A new medium for remote music tuition

Sam Duffy | Patrick Healey

Queen Mary University of London

Abstract

It is common to learn to play an orchestral musical instrument through regular one-to-one lessons with an experienced musician as a tutor. Students may work with the same tutor for many years, meeting regularly to receive real-time, iterative feedback on their performance. However, musicians travel regularly to audition, teach and perform and this can sometimes make it difficult to maintain regular contact. In addition, an experienced tutor for a specific instrument or musical style may not be available locally. General instrumental tuition may not be available at all in geographically distributed communities. One solution is to use technology such as videoconference to facilitate a remote lesson; however, this fundamentally changes the teaching interaction. For example, as a result of the change in communication medium, the availability of non-verbal cues and perception of relative spatiality is reduced. We describe a study using video-ethnography, qualitative video analysis and conversation analysis to make a fine-grained examination of student–tutor interaction during five co-present and one video-mediated woodwind lesson. Our findings are used to propose an alternative technological solution – an interactive digital score. Rather than the face-to-face configuration enforced by videoconference, interacting through a shared digital score, augmented by visual representation of the social cues found to be commonly used in co-present lessons, will better support naturalistic student–tutor interaction during the remote lesson experience. Our findings may also be applicable to other fields where knowledge and practice of a physical skill need to be taught remotely, such as surgery or dentistry.

Keywords: remote music tuition, music education, HCI

Introduction

There are many ways to learn to play a musical instrument. Some individuals learn at a young age, encouraged by parents or group music making in school. Others may learn for the first time as adults perhaps through adult education colleges or private tutors, not having had the opportunity or motivation to play an instrument previously, or returning to learning after a lapse of some years (Taylor and Hallam 2008; Creech et al. 2013; Welch and Ockelford 2009). Some learners do not engage formally with music tuition at all, preferring to teach themselves using books, online materials and personal experimentation and often playing instruments informally with groups of friends, as well as engaging with online music communities (Waldron 2013). The choice of medium depends on the availability of resources locally and the motivation for learning, whether it be to play for personal enjoyment, to join an ensemble or to make music a professional career choice.

A common route is through regular one-to-one lessons with an experienced musician as a tutor. There is a large body of work concerned with one-to-one instrumental teaching techniques and effectiveness (Siebenaler 1997; Creech 2012; Kurkul 2007; Gaunt and Hallam 2009; Nishizaka 2006; West and Rostvall 2003). Here, we are concerned with unpacking how the teaching interaction between a student and tutor unfolds, moment by moment *in situ*, rather than evaluating lesson effectiveness or teaching methods or determining the factors that influence learning. Our aim is to understand how this interaction might be changed when it is mediated by technology such as videoconference, as is necessary for a remote lesson.

There is already a considerable body of work regarding the introduction of technology to the practice of instrumental musical performance and education; for example, systems which deliver expert knowledge for personal learning in place of a tutor or to support home practice (Löchtefeld et al. 2011; Rogers et al. 2014; Yin et al. 2005), or that students and tutors can use together, for example, a haptic jacket that uses vibrations to inform young violin students on their bowing technique (Linden et al. 2011). Some systems are designed both for use in the lesson and by the student at home. For example 'Bagpipe Hero', an interface for a digital chanter that analyses the air control and articulation of Highland bagpipe students, with a visual performance feedback interface that can be used both in the lesson and during personal practice (Menzies and McPherson 2013) or an interface which evaluates the ability of saxophone players to control air pressure (Robine et al. 2007). However these do not directly address the question of how to facilitate live interactive instrumental tuition remotely.

There are many reasons why suitable expert tuition may not be available within reasonable travelling distance, for the timescales required. For pupils who choose to study performance at an undergraduate level and beyond, their choice of conservatoire may be guided by their desire to study with a particular professor. 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. This could necessitate a geographical move to attend the most suitable institution. For younger students, it may not be possible to identify and nurture talent if they live in a geographically isolated or rural location where expert tuition is not locally available, rather than a densely populated urban location. In addition, travel is an inevitable part of

a musician's life, and their teaching commitment represents only a small proportion of their wider professional lives (Presland 2005). Tutors travel to perform, lecture and teach, whilst students travel to study, perform, audition and attend residencies and workshops (Kruse et al. 2013). Managing temporary separation of student and tutor at critical times, for example, when preparing for a career defining audition with a prestigious orchestra, is essential to maintain contact and maximize a student's opportunities.

One solution to these problems is a remote lesson using communication technology. Videoconferencing was initially developed in the 1960s as a tool to augment teleconferencing in distributed meetings. The value of a visual channel over text-based or audio teleconferencing was questioned early in its inception (Noll 1976; Pye and Williams 1977). The belief that technology could be developed which would replicate face-to-face interaction, simply at a distance, contained a fundamental misunderstanding about how people interact when working collaboratively to achieve a task (Hollan and Stornetta 1992). Whilst users were initially enthusiastic about these technologies, they later confessed that they felt uncomfortable using them or believed that it took a face-to-face meeting to really get to know the another person, with no real sense of the origin of these feelings (Ruhleder and Jordan 2001). As a result, it took some time for the technology to become widely adopted, the failure of the technology to meet expected take-up being attributed to psychological and sociological effects – a failure to understand how people communicate and what constitutes the feeling of 'being there' (Edigo 1988). To some extent, the disruption to collaborative work caused by interactive technologies depends on the required outcome of the interaction – the effect on cognitive tasks being far less than the effect on social tasks, or those requiring interpersonal

information or complex joint physical manipulation of an artefact in a shared environment (Whittaker 2003). Collaborative work recurrently involves variable and contingent access, not simply to each other's physical domain and artefacts, but to the emergent activities in which the participants are engaged (Heath et al. 1997).

Videoconference technology is now common in schools and universities, partly due to improved network bandwidth availability, and is being used for many aspects of education. Blended undergraduate courses are common, combining technologically delivered content with traditional lectures, and many blended courses now include many elements that only a few years ago were only to be found in distance education contexts (Thorne and Payne 2005). Many music conservatoires have videoconference facilities. The Manhattan School of Music (MSM) in the United States have worked in partnership with a manufacturer of videoconference equipment, Polycom, to devise and test a Music Mode, which allows users to suppress the speech processing features which have a detrimental impact on music (Orto and Karapetkov 2011). Distance learning programmes use videoconference and blended learning to reach geographically isolated areas, for example, to support students living in remote parts of Australia (Lancaster 2007; Tait and Blaiklock 2005; Crawford 2016). The Satellite Education Program (SEP) in New South Wales recognized the need for something to supplement occasional home visits by music tutors, or residential programs which could require a child to undertake a trip of hundreds of kilometres over rough terrain from their isolated homestead, with a parent or guardian (Anderson 2008).

The motivation for this work was involvement in a project Remote Music Tuition (Duffy et al. 2012), evaluating a prototype designed to enhance a commercial

videoconference system specifically for remote music tuition. This work took place at Aldeburgh Music in Suffolk, an organization dedicated to developing musical artists, in collaboration with British Telecom Research and Development. The prototype provided the tutor with controllable multiple views of the student and was tested with a number of student–tutor pairs and different instrument types (harp, oboe, French horn, violin, cello and piano). Participants were members of Aldeburgh Young Musicians, a programme for children aged 8–18 years, and tutors who were visiting to run ensemble residencies. Each tutor was able to suggest a number of alternative camera views that they would find useful in a video-mediated lesson. Cameras were positioned at the start of each test in collaboration with the tutor and student and each view was mapped to a thumbnail on a tablet control device. Selection of a thumbnail during the lesson placed the view selected into the main viewing area on the videoconference screen (Figure 1).
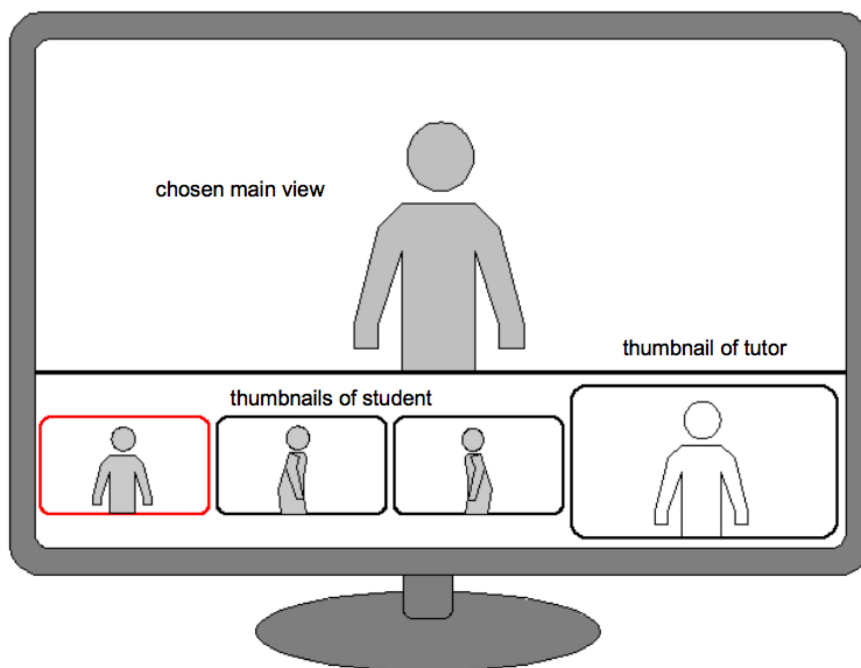

Figure 1: An image from the use case for the Remote Music Tuition prototype.

