and visual feedback in the DAW);
4. Kyle's hand moving in that perceptual field with a specific rhythm, and a relationship to space, tools and memory; and
5. parallel representations of space that make up the complex perceptual field where actions are transformed (screen space, body space, sonic space).

The activity in progress creates a context that each can use to make inferences about what features of the complex perceptual field are being pointed at in any moment. These inferences are then used to decide what should be attended to for subsequent action through digital tools. We draw attention to the requirement for *Commands* to recognise parallel representations of perceptual space. Related to musical timing, parallel representations are used to establish shared understanding and offer opportunities to veto decisions. For instance, Kyle's use of pointing and vocalisation highlights previous and current actions in its use (L10-13). Simultaneously, Kyle can offload memory related to previous position states, and provide a shared representation of space that can be used to set Keir a placement problem. The placement problems offers Keir the opportunity to veto or consent with the current input decisions. To answer this, Keir must transform representations in memory and the complex perceptual field. This highlights how cognition is distributed across both *Commands'* internal cognition, speech, gesture, and computer resources.

During discussion of this video in the follow-up, two features arose, (i) what Kyle was doing in his words; and (ii) how focus was shared in the process of making. Kyle described that using pointing, vocalisation and the tool required thinking about how to map the ideas in the head, to the tool, at a musical level, stating "to create that thing its like stretching it out in my head. There, there, there, and I'm thinking about where the beat goes, in terms of beats per minute". Using Atmos to mix their track *Commands* described a level of shared focus in working, they focused quite specifically on processes and relationships of the tool to ideas, tracking what each other was doing.

Temporal Volumes

A further example of shaping sounds can be seen in figure 7, it represents a scaling gesture that highlights relationships of space and sound. The figure also displays explicit space-location-time referencing. Prior to the segment *Commands* completed work on the synthesiser sound and moved onto spatialising another song element; a bass track. They discussed how to get more musical impact for the song section by routing the bass track to the same or inverse spatial trajectories as the previous instrumentation track. In figure 7 they plan the following actions: slice the chorus section bass clip, place it in space, return and do inverse panning process to the verse clip. During this, ideas are presented for how phases of the track should have different spatial relationships. Musically, this would sound like hearing one part of the song happening in one place until a change of song section where it then suddenly moves to another.

At certain points in time Kyle's hands actively 'sculpt' areas of space; though it is ambiguous whether this is to be interpreted as body-space or sonic-space. Kyle's wide hand and arm gesture (L2-6) expands and contracts to emphasise a spherical/rectangular volume (like opening and closing an accordion), then dual handed panning movements are made maintaining a volume 'held' between hands (L7-9). Going from wide to narrow was controlling the spatial impression of that song part, and then pointing the segment of the song with specific relationships. Discussion of this in the follow-up interview highlighted that the wide gesture inferred to make the sound envelop the listener by coming from all speakers, whereas the return to centre (L6) emphasised a return more controlled subtle panning. Kyle utilises gestures to support the idea to selecting specific aspects of the track to have different spatial relationships based on the compositional phase (chorus vs verse). During this, the specific panning relationships are described by Kyle's dual handed points. This is interesting as Kyle is effectively pointing with a section of the song, not just an individual track. This example highlights how the spatial music composition is individuated and shaped through interaction in space, serving the collaborative goal of consensus via multi-modal stimulus evaluation.

DISCUSSION

Role of Gestural Shaping in Collaborative Spatial Music Composition

Gestural shaping of sounds can occur across many periods of collaborative activity in spatial music composition. This is because they provide an easily available method for communication about spatial concepts that are integral to the production process. During the follow-up interview, Kyle described gestures as like "a visual representation of what we were trying to do with the sound". In our findings, the phases of interaction break down into the following overlapping categories:

- *Talking about musical ideas* - deciding what ideas to implement via the software and instruments (fig. 5).
- *Embodied imagination of musical ideas* - enaction of non-instrumental gestural displays of musical ideas to decide what to take forward into practical enactment via the software and instruments (fig. 7).
- *Enacting musical ideas* - work of inputting and refining ideas to achieved desired effects (fig. 6).

