Co-ordinating Non-mutual Realities: The Asymmetric Impact of Delay on Video-Mediated Music Lessons

Sam Duffy (s.duffy@qmul.ac.uk)
Patrick G. T. Healey (p.healey@qmul.ac.uk)
Cognitive Science Group
School of Electronic Engineering and Computer Science
Queen Mary University of London
London, E1 4NS, United Kingdom

Abstract

During a music lesson, participants need to co-ordinate both their turns at talk and their turns at playing. Verbal and musical contributions are shaped by their organisation within the turn-taking system. When lessons are conducted remotely by video conference, these mechanisms are disrupted by the asymmetric effects of delay on the interaction; in effect a "non-mutual reality" comprised of two different conversations at each end of the link. Here we compare detailed case studies of a co-present and a remote music lesson, in order to show how this effect arises, and how it impacts conduct during the lesson.

Keywords: video mediated communication; conversation analysis; music education; distance learning

Introduction

When a student and tutor come together for the purpose of an instrumental music lesson, intuition suggests that the principal activity would be playing. However conversation is important, not just as a way to analyse musical contributions, but to organise them within the lesson flow. Participants may respond to talk with performance and vice versa, or even spend periods of time exchanging purely musical contributions (Duffy & Healey, 2014). For example the tutor could give a verbal instruction that the student should action through performance, or the student could ask a question that the tutor answers through demonstration with their instrument. Activities are managed conversationally; discussion interleaved with performance, demonstration and musical experimentation, resulting in a rich multi-modal social interaction. The musical contributions include unscripted exchanges of short musical fragments intertwined with lesson dialogue. Analysis of their shape and timing shows that they are managed in ways analogous to conversational turn-taking. For example, a tutor's musical contribution can be used to initiate student self-repair in their performance (Duffy & Healey, 2013). Non-verbal communication such as gaze, or maintaining spatial configurations with respect to each other and the music stand, are also an important part of student-tutor interaction (Duffy & Healey, 2012).

The transition between speakers is an essential part of the organisation of turn-taking in conversation (Sacks, Schegloff, & Jefferson, 1974). The preference for just one person to talk at a time requires participants to work together to minimise gaps and overlaps. Anticipating the possible end of a speaker's turn allows a listener to prepare to take the floor when an opportunity presents itself. Interactive turn-taking phenomena such as backchannels, or making a bid for the floor for a turn at talk, require very precise timing. The timing of the transition between speakers is sometimes referred to as *turn offset*. It is usually reported as positive if there is a pause between speakers, and negative if there is an overlap (Stivers et al., 2009). Longer pauses and overlaps do occur, but the average turn offset in natural speech tends towards a short pause. A positive turn offset in the range of 0-200ms is most likely to be perceived as a smooth turn transition (Stivers et al., 2009; Heldner & Edlund, 2010).

Remote music tuition using video conferencing is a popular way to support music education in geographically remote areas but has also become an important part of urban mainstream conservatoires, for example to manage temporary separation when students or tutors have to travel to perform, or to manage international auditions. However the medium of communication is known to change aspects of conversational turn-taking, and this has important implications for video-mediated remote music tuition (Duffy et al., 2012). Even minor disruptions to the transmission characteristics of the medium of communication, such as the latency and delay associated with video mediated communication, can seriously affect turn-taking (Whittaker, 2003).

Qualitative video analysis and conversation analysis (CA) have been used to examine video-mediated workplace communication (Heath & Luff, 1991) and how participants complete collaborative tasks in video-mediated environments (O'Conaill, Whittaker, & Wilbur, 1993; Heath, Luff, & Sellen, 1997; Ruhleder & Jordan, 2001). However there have been relatively few studies of the detailed effects of video-mediated communication (VMC) on the timing of conversational turn-taking, and the results are inconsistent, driven by subtle differences in experimental set up. For example, some studies include signal delay (Ruhleder & Jordan, 2001) whilst others exclude it (Sellen, 1992); in some cases specifically to isolate other interactional factors. Some studies compare video-mediated interaction (with or without delay) to same-room conditions, whilst others compare it to other lower quality remote communication systems or audio only scenarios (O'Conaill et al., 1993; O'Malley, Langton, Anderson, Doherty-Sneddon, & Bruce, 1996). Studies which have included delay as part of their experimental set-up (O'Conaill et al., 1993; Ruhleder & Jordan, 2001) suggest a further subtle effect; changes in the time of arrival of utterances with

