Spatial Co-ordination in Music Tuition

Sam Duffy & Patrick G. T. Healey
Queen Mary University of London
Media and Arts Technology Programme
Interaction Media and Communication Group
School of Electronic Engineering and Computer Science
London, E1 4NS, United Kingdom

s.duffy@eecs.qmul.ac.uk, ph@eecs.qmul.ac.uk

Abstract

Playing an instrument is a physical skill learned through observation, repetition and rehearsal. Students of orchestral instruments seek one-to-one tuition from expert musicians. However as they become more accomplished, the number of suitable tutors becomes more concentrated, especially for less common instruments. Often a tutor-student relationship develops over several years and temporary separation due to overseas performing, auditioning and teaching commitments is problematic. Some music education organisations use video conferencing as a solution to these problems, however it has long been recognised that interaction mediated by video conferencing is not analogous to a co-present experience. In this paper, ethnographic video analysis is used to study the interactions in co-present and separated instrumental music lessons. We find that the musical score represents more than a physical embodiment of the music - it plays an important role in coordinating activity and interaction. In video mediated lessons a single physical score can no longer be shared and interaction is changed as a result.

Keywords: video conference; interaction; ethnography.

Introduction

A recognised method for learning to play an orchestral musical instrument is through regular one-to-one lessons with an experienced tutor. Playing an instrument is a physical practice requiring dextrous manipulation of a complex tool. Marchand (2010) proposes that the interpretation, understanding, and realisation of practice is based in motor cognition. As the student watches the tutor, visually processed signals are paired with observed actions, gestures, and postures. These may be co-ordinated with verbal instruction and commands from the tutor. However learning a musical instrument cannot be achieved purely by verbal description or observation, students learn a practice by ‘doing’. Through observation followed by repetition and rehearsal, with iterative feedback from the tutor, the student develops motor and kinaesthetic cognition of how to play their instrument. This is a collaborative process, involving the co-ordination of understanding (Clark & Brennan, 1991).

At an undergraduate level of study, music students seek professionally recognised performers as tutors and their choice of where to study could be influenced by resident tutors and professors at an institute. The number of qualified professional tutors in any particular field is finite, but becomes more limited the more accomplished a student becomes, especially for less common instruments. Once a teaching relationship has been established, musicians tour and travel frequently, so temporary separation of tutor and student can occur at critical times, such as prior to an important audition or performance. One solution to these problems is video conferencing. This is popular in geographically remote areas such as Australia (Lancaster, 2007) but is also part of urban mainstream conservatoires such as The Manhattan School of Music in New York. However interaction when video is the medium of communication is not analogous to the co-present experience. The belief at the inception of video conferencing that technology which replicated face-to-face interaction, simply at a distance, would enhance communication, contained a fundamental misunderstanding about how people interact when working collaboratively to achieve a task (Whittaker, 2003; Edigo, 1988; Hollan & Stornetta, 1992). Heath et al (1997) found that the visual focus of collaborative work is likely to be aligned to the focal point of the activity, such as a document or object, rather than face-to-face.

Existing research concerning separated musicians has focused on collaborative performance, the impact of latency and delays and tools to enable distributed ensembles to perform, improvise and compose (Chew et al., 2004; Sarkar & Vercoe, 2007; Hamilton, Iyer, Chafe, & Wang, 2008; Barbosa, 2003; Bryan-Kinns & Healey, 2006). However the activity taking place during music tuition is not the same as performance due to the educational frame of reference.