Whilst tutors found the control device intuitive to use, during test lessons it was usually put to one side after an initial period of experimentation and the conventional main view retained for the rest of the session. Subsequent analysis of video footage of the student and tutor in their separate locations revealed evidence of communication problems that seemed to affect the student to a greater extent than the tutor. Initial findings pointed to a fundamental change in how the participants shared their space and tools in a remote lesson, compared to a 'same room' or co-present lesson. Whilst the majority of musicians were knowledgeable about the impact of technical aspects of video-mediated performance, such as the effect of signal delay (latency) on playing together in ensemble, they were less aware of the effect of the medium on interaction more generally. This reflects a general problem that investigations of mediated communication often focus on technology, rather than affordances and the effect of the medium on communication behaviours (Whittaker 2003). Here, we will examine how the video-mediated environment affects student–tutor communication in a remote instrumental music lesson, compared to the same-room scenario, and suggest ways that technology could be developed to better support the teaching interaction. This work has implications not just for remote music tuition but also for any teaching interaction designed to teach a craft, skill or practice.

Methods

Video footage was obtained from both co-present and remote music lessons and analysed using qualitative video analysis and conversation analysis. Several aspects of the interaction were studied including the exchange of talk and musical fragments, the use of space and the use of tools to co-ordinate collaborative work. In addition, student–

tutor interaction during one co-present clarinet lesson and one remote video-mediated oboe lesson was coded, analysed in detail and compared. We will now outline the methodology for each of these phases of work in more detail. Further details of room set up, equipment and data analysis techniques can also be found in Duffy (2015, chapters 5–7).

Fieldwork and data

Fieldwork was carried out at Aldeburgh Music in the East of England and several music education organizations in London, in order to gain a deeper understanding of the practice of instrumental teaching. As part of this work, five co-present clarinet lessons were filmed at the Junior Schools of two Music Conservatoires in London. The classes were between 30 and 60 minutes in duration and held each Saturday during term time, students having already completed a week of academic study in their secondary school. The choice of instrument for the study was influenced by two factors. First, it was noted during initial observations of lessons including many different instrument types that woodwind instruments are relatively light, and players can easily move about the rehearsal room, move their head to adjust their gaze whilst playing and carry their instrument in one hand, leaving the other free for gestures and other tasks. This is useful for a study exploring the challenges of multimodal interaction through different mediums. Additionally, the author is also a woodwind player, with knowledge and experience of the technical demands of the instrument genre. Students included in the study were preparing to take, or had just passed, their grade 8 performance examination (e.g. through the ABRSM, the exam board of the Royal Schools of Music). They had taken regular music lessons from an early age and planned to study performance at an undergraduate

level. They often studied another instrument, for example, piano or saxophone, and took weekly lessons with other tutors in addition to the lessons observed. They had progressed beyond the purely technical stage of playing and were largely comfortable with the technical challenges of their instrument and exploring musicality and expression.

During the six-month placement at Aldeburgh music for the Remote Music Tuition project, one of the lessons filmed was an oboe lesson, which is also a member of the woodwind family. The student had achieved a level of skill comparable to the students observed in co-present clarinet lessons. This lesson was chosen for a more fine-grained analysis, and comparison with one of the co-present clarinet lessons. The tutor and student were working together on a residency organized by Aldeburgh Young Musicians in order to develop ensemble skills and agreed to take part in a lesson with the videoconference prototype. The student chose the piece of music that they would work on. Whilst student and tutor had not worked on the piece together before, they were both familiar with it, being a recognized part of the oboe repertoire, and each had their own copy of the music. The two rooms used for video-mediated lessons were geographically close together, tests taking place in two adjoining (but separate) suites (Figure 2). During the remote tests, a signal delay was simulated between the rooms, recreating the order of transmission delay that would be experienced during an inter-continental videoconference call (0.9s). The magnitude of the delay was constant, which was useful for our analysis, but in reality, the delay would vary over the duration of the call, depending on the signal journey through different servers and exchanges.
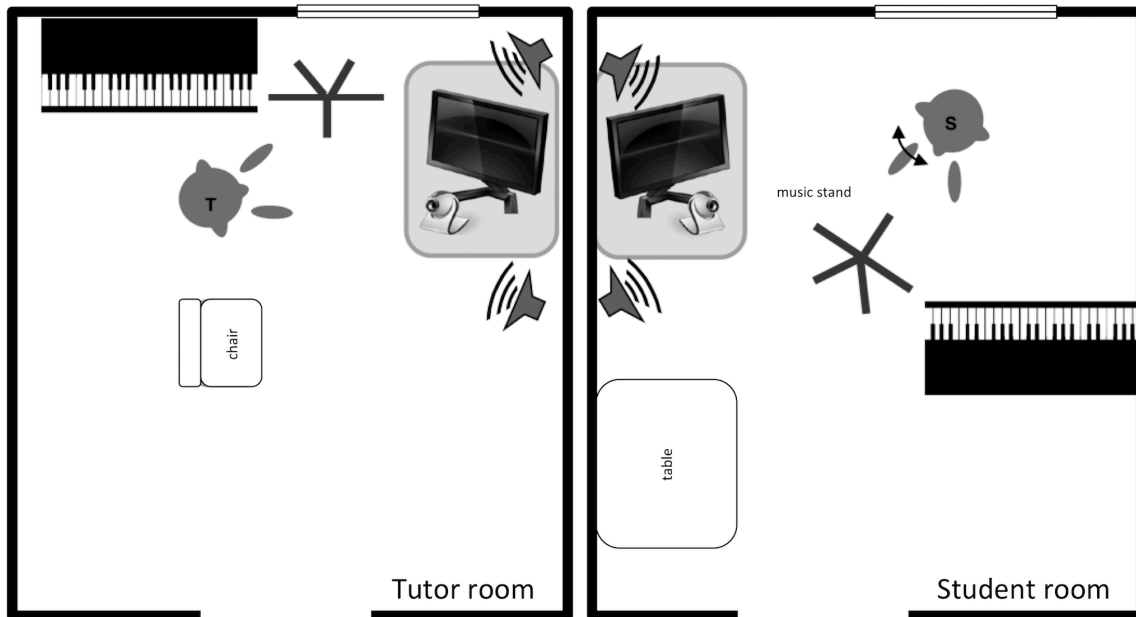
Figure 2: Room set up for the remote oboe music lesson.

Video-ethnography

During initial co-present lesson observations, lesson dialogue indicated that environmental aspects such as time of day, position of the lesson in an overall schedule, tools, space available and room shape all had the potential to influence the teaching interaction. It would be necessary to study this embodied social practice as it naturally occurred, rather than in experimental conditions created by the researcher, or through analysis of participants' recollections subsequent to the event. In a review of a broad interdisciplinary collection of anthropological work on the relationship of music to language, Feld and Fox (1994) noted a trend towards ethnographic studies. This trend has continued and ethnographic methods have been used to analyse many aspects of music

education (Bannister 1992) and performance (Cohen 1993; Morton 2005; Moran 2011). However, whilst the practice of instrumental music teaching has been subject to some examination, it is rarely at the fine-grained level of detail of interaction (Karlsson and Juslin 2008).

If it is necessary to consider how people orient bodily, point to objects, grasp artefacts and in other ways articulate an action or produce an activity, then it is unlikely to be possible to grasp more than passing sense of what happened unless the interaction can be recorded somehow (Heath and Hindmarsh 2002). Video-based investigations extend the possibilities of traditional ethnographic methods for data collection and analysis (Ruhleder and Jordan 1997) allowing situated analysis of the real-time organization of the ways in which an 'expert' and 'apprentice' interactionally organize training on a moment-by-moment basis (Heath et al. 2010).

Conversation analysis

Conversation analysis (CA) is a social science method which has grown over 40 years to become a dominant approach for the study of human social interaction, across the disciplines of Sociology, Linguistics and Communication (Stivers and Sidnell 2012). Having its roots in ethnomethodology (Levinson 1983: pp. 294–96; Heath and Hindmarsh 2002: 5–6), the CA approach assumes that language use and social interaction are organized at a fine-grained level of detail. This makes it a natural choice for the analysis of ethnographic data in order to understand conversationally organized interaction. CA has previously been applied to video-mediated communication (Ruhleder and Jordan 2001), classroom discourse (Lerner 1995; Wells 1993; McHoul 1978),

orchestral instruction (Weeks 1996), public music masterclasses (Szczepek Reed et al. 2013; Reed 2015; Reed and Reed 2014) and instrumental music tuition (Nishizaka 2006).