respect to 'local' sound. Ruhleder and Jordan (2001) suggest that two people having two fundamentally different conversations with each other raises serious questions about what it means to 'share' a conversation in distributed settings. This leads to some interesting questions in terms of remote music lessons. How might the medium change the turn transitions observed in co-present lessons when they are mediated by video conference? How might the transition between 'speakers' be affected by the inclusion of musical contributions?

In order to investigate these questions, a detailed study was made of student-tutor interaction during a co-present and a remote music lesson, using CA and qualitative video analysis. CA has previously been used to examine aspects of instrumental music tuition (Ivaldi, 2014; Nishizaka, 2006; Szczepek Reed, Reed, & Haddon, 2013), as well as the effect of medium on conversational turn-taking. This fine grained analysis of a same-room and a separated lesson allows us to examine both the turn-taking characteristics unique to a music lesson, and how these are affected by the medium of video. This work is part of a larger study of a number of co-present and remote music lessons (Duffy, 2015).

Methodology

Two one-to-one lessons featuring woodwind instruments were observed, filmed and analysed in detail; a co-present ('same room') lesson and a video-mediated remote lesson. The co-present lesson featured a male student studying ABRSM grade 8 clarinet performance and was filmed during one of his regular weekly lessons at the junior school of a London Conservatoire. The female tutor had taught the student for many years. The remote lesson featured a female oboe student taking part in an ensemble residency with Aldeburgh Young Musicians, filmed during a remote music tuition study at Aldeburgh Music in Suffolk (Duffy et al., 2012). The student had been working with the tutor during the residency, but had not previously taken regular lessons with her. Both students had advanced to a similar level of proficiency; they were largely comfortable with the technical challenges of their instrument and capable of exploring musicality and expression. Both tutors were experienced in one-to-one tuition, but not video mediated tuition. The scope of this study was to examine student-tutor interaction, and did not consider teaching effectiveness between conditions.

Conversational turns are defined as the period during which a participant holds the floor, until there is a change in speaker (Sacks et al., 1974, pp.702-703). Turns in the footage from each lesson were coded using ELAN (Brugman, 2004). A separate tier was created for analysis of each of the following types of contribution: student talk, tutor talk, student play and tutor play. This data was exported as a transcript with timecode information so that calculations could be made such as turn frequency, mean turn duration and turn onset in relation to the preceding turn. Pauses between turns were coded as a positive offset, and overlap as a negative offset, similar to the approach used by Stivers et al. (2009). This allowed calculation of a net offset for a period of time or subset of turn types. Backchannels were excluded from the distribution, similar to the approach used by Sellen (1992), since they are not a bid for the floor or intended to initiate a change in speaker. As discussed, a difference between this analysis and existing literature is that we consider the transitions between musical, as well as verbal contributions. As a result, the following categories of turn transition were identified:

1. Talk following talk.
2. Talk following play.
3. Play following talk.
4. Play following play.

Established notation for conversation analysis, as described in the appendix of Sacks et al. (1974), was adapted to analyse musical contributions to lesson dialogue (Table 1).

(0.2s)	Elapsed time (seconds) used to denote pauses or silence
____ (1.4s)	Long single note and duration
- - - - (2.3s)	Individual notes in a musical phrase and phrase duration
↑ - - -	Rising passage of notes
↓ - - -	Falling passage of notes
' ' ' (1.2s)	in-breath in preparation to play, and duration
__ //__	onset of 'talk over play' overlap
{first octave}	Additional information for music notation
[0.6s]	duration of period of overlap
=	Latching (no interval between two pieces of talk)

Table 1: Transcription notation.

