

Chapter 8

Supporting Sonic Interaction in Creative, Shared Virtual Environments



Liang Men and Nick Bryan-Kinns

Abstract This chapter examines user experience design for collaborative music making in shared virtual environments (SVEs). Whilst SVEs have been extensively researched for many application domains including education, entertainment, work and training, there is limited research on the creative aspects. This results in many unanswered design questions such as how to design the user experience without being detrimental to the creative output, and how to design spatial configurations to support both individual creativity and collaboration. Here, we explore multi-modal approaches to supporting creativity in collaborative music making in SVEs. We outline an SVE, LeMo, which allows two people to create music collaboratively. We then present two studies; the first explores how free-form visual 3D annotations instead of spoken communication can support collaborative composition processes and human–human interaction. Five classes of use of annotation were identified in the study, three of which are particularly relevant to the future design of sonic interactions in virtual environments. The second study used a modified version of LeMo to test the support for a creative collaboration of two different spatial audio settings, which according to the results, changed participants’ behaviour and affected their collaboration. Finally, design implications for the auditory design of SVEs focusing on supporting creative collaboration are given.

L. Men (✉)
Liverpool John Moores University, Liverpool, United Kingdom
e-mail: l.men@ljmu.ac.uk

N. Bryan-Kinns
Queen Mary University of London, London, United Kingdom
e-mail: n.bryan-kinns@qmul.ac.uk

8.1 Introduction

Music has long been produced in social and collaborative ways [16, 67], being inherently multi-modal, music making includes not only the produced sound itself but also other presentations such as body posture[25], physical activation of the instrument [7], and written symbols and sketches [40, 66] to manage the joint creation and production of music. Many of these modalities such as body position are promoted by the physical proximity of musicians. Immersive virtual environments (VEs) provide a great opportunity to mimic these multi-modal experiences and to explore radical sonic interaction design spaces for collaborative music making (CMM) [17, 70], such as telepresence for networked performance and composition. Indeed, whilst many screen-based collaborative systems treat users as outsiders looking in [3], VEs offer an opportunity to truly immerse people into interactions. Compared to traditional media, VEs may provide a greater sense of community and more intuitive interactions [68], and offer new forms of human–computer interaction [36] and interpersonal interaction [34]. Furthermore, VEs have some unique advantages over other media to simulate multi-modal senses and enable people to interact in a natural way that is similar to the real world.

However, although VEs have become a hot topic and have been researched in depth and the potential of multi-user immersive virtual reality to promote social activities has been well established (see AlterspaceVR,¹ Venues from Oculus²), little attention is paid to interpersonal interactions in creativity, which includes collaborative sonic interactions, e.g. CMM. This raises many open research questions on how to design user experiences in VEs to support collaborative sonic interactions, such as CMM. In this chapter we will explore two design features of SVEs, trying to understand their roles in supporting collaborative sonic interaction: i) visual annotation and ii) acoustic attenuation.

We will start by reviewing the related work in related areas. Then two studies will be presented, with each exploring one of the two features. Finally, the findings of the two studies will be compared and implications for supporting collaborative sonic interaction in SVEs will be proposed.

8.2 Shared Virtual Environments

The term VE can be traced back to the early 1990s [12], and it emerged as a competitive term to virtual reality (VR). Both are usually equally used to refer to the world created entirely by computer simulation [32]. In the mid-1990s, the development of network technology made it possible to connect many users in the same VE, prompting the shared virtual environments (SVEs) [53]. In addition to “SVEs”, other similar terms being used include multi-user virtual environments, multi-user

¹ AltSpaceVR: <https://altvr.com>.

² Venues: <https://www.oculus.com/experiences/quest/3002729676463989/>

virtual reality [18], collaborative virtual environments (CVEs) [75] and social virtual reality (SVR) [19]. To stay consistent, we will herein use the term SVEs to refer to VE systems in which users experience other participants as being mutually present in the same environment and can interact inter-personally [53]. Whilst single-person VEs concern how to create detailed (visual) simulations, the design of SVEs usually prioritises enabling collaboration between users [41]. By providing a natural medium for three-dimensional collaborative work [6] and allowing multiple people to interact with each other, SVEs are considered emerging tools for a variety of purposes, including community activities [31], online education [51], distributed work and training [42], and gaming and entertainment [45, 47]. Despite this, there is little research in the field of supporting collaborative creativity (such as CMMs), which presents the necessity to explore the design space to support the rich forms of interpersonal interaction inherent in CMMs, and leaves many open questions: whether collaborative creativity in SVEs follows a similar pattern with real-world collaborative creativity or not; how to design the virtual environments support creative collaboration is also unclear, see [2]. For further discussions on these issues, refer also to Chap. 6.

8.2.1 Embodiment in Collaborative Virtual Environments

Our bodies provide continuous and immediate information about our presence, activity, attention, availability, mood, status, location, identity, capabilities and many other factors to ourselves and others, hence using body language explicitly to facilitate communication is recommended [3]. Questions have been voiced in regard to embodiment, including the impact of embodiment on users' social communication and behaviour [68], how the avatars' appearances and behaviours impact users' sense of presence [20, 38, 57, 64] and co-presence [48, 59]. Research suggests that the embodiment plays an important role in conveying presence, location and identity [3, 4], all of which are crucial to the success of collaboration [16, 21]. Social interactions in the real world and in virtual environments are regulated by the same social norms [73]. An appropriate use of embodiment can enhance the sense of telepresence [43], the sense of social presence (the feeling that others are present with the user in the mediated environment) [3, 43] and promote the sense of community [52]. Having embodiment is also beneficial to achieve a better sense of co-worker's locations, actions, intention and construction of workspace awareness, see [24]. The embodiment can also create a strong sense of identification, which is essential in collaboration since it is a fundamental component in creating workspace awareness [24], and it can influence collaboration both positively and negatively in group work situations [21]. Mutually engaging interactions can be significantly increased with proper awareness of the identity of others [16], and in VEs, to a large extent, the identification is shaped by the embodiment. As a result, embodiment decisions are critical and can influence the quality and scope of collaboration in VR [68]. The

avatar might be as basic as a T-shape with eyes to indicate orientation and viewing direction, or as sophisticated as a full 3D body scan of the user [58].

8.2.2 Collaborative Music Making

As previously discussed, music making, as a collaborative activity that relies on common goals, understanding and good interpersonal communication, has long been a key form of collaborative creativity (cf. [16, 67]). Although music making tools for multiple users have become more and more popular with the aid of digital technologies, this field remains fairly unexplored [29]. In 2003, Blaine and Fels [8] explored the design criteria of CMM systems and pointed out the main features including the media used, player interaction, learning curve of systems, physical interfaces and so on. In the same year, inspired by Rodden's Classification Space for collaborative software [49], otherwise known as groupware, Barbosa developed the Networked Music Systems Classification Space [1], which classifies CMM systems in terms of the time dimension (synchronous/asynchronous) and space dimension (remote/co-located). Examples based on tangible user interfaces include reacTable, where multiple users can construct and play the instrument by moving the tangible objects on the table [29], and Jam-O-Drum [9], which enables participants to join collaborative, musical improvisation. The Music Room provides a room-scale experience, allowing people without music expertise to compose original music inside an interactive space [39]. Sync'n'Move enables two users to explore a multi-channel pre-recorded music piece and users can generate an audio content by synchronising their movements using mobile phones as a collaborative interface. Another phone-based system is Daisyphone [13], which provides shared editing of short musical loops. Other examples include BilliArT [11], which offers a co-located music-making experience, and Ocarina [69], which provides a distributed experience. Though many CMMs have been developed, most of them rely on users to be in a relatively fixed position, e.g. in front of a computer [72]. Potentially, the head tracking and spatialised audio provided by VEs can be applied to break this chain and free users. However, this research area is little explored, especially for the collaborative aspect.

8.3 LeMo: An SVE Supporting CMM

To build a basis for exploring CMM in SVEs, we created Let's Move (LeMo³), which enables two users to manipulate virtual music interfaces together in an SVE to create a music loop, see Fig. 8.1. LeMo was programmed in Unity, and models and textures were made in Cinema 4D and Adobe Photoshop, respectively. The run-

³ More information is available at:
<https://sites.google.com/view/liangmen/projects/LeMo>

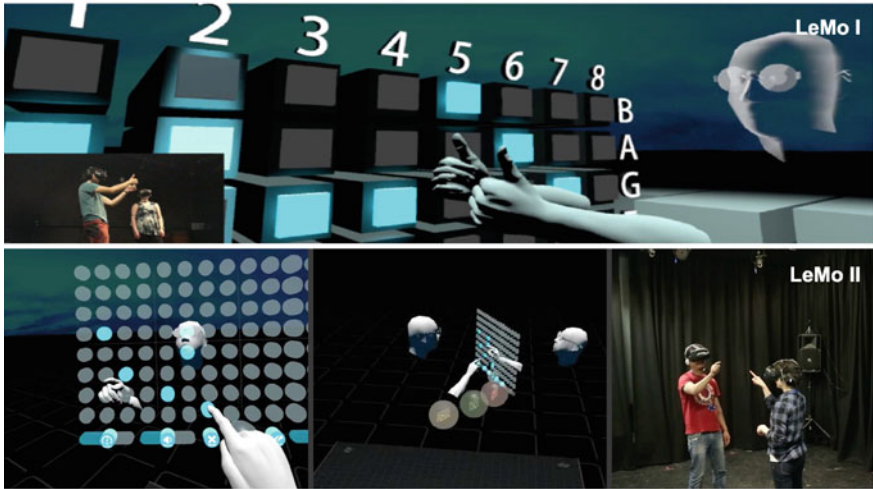


Fig. 8.1 LeMos enable two players to work together on a music loop in VR (reproduced from [36] and [34])

time environment includes two HTC Vive headsets (each with one Leap Motion mounted, see Fig. 8.1c) and two PCs connected and synchronised via a LAN cable. LeMo currently has two major versions: LeMo I and LeMo II (together referred to as LeMos). Both LeMos have three key elements:

- **Music interface**—For producing music. As shown in Fig. 8.2, the *matrix* interface contains a grid of grids/dots. Each row represents the same pitch, forming an octave from bottom to top, see Fig. 8.2. Users can edit notes by tapping the grids/dots. A vertical play-head repeatedly moves from left to right playing corresponding activated notes. In this way, each interface generates a music loop.
- **Avatars**—Each user has an avatar, including a head and both hands, check Fig. 8.1. Avatars are synchronised with users’ real movements in real time, including position and rotation of heads, as well as gestures. So users can not only see their own embodiment but also their collaborator’s.
- **A virtual space in which users co-present.** LeMos provide visual aids for collaboration by synchronising the virtual environment (virtual space and music interfaces) and avatars across a network, providing participants the sense of being in the same virtual environment and manipulating the same set of interfaces.