In the case of spatial music, temporally evolving gestural shapes can provide a form of contextually relevant stimulus, acting as a precursor to adaptation, forming part of negotiation. In addition to stimulation, gestures could help create a context that allows veto and consensus. Gestures are acknowledged and transformed in a social way (figure 6), they act as the means to operate in, and with, a shared problem space. Their indexicality of space, language and music allows momentary references to meaningful objects of perception or cognition, located in a co-constructed problem space.

Articulation of Musical and Temporal Structures

All three examples of action contain events of musical vocalisation timed precisely with gesture. Across examples, sketches of sound position in space are used to articulate musical structures. Similar to speech gestures, the timing of the gestures move through phases: rest, significant movements, modulations, and back to rest. The timing of gestures is related to a mixture of underlying tempo information (fig. 5 L16), discrete tonal separation of the content (fig. 6 L8-11), continuous temporally evolving sonic features (fig. 7 L4-6), or phases of action related to the process of inputting information (fig. 6 L9-12). Across these types, a sonic topography relates compositional action to the digital music objects, where the spatial referent of the gesture is either sonic or interface-based, or both. It is interesting that gesture can operate at the level of individual notes and parameter actions, but also provide a capacity for continuation over phases of these actions. In the case of tonal separation and inputting information, gestures may assist the spatial and musical process of cognitive decomposition. By shaping sound in space *Commands'* can extract specific features from the musical composition to work on them using the GUI, for instance the contour of the sound movement through space. In this way, gesture helps with cognitive offloading to the environment, in an distributed cognition sense [45, 12]. This may be important as the added load of individuals computing parallel versions of space (sonic room space and abstracted 2D GUI space), could make spatial music composition with 2D GUIs a burdensome cognitive process. This idea is reflected in follow-up discussion where Kyle states he used gesture and vocalisation while working to "lock in" his mind an idea in the making, especially given that this was a new tool and work process.

Ambiguity as a Resource for Interaction

In a functional interpretation, gestures are used to shape sound phenomena that either mark current action or describe further development of actions on the composition. But as representations they are not just mere schematics of action, they are tools of cognition, both individual and collaborative. They develop ideas through embodied description activating memory and enabling projection. We highlight ambiguity as a feature of spatial communication that sketches of sound in space use well. As a representational affordance, ambiguity is the property of a medium that allows its use to appropriate artefacts, concepts, relationships, and surroundings [46]. As gestures are not permanent, they are not available for revision and scrutiny in the same way as static visual diagrams like a 2D GUI [50]. By being visible and contextually relevant, but not as concrete as a drawing, the opportunity for interpretation is key. These properties mean gesture can act as a bridge between action and abstract thought.

[Audio playback stops]

1 Kyle What we could do is even [zooms out
2 LPX timeline] is even {1} **slice the**
3 **chorus bass** {2} and have
4 **that** {3} go
5 {4} **everywhere** {5} (0.6)
6 and then go {6} **back** to
7 {7} **verse's**
8 {8} **dun**
9 {9} **duh**
10 [{10} **da**
11 Keir [Yeah
12 Kyle Do that shit
13 Keir Yeah why not
14 Kyle Why not man
15 Keir Why not