The two rooms used for the remote lesson were adjoining suites at the same organisation (see Duffy et al. (2012) for more details). A separate video camera was placed in each suite, in addition to the video conference equipment, in order to capture student and tutor position with respect to the screen and provide a separate audio recording for each location. There was a small delay in visual processing caused by additional software being tested during the lesson. A delay was added to the audio so that audio and visuals arrived synchronised in each location. Audio samples from each room were synchronised and analysed using clearly visible audio transients which did not overlap with local sounds. Whilst the rooms were geographically close, the delay was of the same order as the latency experienced in a typical transatlantic video call (0.9s). This delay was constant, but in reality the magnitude of the delay would vary somewhat over the duration of the call, depending on the signal journey through different servers and exchanges.

Results

First we will look at some general effects of the medium on the lessons analysed. Whilst the co-present lesson was slightly longer than the consolidated sections of the video-mediated class analysed, they both contained similar proportions of instances of turns at talk (73% and 71% table 2) and instances of musical contributions (27% and 29% table 2). However the turn structure within this was quite different. The co-present lesson contained 753 turns in total whilst the

video mediated lesson contained just 234, and the average length of both turns at talk and musical contributions were significantly longer for the video-mediated lesson. Net mean offset for the remote lesson was 337ms, 143ms longer than the co-present lesson offset of 194ms (table 3). These results are consistent with findings that video-mediated conversations are characterised by fewer turns of greater length and reduced overlapping speech (Cohen, 1982; O'Conaill et al., 1993; Sellen, 1992).

Table 2: Turn structure of co-present vs. remote lesson.

	co-present	remote
instrument	clarinet	oboe
total duration (mins)	36	27
number of turns at talk	550	165
as a % of total turns	73%	71%
number of musical contributions	203	69
as a % of total turns	27%	29%
total lesson contributions	**753**	**234**
average length of turns at talk (s)	2.0	4.4
average length of musical contributions (s)	4.7	6.7
total lesson average contribution length (s)	**2.8**	**5.1**

Table 3: Net offset duration (ms) by transition type.

	talk following talk	talk following play	play following talk	play following play	total lesson
co-present					
n	345	129	134	50	658
%	52%	20%	20%	8%	
mean (ms)	287	-61	297	-70	194
remote					
n	64	55	62	4	185
%	35%	30%	34%	2%	
mean (ms)	39	-40	993	124	337

Next we examine the net offset by transition type. In the co-present lesson, the net offset for turn transition type *talk following talk*, representing periods of student-tutor discussion, was 287ms (table 3). This was slightly outside the range of 0-200ms from the literature, but still showed a preference for a pause of the same order of size. For transition type *play following talk*, for example a student performing in response to a verbal instruction from the tutor, the co-present net offset was again in line with the literature (297ms, table 3). In the remote lesson, the net offset for transitions of *talk following talk* decreased to 39ms in line with our expectations from the literature, but the net offset for *play following talk* lengthened considerably to 993ms.

Looking specifically at overlap by participant, the student showed a preference to play over tutor talk (33 instances of student *play over talk* overlap compared to 2 tutor instances - table 4). One explanation for this is that in co-present lessons tutors were found to make long instructional turns to initiate student play, comprised of several utterances separated by pauses, interspersed with backchannels by the student. The backchannels were placed with precision to show attentiveness without making a bid for the floor or disrupting the tutor's turn. Non-verbal cues enabled the student to determine when these turns were complete and they should start to play (Duffy, 2015, pp. 140-148). As the next example shows, in the remote lesson the student found this more difficult.