The Use of Shared Space

Individuals in shared space coordinate their actions through spatial awareness, peripheral monitoring of non-verbal signals and the ability to joint reference; where gaze and gesture around a shared point coordinates the attention of participants (Whittaker, 2003). They use their position relative to each other to create mutually recognised shared space. Kendon (1990) describes how two or more people can organise themselves to create and sustain a shared space, called ‘o-space’, to maintain a common focus of attention. He goes on to describe sustained clusters and patterns as formations, an ‘F formation’ being where participants have equal, direct and exclusive access to their o-space. When two people are performing a collaborative task through the medium of video, they no longer have a concept of negotiated mutual distance (Sellen, 1992) and they cannot easily manage their position relative to each other or objects in their environment, there-

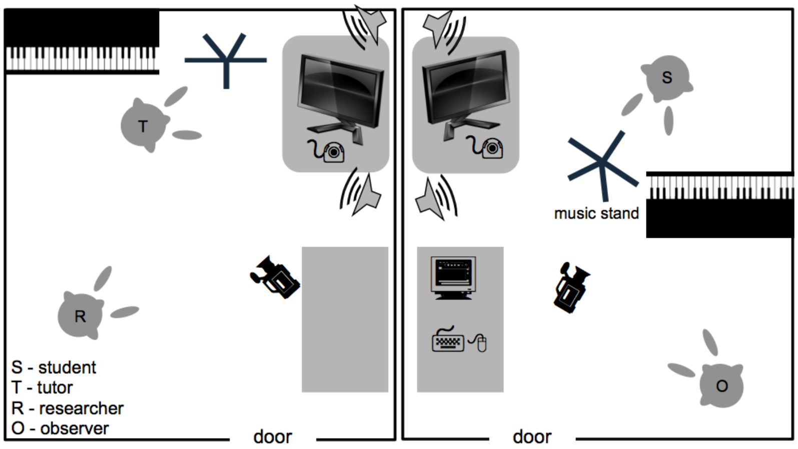

Figure 1: The layout for video lessons.

fore participants cannot use formations to create an o-space. Gestures and gaze are also shown to be less efficient in video mediated communication (Heath & Luff, 1991b). There is a body of work investigating the importance of gestures and non-verbal communication to teaching (Roth, 2001), to performing musicians (Vines, Wanderley, Krumhansl, Nuzzo, & Levitin, 2004; Wanderley, Vines, Middleton, McKay, & Hatch, 2005; Broughton & Stevens, 2009) and even specifically to instrumental music tuition (Kurkul, 2007; West & L. Rostvall, 2003). In this paper we analyse the interactions of co-present musicians in a learning environment and compare them to the interactions seen when student and tutor interact through the medium of video.

Methodology

Ethnography requires a researcher to participate in people's daily lives, watching what happens, listening to what is said and gathering data to understand the issues emerging (Hammersley & Atkinson, 2007). It can offer fine-grained, detailed qualitative insight into how users interact with technology (Whittaker, 2003) and is often used as a tool to assess HCI and CSCW systems, for example the studies by Heath et al (1997; 1991a; 2005). It is the only way to study embodied social practice as it naturally occurs, rather than in conditions created by the researcher.

Co-present Studies We observed co-present music lessons taking place at the educational establishments where the students would normally have their weekly lessons. Three thirty-minute lessons were observed and filmed, two clarinet lessons and a trumpet lesson. The students observed were preparing for Grade 7 or 8 exams[1]. A researcher was present since it was not possible to know in advance how much the participants would move around the room, necessitating repositioning of the camera (see researcher position R1 and R2 in Figure 2), however the researcher took no part in the lessons. The footage was analysed using ELAN and a detailed transcript produced for each class.

Video Mediated Studies The video conference data was obtained from a study run by British Telecom Research and Development[2] to evaluate a video conference prototype, designed specifically to support instrumental music tuition. We were invited to observe tests which involved students and visiting tutors at Aldeburgh Music in Suffolk. Six one-hour lessons using the prototype were observed and filmed over three days, including harp, cello, violin, oboe and french horn. The tutors had a photocopy of the student's music, or their own editions of the score to be worked on. Some of the tutors had previous experience of teaching via video conference and some of the student-tutor pairings had worked together previously. A researcher observed from the tutor room, there already being an observer from the prototype team in the student room. Video footage was obtained from both rooms (see camera positions in Figure 1) and analysed synchronously.