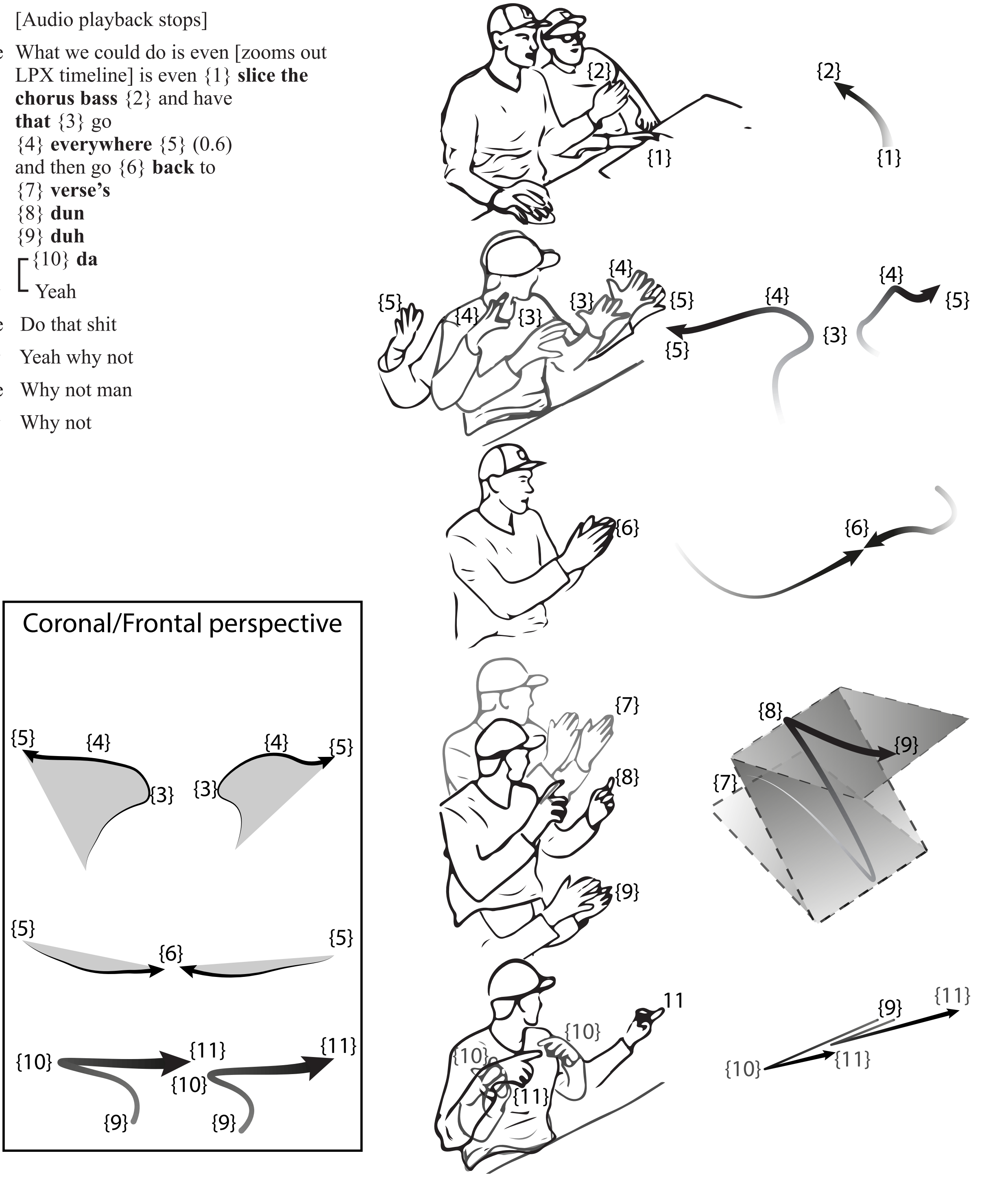


Figure 7. Temporal Volume example where volumetric gestures are deployed with musical information. All gesture trajectory lines drawn in perspective of video, unless within the inset. The inset uses shaded areas to show negative spaces more clearly. The trajectory and volume drawn for gesture moves 7 - 9 is enlarged for detail.

Support of Perception and Communication of Spatio-Temporal Information

In the Gestural Exchange example, *Commands* used a mixture of first and second-order isomorphisms to describe relationships of sound objects [48]. For instance, they described sound both in terms of their perception, but also utilise the abstractions presented in the GUI. As an interactional achievement this is quite interesting, the various gesture spaces used by participants can switch seamlessly between GUI, body, room, sonic and speaker space. This highlights the flexibility of gesture and spatial cognition, where information coherence can be achieved through rhetorical gestural combinations, where each gesture instance transforms space in different ways. This is similar to cases of multiple gesture deployments in verbal repetition, where gesture can be used differently each time to create a novel contribution [33, pp. 151-156]. Finally, this highlights that coherence in communication about gestural-spatial-melodic-temporal relationships is an interactional achievement, rather than merely a given, that relies on access to shared audio-visual media (screens, air).

