ORIGINAL ARTICLE

Virtual environments for creative work in collaborative music-making

Michael F. Schober

Abstract Virtual environments are beginning to allow musicians to perform collaboratively in real time at a distance, coordinating on timing and conceptualization. The development of virtual spaces for collaboration necessitates more clearly specified theorizing about the nature of physical copresence in music-making: how the available communicative cues are likely to affect the nature of visually mediated rehearsal and performance. Pilot data for a project carried out at the New School for Social Research demonstrate some important factors relevant to designing remote spaces for musical collaboration, and suggest that virtual environments for musical collaboration could actually enhance the feeling of being together that creative musical expression requires.

1 Introduction

Music-making is a quintessentially creative activity, both in rehearsal and in performance. In rehearsal, the performer creates an interpretation of a piece, improvising upon some starting point—in Western classical and popular music, from a musical score (within a range of acceptable variability), and in jazz, from an underlying harmonic structure. In performance, creativity manifests itself in how the performance moves and breathes in that particular setting and moment. Collaborative music-making, where two or more performers rehearse and perform together, involves the same kinds of creativity in rehearsal and performance, but adds a large set of additional complexities. In rehearsal, collaborative musicians engage in a creative process of coming up with a joint interpretation or set of potential improvisations; in performance, they initiate and react to their partner's variations in pacing, flow, and mood, co-constructing the particular performance dynamically. This is true even when musical partners are following a score. No two performances are ever alike, and it is creative deviation that makes a performance sing.

Musical collaborators' feelings of copresence—of breathing together, of being able to anticipate each other's moves, of feeling both independent and like one being—are essential for this creative process to occur. And yet in ordinary musical settings there can be multiple impediments to this copresence. Physical constraints can get in the way; if partners cannot see each other, because of their placement in an ensemble or because their instruments are blocking the view, feelings of copresence can be diminished. Partners' level of knowledge about what the other is experiencing, or attentional limitations on their own performance (e.g., needing to fixate on a score), can make it hard or impossible to coordinate creatively.

Virtual reality allows us to imagine new forms of copresence that have the potential to enhance creativity in rehearsal and in performance. Freed from the constraints of the physical world, one could imagine being surrounded by one's cellist partner's bowing, seeing a visualization of the trumpet soloist's improvisation, or feeling one's singer partner's breathing in a

M. F. Schober (✉)
Department of Psychology, F330,
New School for Social Research, 65 Fifth Avenue,
New York, NY 10003, USA
e-mail: schober@newschool.edu

way that is physically impossible in real world settings. Enhanced access to cues could be useful both in rehearsal and performance.

At present, this sort of technology has not been fully developed. Broadband technology is only beginning to allow sufficient bandwidth to allow fully two-way communication with imperceptible lag (see, e.g., Barbosa 2003; Gu et al. 2005; Jung et al. 2000; Konstantas et al. 1998, 1999, for discussion of technical specifications for remote musical coordination). Beyond this, there is a vast amount of additional knowledge needed to understand what kinds of visualizations or sensory representations will actually enhance creative rehearsal and performance, and for which kinds of performers and musical styles. Systematic research on the visual and auditory cues that musicians use to coordinate in rehearsal and performance is underway (e.g., Blank and Davidson 2006; Himberg 2006), but at present far too little is still known. We have nothing like a systematic database or taxonomy of different kinds of musical coordination which would allow us to precisely specify how, for example, the cues that orchestra members use in coordinating with conductors and each other differ from the cues in rock bands, jazz ensembles, or string quartets.

This paper describes the beginnings of a line of research that will form the building blocks for understanding virtual environments for creative work in musical collaboration, allowing us to better understand which aspects of physical copresence are worth enhancing in a virtual environment and which are not. The work starts by looking at low-level forms of virtuality that are technically feasible today: remote video and remote audio connections.

Obviously these are only the tip of the iceberg as far as what virtual environments will allow, but understanding what happens here is an essential first step. The virtual space that is created even with remote video is—at least in the minds of the musicians using it—already at a substantial remove from the true physical copresence that musicians expect. Understanding what happens when visual cues are presented to musicians in a non-physically-copresent way is the first step towards far more elaborate virtual musical spaces.