It was noted during early fieldwork that many of the interactional contributions to the lesson were musical, and whilst there is already an established methodology for transcribing dialogue (Sacks et al. 1974), there is no similar system for representing and analysing musical interaction. Additional notation was devised to represent these musical contributions as they naturally occurred within student–tutor dialogue (Duffy and Healey 2013b). The video transcription tool ELAN (Brugman 2004) was used for this analysis. Qualitative and quantitative video analysis

In addition to qualitative video analysis of the five co-present clarinet lessons and the remote oboe lesson, two films were chosen for further detailed analysis – one of the co-present clarinet lessons and the remote oboe lesson. The onset and duration of all musical and verbal contributions in each were precisely annotated in the footage timecode using ELAN and coded by participant and contribution type (student talk, student play, tutor talk, tutor play) for a more qualitative form of analysis. This was used primarily to investigate the precise impact of latency on turn-taking (see Duffy 2015, chapters 11, 12, 15), which will not be reported here. However, this work also informed some of the findings which will be discussed next.

Findings

Co-present music lessons

The musical score is one of the most important tools for musicians (Bautista et al. 2009). It is essential for dissecting and learning a piece, even if it is then discarded in favour of performing from memory (Lisboa et al. 2004). Participants usually share a

score in the lesson, most often that belonging to the student, and it is a fundamental tool for the coordination of lesson activity and communication. Many non-verbal interactional resources were used by the student and tutor to coordinate collaboration, such as gaze and use of space, but they were often made in relation to the music score. For example, different types of lesson activity were characterized by different spatial configurations of the tutor, student and the music stand holding the shared music (Duffy and Healey 2012). These configurations enabled the participants to monitor both the music, and each other, simultaneously. It was possible for one participant to see where the other was looking on the score, and so determine whether they understood which part of the music was being referenced (Figure 3).

a. student (S) and tutor (T) in detailed work with the score on a music stand

b. the tutor steps back to listen to the student perform a longer section of music

c. the student steps back to play
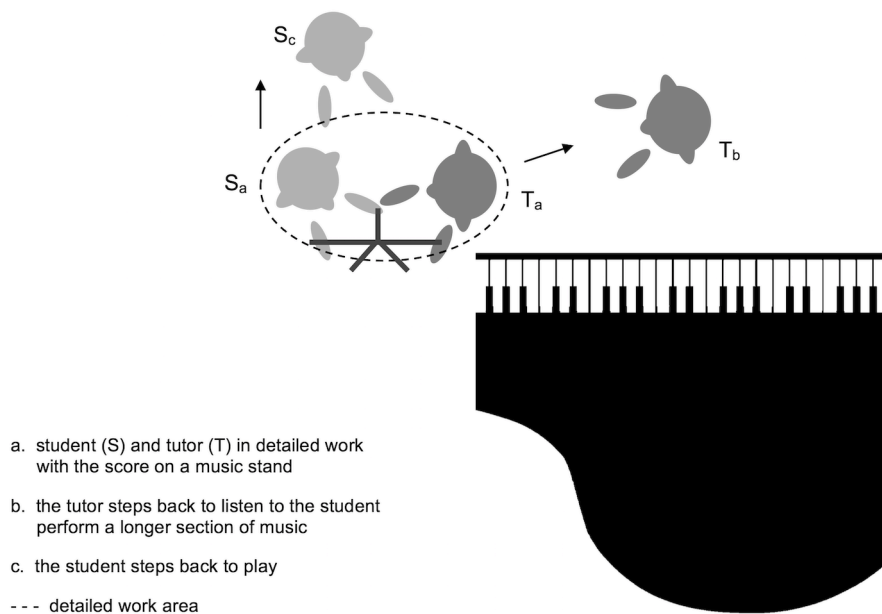
- - - detailed work area

Figure 3: Examples of spatial configuration in a co-present music lesson.

Changes to this configuration provided spatial cues, which helped the student to understand when they should play, and when they might need to stop playing to

accommodate feedback. For example, during detailed work, student and tutor often stood side-by-side in front of the music stand, close to the music (S_a and T_a in Figure 3). When the tutor expected the student to perform a longer extract or whole piece, they moved back from the stand (T_b in Figure 3), adopting a consistent and recognizable listening posture and changing their spatial configuration in relation to the student and music stand so that they could monitor the student's hands on their instrument and the music. The student also stepped back, making room for expressive movement of their clarinet and making it possible to monitor both the music and the tutor at the same time (S_c in Figure 3). The configurations associated with detailed work, performance of a longer section and social discussion were seen across each of the lessons analysed. Even in larger rooms where the tutor was able to pick up the music stand and move it to another part of the room during the lesson, the spatial configurations were recreated in the same way around the music stand's new position (Figure 4). All tutors were observed to adopt a consistent and recognizable listening position and posture in each lesson, influenced by the free space available and position of furniture in the room.
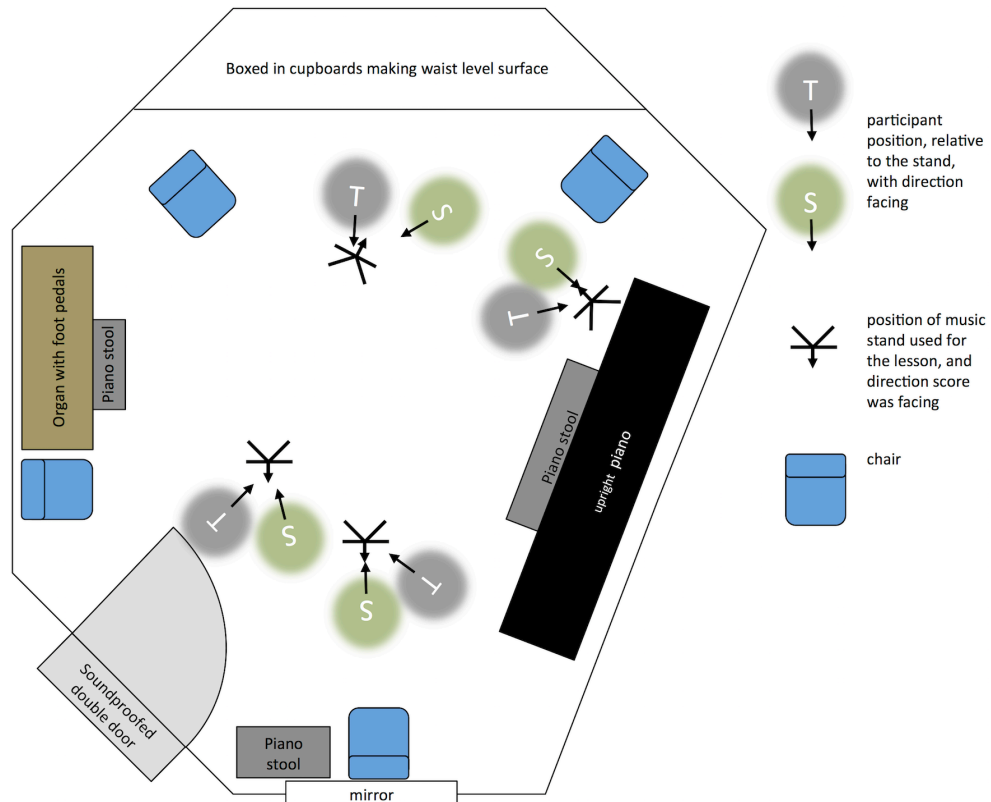
Figure 4: Examples of spatial configurations created around the music stand as it is moved around the room by the tutor. Adapted from images used in Duffy (2015, chapter 10).

Tutors annotated the score during the lesson in order to illustrate and document recommendations, using a variety of words, symbols and drawings. Access to annotations made by the student between lessons provided the tutor with insight into the student's home practice. References to the annotations were integrated into lesson dialogue. In this way, annotations made by the tutor in each lesson were combined with student's annotations during home practice and aggregated week-by-week into a cumulative record of learning (see Figure 5 for an example). Many musicians value the handwritten annotations made by mentors or respected colleagues (Winget 2006). By 'storing' these marks for future reference, the score became an artefact facilitating distributed cognition

(Berg 1999; Hutchins 1995), since points could be carried from one lesson to the next, reducing the need to repeat explanations and reminding both participants what had been discussed previously.

Figure 5: Example of annotations. Study No.76, Music by Iwan Müller (1786–1854). This edition ©1986 by Faber Music Ltd, London WC1B 3DA. Reproduced from 80 Graded Studies for Clarinet, Book 2 by permission of the publishers. All Rights Reserved.