Results and Discussion

Professional musicians interviewed as part of this work believed latency to be the biggest barrier to teaching via video conference, as the delay makes it very difficult to play together (Chew et al., 2004). However analysis of the three co-present lessons showed that synchronous activity (singing or playing together, accompanied playing, the tutor conducting) made up only 11 percent of the lesson on average. This activity was used largely to resolve specific rhythmic problems. Where video conference is used to manage temporary separation of a student-tutor pair, many normal lesson activities can still take place, synchronous tools being saved for the next co-present lesson. The impact of the medium on interaction seemed to be a more significant problem as this affected

[1] Grade 8 from a recognised exam board such as the Associated Board of the Royal Schools of Music is often an entry requirement for music performance undergraduate degrees

[2] As part of the EU FP7 project Together Anywhere Together Anytime ("TA2") http://www.ta2-project.eu/Pages/overview.html

all lesson activities.

Co-present Lesson Interactions

The rooms where lessons took place were small, the space constrained by the piano and the music stand. Nonetheless, in each case, tutors used their position relative to the student and the music stand to communicate their intention to act. This led to the establishment of specific zones within the space, which both participants could be seen to observe. To illustrate this we present vignettes illustrating examples of the use of these zones.

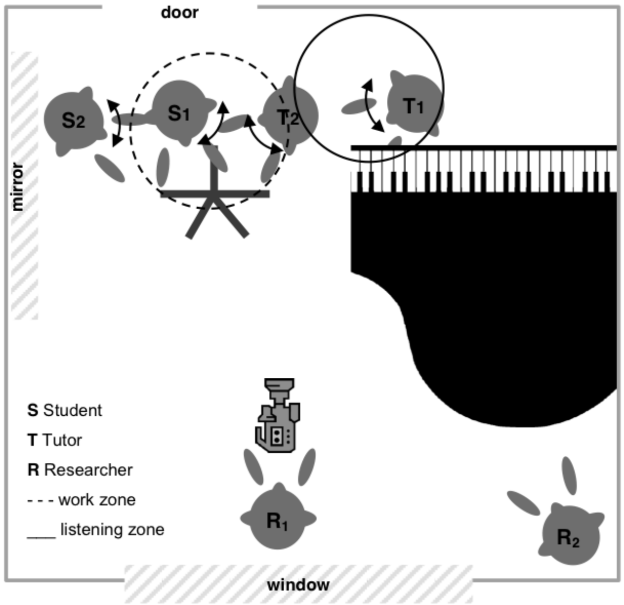

Figure 2: The work zone and listening zone.

The Work Zone The music stand became the focal point of a side-by-side F-formation (Kendon, 1990) as participants shared the student's score, the o-space created between them being designated as a 'work zone' (Figure 2). In one example, towards the end of the lesson the student and tutor briefly relax into social conversation. The student moves back slightly and moves her upper body to face the tutor, rather than the music (position S2 Figure 2). The tutor moves her upper body to face the student, the stand no longer the focus of their o-space. The student rubs her shoulder and moves about, relaxing her muscles (Figure 3). The tutor then puts her left hand on the music stand, between two short utterances, whilst still maintaining eye contact with the student. In this way the tutor holds both the stand and the student, triangulating her position (Healey & Battersby, 2009) as she begins the transition back to the work zone. Finally, the tutor turns her head back towards the music stand, pulling her body round, facing back into the work zone (Figure 4). The student also turns back to the stand (position S1 Figure 2), and swings her clarinet up towards her face, having understood the signal that they are going to go back to work.

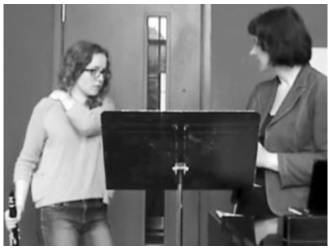

Figure 3: Moving out of the work zone.