2 Remote copresence in chamber music

In recent years musicians have created rehearsals and performances where two or more participants are separated in space and time, sometimes even across continents (e.g., Konstantas et al. 1997, 1998, 1999; Jung et al. 2000; see Barbosa 2003, for discussion of

system requirements for different kinds of collaborative musical tasks). Thus far bandwidth limitations remain such that the performances are not fully two-way in the way that live performances with physically co-present musicians are; typically a second party must follow a first party's lead because of notable signal delay. But such limitations are on their way to being reduced or eliminated so that musicians in different physical environments will be able to collaborate with no perceptible time delay. Consider, for example, the October 2002 Interactive Cross-Continental Jazz performance between Montreal, Canada and Palo Alto, CA, where the delay was at times reduced to 50 ms; this reportedly felt to the musicians "almost like being on the same stage" (see http://www.ultravideo.mcgill.edu/overview/).

The stage is thus set for the beginnings of far more elaborate virtual environments for collaborative musical performance. The sine qua non is that any such environment must allow signals to be passed so quickly that true two-way reactivity can occur. In order to build such environments, however, we will need to theorize the nature of musical collaboration far more clearly.

2.1 Levels of coordination

A key part of traditional chamber and jazz collaborative performance is that the performers must coordinate the timing and expressivity of their performances at the millisecond level. Simply coordinating competently enough that there is nothing obviously wrong—no mistimed attacks, no notes held longer by one party than the other, no obvious differences of opinion about what the right tempo should be or how exactly a phrase should slow down or speed up—is no mean feat, as skilled soloists who attempt to collaborate for the first time quickly discover. Musicians following a score have a script to follow that dictates a good deal about what they should do, sometimes including precise metronomic tempo markings and expressive timing cues. Musicians improvising without a score, say on the basis of a jazz chord progression, have a less detailed and more implicit "script" to follow. But with or without a score, how exactly the script gets enacted by particular performers varies across performers, and even within a performer on different occasions. This is why musicians rehearse a piece again and again before deeming the performance ready for the world. Coordination isn't an all or none phenomenon—performers can be more or less tightly coordinated on timing, dynamics, expressive features, and conceptualizations.

Beyond this, collaborative musicians strive to achieve a fuller sense of copresence or mind-meld, such that the coordination and give-and-take feels effortless and organic, and the pair or group operates as one entity. To use Shütz's (1951) characterization, they strive to achieve a "mutual tuning-in relationship by which the 'I' and the 'Thou' are experienced by both participants as a 'We' in vivid presence." Some performers are better at this than others, and some combinations of performers are better at this than others. Although the vocabulary for talking about this is impoverished, and not every musician feels comfortable discussing the issues, an almost mystical discourse has arisen among musicians who do talk about the shared mental space they work at creating. They talk of getting into the magic zone or getting into "chamber space," and they share their peak experiences, stumbling for words, as if describing food or sex. The give-and-take of high-level collaborative performance, where different partners "drive" (take the initiative) at different points, has the quality of transcendent play.

Of course, such a high-level feeling of copresence is not unique to collaborative musicians. Participants in other joint actions—pairs of ice dancers, actors performing a scene, tennis doubles partners—have related notions of being "in the zone." They also, in various ways, can distinguish between carrying out a joint activity competently enough that it works and entering a kind of shared space in which they achieve what feels like real mind-meld.

Little in current scholarly theorizing quite gets at distinguishing joint action at this very high level from ordinary joint action. The work on "flow" (Csikszentmihalyi 1992; Csikszentmihalyi and Csikszentmihalyi 1988) focuses on how individuals feel a sense of connectedness with and transcendence in the tasks they engage in, but does not focus on pairs' coordination. Theorizing on joint action and copresence (e.g., Blascovich 2002; Clark 1996) does not distinguish between competent and masterful coordination. In any case, far more needs to be understood from a scientific perspective about the basics of musical coordination, let alone the most transcendent levels.

3 Cues in collaborative musical performance

Predicting how musicians will coordinate in virtual environments, and what sorts of spaces will afford what sorts of coordination, requires greater understanding of the cues that musicians pass back and forth when they are physically copresent. Although the music that performers create is acoustic, visual cues are also available, and these cues have been proposed to be all-important (Schütz 1951):

"...making music together occurs in a true face-to-face relationship—inasmuch as the participants are sharing not only a section of time but also a sector of space. The other's facial expressions, his gestures in handling his instrument, in short all the activities of performing, gear into the outer world and can be grasped by the partner in immediacy. Even if performed without communicative intent, these activities are interpreted by him as indications of what the other is going to do and therefore as suggestions or even commands for his own behavior. Any chamber musician knows how disturbing an arrangement that prevents the coperformers from seeing each other can be" (p. 91).