In the case of spatial music, the intertwined technical and aesthetic requirements means offloading computation onto forms of representation is essential and evident. This is also because, sound itself is like an action, it unfolds in time requiring immediate perceptual attention. This puts burden on composers to be able to reflect collaboratively on the compositional structure, the spatio-temporal arrangement of elements, and tonal balance of the mix. During the follow-up interview, Keir described gesture use as part of "solving a puzzle", in relation to the difficulty of using the tool. This means versions of spatial sound sketches and sculptures are used to explicate new knowledge being developed as the composition takes form. The indexical character of gestures allows points of focus for collaborative problem solving, by assisting participants to package complex information for evaluation and action.

IMPLICATIONS FOR DESIGN

The themes discussed previously can allow us reframe design questions. Such as, what features of XR technologies can build on top of our natural skills in shaping sounds for each other? In our findings, sense-making happened in a distributed way across people, objects and actions. This has implications for spatial information design to support shared space creation. The abstraction of decomposing 3D space into 2D GUIs forces a *tooling of space* that creates patterns of use, seen through *Commands*' active interpretation between screen and various egocentric spatial frames. *Commands* begin to think through the sound creation given the GUI abstraction. This symbiosis of space and music, through the tool, could be expanded by spatialised visualisation and interaction using XR technologies. Take for instance embodied virtual or mixed reality, by using our bodies in space, with expanded representational opportunities, working with sound can merge with our understanding of it, in an active process.

By focusing on action in phases of work, we highlight the importance of being able to jointly focus, and act, to develop problems and solutions in the process of work. This means design of systems to support similar activities needs to be sensitive to shared space. Shared space is more than just visual access. It is the relationship of artefacts, space, memory and interlocutors. It needs to be mutually accessible to support effective non-verbal communication in detailed ways [20]. This argument is of particular import to design interventions that technologically mediate social interaction, meaning shared space is either degraded or augmented, such as MR or VR. As targets of perception and action must maintain stable relationships for the process of sense-making to occur. In the case of VR, it is important to support more than just pointing, gaze tracking and shared visual access to tools. Systems need to provide ways to think together through forms of improvised representation or gestural depiction.

CONCLUSION

As a naturally available method of reasoning and communication, gestures provide an understanding of how people communicate, and may allude to how people think about space, and in particular its relationship to musical forms. Also, people can develop complex relationships to aid communication about spatial information. Gesture can "draw" out the relationships of spatial, temporal, and narrative features. This means that gesture can operate across physical and imaginary space. For collaborative spatial music, gesture assists description of a sound scene, and speculation into future states, based on the current shared context. This is a key point, as collaborators must reflect and discuss what exists, but also develop and negotiate what to do next.

Our position in this paper is that the analysis of how the creative process exists as a shared resource can provide insights into how to design support for collaborative spatial music composition. We have worked from the perspective that reasoning is a mundane achievement: it is supported by social, material and linguistic contexts that serve sense-making purposes endogenous to specific activities. We proposed that the conceptualisation of "Shaping Sounds" that can be used to understand co-creative process in spatial music. Using this metaphor, we are able to analyse and interpret how interactional resources support sense-making through distributed cognition and the affordances of gesture as a medium. By isolating how *Commands* shaped understanding of sounds for each other, we drew attention to how social interaction is juxtaposed with material practices, in fluid problem spaces, that are co-constructed with an intrinsic relationship to spatiality.

ACKNOWLEDGMENTS

The authors would like to thank *Commands* for sharing their process and the engineers and staff of Dolby UK. Also, the first author would like to sincerely thank Claude Heath for his time and feedback on the gestural annotation methods used. This work is supported by EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1).