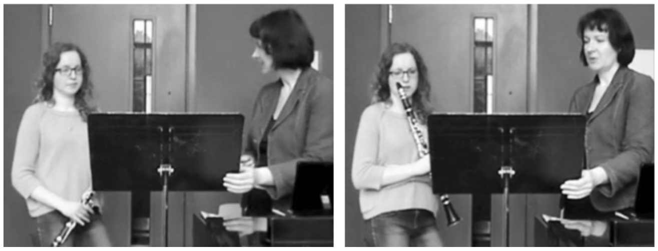

Figure 4: The tutor signals 'back to work'.

The Listening Zone In each case, the tutor defined a listening zone (for example T1 in Figure 2). When the tutor occupied this zone, the student understood that they could play without immediate interruption as the tutor wanted to hear a longer section. The tutor also defined a listening position within the listening zone, for example one tutor stood with feet slightly apart, hands loosely folded in front of her body, shoulders relaxed; attending to, but not bidding for, the floor.

Transitions Between the Zones When a tutor wanted to give detailed feedback on a passage, they moved forward into the work zone (position T2 Figure 2). When they stepped backwards into the listening zone again, the student understood that the specific topic of detailed work was finished and they should play a longer passage again, for the tutor to listen to and assess.

The transition from listening position to interruption could be sudden or more gradual. In some cases, the first indication that a tutor had diagnosed a problem from their listening position was when they lifted their gaze from the score to look at the student's face or instrument. Sometimes the student had already looked up, aware of their error, to see if it had been detected. In other cases the student demonstrated peripheral awareness; looking up in response to the tutor's movement, returning their gaze, and sometimes even stopping playing. The intent to make a more significant interruption could be indicated in advance by the tutor leaning into the piano to pick up a pencil whilst still in the listening zone, signalling intention to write on the score; or picking up their own instrument, indicating that they wished to demonstrate something the student had played incorrectly.

The duration of the tutor's planned interruption was indicated through the extent of her movement into the work zone.

When the tutor intended only a brief interruption she leaned forward into the work zone, without moving her feet, and pointed to the score whilst giving feedback. Then leaned back into her 'home position' having used body torsion to indicate a temporary movement into the work zone, the lower body remaining in the listening zone or 'base position' (Schegloff, 1998).

Control of Musical Turns

When a student is playing, they are emotionally engaged with their performance and concentrating on the score. However the tutor frequently interrupts to provide immediate feedback on an identified problem. Frequent, unexpected interruptions could become frustrating, however student-tutor pairs managed interruptions in order to reduce the impact. In a music lesson, where a verbal utterance is often followed by a musical response, it is reasonable to assume that a musical phrase is analogous to a conversational turn and we should therefore be able to see the characteristics of turn management (Sacks, Schegloff, & Jefferson, 1974) such as transition-relevant places for a new turn, back-channelling (Moran, 2011), turn breakdown and repair.

We could see evidence of musical turn management whilst the student was playing. Musicians have been shown to have awareness of anticipation of other musician's intentions with respect to musical structure (Moran, 2011) and the position between two adjacent musical phrases or ideas was observed to be preferred by the tutor as a transition-relevant place to interrupt. Even if a change in the tutor's posture indicated that they had diagnosed a problem earlier in a phrase, they would often wait for the long note at the end of the phrase before initiating interruption of the student's performance. This provided an ideal opportunity for the tutor to speak, whilst the student had a natural point to finish, before moving on to the next musical idea.

The transition from verbal instruction, to visual observation, to motor cognition (Marchand, 2010) was observed. After verbally explaining a point, in some cases the clarinet tutor then demonstrated it for the student on her own instrument. The student was asked to imitate. If the tutor was not satisfied with the performance she played the phrase again, exaggerating the aspect not yet corrected. The alternating musical turns increased in intensity, the student's technique converging over time with that of the tutor's. In one example a student impatiently copied the tutor's demonstration, starting to play before she had finished demonstrating. The tutor did not detect any noticeable improvement and musically admonished him, interrupting his performance and playing it again herself, taking back the turn.