Generally in conversation one person speaks at a time through a collaborative system of turn taking, and speakers take action to minimize a period of overlap where it

occurs (Sacks et al. 1974). Whilst the primary aim of instrumental tuition is the refinement of musical sound, a large part of lesson time is spent in talk. For example, for the clarinet lesson analysed in detail, 52 per cent of lesson time was taken up by conversation, student talk representing 10 per cent and tutor talk 42 per cent. In a music lesson, musical contributions are woven into the lesson dialogue and also need to be managed within the interactional turn-taking system. The timing of conversational turn-taking is very precise. On average, it is most likely that there will be a short pause between turns of around 200ms (Heldner and Edlund 2010; Stivers et al. 2009). Despite the inclusion of musical as well as verbal 'turns', the average duration between contributions from each participant, in the lesson analysed in detail, was in the same order as that found for naturalistic conversation, 221ms. Short musical fragments are characteristic of lesson interaction, and analysis of their shape and timing revealed that they were managed in a similar way to the verbal turns, through conversational turn-taking mechanisms. In addition, some of these musical contributions were not just organized conversationally, but shared some of the characteristics of verbal turns (Duffy and Healey 2013b). For example, the tutor often delivered an instruction with sequential refinements, through a relatively long turn made up of several utterances and pauses. For example, when instructing a scale to be played 'let's go from (pause) let's start with G major (pause) so let's have it slurred (pause) it's three octaves' or 'let's have (pause) F sharp major (pause) let's have it tongued (pause) so by tongued I mean legato tonguing' (Duffy 2015: 140–48). The tutor's non-verbal cues, such as remaining in place or moving back from the music stand, indicated whether they would make a further refinement to the instruction, or the directive turn was complete and the student could start to play. A

similar phenomenon has been observed in vocal masterclasses (Szczepek Reed et al. 2013). When the tutor spoke during student performance, the volume, duration and timing of the tutor's utterances, in relation to the student's musical phrasing, determined whether they were interpreted as encouragement to continue, or a bid for the floor to provide immediate feedback. A listener may indicate attentiveness to the person talking through short utterances such as 'uh-uh', 'okay' or 'yeah', known as 'backchanelling' (Ward and Tsukahara 2000). Whilst these occur as the speaker is talking, they are not usually a bid for the floor, but rather a display of continued attention and understanding. Student talk during the lesson was characterized by these backchannels, which were made during the tutor's instructional turns or during tutor feedback after student performance.

Errors occur naturally in the learning process, and how teachers deal with inevitable errors in student performance is one of the fundamental components of teaching expertise. Much of the literature in the field of music education that addresses the subject of error in performance is limited to the study of error detection. Error correction requires a further step, knowing what, when and how to bring about positive changes in student performance (Cavitt 2003). During a music lesson, the student presents musical contributions for the tutor to assess and make recommendations for improvement. This implies that the student may make mistakes or play incorrectly. In talk, not every problem is caused by a clearly identified 'mistake' by the speaker, ambiguity and misunderstanding by the listener can also lead to the need for both participants to work together to make a conversational repair. Similarly, the notion of 'error' in the production of a musical contribution was not always found to be straight

forward and was sometimes the product of subjective assessment (Duffy and Healey 2013a: 265–76). Both student and tutor were found to identify errors in the student's performance, with different outcomes. If the student identified a problem, either through their own diagnosis or in response to non-verbal indicators from the tutor, there was a preference for them to attempt to correct the problem themselves, the tutor holding back from taking the floor to encourage student correction, in a way that is analogous to self-repair in conversation (Schegloff et al. 1977). Non-verbal behaviours such as gaze and changes in posture were used by the tutor to encourage a student to self-repair and continue with their performance, despite mutual acknowledgement that a problem had occurred (Duffy and Healey 2013b). A similar phenomenon is found in language tuition (Seo and Koshik 2010). If the tutor identified a problem and the student did not stop to self-repair, there was a preference for the tutor to begin to talk or play over the end of the musical phrase containing the problem, in order to take the floor to lead detailed work. This was often preceded by the tutor stepping forward towards the music, raising their pencil or instrument, in a visual signal that was available to the student.

In summary, the score is identified as a fundamental resource for the coordination of lesson activity, influencing the use of gaze and spatiality as interactional resources and as a dynamic record of cumulative learning over time. Participants use whichever communication resource is best suited to the problem at hand, be it non-verbal, talking or playing. As a result, the system for managing conversational turn-taking needs to accommodate musical, as well as verbal contributions. Non-verbal behaviours in relation to the spatial configuration around the music stand are an important part of the

management of activities such as stopping student performance in order to analyse a problem.

Video-mediated music lessons

Analysis of student–tutor interaction during the video-mediated oboe lesson revealed some significant differences to the co-present lessons analysed. The change in spatial arrangement influenced gaze. In co-present lessons, student and tutor could simultaneously follow the music whilst peripherally monitoring each other's movements. Monitoring the tutor through peripheral gaze whilst playing from the score, the student was aware of tutor actions such as moving towards the music, changing orientation of gaze, adjusting upper body position or changing the way that objects were held (such as a pencil, clarinet or book), which provided evidence to the student that a problem had been detected.

In the video-mediated lesson each participant had their own copy of the music and the score could no longer act as a focus of joint attention. Spatial configuration was no longer anchored by a shared music stand; instead, there were two separate formations, one in each room (see Figure 2). The tutor's score was further away than it had been in the co-present lessons, to the side of the screen on a low stand (Figure 6), and whilst she gestured towards it, she could not gesture over it or point with the same level of precision. The student did not always see her gestures, as her gaze followed discussion of the music in her own score. The relevance of spatial cues is reduced when mediated through a flat screen (Cooperstock 2005). Participants no longer have a concept of negotiated mutual distance and cannot easily comprehend their position relative to each other, or objects in the other participant's environment (Sellen 1992). Now that

simultaneous monitoring of each other and the music was more difficult, participants were frequently required to switch gaze, missing interactional cues available on screen when they were looking at their music, confirming the importance of the shared score to co-present interaction. The student was less able to anticipate interruption of their play by the tutor as a result.
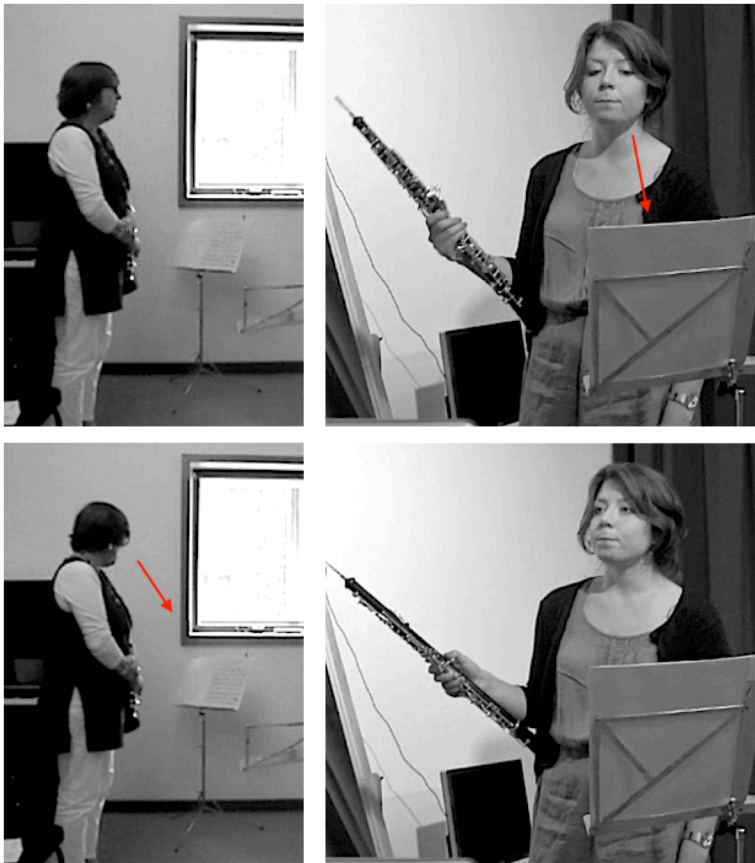

Figure 6: Student and tutor gaze divided between their screen and the music.

Tutors often wrote notes on the student's music during the co-present lessons observed yet hardly any annotation of the score took place, by either participant, in the video-mediated lesson, even though they each had their own paper copy. Any annotations the student had made previously, for example, during home practice, were not available to the tutor.