Table 4: Overlap duration (ms) by activity by participant

			talk over talk	talk over play	play over talk	play over play	total lesson
co-present	student	n	47	1	33	15	96
		mean (ms)	330	65	637	524	463
	tutor	n	42	46	2	9	99
		mean (ms)	317	489	391	436	409
	total	**n**	**89**	**47**	**35**	**24**	**195**
		mean (ms)	**324**	**480**	**623**	**491**	**436**
remote	student	n	12	1	5	2	20
		mean (ms)	780	1,326	615	466	735
	tutor	n	10	16	1	-	27
		mean (ms)	1,113	478	1,234	-	741
	total	**n**	**22**	**17**	**7**	**2**	**48**
		mean (ms)	**877**	**528**	**647**	**467**	**703**

Instructional turns

The tutor asked the student to play a scale. When the tutor paused after her first utterance, the student made physical preparations to play such as stepping back from the screen and raising her hands to the instrument body (transcript 1: line 1 and transcript 2: line 2). However the tutor retained the turn, choosing to demonstrate by playing the scale herself (transcript 1 and 2: line 3). Towards the end of the scale the student raised her clarinet to her mouth again, this time placing the reed in her mouth (transcript 1: line 4 and transcript 2: line 5). However the tutor started a new utterance "so I mean you go C sharp to C sharp" and the student lowered her oboe again. This was the second abandoned attempt to play. At the end of this utterance the student nodded and raised her oboe for a third time. She placed the reed in her mouth and took an in-breath whilst the tutor talked (transcript 1: line 6). The tutor made no further utterances and finally the student moved into playing the scale. From the tutor room footage it was not clear why the student made two preparations to play which could not be followed through. From the student room footage it was clear that the student was placing these actions in the pauses that were interpreted as the end of the tutor's instructional turn. This happened several times, and towards the end of the lesson the student exclaimed "sorry sorry it's hard to know when to play". This example is analysed in more detail in Duffy (2015, pp. 350-359).

Bidding for the floor to provide feedback

Transitions involving turns following play did not follow the literature. Both *talk following play* and *play following play*

```
1. T: can you just play me a scale [starting on top A?]
                                   [((S steps back from screen))]
      [(1.0s)]
      [((S lifts second hand onto instrument body))]

2. T: [u::m in fact]
      [((S steps back from screen))]
      [((T raises oboe))]
      (0.4s)
      ((sucks reed loudly twice))
      (0.9s)
      [((S raises oboe))]

3. T: [↓{A}____(0.6s) _ ] [_ _ _ _ ]_ _ _ _ _ _=
      [(( * ))]           [ ((**)) ]

4. T: =↑{C♯}_ _ _ _ _ _ _ _ [_ _ _ _] _ _ [____]
                            [ ((*)) ]     [((***))]

5. T: [so I mean you go C♯] to [C♯ [it's still A major]]
      [(( ** ))]               [    ((  S nods  ))    ]
                                 [    ((  *  ))      ]

6. T: [it's just A major] [let's just have a listen]
      [ (( *** )) ]       [((S takes an in-breath))]
```

Transcript 1: The tutor initiates a scale - student room audio.

```
1. T: can you just play me a scale starting on top A?

2. T: [(1.0s)]
      [((S steps back from screen))]
      [uuum] in fact
      [((S lifts second hand onto instrument body))]
      (0.4s)
      ((sucks reed loudly twice))
      (0.9s)

3. T: ↓{A}____(0.6s) _[_ _ _] _ [_ _ _ _ _ _]=
                     [((*))]    [  ((**))   ]

4. T: =↑{C♯}_ _ _ _ _ _ _ _ _ _ _ _[ _ _ ____ ]
                                   [   ((*))  ]

5. T: [so I mean you go C♯] [to C♯] [its still A major]
      [((***))]             [((**))] [((S nods))]

6. T: [[it's just A major]] [let's just have a listen]
      [  (( S nods ))  ]
      [      ((*))     ]    [ (( *** )) ]

          *S lifts oboe to playing position
         **S lowers oboe, keeping both hands on keys
        ***S places reed in mouth
```

Transcript 2: The tutor initiates a scale - tutor room audio

tended towards overlap in the co-present lesson, rather than a short pause (-61ms and -70ms table 3). *Talk following play* tended towards overlap in the remote condition (-40ms table 3). *Play following play* tended towards a pause in the remote condition (net offset 124ms) but the proportion of this type of turn was significantly reduced to just 2%, or 4 turns. All but one incidence of *talk over play* overlap was made by the tutor, in both the co-present and remote lesson (table 4), evidencing the tutor's preference to talk over student play when a problem was been diagnosed in order to provide feedback. This did not appear to be as disrupted by the medium as the previous example, perhaps because the length of the note during which the tutor bid for the floor was often of the same order as the duration of the delay (Duffy, 2015, pp. 245-263), so the tutor's interruption still arrived before the student could start the next musical phrase. What is beginning to emerge is asymmetry in the preferences for taking a turn to talk or play between the participants, some of which are disrupted more by the medium than others.