REFERENCES

[1] 2014. *Dolby Atmos Next Generation Audio for Cinema*. Technical Report 3. Dolby Laboratories. `http://www.dolby.com/us/en/professional/cinema/products/`

dolby-atmos-next-generation-audio-for-cinema-white-paper.pdf

[2] 2017. *Dolby Atmos production suite guide.* Technical Report April. Dolby Laboratories.

[3] Saul Albert. 2017. Research Methods: Conversation Analysis. In *The Routledge Handbook of Discourse Processes*. 99.

[4] Martha W Alibali. 2005. Gesture in Spatial Cognition: Expressing , Communicating, and Thinking About Spatial Information. *Spatial Cognition and Computation* April 2013 (2005), 37–41. DOI: http://dx.doi.org/10.1207/s15427633scc0504

[5] Enda Bates and Francis M. Boland. 2016. Spatial Music, Virtual Reality and 360 Media. *AES Conference on Audio for Virtual and Augmented Reality* (2016), 1–8.

[6] Joe Bennett. 2012. Constraint, Collaboration and Creativity in Popular Songwriting Teams. In *The Act of Musical Composition: Studies in the Creative Process*, D. Collins (Ed.). SEMPRE Studies in the Psychology of Music, Farnham, Chapter 6, 139–169.

[7] Joe Bennett. 2014. Collaborative Songwriting – The Ontology Of Negotiated Creativity In Popular Music Studio Practice. *Journal on the Art of Record Production* 1 (2014), 1–8.

[8] Phillip Brooker and Wes Sharrock. 2013. Remixing Music Together : The Use and Abuse of Virtual Studio Software as a Hobby. In *Ethnomethodology at Play. Directions in Ethnomethodology and Conversation Analysis*, Peter Tolmie (Ed.). Ashgate, London, 135–155.

[9] Phillip Brooker and Wes Sharrock. 2016. Collaborative Music-Making with Digital Audio Workstations: The n-th Member as a Heuristic Device for Understanding the Role of Technologies in Audio Composition. *Symbolic Interaction* 39, 3 (2016), 463–483. DOI: http://dx.doi.org/10.1002/SYMB.238

[10] Nick Bryan-Kinns. 2012. Mutual engagement in social music making. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering* 78 LNICST, 0 (2012), 260–266. DOI: http://dx.doi.org/10.1007/978-3-642-30214-5_35

[11] David Byrne. 2012. How music works. (2012). http://site.ebrary.com/id/10602980

[12] Andy Clark. 1998. *Being There: Putting Brain, Body, and World Together Again*. 292 pages. DOI: http://dx.doi.org/10.2307/2998391

[13] Andy Crabtree, Peter Tolmie, and Mark Rouncefield. 2012. *Doing Design Ethnography*. Springer. 205 pages. DOI:http://dx.doi.org/10.1007/978-1-84996-272-8

[14] William J Davies. 2015. Cognition of soundscapes and other complex acoustic scenes. In *Inter Noise*. San Francisco.

[15] Jérémie Garcia, Thibaut Carpentier, and Jean Bresson. 2017. Interactive-Compositional Authoring of Sound Spatialization. *Journal of New Music Research* 46, 1 (2017), 20.

[16] Gabriela Goldschmidt. 1991. The Dialectics of Sketching. *Creativity Research Journal* 4, 2 (1991), 123–143. DOI: http://dx.doi.org/10.1080/10400419109534381

[17] Charles Goodwin. 2003. Pointing as situated practice. *Pointing: Where Language, Culture, and Cognition Meet* (2003), 217–241. DOI: http://dx.doi.org/10.4324/9781410607744

[18] Charles Goodwin and Marjorie Harness Goodwin. 1992. Assessments and Construction of Context. In *Rethinking Context*, Alessandro Duranti and Charles Goodwin (Eds.). Cambridge University Press, 147–190.