Collaborative musicians are indeed trained to watch for each other's signals—to breathe together, literally and metaphorically. But just how crucial are visual cues for competent or masterful performance? Just how crucial is actual physical copresence, as opposed to merely being able to see the other's visual cues (for example, via remote video)? Just how harmed will coperformers be by not being able to see each other at all?

From musicians' perspective, visual cues ought to be particularly useful only in some circumstances. One could imagine that audio cues could partially or fully substitute—for example, a singer's inbreath may be just as audible as visible. Visual cues won't be particularly useful if the musicians are unfamiliar enough with the music that their attention is fully engrossed in reading a score, keeping track of where they are in the chord progression, or mastering their (solo) technique. Visual cues might also not be particularly important for a collaborating pair who have performed together so often that they can predict each other's stylistic moves from prior knowledge, or who have rehearsed a particular piece so well that the precise timing is prestored.

The section that follows lays out cues that are available to musicians when they can see each other in the same physical space, and explores how these cues will vary in video-mediated and audio-mediated settings.

3.1 Cues for physically copresent musicians

As collaborative cues will vary for particular settings and styles of music, let us consider as an initial test case a classical chamber setting with two collaborating musicians, a pianist and a singer. (This can form the basis for comparisons with other settings and styles. It

also has the useful feature that performance accuracy can be assessed in relation to the musical score). In the typical physical setup, the singer stands in the bend of the grand piano facing outwards towards an audience. In rehearsal, the singer will often have a music stand with a score on it; in performance, the singer may perform from memory. Collaborative pianists most often perform with a score.

3.1.1 Auditory cues

The most obvious auditory cues available to physically copresent performers are the sounds each instrument makes. Beyond the obvious cues of producing the correct pitches in the correct rhythm, each partner presents additional audio signals for coordination based both on the musical score and their interpretation: rhythmic and dynamic impulses that signal how the other partner should join in. How a singer or pianist creatively "moves" a phrase can clearly convey their performance intentions and conceptualization.

Other audio cues include the words that performers use in rehearsal to help them coordinate. For example, performers can use language to decide where in the piece to begin and to discuss tempo, dynamics, or other expressive features; they can use language to discuss their creative and interpretive visions of the piece. Other potential auditory cues include sounds of breath intake and sounds of the pages of scores turning. The quality of musical sound being produced by the partner can also be informative; for example, a pianist might hear that a singer is running out of breath and thus might end a sequence a bit more quickly.

Note that the acoustic quality of the audio available to each performer is rather different. Each performer hears the other's instrument from an external (audience) rather than internal (performer's) perspective, with the attending directionality and volume.

3.1.2 Visual cues

3.1.2.1 Jointly attended artifacts During rehearsal, both performers will usually have a score for the music they are working on, and so there will be periods of time that both performers are looking at the musical score rather than at each other. The scores may or may not be identical; both parties may or may not see exactly what the other sees, and page turns may or may not be in the same places. (With other pairings, it is more frequent that solo instruments have just their own part, and the pianist may have a score that shows both parts).

The score forms a point of joint attentional focus—even if it is not the same physical object—that is relatively unaffected by the physical setting or which aspects of the partner are visible. Nonetheless, attention to the score can differentially affect how singers and pianist see each other, because singers and pianists do not have identical views of each other. The typical physical arrangement for a pianist and singer allows the pianist to see the singer peripherally while looking forward at the score, or to focus on the singer with a simple glance upward. For singers looking at their score, the pianist is usually noticeably off to the side, and so looking at the pianist peripherally is less straightforward.

3.1.2.2 Visual cues from the partner The precise visual cues vary depending on the precise physical arrangement of the performers. Most frequently there will be a number of gaze cues, some of which will be intended as communicative and some of which are simply symptoms of the partner's processing. Partners can tend to see if the other is looking at the score, at their hands, at the audience, or at the other partner; occasionally they may look at each other simultaneously.