LeMo I and II have three major differences, which are mainly because LeMo II was built later on the basis of LeMo I, and thus provides more and possibly better functionalities. These differences are:

- **Size of interface matrix of LeMo I is 8*7 while that for LeMo II is 16*8.** So participants can create an 8-beat loop in LeMo I and can create 16-beat loops in LeMo II, see Fig. 8.2.

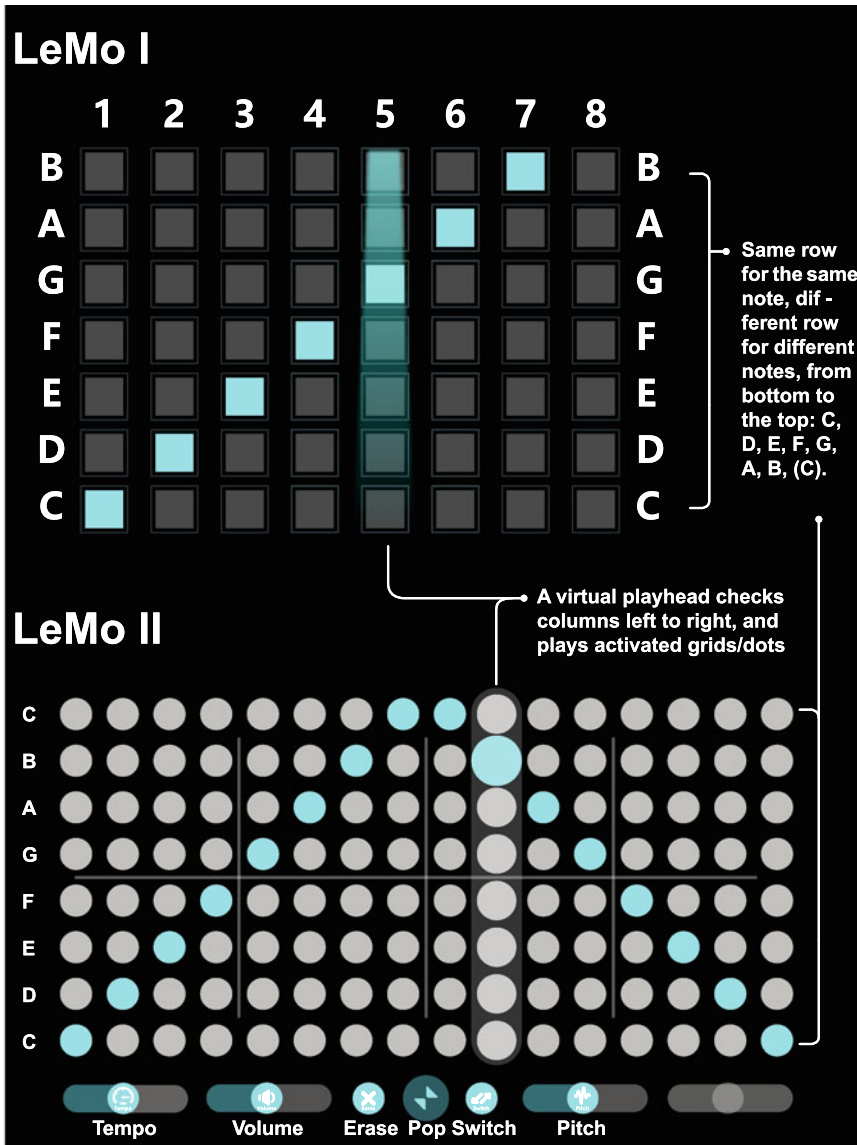


Fig. 8.2 The interfaces of LeMo I and LeMo II

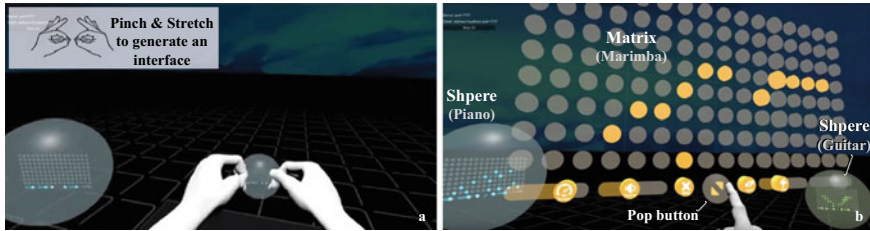


Fig. 8.3 **a** The gesture to generate a new interface; **b** *Matrix* (opened interface) and *sphere* (packed interface), double-click the pop button to switch in between (reproduced from [34])

- While LeMo I only provides one stationary music interface, LeMo II allows users to generate, remove, position and edit up to eight virtual music interfaces. Music interfaces in LeMo II have two modes: *sphere* and *matrix* (Fig. 8.3b), with *sphere* mainly for storage and positioning, and *matrix* for music editing. Users can generate *spheres* with pinch and stretch gesture, see Fig. 8.3a. The *sphere* and the *matrix* form can be switched in between using the pop button at the central bottom of the interface, see Fig. 8.3b. Users can have up to eight music interfaces at the same time,⁴ which means they can have eight music loops at the maximum at the same time.
- Compared with LeMo I, LeMo II allows users to control more music features; users can now use sliders to control tempo, volume and pitch, and use “erase” button and “switch” button to erase or switch among four different instruments, including piano, drum, marimba and guitar, see bottom part of Figs. 8.2 and 8.3b.

8.4 Study I—Visual Approach: 3D Annotation

Writing and sketching are often used in collaboration to exchange ideas, acting as a memory aid, conveying approval, ideas, doubts and so on. In the CMM systems *Daisyphone* and *Daisyfield* in [14], people are given a shared annotation mechanism, which enables collaborators to draw lines that are publicly visible. This has been suggested as an advantage to music making. Taking inspiration from this, the goal of this study is to explore how similar visual cues (e.g. 3D annotations) might impact the creative collaboration when it comes to VR setting. We are interested in exploring how this capability may be used in an SVE to allow collaborative sonic interactions (CMM in this case).

To explore this, LeMo I enables users to draw 3D lines (annotations) by pinching their thumb and index finger together and moving their hands, see left part of Fig. 8.4. These 3D lines are shared and visible to both collaborators, and can therefore potentially be used for communication. To avoid clutter or confusion, users can flip both

⁴ We limit the number to 8 to achieve a proper frame rate.

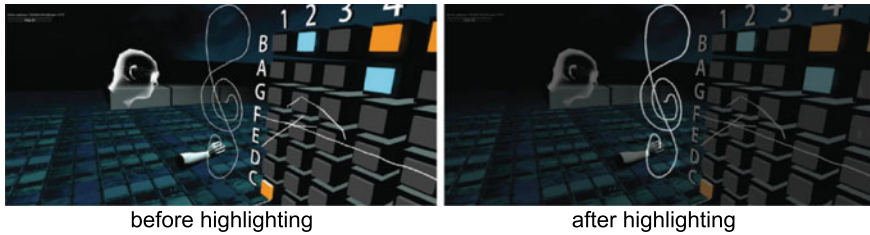


Fig. 8.4 All annotations in subsequent figures have been emphasised by darkening the background and brightening the annotation lines to enhance their legibility outside of VR (from [33])

hands downward to discard all the 3D lines. Users can add or discard lines at any time as they wish.

8.4.1 *Participants and Procedure*

Thirty-two participants (16 pairs) were recruited via group emails at the authors' university and the authors' social media for this study.⁵ Of the participants, 25% had not used VR before, 37.5% of them had tried it only once, nearly a third (28.5%) of them played 2–5 times and nearly 10% played VR frequently. Only two rated themselves as music experts, with the majority rating themselves as novices in musical field. Twelve pairs of participants were familiar with their study partner prior to the study. It took each pair of participants roughly 1 h to finish the experiment, participants received no compensation.

After reading and signing informed consent forms, each pair of participants first received a tutorial on how to use LeMo I and then undertook a task-free trial of LeMo I for 5 min, during which they could change music notes and make annotations, helping them get familiar with LeMo. After that, each pair undertook four sessions of composing music, each lasting 5 min. They were asked to create a music loop that sounds nice to them together. Note only two of these sessions were set for this study, in which participants could make annotations. Participants' annotations were recorded and are highlighted for better readability—see an example in Fig. 8.4. The study ended with a semi-structured interview (around 5 min). Although participants are physically co-located during the experiment, we purposefully did not support nor allow spoken communication. This is because the creative content is in the sound domain and we are interested in how to design systems which foreground the creative uses of sound whilst using complementary modalities to manage the creative process.