Local differences in turn placement

Next we will look at an example which demonstrates the effect of the delay on the placement of a single turn. Audio waveforms from each room illustrate the effect in addition to the transcripts (figure 1). Coloured blocks have been annotated using Logic Pro 9 to highlight the different position of parts of the dialogue shown in transcript 3. The tutor waveform is narrower because the camera in the tutor room was further away from where the tutor was positioned, as a result the waveform has smaller amplitude (vertical height representing volume). This does not affect our analysis. The two audio samples were synchronised using the visual transients in the tutor's utterance "ba ba ba ba ba ba" (line 3 of transcript 3). Transcript 3 shows the difference in turn transition and sequence between the two rooms.

```
Audio from the camera in the student's room
1. T: 'cause the A sharp is always there from the crotchet rest ((*))
2. S: yeah fine
        (0.2s)
3. T: You've just got it so actually you think ba ba ba ba ba ba

Audio from the camera in the tutor's room
1. T: 'cause the A sharp is always there from the crotchet rest ((*))
        (0.8s)
2. S: [yeah fine]
3. T: [You've just got it] so actually you think ba ba ba ba ba ba

* a door slams shut as an observer leaves the room
```

Transcript 3: Turn sequence discrepancy between rooms

There are two main differences between the audio samples. The first relates to the student's utterance "Yeah fine" in line 2 (circled section of the waveform in figure 1). This utterance was made with respect to the tutor's turn in line 1. In the student room audio the response followed straight on, after the noise of a door slamming at the end of the tutor's turn. In the

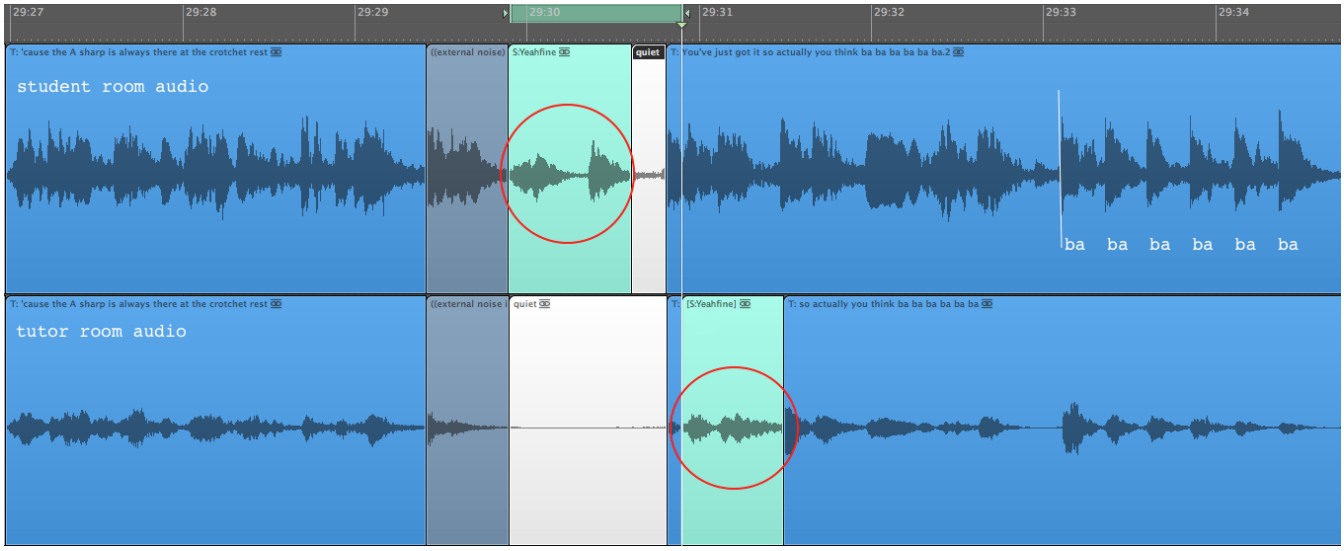

Figure 1: Audio discrepancy between rooms.