The Shared Score

The score played a much greater role in the lesson than a physical embodiment of the music, also providing a shared reference to coordinate joint attention (Whittaker, 2003). In co-present lessons the participants shared a score, usually that which belonged to the student, and both pointed to parts of the score as they spoke as a convenient way to reference without having to mention bar numbers specifically. Tutors gestured over the score as they talked, linking their comments to specific notes and putting phrases into context of the whole piece. For example, in one lesson the tutor pointed to the music 32 times, half of the instances for navigation purposes such as "from here" and half to reference feedback against musical notation.

Direct eye contact was shared for less than 5 percent of the time, and was made up of brief glances (for example one lesson contained 112 instances of shared gaze, the average duration being less than one second). More often both looked at the score, even when in conversation together, whilst exhibiting a high level of peripheral awareness. For example, when one student performed for the tutor and stumbled over a note, the tutor immediately moved in towards the score with her pencil, starting to speak. However the student interrupted before her pencil reached the score saying "change my right?", his gaze not having moved from the music. The tutor stopped moving, looked up at him nodding and said "you read my mind, yeah" then moved back to her listening position.

Whilst the students had all made pencil annotations on their music outside of the lesson, such as marking fingerings, phrase marks, accents and breathing; during the lesson it was the tutor who annotated the score. For example the clarinet tutor used the character 'O', writing it on specific notes to indicate the 'open throat' required to control tone in some registers or 'X' to signify a particular spacing of fingers on the keys. Through annotation, the student's score built into a permanent and cumulative record of the learning imparted by the tutor; a record of how they had developed.

Video Mediated Lesson Interactions

In the video mediated lessons, students demonstrated awareness of their need to monitor both their music and the tutor at the same time, by initially positioning themselves directly in front of their screen so that they could see both the tutor, and their music, without significantly moving their head. However the score and music stand obstructed the main camera view (Figure 5) and in all cases tutors asked the students to turn around so that they could see their hands on the instrument. The students were then turned between 45 and 90 degrees away from their screen (position S in Figure 1) and no longer had peripheral awareness of the tutor when performing.

The Divided Score

Now that the student's gaze was divided between their score and the screen, there was a dramatic impact on turn control. With a separate music stand and score in each room, a shared work zone could no longer be established, and the tutor could not create a listening zone, removing communication through spatiality. Whilst some tutors still adopted a listening position (for example one tutor formally placed her folded hands in her lap - see Figure 6) they now used exaggerated gestures such as a raised arm or a wave to indicate when they

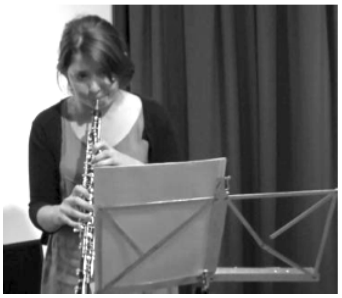

Figure 5: The score obstructs the camera view.

wanted the student to stop playing, and when these were also unseen, resorted to a vocal request. This was sometimes not heard by the student who was absorbed in their playing and not facing the video conference system speakers, and the tutor had to raise their voice. Even when they were able to rapidly switch their attention, students missed cues through looking in the wrong place at the wrong time. From their perspective, they were continually being stopped unexpectedly by a raised voice, requiring a significant twist of their upper body to see the screen (Figure 7) and this quickly led to frustration.

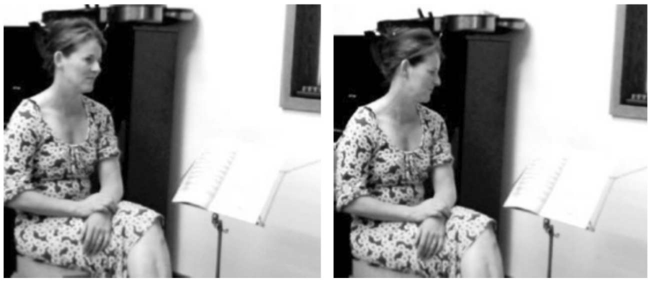

Figure 6: Tutor's gaze divided between screen and score.