[19] Maya Gratier. 2008. Grounding in musical interaction: Evidence from jazz performances. *Musicae Scientiae* (2008), 71–110.

[20] Patrick G. T. Healey and S.a. Battersby. 2009. The interactional geometry of a three-way conversation. *Proceedings of the 31 st Annual Conference of the Cognitive Science Society* (2009), 785–790. http://141.14.165.6/CogSci09/papers/138/paper138.pdf

[21] Patrick G. T. Healey, Joe Leach, and Nick Bryan-Kinns. 2005. Inter-play: Understanding group music improvisation as a form of everyday interaction. In *Proceedings of the 1st International Forum on Less is More-Simple Computing in an Age of Complexity*.

[22] Patrick G. T. Healey and Charlotte R. Peters. 2007. The conversational organisation of drawing. *1st International Workshop on Pen-Based Learning Technologies, PLT 2007* (2007). DOI: http://dx.doi.org/10.1109/PLT.2007.25

[23] Claude Heath and Patrick G. T. Healey. 2011. Making Space for Interaction: Architect's Design Dialogues. In *Gesture Workshop*. Athens, Greece.

[24] Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research*. SAGE.

[25] Christian Heath, Marina Jirotka, Paul Luff, and Jon Hindmarsh. 1995. The Individual and the Collaborative : the Interactional Organisation of Trading in a City Dealing Room. *Journal of Computer Supported Cooperative Work* 3, 1 (1995), 147–165.

[26] Christian Heath and Paul Luff. 1991. Collaborative activity and technological design: Task coordination in London Underground control rooms. *Proceedings of the Second European Conference on Computer-Supported Cooperative Work* (1991), 65–80. DOI: http://dx.doi.org/10.1007/978-94-011-3506-1_5

[27] Christian Heath and Paul Luff. 1992. Collaboration and control: Crisis Management and Multimedia Technology. *Computer Supported Cooperative Work: CSCW: An International Journal* 1 (1-2), 1990 (1992), 69–94.

[28] John Heritage and Geoffrey Raymond. 2005. The Terms of Agreement: Indexing Epistemic Authority and Subordination in Talk-in-Interaction. *Social Psychology Quarterly* 68, 1 (2005), 15–38.

[29] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: Toward a new foundation for human-computer interaction research. *Transactions on Computer-Human Interaction* 7, 2 (jun 2000), 174–196. DOI:http://dx.doi.org/10.1145/353485.353487

[30] Jay Jantz, Adam Molnar, and Ramses Alcaide. 2017. A brain-computer interface for extended reality interfaces. In *SIGGRAPH '17 ACM SIGGRAPH 2017 VR Village*, Vol. 13. 73–84.

[31] Brigitte Jordan and Austin Henderson. 1995. interaction analysis: Foundations and practice. *Journal of the Learning Sciences* 4, 1 (1995), 39–103. DOI: http://dx.doi.org/10.1207/s15327809jls0401_2

[32] Seokmin Kang, Barbara Tversky, and John B. Black. 2015. Coordinating Gesture, Word, and Diagram: Explanations for Experts and Novices. *Spatial Cognition and Computation* 15, 1 (2015), 1–26. DOI: http://dx.doi.org/10.1080/13875868.2014.958837

[33] Adam Kendon. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.

[34] Robin Laney, Chris Dobbyn, Anna Xambó, Mattia Schirosa, Dorothy Miell, Karen Littleton, and Nick Dalton. 2010. Issues and techniques for collaborative music making on multi-touch surfaces. *7th Sound and Music Computing Conference* (2010).

[35] Glenn McGarry, Peter Tolmie, Steve Benford, Chris Greenhalgh, and Alan Chamberlain. 2017. "They're all going out to something weird": Workflow, Legacy and Metadata in the Music Production Process. *Proceedings of the 2017 ACM Conference on Computer-Supported Cooperative Work and Social Computing* (2017).

[36] Sean McGrath, Alan Chamberlain, and Steve Benford. 2016. Making music together : an exploration of amateur and pro-am Grime music production. In *Audio Mostly*, Vol. 2016. 4–6.