tutor room audio there was a 0.8 second pause after the tutor said "Cause the A sharp is always there from the crotchet rest". When the student utterance "Yeah fine" arrived, the tutor had already started talking again, so it overlapped with the start of the tutor's next comment "You've just got it". From the student's perspective, she had replied as soon as she heard the tutor's comment. However her reply was delayed in its return to the tutor by 0.9s. When the student's response arrived, the tutor had already started to talk again having only heard silence, and so she talked over the student's response. As a result, the student's utterance "Yeah fine" was placed within a pause in the tutor's speech, but transformed into an overlap with tutor speech when received in the tutor's room. Several examples of similarly misapplied feedback are reported in Ruhleder and Jordan (2001).

The next example shows how turn sequence can be changed. Examining the student audio first, the musical phrase in line 1 of transcript 4 includes a pause notated in the score before a phrase is repeated. The student makes this pause 0.4 seconds in duration and starts the repeated phrase in line 2. However the tutor appears to talk over this second phrase with "May-maybe a" (line 3). This is unusual, the tutor usually waits until the end of a musical phrase to start talking, the only overlap being with the final note (Duffy & Healey, 2013); here the tutor starts talking mid-phrase. It is also unusual that the student does not stop playing, instead the tutor stops talking and the student continues. The tutor interjects again with "yeah" but the student still continues. The tutor then talks again straight after the last note of the phrase. Now the student stops playing, immediately looking up from the music and at the screen.

Looking at the tutor room audio, shown in the second half of transcript 4, we see that the tutor started the utterance "May-maybe a" during the notated pause in the student's performance (transcript 4: line 2a). However the delay in transmission of this utterance to the student room, meant that it arrived after the student had started to play the next phrase in line 2. In the tutor's room "May-maybe a" was interrupted by the student starting her second phrase in line 3a and the tutor stopped talking. Her next utterance "yeah" in line 4a started during a long note played by the student, which could be interpreted as a bid for the floor. However when this utterance arrived in the student room, the long note was already complete and the student had moved on to the next phrase. From the tutor's perspective she had tried unsuccessfully to take the floor at the end of the first phrase.

```
Audio from the camera in the student's room
1. S: ↑_ _ _ _ _____↓ _ _ _ _ ↑_____
        (0.4s)
2. S: ↑_ _ _[_ ___]___ ↓ _ [ _ ] _ _ ↑_____=
3. T:       [May-maybe a]     [yeah]
4. T: =Just a thought maybe make the four a little slower

Audio from the camera in the tutor's room
1a S: ↑_ _ _ _ _____ ↓ _ _ _ _ ↑_____
        (0.2s)
2a T:  May-may[be a
3a S:         [↑_ _ _ _[__]__ ↓ _ _ _ [_ ↑_____
4a T:                  [yeah]          [Just a thought maybe make
                                        the four a little slower
```

Transcript 4: Relative position of tutor interruption.

Discussion

The short fragments of music which occur during an instrumental lesson have been shown previously to be managed conversationally, and share some characteristics with turns at talk. Here we see that participants exhibit different preferences for how they manage transitions between verbal and musical contributions. The tutor more often leads lesson