⁵ The Queen Mary Research Ethics Committee granted ethical approval to carry out the study within its facilities (Ethical Application Ref: QMREC1592).



Fig. 8.5 Presence annotation: “XiaoB” (a) and “it me” (b), from [33]

8.4.2 Annotation Categories

Seventy-eight annotations were post-hoc identified and categorised by the researchers according to the annotations for Mutual Engagement classification scheme (referred to as aME classification) in distributed music making: presence, making it happen, quality, social and localisation [14]. We use aME classification scheme as a starting point for understanding the use of annotations in LeMo I. The following subsections report on the kinds of annotations participants used when making music together in LeMo I, and later sections reflect on these annotations and the utility of the aME classification scheme for SVEs.

8.4.2.1 Presence

The concept of presence has been defined and interpreted in different ways, e.g. [26, 62, 63, 71]. Presence is a subjective experience [26, 61] which can greatly affect collaboration [22, 50]—having knowledge of oneself and those we are working with is important in collaboration. An earlier study found many participants in distributed music making used annotations as a way to express and query presence, helping participants know about each other’s existence [14]. In this study only two users used annotations to convey presence. One wrote “XiaoB” (the participant’s name) and the other wrote ‘it me’ to tell the collaborators their presence and identity, see Fig. 8.5. The reason that much fewer people used annotations to convey presence could be that the avatars provided a sense of presence and identity not available in the original Daisy studies in [14]. Avatars intuitively show the collaborators where they are, what they are doing and where they are looking. Another reason might be because the collaborators were co-located and that they had previously met in the real world before entering the virtual realm.

8.4.2.2 Making It Happen

Annotations were also used to support the process of collaborative music making in four ways explored below:

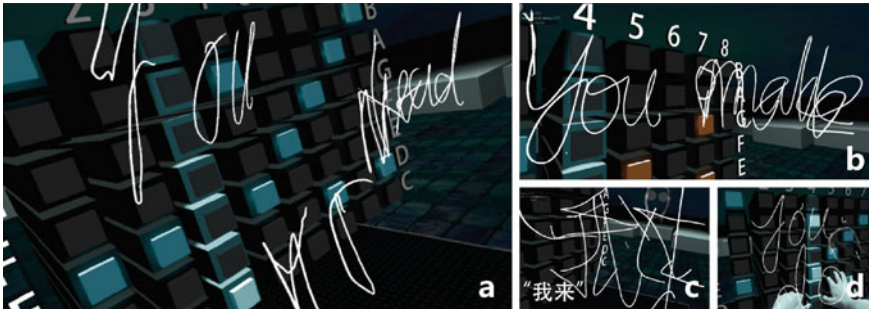


Fig. 8.6 Turn taking annotations: “You go ahead” (a); “you make” (b); “I make” written in Chinese (c); “you do”(d) (reproduced from [33])

- *Turn Taking:* Although LeMo I allows simultaneous editing of the shared musical loop, at some points participants took turns to contribute the musical notes and used annotations to manage the process. As shown in Fig. 8.6, participants wrote “Let me” or “you do” to switch who had the active role. By doing so, the active person could either require or give away full control of the music interface until they agree to a turn change—note that there was no explicit ownership control of the musical interface, so in these cases participants were self-managing their access to the shared musical loop.
- *Composition Thoughts:* Some annotations emerged that were expressing composition ideas at different levels, covering the highest level—music style, the medium level—patterns formed of notes, and the most specific level—single notes. By drawing lines aligning with possible notes on the grid, Fig. 8.7b, c, d, e sketch out participants’ composition ideas. These are more specific communication compared with annotations revealing musical ideas (e.g. “Chinese style?” in Fig. 8.7a). These annotations were usually drawn before activating the corresponding buttons to make and share a plan, possibly so that the partner could help to construct the sequence of notes. Occasionally, these compositional sketches were drawn afterwards (e.g. Fig. 8.7e) and were used to demonstrate a musical idea. In both cases, this kind of annotation may have helped participants to better formulate and understand the collaborative music plan/idea. More directed use of annotations in composition is illustrated in Fig. 8.7f where the participant made three dot markers near the column reference system (B, G and D specifically), asking the partner to make notes in these three columns, which resulted in the partner adding these notes to the shared musical loop. A similar case is shown in 7 h, in which the partner was asked to make notes in rows C, E and G. Participants also directly wrote the reference to ask partners to change specific notes, see Fig. 8.7 h, i, j, k.
- *Area and Position Arrangement:* Annotations were also used to divide the working area and to manage participants’ work focus in the VE. Fig. 8.8a shows an example in which participants drew a horizontal line to divide the music interface into two parts, each for one participant. The pair was composed within their own working

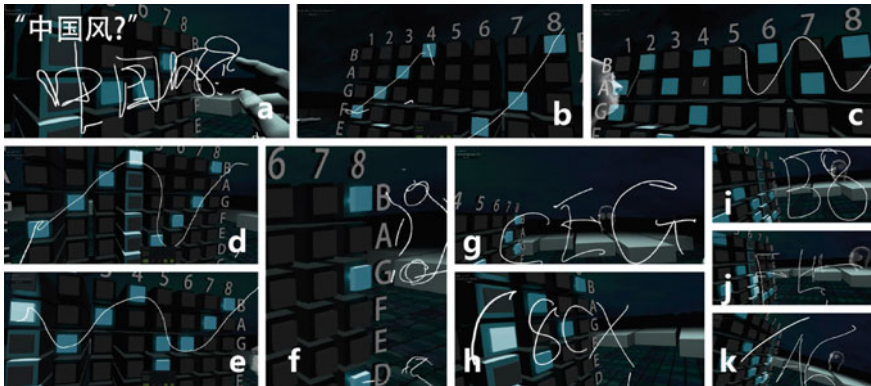


Fig. 8.7 “Chinese style?” written in Chinese (a); Patterns formed of notes (b, c, d, e); Note markers (f); References of notes (g, h, i, j, k) (from [33])

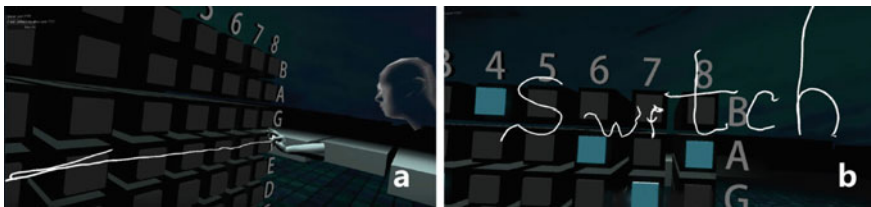


Fig. 8.8 Annotations for working area arrangement (from [33])

area after the line was drawn, and later on, a word “Switch” was written to ask to switch positions (i.e. to swap from top to bottom and vice versa), see Fig. 8.8b. These annotations may have contributed to participants’ working areas and space management.

- *Confusion Expressions*: Participants used annotation to write “what” or to draw a question mark to presumably express confusion about their partners’ activities given that such annotations were made directly after their partners changed notes, drew, wrote or made gestures. Fig. 8.10 illustrates typical indicators of confusion.

8.4.2.3 Quality

When creating the music loop, reflecting and exchanging the ideas of the quality of the piece is crucial to smooth the cooperation and ensure a final output with good quality. In LeMo I, participants used annotations to express and exchange their judgments of the quality. These annotations are usually short words or simple shapes, either positive (e.g. “OK”, “Nice”, “Cool”, “Good” and heart shape) or negative (e.g. “No”), as illustrated in Fig. 8.9. Some of the confusion expressions such as “?” were probably indicators of queries of quality, not just queries about

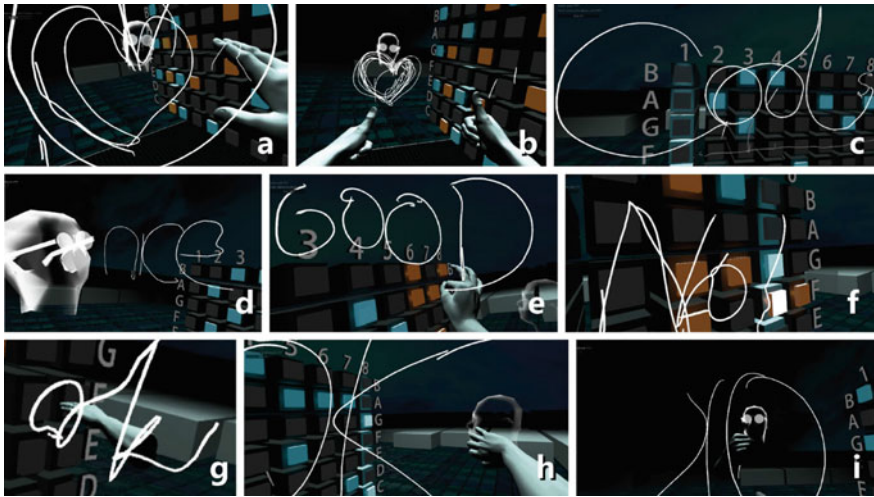


Fig. 8.9 Quality Annotations (from [33])



Fig. 8.10 Confusion annotations (from [33])

the process. It is also interesting to note that positive words may convey different meanings when temporal relationships change. For example, a “yes” written shortly after a note addition means the writer’s satisfaction with the addition while an “OK” write much later with a certain addition has fewer relation with the addition and means more satisfaction about the whole piece. These emerging annotation-based judgments help collaborators exchange feelings about the piece being made, reduce the idea variation and strengthen the cooperation on the activity.