Body movements (e.g., "conducting" arm movements, large breath intakes), nods, and looks can be informative about one or the other partner's readiness to begin and their conception of the piece. When both parties need to begin music simultaneously after silence, in-breaths that signal the start of a phrase may be essential for the attacks to be coordinated. Facial and bodily affective expression may also be informative about a partner's performing intentions (fierceness, calm, joy, sorrow expressed for audience effect; a bodily lean that conveys a rhythmic or phrasal impulse), and other kinds of looks might be informative about a partner's having trouble due to nerves or fatigue.

3.2 Visual cues for virtual performers

Would all of the cues available to physically copresent performers be sufficiently available if the performers are video-mediated, or otherwise visually represented in a virtual space? Are there physiological cues that depend on being in the same space? This depends on just how much information is cut off through visual mediation. If the visual representations degrade the images of each other that performers see (e.g., shrinking them too far, reducing clarity, leaving some bodily or facial movements invisible), this could suffi-

ciently reduce the sense of shared presence and thus hinder creative musical coordination. If the visual representation actually omits the partner cues that are potentially essential for coordination—shared gaze, physical movement, facial expression—creative collaboration could be harmed. Of course, this only matters if what is degraded cannot be compensated for with other (auditory) cues.

If visual mediation simply *feels* unnatural, this could also distract enough from musicianship that musical coordination will be disturbed. Presumably, experience in virtual spaces of this sort could reduce the sense of unnaturalness.

3.3 Cues for audio-mediated performers

Performers with no access to visual cues of any sort must rely on audio cues instead. The extent to which this degrades performance should depend on the extent to which audio cues can replace visual cues.

For example, having to rely only on audio cues might be especially disturbing at particular stages of rehearsal or performance. Experienced performers who can decenter enough from their solo performance to be able to benefit from visual cues might suffer more from having to rely only on audio cues; inexperienced performers who are still focusing on the musical score might not mind so much. Also, independent of the technical (vocal or pianistic) difficulty for each partner, some pieces are harder to coordinate than others, with more complicated timing, attacks and tempos. Performances of such pieces should suffer more if only auditory cues are available. Finally, pairs of performers may become so used to each other's general style of performance or so well rehearsed on a specific piece that the visual cues are no longer paramount; they may be able to rely solely on sensitive listening to audio cues.

4 Pilot study

A project at the New School for Social Research has been starting to explore musicians' sense of copresence in simulated remote spaces of the future, where bandwidth and signal delay are not at issue. The methods of the project could also be used to explore the effects of various signal delays on the success of collaborative performance.

In the project, two relatively soundproof rooms are connected via video and audio cabling. A camera and microphones in each space pick up video and audio signals and project them onto a large monitor and high quality speakers in the other room. In the first iteration of the project, each room contained a grand piano. In the pianist's room, the camera and monitor were placed in the crook of the piano exactly where a singer would ordinarily stand, with the monitor placed so that the pianist seated at the keyboard could see it. In the singer's room, the camera and monitor were placed where the pianist would ordinarily sit, so that the singer standing in the crook of the piano could see just those parts of the pianist that she would ordinarily see. Through numerous iterative attempts it became clear that the camera and monitor needed to be closely aligned, so that when one party looked directly at the monitor the other party would experience their partner looking directly at them; even very small misalignments proved extremely distracting (Fig. 1).

With this setup, we could now compare the experience of rehearsing and performing in the same physical space (full physical copresence) with video-mediated copresence (video-and-audio copresence without physical copresence). We could also compare this with audio-only copresence by keeping the same audio setup (microphones and speakers) but removing the cameras and monitors. Again, this is many steps removed from the potential kinds of virtual environ-

Fig. 1 Singer's and pianist's video-mediated views of each other (minus time stamp)
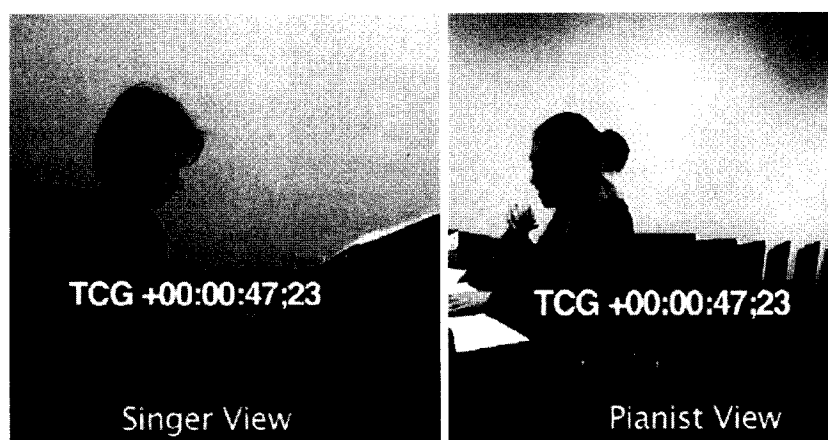
ments we intend to explore; it is simply a first step on the path toward more elaborate musical CVEs.