Tutors also struggled with dividing their gaze between the screen and the score (Figure 6). Often they would discuss a phrase looking down at their score, as the student looked at their own separate score. Neither party could monitor their video screen at the same time. Previously the tutor could use peripheral awareness of the student's gaze as evidence of continued attention and the student could use indicative gestures to confirm their understanding of the feedback and it's relation to the score (Clark & Brennan, 1991). Navigation became problematic as a result, requiring detailed reference to page and bar numbers to establish precisely where in the score feedback related to, or for the tutor to establish where they would like the student to play from. For example, as shown in the following extract of dialogue.

Tutor: that wasn't quite right, let's try it again...
[the student starts to play]
Tutor: ...from the beginning of the bar.
[student stops, looks up]
Student: from the, sorry? From the?
Tutor: from the beginning of the bar.
[the tutor starts playing the phrase that she wants to hear, the student is looking hesitant]
Tutor: can you play it from the beginning of the bar and stop on the B and the E?
Student: OK [with instrument raised to playing position]
Tutor: Do you see where I mean?
[the student looks at the music intensely, wiggling her fingers on the fret board]
Student: "uh hum" [hesitantly]

The tutor could no longer directly annotate the student's score and it was noted that, in comparison to the co-present lessons, notes and annotations were not frequently made by either participant. One tutor made reference to a student's annotations where they were available on their photocopy of the student's music, confirming their value.

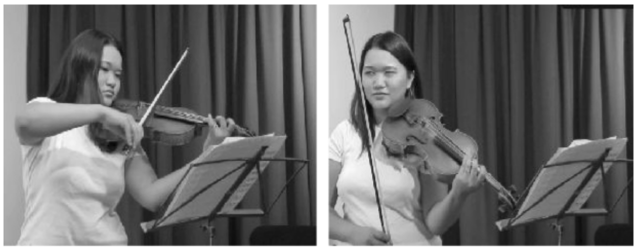

Figure 7: The student must switch gaze from score to screen.

Conclusions and Further Work

Ethnographic analysis of co-present lessons provided a useful framework to assess the effectiveness of video mediated communication to teach a practice based skill. The importance of the shared score to lesson interaction was evidenced by problems managing interaction such as turn control when participants were separated and could no longer share the same physical representation of the music.

A further study is planned where the instrument class will be confined to woodwind (for example clarinet or oboe) and an additional camera will be placed behind the participants to capture gesturing over the score in more detail. The score will be photographed and the annotations discussed with participants during post-observation interviews. Technological solutions to the problem of interaction lost through the divided score will be suggested. These are likely to involve an interactive visual layer over a digitised representation of the physical score, which shows the separated participants where each person is gesturing on the music. Ideally both participants should be able to mark their layer in a way which allows the student to take an annotated copy away, and return with it for the next lesson. There should be a way for the tutor to communicate intent to interrupt the student's performance through visualisation of gestures on the music. In this way some of the functions of the shared score can be introduced to the separated lesson.

Acknowledgments

The author would like to thank Doug Williams at British Telecom Research and Technology Future Content Group; Jonathan Reekie, Bill Lloyd and Marie Bennell at Aldeburgh Music; Ruth Aldred at Manchester Grammar School and Nicola Baigent at Trinity School of Music. Thank you to all of the students and tutors who kindly allowed their lessons to be filmed and the reviewers for their helpful feedback. The Media and Arts Technology programme is supported by an EPSRC Doctoral Training Centre EP/G03723X/1. Arts Council England supported Aldeburgh Music's purchase of equipment.

References

Barbosa, A. (2003, December). Displaced Soundscapes: A Survey of Network Systems for Music and Sonic Art Creation. *Leonardo Music Journal*, *13*, 53–59.

Broughton, M., & Stevens, C. (2009, April). Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*, *37*(2), 137–153.