8.4.2.4 Social

Beyond music making and process management, annotations were also used for non-task-related purposes, as illustrated in Figs. 8.11 and 8.12. As shown in Fig. 8.12, one participant started detailed steps of a social drawing activity, their partner then saw this and joined in with the drawing activity and they finished the drawing together. It is interesting to note that in total five human doodles appeared, two of which were drawn collaboratively. The possible reasons for its frequent emergence could be that

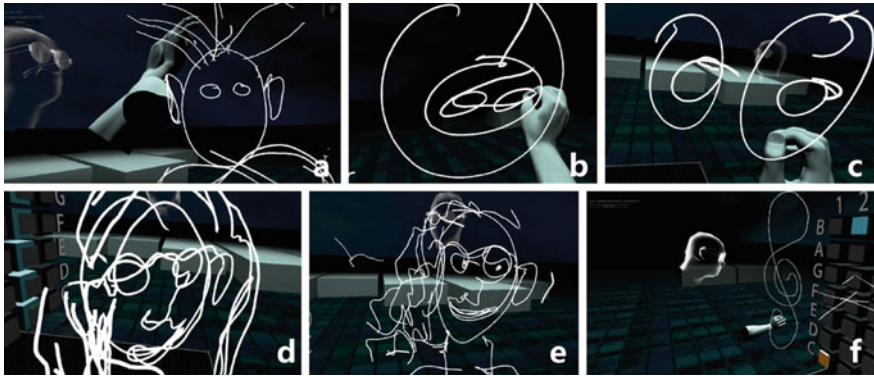
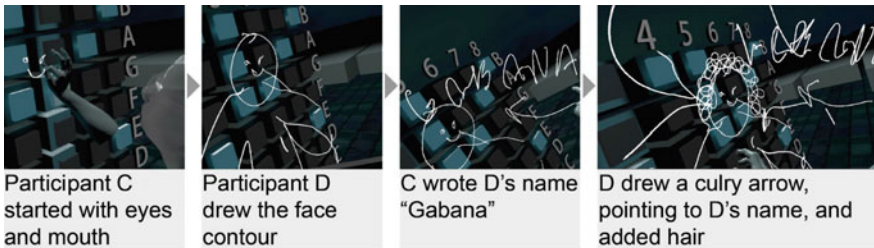


Fig. 8.11 Annotations for social purposes (from [33])



Participant C started with eyes and mouth
 Participant D drew the face contour
 C wrote D's name "Gabana"
 D drew a curly arrow, pointing to D's name, and added hair

Fig. 8.12 Annotations for social purposes (reproduced from [33])

participants were inspired unknowingly by the kinetic avatar or people just naturally love to draw faces. Although social annotations did not contribute to the music directly, making these lighthearted drawings, as a social interaction, contributes to a close relationship between the collaborators.

8.4.2.5 Localisation

Bryan-Kinns [14] identified the frequent use of annotation as a localisation cue (mainly by drawing arrows), but in LeMo I we only found one similar case, in which the participant drew an arrow, and from the review of the interaction successfully obtained their partner's attention, as illustrated in Fig. 8.13. However, in this case the arrow may have been more to attract attention to the activity rather than to highlight a specific part of the joint creation. The reason that annotations are not used for localisation in LeMo I could be that participants could simply draw each other's attention to a certain location by waving their hands and then pointing to that location.



Fig. 8.13 A participant drew an arrow (a), and this successfully drew their partner’s attention to the intended area (b) (from [33])

8.4.3 Interviews

Post-task interviews with participants revealed more reflective insights into the use of the annotations. The interviews were transcribed (around 5,000 words) and a thematic analysis was undertaken, see more information about thematic analysis in [10, 74]. The thematic analysis started with a reading through of the transcript, then an inductive analysis of the data was performed, and relevant patterns were collapsed into codes. Next, these codes were combined into overarching themes, which were then reviewed and adjusted until they were appropriate for the codes. In total, 41 codes and 4 overarching themes emerged from the thematic analysis. Two themes were directly related to annotation: (i) annotation’s usefulness, and (ii) annotation’s problems.

Many participants described that they had a positive feeling when they could write something to support their communication. They reported annotations were used to make “signs and symbols” to support composition, or to “create drawing together [...] like a physical warm up”. Participants also reported that annotations exceeded vocal communication in some ways, “with the lines, [they] could just circle the notes to say that was [note] G and go back to [note] C, from that perspective, drawing was more effective”. Many participants reported that they successfully understood each other’s intentions via the annotations, e.g. one participant drew a line and “used the line to affect the partner”, guiding their partner to move notes to lower positions, the partner fully understood and reported they “did the changes”. Other examples mentioned are showing satisfaction by “writing an OK” or using “Hi” for greetings.

Meanwhile, writing and reading in 3D space were reported by participants to be quite different from the real world and these differences caused inconveniences and problems. For instance, the 3D nature of the annotations reduced their readability, it only “makes sense to [them] from [their] perspective[s], because it was 3D”. For ease of identifying the annotations, “[they] need to stand where the person wrote it stood”. Furthermore, making annotations was reported as being time-consuming, and “when [they] finish[ed] it, it [did] not make sense” anymore. Also, the low accuracy of movement tracking led to annotations being drawn at quite a large size, which then led to a limitation of “how much [they] [could] write”. Finally, participants reported that it was hard to notice each others’ annotation activities, a participant

“waved hands to [their partner], but [the partner] did not see”, the participant “had to wave hands [closer], directly in front of [the partner]” to draw their attention to the annotations so as to get the annotations read. This was probably due to the narrower field of view (FOV) in VR *vs* real life as the FOV is about 100 horizontal degrees with HTC Vive *vs* about 200 degrees binocular FOV in real life, see [28, 30].

8.4.4 Reflection of Study I

Similar to Bryan-Kinns’ findings [14], most of the annotations that emerged in the use of LeMo I fall into three types: making it happen, quality and social. However, unlike the aME classification, presence and localisation appear to be well managed through avatar interaction. This similarity suggests that 3D annotations can function similarly in an immersive collaborative music-making system as they can in a 2D non-immersive CMM system. However, much fewer annotations are used to convey presence compared with the findings of Bryan-Kinns [14] which may be because avatars already support this well, or it may also have been due to the physical collocation of participants with LeMo I compared to the Daisy* studies which were distributed remotely. The length of the musical loop in LeMo I is 8 beats, whereas in the Daisy* studies the length was 48 beats which may have affected the kinds of annotation produced as the LeMo I loop was simpler and required less temporal organisation. Regardless of these issues, the use of aME to classify annotations in a study of CMM indicates that the annotation classification scheme applies to media beyond the Daisy* systems it was previously used to evaluate [14].

For sonic interaction design of VEs, the findings of this exploratory study indicate that 3D graphical annotations of a virtual environment can support a music making as a tool for communication where the co-produced sound is prioritised over other modalities—CMM in our case. We specifically prevented conversation during the creative process to allow us to explore how to support collaboration without interrupting or interfering with the music being created by collaborators. The step sequencer used in LeMo I was intentionally simple to allow initial exploration of the role of annotations without conflating this with the complexity of an interface. For richer and more complex sonic creation and exploration in VR, we suggest that annotations could usefully support communication about the process, quality and also social aspects of interaction without compromising the joint product being produced. It may facilitate a foregrounding of the creative sound product to such an extent that the sounds created are able to use the full width of the sound domain at the exclusion of all other parts of the human–human interaction necessary for collaboration.

Whilst the annotations of LeMo I supported co-creation of music, they did generate some issues. More specifically, making annotations and viewing them were reported to be very different from real life, daily experiences. Participants needed to get used to controlling strokes by pinching and releasing fingers. Besides, compared with writing or drawing with a real pen, the LeMo I has a less accuracy in supporting these. To increase the readability of written contents and sketches, par-

ticipants tended to write or draw in a bigger size, which resulted in a limitation of how much they could write/draw. But on the positive side, the larger size made it possible to write and draw together, which expanded the range of annotating action, making it less personal but more social-friendly and more accommodating to multiple people. Another unexpected problem found in this study was that 3D annotations can, of course, be viewed from many angles, so written text is often reversed for a participant's collaborator, especially if they write in the space between themselves. This clearly decreases the readability of the annotations. Some participants wrote in reverse to try to compensate for this issue, see an example shown in Fig. 8.9h and i. Future development of the use of annotations in VR would need to explore how this mirroring issue could be addressed.

8.5 Study II—Audio Approach: Augmented Acoustic Attenuation

Sound attenuates as a result of diminishing intensity when travelling through a medium. Acoustic attenuation is one of the primary cues for sound localisation of distance; it enables humans to use their innate spatial abilities to retrieve and localise information and to aid performance, see [5]. Whilst augmenting the acoustic attenuation of a real medium (e.g. the air) is difficult, this can be easily done in VEs with the aid of audio simulation (refer to Chap. 3 for modularity in the auralisation). Research has begun to investigate the impacts of spatialised sounds on user experience in VR, see [27]. However, little research explores how the spatialisation of sound may affect or aid collaboration in a VR context. Considering sound is both the primary medium and the final output of the creative task [34], by affecting sound, different settings of acoustic attenuation can possibly affect the collaboration differently. With the ability to modify the simulated acoustic attenuation in an immersive virtual environment, we can possibly create sonic privacy by augmenting acoustic attenuation, and then use sonic privacy as personal space to support individual creativity in CMM. Supporting individual creativity is important as it contributes to the group creativity [37, 44, 46, 60].