This setup allows for a number of different measures of the sense of presence. First, one can ask the performers to report about their experience. More intriguingly, one can assess the success of their collaboration on several dimensions. On the basis of the musical score, one can measure the accuracy of a pair's timing, at the millisecond level, with sound editing software: the extent to which notes that are supposed to start and end simultaneously actually do start and end simultaneously. One can assess the accuracy of the pitches produced. One can also solicit ratings (from the performers, from other trained musicians, and from a naïve audience) of the quality of the pair's conceptual coordination: the extent to which they conceived of phrases the same way.

Here I report on pilot data from two pairs of chamber performers who were strangers to each other before the test sessions. Both pairs were asked to rehearse and perform an unfamiliar art song, John Duke's (1967) "The White Dress." The piece was chosen not only because it was likely to be unfamiliar, but also because the vocal and piano lines are not of great technical difficulty. Nonetheless, the piece has its tricky parts; the meter varies between 3, 4, and 5 beats per measure, and the intervallic and harmonic changes are not necessarily predictable on an initial sight reading.

Both pairs were first given 10 min of solo rehearsal time in separate rooms (each room had a piano in it). They then had 10 min of rehearsal time in the same room (full physical copresence). Then they had a sequence of three "performances" (full run-throughs of the piece): first in the same room (full physical copresence), then in separate rooms but video- and audio-mediated, then in separate rooms with only audio mediation. (Obviously in a full-scale study some pairs would rehearse under video- or audio-mediation, and the ordering of performances would be counterbalanced). All rehearsals and performances were videotaped for subsequent analysis.

5 Results

The pilot results from the two pairs provide a set of intriguing observations about visual and auditory cues in rehearsal and performance. Overall, Pair A (Female singer, Male pianist) had a great deal of trouble sight reading the music and coordinating at all; Pair B (Male singer, Female pianist) started at a much higher level

and were able to give passable performances by the end of the session.

5.1 Rehearsals

During the joint rehearsal periods in the same room, both pairs ended up moving much closer to each other, with the singer standing facing the piano to the pianist's right. Apparently neither pair felt sufficiently copresent when they were placed in the typical performance setup (singer in crook of piano).

Pair A's rehearsal proceeded with great difficulty, with a great many starts and stops. The singer in particular had trouble finding the pitches and rhythms; the pianist sometimes forged ahead in hopes the singer would keep up, but this led to frequent breakdowns. At times the pianist stopped playing his part and simply played the singer's pitches in order to help her along. Both partners fixated steadfastly on their own scores, with only very occasional direct glances at each other during moments of breakdown (never during music-making).

Close attention to one early breakdown sequence is informative about how the pianist and singer managed to coordinate on where they were and when they would start. Here is a verbal and action transcript of the relevant moments for the singer (S) and pianist (P):

20:39	S:	[Fails to enter at *I shall* and stops singing]
	S:	"We're changing – changing three and four" [explains that metrical confusion caused breakdown]
	P:	[Plays singer's pitch]
	S:	"Oh"
	S:	[steps back]
20:53	P:	Says something unintelligible [looking at S's face 2 seconds]
20:55	S	"*motion?*" [proposes text/pitch as starting place]
20:58	S	[glances overtly at P's score]
20:58	P:	"*Yeah*" [overlapping speech]
	S:	"*Same-*" [overlapping speech]
	P:	"Oh yeah. The same" [points at his score to agree to proposed starting place]
	S:	"*Motionless?*"
21:04	P:	[begins playing]
	S:	[glances at P's hand 3 times in 6 seconds, perhaps to verify where he is]

A few things are notable in this sequence. First, agreeing on where to start in again takes some negotiation; the singer's proposal, although it turns out to be accepted, requires several turns of talk before the pianist starts again, and the singer must repeat "*Motionless?*" before she is convinced that they have agreed on where to start. Second, in addition to their verbal negotiation, both parties engage in visible

behaviors as they coordinate on where to start: the singer glances at the pianist's score, and the pianist points at his score as if to verify that they are in the same place. But neither of these visible behaviors is accompanied by the partner's looking at them. Even if the partners can see each other's visible cues peripherally, there is no evidence of any uptake by the partner. There is never any mutual gaze during this sequence; in fact, there is virtually no mutual gaze during the entire rehearsal (see Fig. 2).