When musicians play together, they seek to co-ordinate their playing and their tolerance of latency is much lower than for speech (Chafe et al. 2004). However, since it was observed that many activities in the co-present lessons relied more on conversational style exchanges of musical fragments than synchronous ensemble playing, a more significant effect of signal delay was the impact on turn-taking. Interactive phenomena such as coordinating repair, and managing interruption and overlap require very precise timing; even minor disruptions to transmission can seriously affect them (Whittaker 2003). The student made far fewer backchannels in the video-mediated lesson, consistent with findings by O'Conaill et al. (1993) and O'Malley et al. (1996) that listeners in video-mediated meetings with a significant delay produced fewer backchannels. The student appeared to interrupt the tutor several times, but this was found to be due to their inability to predict when the tutor's instructional turn was complete. The visual cues which indicated that the tutor was approaching the end of their directive were no longer available, and the delay meant that by the time the student had actioned the instruction, through talk or play, the tutor had already continued with the next utterance in their turn. This led to frustration for the student, evidenced by the statement 'Sorry I-it's hard to know when to play' (Duffy 2015: 355–59). The tutor's preference for making a bid for the floor on detection of a problem in the student's play was also disrupted. From the perspective of the tutor's location, they made their bid in the same way as seen in co-present lessons, preceding it with non-verbal cues and talking over the end of a musical phrase as they heard it. However, these visual cues were no longer so available to the student, and due to the delay, by the time the interruption had arrived they had already carried on to playing the next phrase. The student's evident frustration with the delayed

interruption evidences the significance of the timing of the tutor's bid for the floor to provide feedback.

Implications for design for remote music lessons

One way to approach this problem is to use increasing network bandwidth, technology and computing power to produce a more immersive experience which aspires to convey co-presence through the transmission of life-size video, spatialized audio and vibrosensory information (Cooperstock 2005). Whilst a sensation of increased awareness of the remote partner may be achieved, the interaction is still fundamentally changed. Even when musicians are together, physical constraints can reduce musical collaborators' feelings of co-presence, for example, if partners cannot see each other because of their placement in an ensemble or because instruments are blocking their view. Schober (2006) suggests that this collaborative space can be reimagined through augmented reality, for example, creating enhanced access to the cues musicians use to co-ordinate their performance. However, the complexity of hardware and software required to achieve these solutions is expensive, in need of ongoing technical support and beyond the resources of many music education organizations.

Another response is to prioritize existing bandwidth to deliver the highest possible quality audio, at the expense of a visual representation. Experienced musicians can diagnose the cause of problems in a student's performance, without necessarily seeing the physical execution, especially when they already have an established teaching relationship with the student (Duffy et al. 2012: 337). However, musical performance, like speech and dance, involves a multi-sensory experience in which sound and visual stimulation interact over time (Vines et al. 2006). Whilst experienced tutors may be able

to guess the underlying physical cause of problems in their student's performance from audio alone, both participants will still lose the non-verbal cues that facilitate the turn-taking required to work on the problems together.

Some studies have evaluated remote teaching programmes using existing technology such as video-conference or Skype (Brändström et al. 2012; Kruse et al. 2013; Denis 2016). Whilst they can provide useful practical advice for tutors and students, they often fall short of recommending technological improvements. There is an opportunity to create a unique collaborative interactive space that blends digital artefacts with natural work practice in remote instrumental music tuition.

Sharing interactional cues through an interactive digital score

Given the importance of the shared score to coordinate lesson interaction, and the importance of peripheral monitoring rather than 'face-to-face' communication, it may be more useful to incorporate social cues into a digital representation of the music, rather than develop enhancements to the current video-conference set up with a separate screen. Enhancing the size and clarity of the video screen will not necessarily reintroduce interactional cues to a video-mediated lesson, or help with the sharing of precise locations in the score. Sharing interactional cues on the musical score itself may reduce the need to switch attentional focus back and forth between the music and your partner (Schober 2006: 93).

There are many advantages associated with digital scores. Paper incurs wear and tear, especially when being carried to and from classes regularly, and whilst pencil can be erased, a score can quickly become marked with eraser scrubs and repeated annotation changes. Paper scores come in a range of physical forms, from a single A4 sheet to large

heavy books and collections. A review of hands-free page turning devices noted the difficulty in designing for such a range of music formats (Wolberg and Schipper 2012). Moving between digital pages is much easier (but not without problems, see next). A digital music library is easier to store and catalogue, facilitating retrieval of specific pieces. A large library can be stored on one device and back-ups can be made. The digitization of sheet music also opens many new fields such as electronic music management systems for orchestras and ensembles (Bellini et al. 1999; Winget 2008). The Muse, a digital music stand designed for symphonic musicians, accommodated the tools used during a rehearsal such as a tuner (Graefe et al. 1996). Since many musicians now use applications in their phones which provide high-quality digital tuners and metronomes at low cost, the stand could incorporate functionality to download apps for use in the display provided, which will then have a familiar format and evolve as the applications are updated.

None of the students observed for this work used a digital score in their lesson. There are a number of reasons why digital music and display devices are not more widespread in music education. A4 sized scores cannot be represented in full on many computer screens or tablet display devices and it is difficult to read music accurately when it is scaled down in size (Laundry 2011: 33–34). There is very little literature on individual musician adoption of digital scores or preferred display devices, but musicians report difficulties reading scores from tablet devices, such as navigation, reduced page size and managing device weight on the light-weight portable music stands favoured for travel to lessons, rehearsals or performances on public transport. Owners of paper scores can scan them and create PDFs to be displayed on an iPad or tablet (the copyright

implications of this will not be considered here); however, there is an important

distinction between simple digital image data such as PDF, and digital semantic data that

have the ability to associate meaning with the symbols (Lin and Bell 2000). Navigation to

a specific bar number, to return to a sign, make a repeat, or turn forward to a coda is

problematic when swiping through pages one-at-a-time on a PDF. Navigation of digital

sheet music is already an area of interest, amongst the research community (Laundry

2011; Jin 2013; Ringwalt et al. 2015) and commercially. For example, Musescore1 have

developed open source music notation software which allows digital scores to be created,

shared and displayed on devices such as smartphones and tablets, through sophisticated

recognition of the semantic meaning of the musical symbols. The format also adjusts for

efficient viewing on small screens. However, this is currently only available for scores

which members of the Musescore community have created and uploaded, and the number

of pieces that can be accessed in this way is restricted by copyright law. Some scores are

distributed in electronic formats by publishers, but this seems to be motivated towards

popular music, such as distribution of sheet music sold as part of the commercial

merchandise accompanying big studio animated film releases, rather than classical

repertoire. Ideally an interactive digital score device for remote music tuition would

incorporate fully digitized music but in reality, considering the cost, availability and

practical issues, the device should be able to create and manage digitized paper scores.

Digital annotations could be made through the use of a stylus. There is an

opportunity to add functionality in this area. The movement of the stylus in three-

dimensional space could become a gestural interface, through the use of an embedded

accelerometer and gyroscope, for example, as used to represent beat gestures in music

lessons (Bevilacqua et al. 2007). Digital markings are not permanent and could be selectively hidden from view. A 'clean' score could be produced when the student preferred not to have their workings visible. The student could choose to see annotations made by the tutor, or just their own notes. The tutor could be given access to the student's score and annotations remotely, enabling them to monitor their student's personal practice. Annotations made and stored electronically could be backed up and retrieved if the score or device was lost. Recordings of these gestures, along with annotations and lesson audio, could prove effective for subsequent lesson review, in a similar way to the use of interactive whiteboard technology for recording and transmission of lectures, where hand-written annotations can be viewed by remote students in real time, or subsequent to the lesson along with a recording of the lecture (Anderson et al. 2003).

Our ability to focus on one area whilst passively attending to another activity in the edge of our visual field allows us to focus on a computer screen where the main activity is being completed whilst peripherally monitoring toolbars arranged at the side of our screen for progress reports, status updates and new messages (Dourish 2001: 12). Whilst the social cues should not distract from play or listening, the student is used to peripherally monitoring the co-present tutor and they may also be able to monitor representations of the tutor's actions on the edges of their digital score in a similar way. Representations of gaze, breathing and spatiality could be captured, transmitted and displayed in this way. For wind instruments such as flute, clarinet and saxophone, analysis of breathing may be able to differentiate the large in-breath a student takes in preparation to play from normal breathing. Whilst the nature of the spatial configurations maintained in the remote lesson was different from those observed in co-present lessons,

they still may contain useful interactional context. For example, in video-mediated lessons tutors were observed to adopt a consistent listening pose, even though it was less visible to the student (Duffy and Healey 2012). Representation of spatial configuration and posture may enable the student to determine transitions between detailed work and performance of a longer section of the music.