[37] F Melchior, A Churnside, and S Spors. 2012. Emerging Technology Trends in Spatial Audio. *SMPTE Motion Imaging Journal* 121, 6 (2012), 95–100. DOI: http://dx.doi.org/10.5594/j18221

[38] Keith Murphy. 2003. Building Meaning in Interaction: Rethinking Gesture Classifications. *Crossroads of Language Interaction and Culture 2003 UC Regents CA Vol 5. p. 27-47 2005* 5, 1972 (2003), 1–10. DOI: http://dx.doi.org/10.1007/s13398-014-0173-7.2

[39] Keith M. Murphy. 2005. Collaborative imagining: The interactive use of gestures, talk, and graphic representation in architectural practice. *Semiotica* 2005, 156 (2005), 113–145. DOI: http://dx.doi.org/10.1515/semi.2005.2005.156.113

[40] Shahin Nabavian and Nick Bryan-Kinns. 2006. Analysing group creativity: A distributed cognitive study of joint music composition. In *Proceedings of cognitive science*. 1856–1861. http://www.eecs.qmul.ac.uk/{~}nickbk/papers/AnalysingGroupCreativityA

[41] Matt Rahaim. 2008. Gesture and melody in Indian vocal music. *Gesture* 8, 3 (2008), 325–347. DOI: http://dx.doi.org/10.1075/gest.8.3.04rah

[42] Francis Rumsey. 2001. *Spatial Audio*. 240 pages.

[43] Francis Rumsey. 2016. Virtual Reality: Mixing Rendering, Believabiity. *J. Audio Eng. Soc* 64, 12 (2016), 1073–1077. http://www.aes.org/e-lib/browse.cfm?elib=18538

[44] R. K. Sawyer and S DeZutter. 2009. Distributed Creativity: How Collective Creations Emerge From Collaboration. *Psychology of Aesthetics, Creativity, and the Arts* 3, 2 (2009), 81–92. DOI: http://dx.doi.org/10.1037/a0013282

[45] Mike Scaife and Yvonne Rogers. 1996. External cognition: How do graphical representations work? *International Journal of Human Computer Studies* 45, 2 (1996), 185–213. DOI: http://dx.doi.org/10.1006/ijhc.1996.0048

[46] Jean-Baptiste Thiebaut and Patrick G. T. Healey. 2007. Sketching Musical Compositions. *Cognitive Science Society Journal* (2007), 1079–1084.

[47] Jean-baptiste Thiebaut, Patrick G. T. Healey, and Nick Bryan-Kinns. 2008. Drawing Electroacoustic Music. In *ICMC*.

[48] Barbara Tversky. 2014. The Cognitive Design of Tools of Thought. *Review of Philosophy and Psychology* 6, 1 (2014), 99–116. DOI: http://dx.doi.org/10.1007/s13164-014-0214-3

[49] Barbara Tversky. 2017. Gestures can create diagrams (that are neither imagistic nor analog). *Behavioral and Brain Sciences* 40 (2017), e73. DOI: http://dx.doi.org/10.1017/S0140525X15003088

[50] Barbara Tversky, Azadeh Jamalian, Valeria Giardino, Seokmin Kang, and Angela Kessell. 2013. Comparing Gestures And Diagrams. In *Proceedings of the 10th International Gesture Workshop*.

[51] Robin Wolff, Dave J. Roberts, Anthony Steed, and Oliver Otto. 2007. A review of telecollaboration technologies with respect to closely coupled collaboration. *International Journal of Computer Applications in Technology* 29, 1 (2007), 11–26. DOI: http://dx.doi.org/10.1504/IJCAT.2007.014056

[52] Anna Xambó, Robin Laney, Chris Dobbyn, and Sergi Puig Jordà. 2012. Towards a taxonomy for video analysis on collaborative musical tabletops. In *Proceedings of BCS HCI 2012 Workshops: Video Analysis Techniques for Human-Computer Interaction*. 1–4.