## Introduction

Learning is inherently a task that can be done alone, but arguably can never be done without feedback. There are countless examples of learning without human intervention. For example, birds can learn to use simple tools to get food, and children can learn what types of foods they like through simple sensory feedback. Kolb (1984) used the term experiential learning to describe the cycle of experimentation, experience, reflection, and conceptualisation that typifies much of our learning. While learning can take place without the intervention and feedback from a teacher, the assistance of a teacher can obviously increase the productivity of learning many fold by guiding a learner to resources or experiences they may not have otherwise had, and to assist in the reflection process.

While it is most common for the teacher and learner to be co-located, this is not always possible for reasons of logistics, economics, or social reasons. Examples are K-12 lessons delivered to students in remote locations such as achieved by the Alice Springs School of the Air (http://www.assoa.nt.edu.au/the-school/), and adult learners such as engineering technicians receiving instruction on repairs on equipment far away from the person with the expertise about that equipment.

When one considers the three learning domains – cognitive, affective and psychomotor (Krathwohl, Bloom, & Masia, 1973), the challenges that face distance education are somewhat different for each domain. While the cognitive and affective domains may be less dependent on real time feedback and are often successfully achieved in an asynchronous environment, learning skills in the psychomotor domain arguably more suited to synchronous teaching and learning because often the number of tasks is great and duration of a task is very short. Take for example the task of learning how to print the alphabet. The time taken to print each character is very short as would be any feedback from a teacher on how to form the letter. An asynchronous environment would achieve very slow progress in learning such as skill.

It follows that there are tasks involving psychomotor skills to be learnt and in some circumstances it is not practicable for the teacher and learner to be in the same location. We came together from different disciplines, engineering education, music psychology and music education, to explore a particular teaching task that is usually only attempted in a face-to-face setting. The task is learning how to play a musical instrument. While some people learn without the help of a teacher, becoming an accomplished musician typically involves a teacher. Learning an instrument is a complex task that requires complex mix of visual, auditory and tactile senses. Both teacher and student need to observe aspects such as posture, dynamic movements of major and minor limbs, sound frequency and amplitude, timing, sequence of events, etc of each other, as well as the student being aware of these characteristics within themselves. Because of the complexity of the task, lessons are usually face-to-face and since explaining verbally a particular aspect is often so complex so to as make it totally impractical, physical demonstration of a particular aspect by the teacher is often used.

It is postulated that many students miss out on the opportunity to learn an instrument due to lack of a teacher to conduct a face-to-face lessons. Our discussions with secondary school music departments in large regional centres of Victoria reveal that it is common not to be able to source teachers of a range of classical instruments and voice. In both urban and regional areas, music teachers commonly spend a significant portion of their day driving between schools to give lessons because there is insufficient demand to give a full days tuition at one location.

With the advent of the Internet in the late 20$^{th}$ century, there has been modest interest in on-line tuition and collaboration in the realm of instrumental music. None have resulted in ongoing use of the systems. The most advanced is the LOLA (LOw LAtency audio visual streaming) system which attempted to solve the latency problems by designing and selecting all components to specifically reduce latency of transmissions (Allochio, Claudio, Buso, Nicola, & Drioli, 2012). Table 1 outlines features, technical and pedagogical issues of the systems reviewed.

Table 1. Published literature on research and deployment of music tuition.

| Author | System Name | Year | Features | Technical Problems | Pedagogical Problems |
|---|---|---|---|---|---|
| Ruippo, (2003) | Sibelius Academy | 2003 | Combination of asynchronous and synchronous technology | Slow connection speed | Greater degree of preparation required from teachers and they must ensure that students are engaged. General lack of ICT skills among music teachers. |
| Tait & Blaiklock, (2005) | Riverina Conservatorium of Music | 2005 | Video conferencing system | Equipment compatibility, require high connection speeds | Lack of 360 degree view (especially for piano and harp) |
| Lancaster, (2007) | Ultra-video-conferencing system | 2007 | Low-latency IP based video conferencing system | Network latency | None documented |
| Duffy et al., (2012) | Remote Tuition System | 2012 | Video conferencing system, 1 camera at teacher's location, 3 cameras at student's location | Video system delays | Teachers unable to get adequate views of students playing harp and piano |
| Allochio et al, (2012) | LOLA | 2012 | Low-latency IP based video conferencing system | Network latency | None documented |

There are many challenges to reproducing an authentic communication experience with a video conferencing system. Of utmost importance is the latency or time delay involved in signal transmission. Lester & Thronson (2011) have comprehensively documented issues of latency and indicate that the delay in signal transmission of touch from a persons finger to their brain is in the order of 30 ms. When a sound signal travels 20m in air from a source that can also be visualised, the latency of the audio signal is 60 ms and can just be detected by a human. In an audio-only system such as telephony, designers attempt to keep latency below 150 ms, and if it reaches 200 ms it is quite noticeable to users. Apart from latency in face-to-face communication human aural and visual sensory systems can recreate the three dimensional environment and for example detect the location of a sound in a room. Such information cannot be determined by receiving the same sound from a single microphone and speaker (Cooperstock, 2011). He also reports on the issues of how many visual clues are used in communication.