Being able to access the focus of the other participant's gaze on their music, at the same location in the music in front of you, could provide visual feedback that the precise location in the music being discussed is commonly understood. Advancements in eye tracking and sight readings could make this possible (Wurtz et al. 2009; Bigand et al. 2010; Penttinen and Huovinen 2011). Gestures on and over the score could be represented as meta-data on a transparent layer, displayed over the digitized score. They could be less permanent than annotations, fading soon after they have been made, to avoid cluttering the workspace and provide a dynamic, live experience. The tutor could use the representation layer to communicate when they want the student to stop playing, by touching the score at the place in the music where they first noticed a problem. This could be rendered as a 'fingerprint' in the same place on the student's score, alerting them to the tutor's desire for a discussion and indicating where the problem has occurred. 'Shadows' of hands moving over the score might be more helpful than an abstract representation of the other participant's gestures on and over the music (Figure 7).
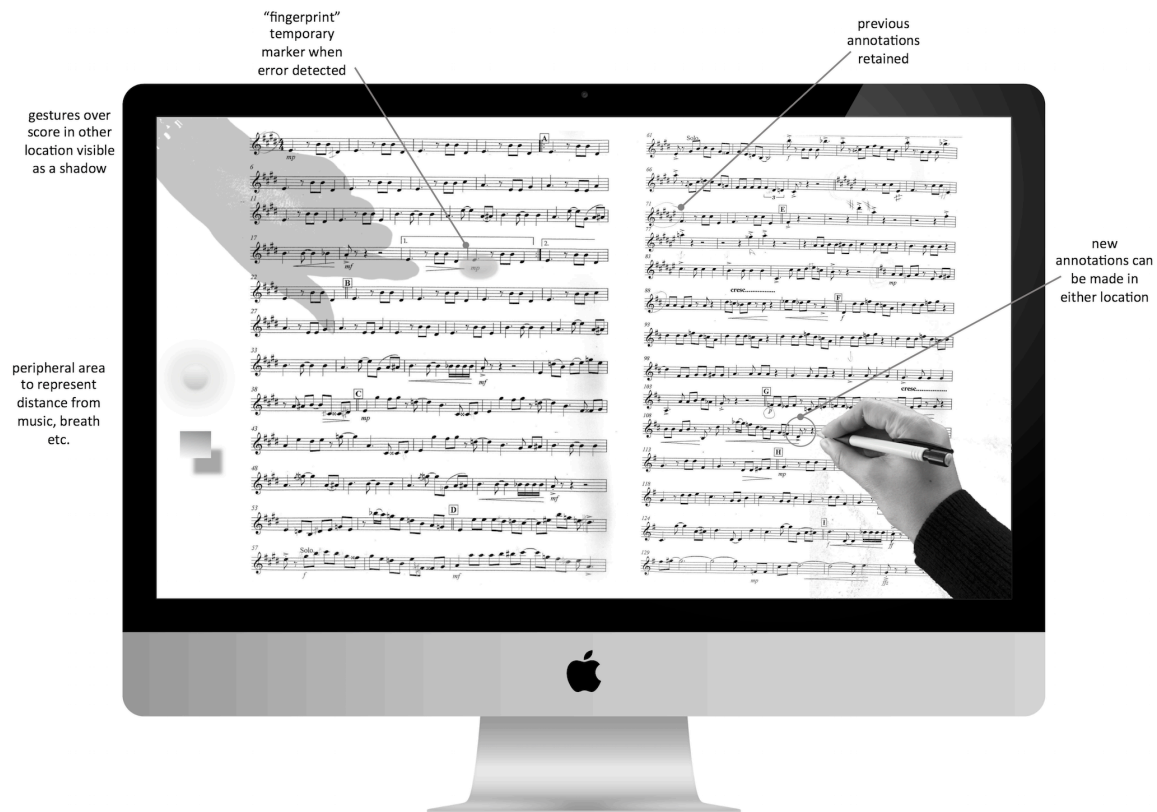
Figure 7: Representation of interactional cues shared on a digitized score, shared remotely.

Microphones and speakers would still be required to transmit audio, and considerations of feedback, echo cancellation and the other advances made in videoconferencing technology would still apply. However, audio takes up a smaller proportion of the total signal bandwidth for high definition video than the visuals, and the proposed interface display may not require as much bandwidth as the high definition visuals associated with a commercial video-conferencing system. This could be useful for organizations without access to high-speed Internet connections.

Final thoughts

A theme that emerges from some of the previous studies of remote music tuition is the way that the participants adapt the technology and the set-up to solve problems

themselves. For example, in Anderson (2008), the tutors found ways to use the equipment available to better share specific aspects of instrument physicality, or when connectivity was problematic, relied on their general knowledge to anticipate what a student might do, in a bid to keep the lesson moving (Anderson 2008). In another study examining remote piano lessons, the tutor suggested that the student use a USB to MIDI interface with a digital keyboard in their home to connect digitally with a Yamaha Disklavier on the tutor's campus 225 miles away. The student's playing was performed by the campus piano, complete with 'ghostly' keys and foot pedals that moved independently. The tutor commented, 'The connection worked almost instantly, and we were able to have a virtual lesson that was almost like being there' (Kruse et al. 2013). Further studies examining how established student–tutor pairs adapt to the online environment over time are required, as these may inspire further design ideas for new remote technology.

This work is limited to remote tuition mediated by videoconference, which is most likely to be available in established music education organizations due to cost, space and technical and bandwidth requirements. However many of the findings could also be applicable to the use of other, more widely accessible online platforms such as Skype, Google Hangouts and Apple's FaceTime. There are some differences between videoconference and these online platforms. Rather than a dedicated studio or video suite, the student is more likely to be at home using personal equipment. The microphones and cameras integrated with laptops or desktops, which are more likely to be found in the home, are likely to have a much lower quality than commercial videoconference equipment and not all homes will have access to a fast broadband network. Even if the tutor is in a studio with high-quality technology, the weakest point in the network will

limit the efficiency of the communication, for both participants. The student may not have the physical space required to sit or stand comfortably in front of a music stand, with a separate table to display their laptop at a suitable height. They may not understand the importance of posture without a co-present lesson where the tutor could reposition them, and the tutor may not be able to detect problems via the laptop webcam. Lessons in the home, rather than being situated in the context of an educational organization, may not have the same access to exam support such as piano accompaniment, music theory tuition or the opportunity to meet other students or take part in regular ensembles. Private online music tuition through platforms such as Skype have been studied (Kruse et al. 2013; Pike and Shoemaker 2013; Brändström et al. 2012; Denis 2016), but they often rely on questionnaires rather than the fine-grained interactional analysis found here. Further work could be done in this area to understand the additional social and interactional challenges posed, teaching strategies employed by tutors using these technologies and the opportunities for social interaction through online music communities integrated with these platforms.

Another limitation of this work is that it only considers advanced students who already have considerable technical skills, and their lessons predominantly focus on expression and musicality. Whilst technical topics were covered, students had enough knowledge to interpret and implement the recommendations discussed. Less advanced students may need help in a more physical form, which will present a different challenge. Different instrument types will also present different challenges and further work could include different instrument groups such as strings or brass.

In consideration of an alternative to videoconference to support remote teaching of a physical skill, an entirely different conceptualization is required – one which considers the importance of sharing tools in a shared environment, and the non-verbal resources used to coordinate collaborative work. The interactive digital score proposed here is a response to detailed analysis of naturalistic student–tutor interaction. The same approach could be applied to other fields where video technology is used to transfer knowledge and practice of a physical skill between geographically remote locations, such as surgical techniques (Hills and Jensen 1998; Luk et al. 2008) or dentistry (Fakhry et al. 2007), as well as enhancing the increasingly popular field of online music tuition.

Acknowledgements

REFERENCES

Anderson, A. (2008), 'Music lessons via satellite', in Australian Council for Computers in Education Conference (ACEC '08). Canberra.

Anderson, R., Beavers, J., VanDeGrift, T. and Videon, F. (2003), 'Videoconferencing and presentation support for synchronous distance learning', in Proceedings of the 33rd Conference for Annual Frontiers in Education (FIE '03), Boulder, CO: IEEE, pp. 13–18.

Bannister, R. (1992), 'Difficult but sensitive: Participant observation research in music education', *British Journal of Music Education*, 9:2, pp. 131–41.

Bautista, A., Perez Echeverria, M. P., Pozo, J. I. and Brizuela, B. M. (2009), 'Piano students' conceptions of musical scores as external representations: A cross-sectional study', *Journal of Research in Music Education*, 57:3, pp. 181–202.

Bellini, P., Fioravanti, F. and Nesi, P. (1999), 'Managing music in orchestras', *Computer*, 32:9, pp. 26–34.

Berg, M. (1999), 'Accumulating and coordinating: Occasions for information technologies in medical work', Computer Supported Cooperative Work (CSCW), 8:4, pp. 373–401.

Bevilacqua, F., Guédy, F., Schnell, N., Fléty, E. and Leroy, N. (2007), 'Wireless sensor interface and gesture-follower for music pedagogy', in *Proceedings of the 7th International Conference on New Interfaces for Musical Expression – NIME '07*, New York, NY: ACM Press, pp. 124–29.