8.5.1 Hypotheses

Research has suggested users should be allowed to work individually in their personal spaces at their own pace, cooperatively work together in the shared space and smoothly transition between both of the spaces during collaboration [23, 56, 65]. In a previous study [34], following this implication, we built three different spatial configurations (public space only, public space + publicly visible personal space, public space + publicly invisible personal space), and tested different impacts

of these spatial configurations on collaborative music making in SVEs. The results show adding personal space to be helpful in supporting collaborative music making in SVE, since it provides a chance to explore individual ideas, and provides higher efficiency in making notes. However, several negative impacts also showed up along with the addition of personal space, e.g. longer average distance between participants, reduced group territory and group edits [34]. We believe this might be due to: (i) the separated stationary locations of the personal spaces forced users to leave each other to access, causing a longer distance between participants and less collaboration; (ii) the rigid boundary between public space and personal space made users more isolated, resulting in a higher sense of isolation. Thus allowing users to access personal space without leaving each other far away might eliminate these disadvantages.

To make the shift between personal and public spaces more fluid, inspired by the implication that the separation between public and personal workspace should be gradual rather than too rigid [23], the attenuation feature can possibly be applied to form a gradual personal space, enabling a fluid transition between personal space and public space. This is because sound is both the primary medium of collaborative tasks and the final work of CMM [33], thus by manipulating acoustic attenuation, we can produce sonic privacy. Thus H1 was developed.

H1: Attenuation can play a similar role to personal space with rigid form in CMM in SVE, providing collaborators a personal space and supporting individual creativity during the collaboration.

Additionally, an acoustic attenuation, rather than a personal space with rigid separation from public space, enables a gradual shift between personal and public workspace, which may possibly increase the fluidity of the experience and support collaboration better, cf. [23]. Thus we developed H2.

H2: Acoustic attenuation provides a fluid transition (no hard borders nor rigid forms) between personal and public spaces, which introduces less negative impacts on collaboration compared with personal space with rigid form in [34].

8.5.2 *Independent Variable*

Spatial configuration is an independent variable in this experiment. Two spatial configurations were designed as the independent variable levels, as shown in Fig. 8.14, including the following:

- **Condition 1: Public space only** (referred to as C_{pub}): where players can generate, remove or manipulate music interfaces, and have equal access to all of the space and the music interfaces. As no personal space is provided, a shift between public and personal space does not exist, i.e. users cannot shift to personal space.
- **Condition 2: Public space + Augmented Attenuation Personal Space** (referred to as C_{aug}). In addition to C_{pub} , the sound attenuation is augmented. The volume of audio drops much faster, creating a sonic privacy, which can be seen as a personal

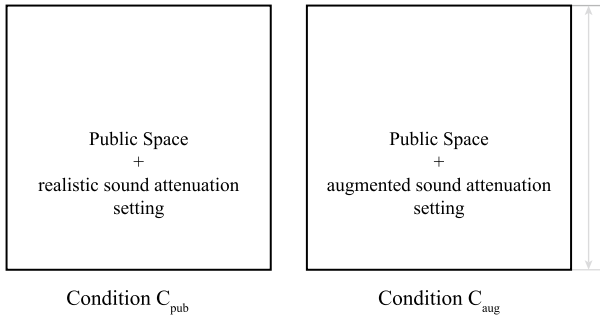


Fig. 8.14 Top view of the two experimental condition settings

space. As the volume changes gradually with the changes of distance, the shift between personal space and public space is gradual.

8.5.3 *Dependent Variables*

To identify how users use the space and the effect of adding augmented sound attenuation as personal space, dependent variables were developed. The Igroup Presence Questionnaire (IPQ) was used to inform the design of questions about sense of collaborator's presence [54]. The IPQ measures the sense of presence using one general measurement—sense of being there, plus three sub-scales covering spatial presence, involvement and experience realism. Questions about output quality, communication and contribution were adapted from the Mutual Engagement Questionnaire (MEQ) [15]. The MEQ is formed of two parts: (i) participant ratings of the quality of the musical outcome and their interaction with musical interface; (ii) participant choices between different conditions when being provided a series of statements covering the music quality, enjoyment, involvement and frustration. The rest of the questions were designed to question people's preference for conditions. The questionnaire included measures on:

- *Presence*: (i) Sense of self-presence, (ii) sense of co-worker's presence and (iii) sense of collaborator's activities.
- *Communication*: quality of communication, which may vary as the visibility of spaces can possibly affect the embodiment and nonverbal communication.
- *Content assessment*: the satisfaction of the final music created reflects the quality of collaboration, cf. [15, 16].
- *Preference*: preference of the conditions, to see if users have subjective preferences towards the settings.
- *Contribution*: (i) the feeling of self's contribution; (ii) the feeling of others' contribution.

Table 8.1 Results of Post-Session Questionnaire and the results of Wilcoxon Rank-Sum Test (two-tailed)^a

Questions (Measure)	C _{pub}		C _{aug}		C _{pub} vs C _{aug}	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>	<i>W</i>
PSQ1 (support for creativity)—I think the space setting in this session was extremely helpful for creativity	8.55	1.44	8.77	1.34	0.5695	259
PSQ2 (support for creativity)—I feel like the space setting in this session was extremely helpful to support the development of my own ideas	7.82	1.92	8.35	1.50	0.5211	255.5
PSQ3 (preference)—I enjoyed the space setting of this virtual world very much	8.27	1.61	8.65	1.60	0.2622	233
PSQ4 (sense of collaborator’s presence)—I always had strong feeling that my collaborator was there, collaborating with me together, all the time	8.91	0.92	8.54	1.68	0.7961	298.5
PSQ5 (content assessments)—How satisfied are you with the final piece of loop music you two created in this session	8.64	1.09	8.50	1.36	0.7644	300.5
PSQ6 (communication quality)—How would you rate the quality of communication between you and your collaborator during the session	8.68	1.09	8.50	1.36	0.7644	300.5
PSQ7 (sense of collaborator’s activity)—I had a clear sense what my collaborator was doing	8.73	1.20	7.96	1.54	0.08094	368.5
PSQ8 (amount of contribution)—The amount of your contribution to the joint piece of music is	8.41	1.44	8.15	1.46	0.4776	320
PSQ9 (amount of contribution)—The amount of your collaborator’s contribution to the joint piece of music is	8.18	1.26	8.23	1.39	0.8486	276.5
PSQ10 (quality of contribution)—What do you think of the quality of your contribution to the joint piece of music is	8.05	1.70	7.81	1.41	0.319	333.5
PSQ11 (quality of contribution)—What do you think of the quality of your collaborator’s contribution to the joint piece of music is	7.73	1.52	8.19	1.20	0.3496	241.5

^a Note: statistics in this table are calculated based on the data collected from third and fourth session to better counterbalance the learning effect

These measures are grouped into a Post-Session Questionnaire (PSQ, see items in Table 8.1).

8.5.4 Participants and Procedure

Fifty-two participants (26 pairs) were recruited through group emails at the authors’ university for this study.⁶ Each participant was compensated 10 GBP for their time (roughly 1 h). Participants’ rating of musical theory knowledge is 3.92 (SD = 2.50) on a 10-point Likert scale, where higher values indicate increased knowledge; 24

⁶ The Queen Mary Research Ethics Committee granted ethical approval to carry out the study within its facilities (Ethical Application Ref: QMREC2005).

participants play one or more instruments, and the remaining 28 do not. Twenty participants had tried VR 2–5 times before, 20 had only tried once and the remaining 12 had no VR experience previously. Thirty-seven participants knew their collaborators very well prior to the experiment; three met their collaborators several times, and the remaining 12 did not know their collaborators at all prior to the experiment.

The experiment started with participants reading the information form and signing the consent form. Then they first received an explanation of the music interface of LeMo II (see Fig. 8.2), with all of the interaction gestures supported in LeMo II demonstrated by an experimenter. Next, a trial (roughly 5–15 min) session was carried out, where participants could try all of the possible interactions. The trial ended once participants were confident enough of all available interactions. The length of time of the tutorial session was flexible to ensure participants with diverse musical knowledge could grasp LeMo II. Participants were then asked to have four sessions of collaboratively composing music that was mutually satisfying and compliments an animation loop. Two of these sessions were set for this study; each covered a condition ($C_{\text{pub}}/C_{\text{aug}}$), and the sequence of conditions was fully randomised to counterbalance the learning effect. We set each session as 7 min because based on our pilot study and a previous study [33], we found 7–8 min were sufficient for the task. In total, four visual, silent animation loops were introduced to trigger participants' creativity; each to be played in one experimental session on four virtual screens surrounding the virtual stage. These clips were played in an independently randomised sequence to counterbalance impacts on the study. Each session ended with a Post-Session Questionnaire (PSQ, see Table 8.1). After all the four sessions finished, a short interview was carried out.

8.5.5 Results

Wilcoxon Rank-Sum tests were run to compare the ratings of C_{pub} with C_{aug} collected by PSQ, see results in Table 8.1. No significant effect was found between C_{aug} and C_{pub} . Post-task interviews revealed more reflective insights. Around 41,000 words of audio recorded interview responses were transcribed and a thematic analysis of the transcription was undertaken (more details about the thematic analysis in Sect. 8.4.3). As shown in Fig. 8.15, in total, 439 coded segments, 15 codes and 3 overarching themes emerged from the thematic analysis: (i) learning effects; (ii) preferences, advantages and disadvantages of conditions; and (iii) advantages, disadvantages of LeMo II and suggestions for improvements. Next, we will only cover the former two themes as the final one is not directly related to the scope of this chapter.