All commercial videoconferencing systems favour quality and low latency of the audio signal over the video signal. These systems also filter the audio signal to remove both low and high frequency components of the audio signal not typically featured in voice signals to reduce the processing time and transmission bandwidth required. This results in very poor reproduction of music signals that require the complete range of frequencies.

Our overall observations from the literature are that there have been significant efforts to reduce latency by many researchers. Less effort has been spent on adjusting teaching approaches to cope with the shortcomings of videoconferencing systems or simple changes to videoconferencing to favour music rather than voice transmission. We therefore decided to take a more holistic approach favouring development of a system that used consumer electronics and modified teaching approaches. While we chose to examine music tuition, one-on-one or small group collaboration in engineering education faces the same challenges in a videoconferencing environment.

## Our approach

An action research methodology was utilised for this project. Members of the project team included the co-researchers, associated personnel (including online instrumental teachers) and, postgraduate research students, as well as the school music teachers/supervisors were be involved in a cyclic pattern of planning, implementing, observing, and reflecting, before re-entering the cycle again. Hardware and software infrastructure testing and/or development was both laboratory-based—at the University—as well as field-based— at the regional school location. School music teachers and student participants were recruited to take part in trialing of online tuition to be given by city-based instrumental teachers that allowed for the testing of equipment and online connectivity. The data collected was based principally on observation (recorded in the form of field notes by school music teachers and specialist instrumental teachers and as video recordings) of student-teacher interactions. In addition, student participants were asked to document their experiences of online tuition after each session through student journals. The project team (together with school music teachers where appropriate via teleconferencing) were involved in post-tuition reviews of the video recordings, student journals and field notes and in group discussion. Concurrently videoconferencing and other digital streaming software and hardware was tested and promising solutions tested by the teachers and students. Development of a means of communicating music scores and annotations was undertaken in an associated project. The research was conducted with ethics approval from the Human Research Ethics Committee of the University.

## Observations and discussion

### Instrumental music tuition

As anticipated the principal advantage of video-conference-based tuition was that this medium allowed students in this regional setting to access specialist tuition that otherwise would not have been available in their local area. This point was well made by the Director of Music at the regional school location – who acted as the on-site supervisor – who commented that this medium was the only way that Year 12 students, some of whom were aspiring to undertake further musical training, could access tertiary level teachers with the obvious expertise to enrich student learning. One of the students whose chosen instrument – the oboe – requires highly specialised teaching remarked: “Having lessons each week using the Internet and the programs was amazing. I was able to learn so much from a very accomplished teacher, and for an oboe player in a regional area, there are not many oboe teachers. So it was really useful to learn from someone who had a different perspective and teaching style to the one [local] oboe player I currently (and have always) learnt from.” The same student summarised her experience in the project, stating “Overall it has been an amazing experience, especially being an oboe player and only having learnt from one person before, to learn about new techniques and approaches.”

Even with an instrument such as drums, where there is likely to be a greater pool of teachers in more rural areas, the student concerned stated that he was grateful to have had the opinions of a someone other than local teacher: “Although there were a few issues with the lessons which could only really be overcome by having lessons in the same room, it was worthwhile to have another view on my VCE syllabus.” His online teacher certainly recognised that at the end of the program, this student has gained from the online experience: “This being the last lesson of term, possibly the last between us, I felt that many learning outcomes had been achieved and that [my student’s] playing has improved. Together with his teacher at school ... , I feel he is able to see an overall picture of what lies ahead and how he can improve. The online learning experience, while having technical hiccups, has certainly been a positive experience and one that I can see a future in.” This was endorsed by the on-site supervising teacher who remarked: “Feedback from ... who was [the student’s] on-site drum teacher ... was that the online tuition added value to [his] learning of drums. {The online teacher] covered some different aspects that [the on-site teacher] didn't, so [the student] received a comprehensive education.”

In the case of the contemporary voice student, the on-site supervising teacher commented: “... with [the vocal student], her teacher ... said that it was really good that [she] could work on a particular area – i.e. scat singing / jazz genre – that added to her learning. It brought an additional dimension that the on-site teacher could not provide.” The on-site supervising teacher commented further: “The availability of the online teacher allowed for learning of additional aspects not covered by the on-site teacher. This was not because of any lack of knowledge or perceptiveness on the part of the on-site teacher, but because there was an additional perspective available – in this case from a tertiary level teacher compared with a secondary level teacher. It wasn't just the extra tuition time per week from the online teacher.”

The overall advantages of the video-conferencing medium was summed by the online oboe teacher as follows: “Obviously, the main strength of this program is that one is able to eliminate distance/cost as a deterrent to high quality individual music lessons. It was also extremely helpful to me, as I assume it was to those undertaking the research, to have the program run over such an extended period of time (as opposed to a token few weeks). This allowed a relationship to be built and, while the program was never exclusively about measuring actual improvement in artistic outcomes for the student, I felt that the research (problems etc) and outcomes were able to be measured in as close to a real-life situation as possible where artistic improvements are expected, and are being achieved.”