Bigand, E., Lalitte, P., Lerdahl, F., Boucheix, J-M., Gerard, Y. and Pozzo, T. (2010), 'Looking into the eyes of a conductor performing Lerdahl's "Time after Time"', *Musicae Scientiae*, pp. 275–94.

Brändström, S., Wiklund, C. and Lundström, E. (2012), 'Developing distance music education in Arctic Scandinavia: Electric guitar teaching and master classes', *Music Education Research*, 14:4, pp. 448–56.

Brugman, H. (2004), 'Annotating multimedia/multi-modal resources with ELAN', in *Proceedings Fourth International Conference on Language Resources and Evaluation (LREC '04)*, Portugal, pp 2,065–68.

Cavitt, M. E. (2003), 'Descriptive error music analysis of in correction instrumental rehearsals', *Journal of Research in Music Education*, 51:3, pp. 218–30.

Chafe, C., Gurevich, M. and Leslie, G. (2004), 'Effect of time delay on ensemble accuracy', in *Proceedings of the International Symposium on Musical Acoustics (ISMA'04)*. Nara, Japan.

Cohen, S. (1993), 'Ethnography and popular music studies', *Popular Music*, 12:2, pp. 123–38.

Cooperstock, J. (2005), 'Interacting in shared reality', in *HCI International 2005*, Las Vegas, pp. 1–7.

Crawford, R. (2016), 'Rethinking teaching and learning pedagogy for education in the twenty-first century: Blended learning in music education', *Music Education Research*, 19:2, pp. 195–213.

Creech, A. (2012), 'Interpersonal behaviour in one-to-one instrumental lessons: An observational analysis', *British Journal of Music Education*, pp. 1–21. 29:3, pp.387–407.

Creech, A., Hallam, S., Mcqueen, H. and Varvarigou, M. (2013), 'The power of music in the lives of older adults', *Research Studies in Music Education*, 35:1, pp. 87–102.

Denis, J. (2016), 'Band students' perceptions of instruction via videoconferencing', *Journal of Music, Technology & Education*, 9:3, pp. 241–54.

Dourish, P. (2001), 'A history of interaction', in *Where the Action Is: The Foundations of Embodied Interaction*, Cambridge, Massachusetts: MIT Press, pp. 1–23.

Duffy, S. (2015), 'Shaping musical performance through conversation', Doctoral dissertation, Queen Mary University of London,

http://qmro.qmul.ac.uk/xmlui/handle/123456789/9531. Accessed 26 January 2016.

Duffy, S., Williams, D., Stevens, T., Kegel, I., Jansen, J., Cesar, P. and Healey, P.G.T. (2012), 'Remote music tuition', in *Proceedings of 9th Sound and Music Computing Conference (SMC 2012)*, Copenhagen, pp. 333–38.

Duffy, S. and Healey, P. G. T. (2012), 'Spatial co-ordination in music tuition', in N. Miyake, D. Peebles and R. P. Cooper (eds), Proceedings of the 34th Annual Conference of the Cognitive Science Society, Sapporo: Cognitive Science Society, pp. 1512–17.

——— (2013a), 'Music, speech and interaction in an instrumental music lesson: An ethnographic study of one-to-one music tuition', in M. Orwin, C. Howes and R. Kempson (eds), Language, Music and Interaction, London: College Publications, pp. 231–80.

——— (2013b), 'Using music as a turn in conversation in a lesson', in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, Berlin: Cognitive Science Society, pp. 2231–36.

Edigo, C. (1988), 'Videoconferencing as a technology to support group work: A review of its failure', in *Proceedings of the 1988 ACM conference on Computer-Supported Cooperative Work*, Portland, Oregon: ACM, pp. 13–24.

Fakhry, A., Cooper, S., Slach, N. and Krenz, S. (2007), 'Video-assisted clinical instruction in dentistry. Overview and applications', *European Journal of Dental Education*, 11:4, pp. 230–37.

Feld, S. and Fox, A. A. (1994), 'Music and language', *Annual Review of Anthropology*, 23:1, pp. 25–53.

Gaunt, H. and Hallam, S. (2009), 'Individuality in the learning of musical skills', in S. Hallam, I. Cross and M. Thaut (eds), The Oxford Handbook of Music Psychology, Oxford: Oxford University Press, pp. 274–94.

Graefe, C., Wahila, D., Maguire, J. and Dasna, O. (1996), 'Designing the muse: A digital music stand for the symphony musician', in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Common Ground – CHI '96, Vancouver: ACM, pp. 436–41.

Heath, C. and Hindmarsh, J. (2002), 'Analysing interaction: Video, ethnography and situated conduct', in T. May (ed.), Qualitative Research in Action, London: SAGE Publications Ltd., pp. 99–121.

Heath, C., Hindmarsh, J. and Luff, P. (2010), Video in Qualitative Research: Analysing Social Interaction in Everyday Life, London: SAGE Publications Ltd.

Heath, C., Luff, P. and Sellen, A. J. (1997), 'Reconfiguring media space: Supporting collaborative work', in K. Finn, A. J. Sellen and S. B. Wilbur (eds), Video-*Mediated Communication*, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 323–47.

Heldner, M. and Edlund, J. (2010), 'Pauses, gaps and overlaps in conversations', *Journal of Phonetics*, 38:4, pp. 555–68.

Hills, J. W. and Jensen, J. F. (1998), 'Telepresence technology in medicine: Principles and applications', *Proceedings of the IEEE*, 86:3, pp. 569–80.

Hollan, J. and Stornetta, S. (1992), 'Beyond being there', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY: ACM Press, pp. 119–25.

Hutchins, E. (1995), 'How a cockpit remembers its speeds', *Cognitive Science*, 19:3, pp. 265–88.

Jin, Z. (2013), *Formal Semantics for Music Notation Control Flow*, Masters dissertation, Carnegie Mellon University. http://www.cs.cmu.edu/~music/mat/thesis/Formal_Semantics_for_Music_Notation_Control_Flow.pdf. Accessed 5 June 2017.

Karlsson, J. and Juslin, P. N. (2008), 'Musical expression: An observational study of instrumental teaching', *Psychology of Music*, 36:3, pp. 309–34.

Kruse, N. B., Harlos, S. C., Callahan, R. M. and Herring, M. L. (2013), 'Skype music lessons in the academy: Intersections of music education, applied music and technology', *Journal of Music, Technology & Education*, 6:1, pp. 43–60.

Kurkul, W. W. (2007), 'Nonverbal communication in one-to-one music performance instruction', *Psychology of Music*, 35:2, pp. 327–62.

Lancaster, H. (2007), 'Music from another room: Real-time delivery of instrumental teaching', in *Music in Australian tertiary institutions: issues for the 21st century (NACTMUS 2007)*, Brisbane: Griffith University, pp. 1–18.

Laundry, B. A. (2011), *Sheet Music Unbound: A Fluid Approach to Sheet Music Display and Annotation on a Multi-touch Screen*, Masters Thesis, University of Waikato. http://researchcommons.waikato.ac.nz/handle/10289/5956 . Accessed 5 June 2017.

Lerner, G. H. (1995), 'Turn design and the organization of participation in instructional activities', *Discourse Processes*, 19:1, pp. 111–31.

Levinson, S. C. (1983), 'Conversational structure', in *Pragmatics*, Cambridge: Cambridge University Press, pp. 284–370.

Lin, K. and Bell, T. (2000), 'Integrating paper and digital music information systems', in *International Symposium on Music Information Retrieval (ISMIR 2000)*, Plymouth, Massachusetts.

Linden, J. Van Der, Johnson, R. and Bird, J. (2011), 'Buzzing to play: Lessons learned from an in the wild study of real-time vibrotactile feedback', in *CHI 2011*, Vancouver: ACM, pp. 533–42.

Lisboa, T., Chaffin, R., Schiaroli, A. G. and Barrera, A. (2004), 'Investigating practice and performance on the cello', in *Proceedings of the 8th International Conference on Music Perception & Cognition*, Illinois: ICMPC8, pp. 161–64.

Löchtefeld, M., Gehring, S., Jung, R. and Krüger, A. (2011), 'guitAR: Supporting guitar learning through mobile projection', in *CHI 2011*, Vancouver: ACM, pp. 1447–52.

Luk, R., Ho, M. and Aoki, P. M. (2008), 'Asynchronous remote medical consultation for Ghana', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, Florence: ACM, pp. 743–52.

McHoul, A. (1978), 'The organization of turns at formal talk in the classroom', *Language in Society*, 7:2, pp.183–213.

Menzies, D. W. H. and McPherson, A. P. (2013), 'A digital bagpipe chanter system to assist in one-to-one piping tuition', in *Proceedings of 10th Sound and Music Computing Conference (SMC 2013)*, Stockholm.

Moran, N. (2011), 'Music, bodies and relationships: An ethnographic contribution to embodied cognition studies', *Psychology of Music*, 41:1, pp. 5–17.