It is as if the singer and pianist both are making visible gestures *available* to the other, but they are not checking at all whether their partner is noticing their gestures. This suggests that technical physical copresence does not automatically lead to full visual (or psychological) copresence, and that visual cues can go unnoticed. It also suggests that Pair A would likely not take advantage of visual cues in a virtual visually mediated setting since they neglected to use them to coordinate in the face to face setting.

Pair B's joint rehearsal is notable for other reasons. The pair made it through two full run-throughs of the song with far fewer stops and starts than Pair A. From the start, the singer took charge of the rehearsal, saying "Why don't I just start singing and you follow me, right?" He also set the tempo through much of the rehearsal by conducting with the arm that was not holding his score. As in Pair A, the singer in Pair B moved closer to the pianist, just behind her and to her right. It is unclear how much of the singer's conducting the pianist could see through her peripheral vision; evidently the singer become dissatisfied with how much the pianist could see, because by the end of the first run-through the singer had moved into the pianist's field of vision and was conducting in her face (see Fig. 3).

By this time, the singer had abandoned his own score so that the pianist and singer could jointly focus on a single score (the pianist's). This suggests just how important shared focus of attention—a jointly attended artifact—can be to the sense of copresence. Note that the videomediated condition, as set up in this experiment, would not allow this flexibility of positioning. To the extent that this pair depended on these sorts of visual cues, they should have more trouble in a virtual space that does not allow them.

Despite the singer's visual intrusions into the pianist's field of vision, the pianist remained focused on her score. The only times she ever looked at the singer were when they interrupted the music-making to talk. For example, the pianist looked directly at the singer when she proposed before the second run-through that the piece ought to be a bit slower than they had taken it the first time through. In the context of all the non-looking, this is quite notable; it makes sense that it should happen at a moment where the pianist presents a potentially confrontational opinion and asserts some control over the rehearsal situation.

The singer continued to conduct throughout the second run-through, most notably during a slowdown (ritardando) section, which both parties need to coordinate. The singer was dissatisfied, and even counted out loud during the ritardando to make clear how he would like it to go (note the redundancy of the visual conducting cues and the auditory counting cues; clearly one could substitute for the other if necessary). There was eventually explicit discussion about the ritardando at the end of the piece after the pair failed to coordinate on it; again, note that such explicit discussion does not depend on visual cues and could happen in a purely audio setup.

What is curious about the singer's conducting is that he continues to do it even after he moves to a position behind the pianist where she certainly cannot see it. This suggests that he conducts in part for himself rather

Fig. 2 Pair A focus on their scores and do not look at each other

Fig. 3 The singer in Pair B conducts in the pianist's field of vision

than simply to control the pianist; nonetheless, it is clear that sometimes the conducting is also intended to be communicative.

5.2 Performances

Pair A's performances are unimpressive; they are unable to get through the piece without stopping, and without the pianist's abandoning his part to play the singer's vocal line.

Pair B's performances are good enough that they can be evaluated for both pitch accuracy and rhythmic accuracy. The metric used here was the percentage of measures (out of 102 for the pianist, 75 for the singer) that contained no errors of commission or omission.

As Fig. 4 shows, the singer in Pair B was far more accurate than the pianist in general; this is unsurprising as the pianist had far more pitches to manage. Over the course of the three performances, the pianist improved notably between the first (physically copresent) and second (videomediated). The very fact of this improvement despite the video mediation suggests that there are practice effects that are independent of the visual cues that are being passed. Given that the pianist never looked at the singer during the performances, it makes sense that the reduction and absence of visual cues would do little to affect her performance.