8.5.5.1 Learning Effects

Members of 18 groups mentioned the effect of the session sequence. Specifically, 43 coded segments contributed by 27 participants were related to learning effects. For

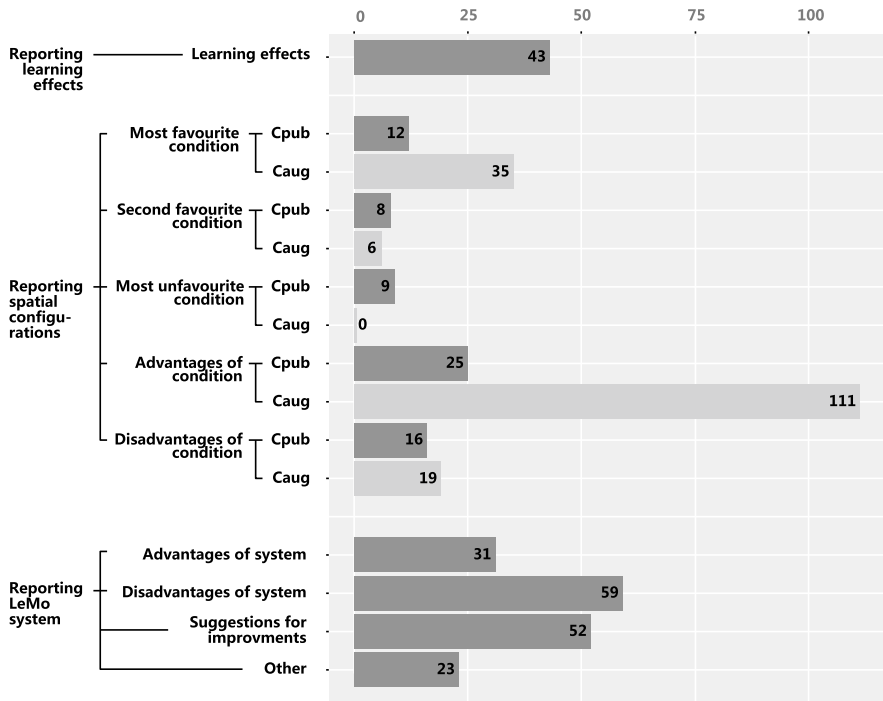


Fig. 8.15 Ingredients of all the coded segments of the interview; number of coded segments are shown in the bars

example, Participant 15A (participant A in group 15, referred to as P_{15A}) reported the sequence is an “important factor”. The first session was felt to be hard as they were “just being introduced to [the system and they were] still adjusting” to it (P_{5A}), trying to “[figure] out how the system was working” (P_{16A}), as they “were progressing into latter sessions, [they] felt easier to communicate and use gestures to manipulate the sound, being able to collaborate more, more used to the system” (P_{5B}), these changes led to a higher level of satisfaction and more enjoyment in later conditions. To better counterbalance the impact of sequence, Table 8.1 only includes data collected from the latter two sessions (note: as aforementioned, there were four sessions that were randomly sequenced, and two of which were related to this study).

8.5.5.2 C_{pub}—Simple but can be Chaotic

With no personal space, participants had to hear all the interfaces throughout the session. In total, 16 coded segments are about the disadvantages of C_{pub}; some exemplars are: “a bit troubling”—P_{11B},⁷ “music always very loud”—P_{9A}, “it was global music,

⁷ P_{11B} refers to participant B in group 11, similarly, P_{9A} indicates participant A in group 9.

and there was someone annoying”—P_{2A}, “you are not going to say anything” to avoid being “rude”—(P_{2A}). It was easier if there is something helpful “to perceive what I was doing, and not get confused with what [the collaborator] was doing” (P_{15B}), it was too “chaotic” (P_{20A}), “too confusing” (P_{22A} and P_{22B}), “annoying” (P_{25B}). They “can not concentrate” (P_{25B}) while “everything [is] open and quite noisy” (P_{26B}), and they “don’t have the tranquillity to operating [their] sounds or the everything’s come mixed, which is difficult to manage” (P_{22A}).

There were 25 coded segments from 14 participants reporting the positive side of the C_{pub}; some examples are: (i) pieces created in “personal space” might clash in a musical way (P_{1A}), “better to work when knowing how it sounds all together” (P_{17B}), music pieces might match better; (ii) better for providing help to the other collaborator, as reported by P_{4A}, saying that they needed someone to lead them and thus the ability to hear all the work all the time was helpful; (iii) “space wise”, compared with having to work closer to “hear the sound well” (P_{12A}) in C_{aug}, C_{pub} does not have this space constraint, they could chose to work “anywhere” (P_{24A}); iv) “easier” to understand the condition (P_{6B}), fewer confusions when simply being able to hear all the things all the time (P_{13A}); v) “collaborative wise” (P_{13A}), less separation, better collaboration compared with “personal space” was provided (P_{3B}, P_{18A} and P_{18B}).

8.5.5.3 Preference on C_{aug}

There were 35 coded segments contributed by 24 participants favouring condition C_{aug}, higher than 12 segments contributed by 11 participants for C_{pub}. There are 111 coded segments contributed by 33 participants from 25 groups reporting the advantages of this condition, much higher than the number of segments reporting other conditions’ advantages. These reports reveal some insights behind the popularity of C_{aug}. C_{aug}’s advantages reported by participants can be grouped into 4 groups:

- *Higher team cohesiveness and lower sense of separation.* Participants reported that, without the rigid personal space, they had to “work with the other person” (P_{6A}). With no rigid personal space, C_{aug}’s “forces [them] to collaborate more the most because [they] had to stay very close to compose music ” (P_{9B}).
- *An appropriate environment for creativity, more consistency and convenience.* As described by participants, it was “a middle point between personal space and no personal space” (P_{6A}), without even triggering something, “[they] could decide in a continuous way if [they] were able to listen to the other sound sources or not”, and “to what extent [they] want to isolate [themselves]” (P_{16A}). Compared with having to hear all sounds in C_{pub}, this provided them a “less stressing” (P_{4A}) context, and they can selectively move away to avoid “getting interrupted with the other” (P_{5B}) and overlapping music. Being able to still “hear a bit of it in the background but not completely” (P_{20A}) was reported good as this kept them “up to date” (P_{9A}) and helped them to “tailor what [the participant] was making” (P_{22B}) to match the co-created music and to make something new and see if it “fit with”

- (P_{20A}) the old. C_{aug} provided them with “a little bit of personal space” although not a quite a “defined thing” (P_{6A}), which provided the possibility “to work on something individually” but also being able to “share work quite easily” (P_{20A}).
- *Easier to identify sounds.* Participants reported it was easier to “locate the source of the sound” (P_{16A}) and “perceive what [they were] doing” (P_{15B}), which helped them “understand instruments better” (P_{7B}) and “not get confused” (P_{15B});
 - *More real.* Interestingly, instead of C_{pub}, which simulates the real-world sound attenuation, C_{aug} was reported to be “real”. “If you want to hear something, you just come closer, like in the real world” (P_{11B} and P_{11B}), “it was good like we were feeling like the real-time experience (P_{26B})”.

It should also be noted that, along with these reports about advantage, there are 19 segments reporting C_{aug}’s limitations, including: (i) a preference “to hear all the instruments all the time” in C_{pub} (P_{26B}), (ii) C_{aug} might lead to “another type of compositions” and “influence the piece” (P_{16B}) and (iii) without being able to hear all sounds led to a feeling of separation (P_{18A}).

8.5.6 Discussion

The issues from having no personal space are clear. Especially for the music-making task in this study, participants reported that without personal space, the auditory background could be messy to develop own ideas, and their creativity required a quieter and more controllable environment, which could be provided by personal space. Providing such an environment is crucial considering individual creativity is an important part of the collaborative creativity. Having personal space was reported to be “an added advantage” because it promoted their own creativity, which can then be combined and contributed to the joint piece. This matches the findings in [34], that providing personal spaces is helpful as it provides a chance to explore individual ideas freely, which then added an interesting dynamic to the collaborative work. However, adding personal space indeed brought a few impacts, next we discuss the impacts of using acoustic attenuation as personal space and its characteristics.

8.5.6.1 Impacts of Adding Acoustic Attenuation as Personal Space

As mentioned above, in the previous study [34], we found the addition of personal space located on the opposite side of the public space led to a shrunken size of group territory, fewer group note edits, a larger size of personal territory, more personal note edits, a larger average distance between collaborators and fewer times of paying attention to collaborator. We argued that these negative impacts are mainly due to the personal spaces distributed on the opposite side of the group space resulting in a larger distance between participants. So we proposed personal space with different features (e.g. gradual boundary—C_{aug}) might reduce these negative effects. In many

ways, C_{aug} is quite similar to C_{pub} , e.g. both do not have a visual boundary for spaces, so not surprisingly, no significant differences were found in most of the statistical measures, see Table 8.1, and most previously identified disadvantages brought by adding rigid personal spaces have been successfully eliminated; more detailed results are available in [35]. By making the personal space invisible and gradual, the isolation and difficulty of coordinating that introduced by the additional personal space was minimised. For example, in the interview, participants reported C_{aug} provided a proper level of group work as working context, making easier to create new that matches the old.