Morton, F. (2005), 'Performing ethnography: Irish traditional music sessions and new methodological spaces', *Social & Cultural Geography*, 6:5, pp. 661–76.

Nishizaka, A. (2006), 'What to learn: The embodied structure of the environment', *Research on Language & Social Interaction*, 39:2, pp. 119–54.

Noll, A. (1976), 'Teleconferencing communications activities', *Communications Society*, 14:6, pp. 8–14.

O'Conaill, B., Whittaker, S. and Wilbur, S. (1993), 'Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication', *Human Computer Interaction*, 8:4, pp. 389–428.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G. and Bruce, V. (1996), 'Comparison of face-to-face and video-mediated interaction', *Interacting with Computers*, 8:2, pp. 177–92.

Orto, C. and Karapetkov, S. (2011), 'Music performance and instruction over high-speed networks', California, http://www.polycom.com/global/documents/whitepapers/music_performance_and_instruction_over_highspeed_networks.pdf. Accessed 7 April 2015.

Penttinen, M. and Huovinen, E. (2011), 'The early development of sight-reading skills in adulthood: A study of eye movements', *Journal of Research in Music Education*, 59:2, pp. 196–220.

Pike, P. D. and Shoemaker, K. (2013), 'The effect of distance learning on acquisition of piano sight-reading skills', *Journal of Music, Technology & Education*, 6:2, pp. 147–62.

Presland, C. (2005), 'Conservatoire student and instrumental professor: The student perspective on a complex relationship', *British Journal of Music Education*, 22:3, p. 237.

Pye, R. and Williams, E. (1977), 'Teleconferencing: Is video valuable or is audio adequate?', *Telecommunications Policy*, 1:3, pp. 230–41.

Reed, D. J. (2015), 'Relinquishing in musical masterclasses: Embodied action in interactional projects', *Journal of Pragmatics*, 89, pp. 31–49.

Reed, D. and Reed, B. S. (2014), 'The emergence of learnables in music masterclasses', *Social Semiotics*, 24:4, pp. 446–67.

Ringwalt, D., Dannenberg, R. B. and Russell, A. (2015), 'Optical music recognition for interactive score display', in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '15)*, Baton Rouge: ACM, pp. 95–98.

Robine, M., Percival, G. and Lagrange, M. (2007), 'Analysis of saxophone performance for computer-assisted tutoring', in *Proceedings of the International Computer Music Conference (ICMC '07)*, Copenhagen: International Computer Music Association , pp. 381–84.

Rogers, K., Röhlig, A., Weing, M., Gugenheimer, J., Könings, B., Klepsch, M., Schaub, F., Rukzio, E., Seufert, T. and Weber, M. (2014), 'P.I.A.N.O.: Faster piano learning with interactive projection', in *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS 2014): Children and Learning*, Dresden: ACM, pp. 149–58.

Ruhleder, K. and Jordan, B. (1997), 'Capturing complex, distributed activities: Video-based interaction analysis as a component of workplace ethnography', in A. S. Lee, J. Liebenau and J. I. DeGross (eds), Informatio*n Systems and Qualitative Research*, Boston, MA: Springer US, pp. 246–75.

——— (2001), 'Co-constructing non-mutual realities: Delay-generated trouble in distributed interaction', Computer Supported Cooperative Work (CSCW), 10:1, pp. 113–38.

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

Schegloff, E. A., Jefferson, G. and Sacks, H. (1977), 'The preference for self-correction in the organization of repair in conversation', *Language*, 53:2, pp. 361–82.

Schober, M. F. (2006), 'Virtual environments for creative work in collaborative music-making', *Virtual Reality*, 10:2, pp. 85–94.

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

Seo, M.-S. and Koshik, I. (2010), 'A conversation analytic study of gestures that engender repair in ESL conversational tutoring', *Journal of Pragmatics*, 42:8, pp. 2219–39.

Siebenaler, D. J. (1997), 'Analysis of teacher-student interactions in the piano lessons of adults and children', *Journal of Research in Music Education*, 45:1, pp. 6–20.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, JP., Yoon, K-E. and 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, pp. 10587–92.

Stivers, T. and Sidnell, J. (2012), 'Introduction', in J. Sidnell and T. Stivers (eds), *The Handbook of Conversation Analysis*, West Sussex: Wiley-Blackwell, pp. 1–8.

Szczepek Reed, B., Reed, D. and 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, pp. 22–46.

Tait, H. A. and Blaiklock, B. A. (2005), 'Outreach, piano pedagogy and the future with technology', in *Proceedings of the 7th Australasian Piano Pedagogy Conference (APPC 2005)*, Adelaide.

Taylor, A. and Hallam, S. (2008), 'Understanding what it means for older students to learn basic musical skills on a keyboard instrument', *Music Education Research*, 10:2, pp. 285–306.

Thorne, S. and Payne, J. (2005), 'Evolutionary trajectories, Internet-mediated expression, and language education', *CALICO Journal*, 22:3, pp. 371–97.

Vines, B., Krumhansl, C. L., Wanderley, M. M. and Levitin, D. J. (2006), 'Cross-modal interactions in the perception of musical performance', *Cognition*, 101:1, pp. 80–113.

Waldron, J. (2013), 'User-generated content, YouTube and participatory culture on the Web: Music learning and teaching in two contrasting online communities', *Music Education Research*, 15:3, pp. 257–74.

Ward, N. and Tsukahara, W. (2000), 'Prosodic features which cue back-channel responses in English and Japanese', *Journal of Pragmatics*, 32:8, pp. 1177–207.

Weeks, P. (1996), 'A rehearsal of a Beethoven passage: An analysis of correction talk', *Research on Language & Social Interaction*, 29:3, pp. 247–90.

Welch, G. and Ockelford, A. (2009), 'The role of the institution and teachers in supporting learning', in S. Hallam, I. Cross and M. Thaut (eds), Oxford Handbook of Music Psychology, Oxford: Oxford University Press, pp. 307–19.

Wells, G. (1993), 'Reevaluating the IRF sequence: A proposal for the articulation of theories of activity and discourse for the analysis of teaching and learning in the classroom', Linguistics and Education, 5:1, pp. 1–37.

West, T. and Rostvall, A. (2003), 'A study of interaction and learning in instrumental teaching', International Journal of Music Education, 40:1, pp. 16–27.

Whittaker, S. (2003), 'Theories and methods in mediated communication', in A. C. Graesser, M. A. Gernsbacher and S. R. Goldman (eds), The Handbook *of Discourse Processes*, Mahwah, New Jersey: Lawrence Erlbaum Associates Inc, pp. 243–86.

Winget, M. (2006), 'Heroic frogs save the bow: Performing musician's annotation and interaction behaviour with written music', in *Proceedings of The International Society for Music Information Retrieval (ISMIR 2006)*, Victoria, Canada.

——— (2008), 'Annotations on musical scores by performing musicians: Collaborative models, interactive methods, and music digital library tool development', *Journal of the American Society for Information Science and Technology*, 59:12, pp. 1878–97.

Wolberg, G. and Schipper, I. (2012), 'Page turning solutions for musicians: A survey', Work (Reading, Mass.), 41:1, pp. 37–52.

Wurtz, P., Mueri, R. M. and Wiesendanger, M. (2009), 'Sight-reading of violinists: Eye movements anticipate the musical flow', *Experimental Brain Research*, 194:3, pp. 445–50.

Yin, J., Wang, Y. and Hsu, D. (2005), 'Digital violin tutor: An integrated system for beginning violin learners', in *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA '05)*, Singapore: ACM, pp. 976–85.

Contributor Details

Sam Duffy is currently a Research Associate at The Knowledge Lab, part of the UCL Institute of Education. Her research interests include music education and social music making, the impact that can be achieved through music in a community and how technology can be used to transform these interactions. After completing her doctoral degree at Queen Mary University of London in 2015, Sam was part of the team that developed 'Steve Reich's Clapping Music App'. The associated research and outreach work in schools won the 'Involve' award for Public Engagement at the Queen Mary Engagement and Enterprise Awards in 2017. She has collaborated with a number of orchestras and music organisations including Aldeburgh Music, The London Sinfonietta

and the London Symphony Orchestra's Outreach programme, LSO Discovery. Web
address: https://qmul.academia.edu/SamDuffy

Contact: Cognitive Science Group, Queen Mary University of London, Mile 49. End
Road, London E1 4NS, United Kingdom.

E-mail: s.duffy@qmul.ac.uk

Patrick Healey's research applies models of human communication – drawn mainly from
Psychology and Sociology – to understanding how digital technologies transform social
interaction. His work uses technology both as an experimental tool for the study of
interaction and as an application area for testing and developing theories of interaction.
Contact: Cognitive Science Group, Queen Mary University of London, Mile End Road,
London E1 4NS, United Kingdom.

E-mail: p.healey@qmul.ac.uk