The findings for rhythmic errors are slightly more complicated. As Fig. 5 shows, video mediation did seem to bring a slight decrement in rhythmic accuracy compared to full physical copresence. As the video mediation took away the possibility of conducting (which continued during the physically copresent performance), the reduction in accuracy may have to do with the loss of the conducting cue. Nonetheless, with
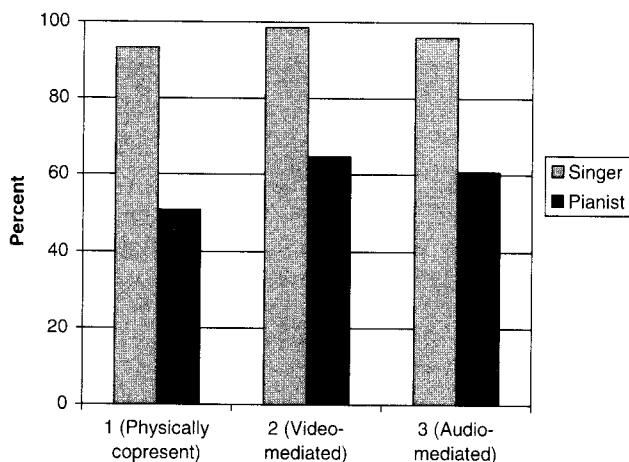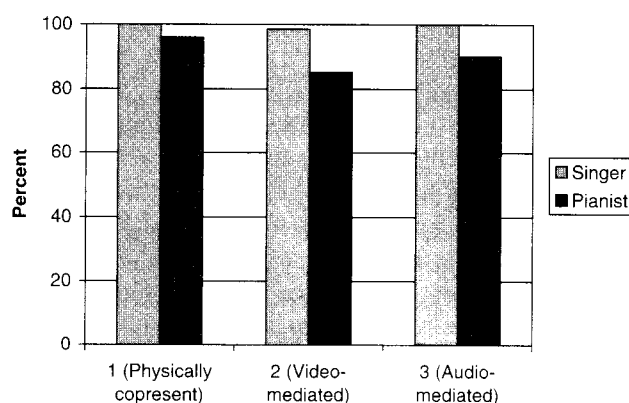
Fig. 5 Percentage of measures with no major rhythmic errors (out of 102 for pianist, 75 for singer)

additional practice—and despite the loss of additional visual cues—rhythmic accuracy improved in the audio-mediated condition.

Informal self-reports from the participants suggest that they did not find either video or audio mediation particularly troubling, and they were intrigued at the prospect of remote rehearsal and performance. They were quite aware of the coordination troubles they had had, but did not seem to feel that the remote settings affected the quality of their coordination one way or another.

6 Discussion

Obviously, a large sample of performers in different genres of music would have to be tested in order to be able to make generalizable claims about the broader effects of even the simple kinds of visual and audio mediation studied here on collaborative music rehearsal and performance. (Such studies are in the works). And much more work would need to be done on different kinds of visual representation of musicians in more elaborate virtual environments to understand what kinds of information about one's partner is particularly effective for facilitating creative collaboration. But these pilot data, as well as informal tests with experienced chamber musicians during the setup of the rooms, allow some plausible conclusions.

First, the video-and-audio situation can instantly feel remarkably like the fully copresent situation. At one level this should not be surprising; many of the same sets of visual and auditory cues—intakes of breath, body leans, eyebrow raises—are, after all, available. But one might expect some delay in feeling comfortable with the remote situation, and this is not what our participants reported. This is encouraging

Fig. 4 Percentage of measures with no pitch errors (out of 102 for pianist, 75 for singer)

news for imagining more elaborate virtual environments for coordinated music-making.

Nonetheless, there may be decrements in performance—as in Pair B's rhythmic coordination—resulting from the lack of real physical copresence that are independent of the performers' level of comfort. This should vary depending on how much coperformers rely on visual cues of each other (as opposed to focusing on a score) as they rehearse and perform. But note that even Pair B's decrements were not large, and on another dimension of coordination (pitch accuracy) Pair B *improved* across performances (a practice effect) despite visual mediation. Of course, we cannot know if they would have improved more if they had been fully copresent; that is what a larger study could show. But given how rarely the pianist in this pair looked at the singer at all, any effects of mediation are likely to be small.