8.5.6.2 Providing Personal Space with Fluid Boundary

Although no significant differences were found in PSQ2, see Table 8.1, which questioned the support each condition gave to individual creativity, C_{aug} has a higher mean rating. The thematic analysis revealed more insights. C_{aug} provides both “an appropriate background” with which participants felt “less stressed” and were able to “tailor” the individual composing to match the co-work, and a space personal enough to “work on something individually”. No major differences were found between C_{pub} and C_{aug} in PSQ, indicating C_{aug} provides a very mild solution, with limited impacts on people’s collaborative behaviour introduced, whilst still providing sufficient support for individual creativity during collaboration, thus H1 is validated.

Compared with natural attenuation in C_{pub} , C_{aug} ’s augmented sound attenuation setting forced or prompted people to work more closely in order to hear each other’s work, as reported by some participants. Compared with adding personal space with visible rigid boundary, by enabling participants to “decide in a continuous way” (P_{16A}) if they want to hear other’s work, an invisible gradual boundary in C_{aug} led to less separation, and higher consistency between personal and public space. H2 is therefore supported. This finding also echos the implication proposed in [23] that there should be many gradations between personal and public space to enable people fluidly shift in between. Popularity—the code “advantage of C_{aug} ” has 111 coded segments, and the code “most favourite— C_{aug} ” has 35 coded segments, both are greater than what C_{pub} gets. All indicate C_{aug} is the most popular condition. The popularity is also partially verified by that C_{aug} has the highest mean in preference measure (PSQ3 in Table 8.1). We believe the reasons behind this popularity are mainly due to its unique advantages, which as reported by participants, include: (i) an appropriate environment for creativity, (ii) easier to identify sounds and (iii) perceived as more “real” (although it should be noted that C_{pub} is more similar to real-world audio attenuation). These features of C_{aug} made it provide better support for collaborative creativity and therefore led to its popularity.

Table 8.2 Comparison between the two routes

	3D Annotation	Augmented Attenuation
Modality	Visual	Auditory
Type of interaction	Explicit interaction—active drawing	Implicit interaction—passive body movement
Supports for collaboration	Supporting communication between users	Supporting development of individual creativity
Characteristic	No influence introduced on audio channel, users hear roughly the same audio ^a	Influence introduced on audio and composition, users do not hear the same audio
Applications	Wider range of application, not restricted to audio tasks, audio tasks requiring precise audio output, or users with hearing/speech impairment	Application restricted to auditory tasks with no requirement for precise audio outputs

^aStrictly speaking, what users hear still slightly differs unless the realistic spatialisation of audio is disabled

8.6 General Discussion

The two studies have explored 3D annotation and augmented acoustic attenuation’s role in CMM. This section compares the two approaches against each other, seeking the potential differences and finding out the usage scenarios. The comparison results are summarised in Table 8.2.

8.6.1 Modality and Interaction Type

3D annotation is a visual approach, while augmented attenuation is an audio approach. This fundamental difference led to their unique advantages and disadvantages, which then determine their scope of usage scenarios. Specifically, the visual approach can fully avoid influencing the audio channel, leaving that modality purely for composers to hear the project they are working on. While on the contrary, the audio approach imposes unavoidable effects on how the audio sounds, as the privacy is produced by augmenting the acoustic attenuation of the medium of the sound.

Unlike 3D annotation, which requires explicit interaction to make 3D lines, the augmented attenuation in Study II only relies on users’ passive listening and active physical locating in space. Explicit interaction is consciously deciding to interact, e.g. clicking a button. It is what we normally think about when we’re interacting with a computer [55]. Compared with explicit interaction, implicit interaction does not require users to perform conscious actions; the interaction is mainly the movement (e.g. head movement and eye movement) of the user. As a result, the 3D annotation introduced a higher learning cost.

8.6.2 Key Support for Collaboration

The 3D annotation helps people to warm up at the beginning, supports the non-vocal communication and provides help for collaborators to understand each other's attention. In other words, it supports the social aspects of the collaboration by intensifying the links between collaborators. While the augmented attenuation gives collaborators the choice to be separated, hence provides support for individual creation. With this flexibility, users have the choice to develop their own work and to switch fluidly between working on own and teamwork.

8.6.3 Characteristic and Application

3D annotation completely avoids impacts on the auditory channel. This supportive measure suits where the sonic output comes with stringent requirements, and users must be able to hear exactly the same final output during their working. Its application is not limited to sonic task because it provides support to communication, which is required by many collaborative tasks in SVEs. In contrast, the augmented attenuation has a narrower application range. It provides better support for individual activity, with still enough context of group work and the cost of hearing (slightly) differed output, making it only appropriate to audio related-tasks with no rigid requirements, e.g. people are improvising music for fun.

These two supportive features do not necessarily contradict each other, and could be applied simultaneously. To manage the simultaneous use, a manipulation system might be needed. For example, the transparency of the visual 3D annotation and the degree of augmentation of attenuation can be adjusted to modify their impacts (visibility/audibility), fitting collaborators' needs during different stages of the collaborative composing. When only one feature is needed, the other can be adjusted to zero, wiping out its impacts entirely.

8.7 Conclusions and Future Work

In this chapter, two different approaches to support collaborative sonic interaction in SVEs have been presented, one exploited visual modality and the other exploited audio modality. The results of both studies have been presented and reflected upon. A comparison between the two approaches has been made. Next, following the findings and discussion above, we propose six implications for supporting collaborative sonic interaction in SVEs, e.g. CMM.

1. Adding a system that supports 3D annotation may be considered to aid collaborator's communication, especially if co-produced sound has to be prioritised over other modalities to avoid any impacts.

2. For audio-related tasks in SVEs, adding personal space should be considered, as it provides sonic privacy and essential support for the development of individual creativity, which forms a key part of the collaborative creativity. This is especially essential when the output of the task is vulnerable (e.g. audio), and co-workers need a space where they can think of own ideas and develop own work.
3. For audio-related tasks (e.g. collaborative music making), manipulating acoustic attenuation as personal space can be an effective way to allow users to shift between personal and public working space continuously by adjusting their relative distance. With light-weight form, it introduces mild impacts compared with the prominent negative impacts introduced by rigid personal space [34].
4. The level of privacy can be adjusted by manipulating the level of augmentation. For instance, in C_{aug} of Study II, participants adjusted their distance between themselves and collaborators to achieve different levels of being personal (herein referred to as “personalness”). Instead of changing positions, adjusting the sound attenuation rate with distance can impact the level of “personalness” and therefore producing a varied level of personalness. Potentially, adding a method allowing users to adjust the level might be useful so users can shift between having a “very personal and isolated” space and a “very public” space.
5. Augmented attenuation can be exploited for creative audio privacy, which can be then used to promote individual creativity during the collaboration. However, augmented attenuation introduces differences in what collaborators hear, making it only applicable to contexts with no rigid requirements on audio outputs.
6. We suggest that augmented attenuation and 3D annotation could be applied together or chosen with a flexible switch so that users can choose the feature fitting their needs during different stages of the collaborative composition.

Future works concern an exploration of how multi-modal approaches can be applied simultaneously, and designing and applying tools based on other modalities to support collaborative sonic interaction in SVEs, such as visual modality. For each modality, it could be interesting to test how that sensory cue can be augmented/depressed to adjust the level of its influence.

Acknowledgements This work is partially supported by EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1). We would also like to thank Ms Louise Bryce and Ms Danqi Zhao for their help in proofreading this chapter.

References

1. Barbosa, Á.: Displaced soundscapes: A survey of network systems for music and sonic art creation. *Leonardo Music Journal* **13**, 53–59 (2003).
2. Basdogan, C., Ho, C.-H., Srinivasan, M. A., Slater, M.: An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)* **7**, 443–460 (2000).

3. Benford, S., Bowers, J., Fahlén, L. E., Greenhalgh, C., Snowdon, D.: User embodiment in collaborative virtual environments in Proceedings of the SIGCHI conference on Human factors in computing systems (1995), 242–249.
4. Benford, S., Greenhalgh, C., Rodden, T., Pycock, J.: Collaborative virtual environments. *Communications of the ACM* **44**, 79–85 (2001).
5. Billinghamurst, M., Kato, H.: Collaborative augmented reality. *Communications of the ACM* **45**, 64–70 (2002).
6. Billinghamurst, M., Poupyrev, I., Kato, H., May, R.: Mixing realities in shared space: An augmented reality interface for collaborative computing in Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on **3** (2000), 1641–1644.
7. Bin, S. A., Bryan-Kinns, N., McPherson, A., et al.: Hands where we can see them! investigating the impact of gesture size on audience perception. *International Computer Music Conference* (2017).
8. Blaine, T., Fels, S.: Contexts of collaborative musical experiences in Proceedings of the 2003 conference on New interfaces for musical expression (2003), 129–134.
9. Blaine, T., Perkis, T.: 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 (Association for Computing Machinery, New York City, New York, USA, 2000), 165–173.
10. Braun, V., Clarke, V.: Using thematic analysis in psychology. *Qualitative research in psychology* **3**, 77–101 (2006).
11. Bressan, F., Vets, T., Leman, M.: A multimodal interactive installation for collaborative music making: From preservation to enhanced user design in Proceedings of the European Society for Cognitive Sciences Of Music (ESCOM) Conference, Ghent University (2017), 23–26.
12. Brooks Jr, F. et al.: Research directions in virtual environments. *Computer Graphics* **26**, 153 (1992).
13. Bryan-Kinns, N.: Daisiphone: the design and impact of a novel environment for remote group music improvisation in Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (2004), 135–144.
14. Bryan-Kinns, N.: Annotating distributed scores for mutual engagement in daisiphone and beyond. *Leonardo Music Journal* **21**, 51–55 (2011).
15. Bryan-Kinns, N.: Mutual engagement and collocation with shared representations. *International Journal of Human-Computer Studies* **71**, 76–90 (2013).
16. Bryan-Kinns, N., Hamilton, F.: Identifying mutual engagement. *Behaviour & Information Technology* **31**, 101–125 (2012).
17. Bryan-Kinns, N., Healey, P. G.: Decay in collaborative music making in Proceedings of the 2006 conference on new interfaces for musical expression (2006), 114–117.
18. Carlsson, C., Hagsand, O.: DIVE A multi-user virtual reality system in Proceedings of IEEE Virtual Reality Annual International Symposium (1993), 394–400.
19. De Simone, F. et al.: Watching Videos Together in Social Virtual Reality: An Experimental Study on User's QoE in 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (2019), 890–891.
20. Draper, J. V., Kaber, D. B., Usher, J. M.: Telepresence. *Human Factors* **40**, 354–375 (1998).
21. Ellemers, N., Rink, F. in *Social Identification in Groups* 1–41 (Emerald Group Publishing Limited, 2005).
22. Frécon, E., Greenhalgh, C., Stenius, M.: The DiveBone—an application-level network architecture for Internet-based CVEs in Proceedings of the ACM symposium on Virtual reality software and technology (1999), 58–65.
23. Greenberg, S., Boyle, M., LaBerge, J.: PDAs and shared public displays: Making personal information public, and public information personal. *Personal Technologies* **3**, 54–64 (1999).
24. Gutwin, C., Greenberg, S. in *Team cognition: Understanding the Factors that Drive Process and Performance* (eds Salas, E., Fiore, S. M.) 177–201 (Washington, DC, US: American Psychological Association, 2004).