flow, placing more of the responsibility for turn placement onto the student in their responses. The tutor is also more likely to bid for the floor during student play, whereas the student rarely interrupts the tutor in talk or play. Differences in preferences have also been reported in turn-taking associated with the roles of the teacher and students in a classroom (McHoul, 1978). The signal delay associated with VMC disrupts these preferences, exhibiting a greater effect on the student. Ruhleder and Jordan (2001) suggest that the mechanisms which are most affected by signal delay are conversational turn-taking, sequence organisation and repair; affecting trust and confidence between the participants. The phenomenon analysed here may explain student frustrations previously reported during remote music lessons (Duffy & Healey, 2012). This study highlights a number of opportunities for further work. For example, it is not known if participants could acclimatise to aspects of the disruption to lesson interaction over time. A longitudinal study is recommended which follows student-tutor pairs taking both co-present and remote lessons. In this way, any effect caused by change in participants across conditions will also be controlled. There may also be different, more effective, ways to represent the naturalistic teaching interaction remotely through alternative technologies (Duffy & Healey, 2017).

Acknowledgments

This work is supported by the Media and Arts Technology programme, EPSRC Doctoral Training Centre EP/G03723X/1.

References

Brugman, H. (2004). Annotating multimedia/multi-modal resources with ELAN. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Portugal.

Cohen, K. (1982). Speaker interaction: video teleconferences versus face-to-face meetings. In *Proceedings of teleconferencing and electronic communications* (pp. 189–199).

Duffy, S. (2015). *Shaping Musical Performance Through Conversation*. Doctoral thesis, Queen Mary University of London.

Duffy, S., & Healey, P. (2012). Spatial Co-ordination in Music Tuition. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society* (pp. 1512–1517). Sapporo: Cognitive Science Society.

Duffy, S., & Healey, P. (2013). Using Music as a Turn in Conversation in a Lesson. In *Proceedings of the 35th annual conference of the cognitive science society* (pp. 2231–2236). Berlin: Cognitive Science Society.

Duffy, S., & Healey, P. (2014). The Conversational Organisation of Musical Contributions. *Psychology of Music*, *42*(6), 888–893.

Duffy, S., & Healey, P. G. (2017). A New Medium for Remote Music Tuition. *Journal of Music, Technology and Education*, *10*(1), (in press).

Duffy, S., Williams, D., Stevens, T., Kegel, I., Jansen, J., Cesar, P., & Healey, P. (2012). Remote Music Tuition. In *Proceedings of the 9th sound and music computing conference* (pp. 333–338). Copenhagen: smcnetwork.org.

Heath, C., & Luff, P. (1991). Disembodied conduct: communication through video in a multi-media office environment. In J. S. Robertson, Scott P., Olson, Gary M. and Olson (Ed.), *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology* (pp. 99–103). New Orleans: ACM.

Heath, C., Luff, P., & Sellen, A. J. (1997). Reconfiguring media space: Supporting collaborative work. *Video-mediated communication*, 323–347.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555–568.

Ivaldi, A. (2014). Students' and teachers' orientation to learning and performing in music conservatoire lesson interactions. *Psychology of Music*, *44*(2), 202–218.

McHoul, A. (1978, December). The Organization of Turns at Formal Talk in the Classroom. *Language in Society*, *7*(2), 183–213.

Nishizaka, A. (2006). What to Learn: The Embodied Structure of the Environment. *Research on Language & Social Interaction*, *39*(2), 119–154.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations Over Video Conferences : An Evaluation of the Spoken Aspects of Video-Mediated Communication. *Human Computer Interaction*, *8*, 389–428.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interacting with Computers*, *8*(2), 177–192.

Ruhleder, K., & Jordan, B. (2001). Co-Constructing Non-Mutual Realities: Delay-Generated Trouble in Distributed Interaction. *Computer Supported Cooperative Work (CSCW)*, *10*(1), 113–138.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735.

Sellen, A. J. (1992). Speech patterns in video-mediated conversations. In *CHI '92* (pp. 49–59). Monterey: ACM.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587–92.

Szczepek Reed, B., Reed, D., & Haddon, E. (2013). Now or Not Now: Coordinating Restarts in the Pursuit of Learnables in Vocal Master Classes. *Research on Language & Social Interaction*, *46*(1), 22–46.

Whittaker, S. (2003). Theories and Methods in Mediated Communication. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *The handbook of discourse processes* (pp. 243–286). Lawrence Erlbaum Associates Inc.