Bryan-Kinns, N., & Healey, P. (2006). Decay in collaborative music making. In *Proceedings of the 2006 conference on new interfaces for musical expression* (pp. 114–117). Paris: NIME.

Chew, E., Zimmermann, R., Sawchuk, A. A., Kyriakakis, C., Papadopoulos, C., François, A. R. J., et al. (2004). Musical interaction at a distance: Distributed immersive performance. In *Music network 2004* (pp. 1–10). Barcelona: DIP Publications.

Clark, H., & Brennan, S. (1991). Grounding in communication. *Perspectives on socially shared cognition*, *13*(1991).

Edigo, C. (1988). Videoconferencing as a technology to support group work: A review of its failure. In *Proceedings of the 1988 conference on computer-supported cooperative work* (pp. 13–24). Portland, Oregon: ACM.

Hamilton, R., Iyer, D., Chafe, C., & Wang, G. (2008). To the Edge with China : Explorations in Network Performance. *Computer*, 7–8.

Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice* (Third ed.). London: Routledge.

Healey, P., & Battersby, S. (2009). The interactional geometry of a three-way conversation. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 785–790). Amsterdam: COGSCI.

Heath, C., & Lehn, D. (2005). Interaction and interactives: collaboration and participation with computer-based exhibits. *Public Understanding of Science*, 1–23.

Heath, C., & Luff, P. (1991a). Collaborative activity and technological design: Task coordination in London Underground control rooms. In *Proceedings of the second conference on european conference on computer-supported cooperative work* (pp. 65–80). Amsterdam: Kluwer Academic Publishers.

Heath, C., & Luff, P. (1991b). Disembodied conduct: communication through video in a multi-media office environment. *Proceedings of ACM CHI 91 Human Factors in Computing*, 99–103.

Heath, C., Luff, P., & Sellen, A. (1997). Reconfiguring Media Space. *Video-mediated communication*, 323–347.

Hollan, J., & Stornetta, S. (1992). Beyond being there. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92*, 119–125.

Kendon, A. (1990). *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge: CUP.

Kurkul, W. W. (2007, February). Nonverbal communication in one-to-one music performance instruction. *Psychology of Music*, *35*(2), 327–362.

Lancaster, H. (2007). Are we (virtually) there yet? Face-to-face v. virtual learning landscapes in musical instrumental teaching. *CAUCE 2007*, 1–16.

Marchand, T. H. (2010, May). Embodied cognition and communication: studies with British fine woodworkers. *Journal of the Royal Anthropological Institute*, *16*, S100–S120.

Moran, N. (2011, May). Music, bodies and relationships: An ethnographic contribution to embodied cognition studies. *Psychology of Music*.

Roth, W.-M. (2001, January). Gestures: Their Role in Teaching and Learning. *Review of Educational Research*, *71*(3), 365–392.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 696–735.

Sarkar, M., & Vercoe, B. (2007). Recognition and prediction in a network music performance system for Indian percussion. In *Proceedings of the 7th international conference on new interfaces for musical expression* (Vol. 2, pp. 317–320). New York, New York, USA: ACM.

Schegloff, E. (1998). Body torque. *Social Research*, *65*(3).

Sellen, A. (1992). Speech patterns in video-mediated conversations. In (pp. 49–59). Monterey: ACM.

Vines, B., Wanderley, M., Krumhansl, C., Nuzzo, R., & Levitin, D. (2004). Performance gestures of musicians: What structural and emotional information do they convey? *Gesture-based communication in human-computer interaction*.

Wanderley, M., Vines, B., Middleton, N., McKay, C., & Hatch, W. (2005, June). The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *Journal of New Music Research*, *34*(1), 97–113.

West, T., & L. Rostvall a. (2003, May). A Study of Interaction and Learning in Instrumental Teaching. *International Journal of Music Education*, *40*(1), 16–27.

Whittaker, S. (2003). Theories and Methods in Mediated Communication. *The handbook of discourse processes*(973), 243–286.