25. Healey, P. G., Leach, J., Bryan-Kinns, N.: Inter-play: Understanding group music improvisation as a form of everyday interaction. *Proceedings of Less is More-Simple Computing in an Age of Complexity* (2005).
26. Heeter, C.: Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments* **1**, 262–271 (1992).
27. Hendrix, C., Barfield, W.: The sense of presence within auditory virtual environments. *Presence: Teleoperators & Virtual Environments* **5**, 290–301 (1996).
28. Hunt, C.: Field of view face-off: Rift vs Vive vs Gear VR vs PSVR <https://www.vrheads.com/field-view-faceoff-rift-vs-vive-vs-gear-vr-vs-psvr>.
29. Jordá, S.: On stage: the reactable and other musical tangibles go real. *Int. J. Arts Technol.* **1**, 268–287 (2008).
30. Lab, R. L.: Field of View for Virtual Reality Headsets Explained <https://vr-lens-lab.com/field-of-view-for-virtual-reality-headsets/>.
31. Lea, R., Honda, Y., Matsuda, K.: Virtual society: Collaboration in 3D spaces on the Internet. *Computer Supported Cooperative Work (CSCW)* **6**, 227–250 (1997).
32. Luciani, A. in *Enaction and enactive interfaces : a handbook of terms* 299–300 (Enactive Systems Book, 2007).
33. Men, L., Bryan-Kinns, N.: LeMo: Supporting Collaborative Music Making in Virtual Reality in 2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE) (2018), 1–6.
34. Men, L., Bryan-Kinns, N.: LeMo: Exploring Virtual Space for Collaborative Creativity in *Proceedings of ACM conference on Creativity and cognition 2019* (2019).
35. Men, L., Bryan-Kinns, N., Bryce, L.: Designing spaces to support collaborative creativity in shared virtual environments. *PeerJ Computer Science* **5**, e229 (2019).
36. Men, L., Bryan-Kinns, N., Hassard, A. S., Ma, Z.: The impact of transitions on user experience in virtual reality in *Virtual Reality (VR)*, 2017 IEEE (2017), 285–286.
37. Men, L., Zhao, D.: Designing Privacy for Collaborative Music Making in Virtual Reality in *Proceedings of the 16th International Conference on Audio Mostly (Trento, Italy, 2021)*.
38. Minsky, M.: *Telepresence*. Omni, 45–51 (1980).
39. Morreale, F., Angeli, A. D., O’Modhrain, S.: Musical Interface Design: An Experience-oriented Framework in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)* (2014), 467–472.
40. Nabavian, S., Bryan-Kinns, N.: Analysing group creativity: A distributed cognitive study of joint music composition. *Proc. of Cognitive Science*, 1856–1861 (2006).
41. Nassiri, N., Powell, N., Moore, D.: Human interactions and personal space in collaborative virtual environments. *Virtual reality* **14**, 229–240 (2010).
42. Nedel, L. et al.: Using Immersive Virtual Reality to Reduce Work Accidents in Developing Countries. *IEEE Computer Graphics and Applications* **36**, 36–46 (2016).
43. Nowak, K. L., Biocca, F.: The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments* **12**, 481–494 (2003).
44. Ochse, R., Ochse, R.: *Before the gates of excellence: The determinants of creative genius* (CUP Archive, 1990).
45. Oculus: Toybox Demo for Oculus Touch <https://www.youtube.com/watch?v=iFEMiyGMa58>.
46. Pirola-Merlo, A., Mann, L.: The relationship between individual creativity and team creativity: Aggregating across people and time. *Journal of Organizational behavior* **25**, 235–257 (2004).
47. Plante, C.: A PlayStation VR demo made me cry from laughing <https://www.theverge.com/2016/3/16/11246334/playstation-virtualreality-social-vr-demo>.
48. Rice, R. E.: Media appropriateness: Using social presence theory to compare traditional and new organizational media. *Human Communication Research* **19**, 451–484 (1993).
49. Rodden, T.: A survey of CSCW systems. *Interacting with computers* **3**, 319–353 (1991).
50. Romano, D. M., Brna, P., Self, J. A.: Collaborative decision-making and presence in shared dynamic virtual environments in *Proceedings of the Workshop on Presence in Shared Virtual Environments*. BT Labs, Martlesham Heath (1998).

51. Roussos, M. et al.: NICE: combining constructionism, narrative and collaboration in a virtual learning environment. *Computer Graphics-New York- Association for Computing Machinery* **31**, 62–63 (1997).
52. Rovai, A. P.: Building sense of community at a distance. *The International Review of Research in Open and Distributed Learning* **3** (2002).
53. Schroeder, R.: *The social life of avatars: Presence and interaction in shared virtual environments* (Springer Science & Business Media, 2012).
54. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments* **10**, 266–281 (2001).
55. Serim, B., Jacucci, G.: Explicating "Implicit Interaction": An Examination of the Concept and Challenges for Research in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (2019), 417.
56. Shen, C., Everitt, K., Ryall, K.: UbiTable: Impromptu face-to-face collaboration on horizontal interactive surfaces in International Conference on Ubiquitous Computing (2003), 281–288.
57. Sheridan, T. B.: Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments* **1**, 120–126 (1992).
58. Sherman, W. R., Craig, A. B.: *Understanding Virtual reality: Interface, Application, and Design* (Morgan Kaufmann, 2018).
59. Short, J., Williams, E., Christie, B.: *The social psychology of telecommunications. The Social Psychology of Telecommunications* (1976).
60. Simonton, D. K.: Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological review* **104**, 66 (1997).
61. Slater, M., Usoh, M., Steed, A.: Depth of Presence in Virtual Environments. *Presence* **3**, 130–144 (1994).
62. Slater, M.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Phil. Trans. R. Soc. B* **364**, 3549–3557 (2009).
63. Slater, M., Wilbur, S.: A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments* **6**, 603–616 (1997).
64. Steuer, J.: Defining virtual reality: Dimensions determining telepresence. *Journal of Communication* **42**, 73–93 (1992).
65. Sugimoto, M., Hosoi, K., Hashizume, H.: Caretta: a system for supporting face-to-face collaboration by integrating personal and shared spaces in Proceedings of the SIGCHI conference on Human factors in computing systems (2004), 41–48.
66. Thiebaut, J.-B., Healey, P. G., Bryan-Kinns, N.: Drawing Electroacoustic Music. in *ICMC* (2008).
67. Titon, J. T., Slobin, M.: *The music-culture as a world of music. Worlds of music: an introduction to the music of the world's peoples.* New York: Schirmer Books (1996).
68. Wallace, P., Maryott, J.: The impact of avatar self-representation on collaboration in virtual worlds. *Innovate: Journal of Online Education* **5**, 3 (2009).
69. Wang, G.: Designing Smule's Ocarina: The iPhone's Magic Flute. in *NIME* (2009), 303–307.
70. Weinberg, G.: Interconnected musical networks-bringing expression and thoughtfulness to collaborative music making in Massachusetts Institute of Technology Media Laboratory (2003).
71. Witmer, B. G., Singer, M. J.: Measuring presence in virtual environments: A presence questionnaire. *Presence* **7**, 225–240 (1998).
72. Wozniowski, M., Bouillot, N., Settel, Z., Cooperstock, J. R.: Large-Scale Mobile Audio Environments for Collaborative Musical Interaction. in *NIME* (2008), 13–18.
73. Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., Merget, D.: The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior* **10**, 115–121 (2007).
74. Yin, R. K.: *Case study research and applications: Design and methods* (Sage publications, 2017).
75. Zhang, X., Furnas, G. W.: The effectiveness of multiscale collaboration in virtual environments in CHI'03 Extended Abstracts on Human Factors in Computing Systems (2003), 790–791.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