All in all these findings suggest that the sense of remote copresence is established through measurable sensory cues (here, visual and audio), and that manipulating these cues can affect how copresence is achieved. Disentangling all the factors that make a difference—not only for competent but also for masterful coordination—will require a good deal of further research. One might imagine, for example, that the effects of mediation on either rehearsal or performance can be affected by

- Individual technical mastery and skill level—performers who are less attentionally tied to their scores should have more resources available for making use of visual cues, and thus be more likely to be affected by their absence. On the other hand, highly skilled performers may be more able to substitute audio for video cues without decrement.
- A pair's familiarity with each other and experience performing a particular piece could make visual cues less necessary.
- Individual personality differences could make visual cues more or less important to a musician, which would thus affect who will be influenced most by the loss of visual cues. As Blascovich (2002) proposed, there are individual differences in susceptibility to social presence effects that should correspond with differences in levels of empathy and suggestibility.

In the longer run, it will be important to understand how findings from this kind of study will scale up to far more elaborate virtual environments. Do representations of one's musical partner need to look human? Would it be more useful—for training purposes in rehearsal, or even in performance—to see a representation of one's partner's breathing, heart rate, or galvanic skin response than to see the partner's face? to see a visual representation of one's partner's emotions that does not rely on the partner's face? to see information about one's partner amidst the musical score that one is reading rather than having to switch attentional focus back and forth between one's score and one's partner's cues? To what extent would it be helpful for each performer to control or manipulate how their avatars are displayed for their partners (as suggested in the remote acting rehearsals in Slater et al. 2000)? Might it be helpful for musical partners to use VR technology to augment reality even in physically copresent situations—say, representing each other's breathing? Which real-time VR representations of aspects of musical partner's performance can already be implemented with today's networks, given that they require less bandwidth than full-scale video networking?

The set of factors that are likely to be relevant for answering these kinds of questions is large. I propose that simulated experimental setups that do not require enormous amounts of development effort to get off the ground, like the one used in this pilot study, will allow careful manipulations that can help provide guidelines for the development of interfaces for virtual musical environments that promote both competent and masterful creative coordination at a distance. At the same time, studies like these have the added benefit that they can help clarify our understanding of the nature of musical coordination more generally—how pairs of people come to feel that "we-ness" that makes their performances take flight.

References

Barbosa A (2003) Displaced soundscapes: a survey of network systems for music and sonic art creation. Leonardo Music J 1(13):53–59

Blank M, Davidson J (2006) "Nod when you're ready": an investigation into the gestures used by co-performers during rehearsal and performance. Paper presented at the Second International Conference on Music and Gesture. Manchester

Blascovich J (2002) Social influence within immersive virtual environments. In: Schroeder R (ed) The social life of avatars: presence and interaction in shared virtual environments. Springer, Berlin Heidelberg New York, pp 127–145

Clark HH (1996) Using language. Cambridge University Press, Cambridge

Csikszentmihalyi M (1992) Flow: the psychology of happiness. Random House Limited, London

Csikszentmihalyi M, Csikszentmihalyi I (1988) Optimal experience: psychological studies of flow in consciousness. Cambridge University Press, Cambridge

Gu X, Dick M, Kurtisi Z, Noyer U, Wolf L (2005) Network-centric music performance: practice and experiments. IEEE Commun Magazine 43(6):86–93

Himberg T (2006) The role of visual and auditory communication channels in musical interaction. Paper presented at the Second International Conference on Music and Gesture, Manchester

Jung B, Hwang J, Kim GJ, Lee E, Kim H (2000) Incorporating copresence in distributed virtual music environment. In: Proceedings of ACM symposium on virtual reality systems and technology (Seoul, Korea, October 22–25, 2000), VRST '00. ACM Press, New York, pp 206–211

Konstantas D, Orlarey Y, Gibbs S, Carbonel O (1997) Distributed musical rehearsal. In: Proceedings of the international computer music conference (ICMC 97). The International Computer Music Association, San Francisco, pp 279–282

Konstantas D, Gibbs S, Orlarey Y, Carbonel O (1998) Design and implementation of an ATM based distributed musical rehearsal studio. In: Proceedings of ECMAST'98, 3rd European conference on multimedia applications, services and techniques, Germany

Konstantas D, Orlarey Y, Carbonel O, Gibbs S (1999) The distributed musical rehearsal environment. IEEE MultiMedia 6(3):54–64

Schütz A (1951) Making music together: a study in social relationship. Soc Res 18:76–97

Slater M, Howell J, Steed A, Pertaub D, Garau M (2000) Acting in virtual reality. In: Churchill E, Reddy M (eds) Proceedings of the third international conference on collaborative virtual environments (San Francisco, CA). CVE '00. ACM Press, New York, pp 103–110