Nevertheless, there were some problems and therefore shortcomings with the video-conference medium experienced by the teachers and students participating in the project. Foremost among these was the problem of latency which is fully discussed below. In addition, there were problems, particularly in the initial stages of the project with technical aspects such as audio quality, video quality and internet connectivity, as well as placement and operation of equipment in the respective studio spaces. Many of this latter set of problems were essentially hardware and software issues, most of which were at least partially remedied during the course of the project.

One aspect that proved to be highly significant was the choice of the video-conferencing software. The project commenced with *VSee* software (VSee Lab, LLC) which to date has had telemedicine as its primary market. However, despite local technical support, there were on-going problems particularly with sound quality. One of the online teachers commented in his journal entry for the week: “The first three or four lessons were incredibly difficult due largely to problems with *VSee*. “There were problems at both ends ... at the technical level but once established our teacher-student relationship, both of us adapted to accommodate the technology. The key aspect for me was adapting to the different audio environment. Processing in music, particularly with drums, is the relationship of the performance to time. There is also the problem of the spatial direction of the sound. I tended to overly compensate with physical movement – i.e. exaggerate my gestures – in order to try to overcome timing issues – e.g. I gave a large gesture to convey the instruction to the student to stop.” This highlights that fact that most video-conference software has been designed from an audio perspective principally for speech rather than music. The frequency response is therefore rather limited which in the case of music makes perception of finer details of timbre and dynamics particularly difficult.

Towards the midway point of the project, the decision was made to transfer video-conferencing software from *VSee* to *Zoom* (Zoom Video Communications, Inc). The reaction of the online teachers, the instrumental music students and the supervising teacher was unanimous in finding that *Zoom* provided superior video images and superior audio. The situation was well summed up by the online oboe teacher with the comment, "The Zoom software is working more than adequately, compared with VSee, and any problems (such as sound delay) can be easily worked around in lessons. [My student] is progressing very well and I feel as though the consistency of the software has allowed us both to focus more exclusively on the craft of playing the oboe, rather than being distracted by the quality (or otherwise) of the software." Some weeks later, the same teacher was even more impressed with the superior performance of *Zoom*: "For the first time, I felt as though I knew exactly what [my student] looked like! This has obvious advantages in that I am able to more clearly see the way she sets up when playing (and fix any issues accordingly). I could see her embouchure very clearly and we were able to work (for essentially the first time) on correct set up of the embouchure, working specifically on the relationship between the top and bottom jaw (they should be essentially in line vertically when looking at the student side-on). The improved vision has also had a less obvious benefit in that I feel it has improved the relationship between teacher and student. While this bond takes some time to develop, and would naturally be expected to improve at around this stage even in normal face to face lessons, I noted significant improvement today. I put this down in large part to the vision and sound being clearer. We are able to converse more freely and are also able to react and adjust to the various visual cues which one relies upon in all human interaction. In this way, *Zoom* is further advanced than *VSee* in attempting to replicate face-to-face communication."

Another significant aspect was the level of technical support available to the teachers and students at their respective sites. Neither the teachers nor the students entered the project with any particular technical background, skill or knowledge about video-conferencing. In the case of the online teachers, a half-day workshop was arranged prior the online tuition commencing in which two of the three teachers were able to participate. For the students, there was no formal induction but at both video-conference sites, there was technical assistance available which contributed significantly in solving many of the practical difficulties that arose. For example, the support given at the regional school site, although not always in the form of 'stand-by' personnel, was nevertheless planned for in terms of 'user support materials:' I am not very tech-savvy so at first I was a little nervous about using the equipment. But the school made it very easy, printing out instruction sheets and laminating them so they were always with the computer we used. These sheets went through very clearly the steps of how to log onto the computer (passwords, usernames etc) how to turn the microphones on and get into the program. The instructions were very explicit and included pictures and screenshots which meant I felt confident in being able to set everything up. Most sessions I would go in, set up, join the meeting, [my teacher] would be there on the other end and [for the most part] everything would be fine." This was supported by a statement from the contemporary voice student who remarked, "The new setup [Zoom replacing VSee] as of last week, I felt was a lot more beneficial as it felt more like a lesson and more face-to-face than when we first started."

Unlike, the situation with *VSee*, there was provision for using an audio codec for *Zoom* that would specifically cater for the wider frequency and dynamic ranges of musical instruments and thereby enhancing the sound quality being transmitted to the remote location.

Nevertheless, there were frequently problems with audio that related to aspects other than software that included microphone placement, sound distortion, echo, hardware malfunction, etc. However, given the experimental nature of the project, overcoming such problems was generally taken as being 'par for the course'. The contemporary voice teacher explained one of several such situations as follows: “We had serious echoing problems to begin with, which seemed to be happening at both ends, making it impossible to communicate. It took about 15-20 minutes to get things working well enough to be able to proceed with the lesson. The technical assistant at their end experimented with speaker placement and volume, and checked the mic, and at our end [our technician] changed to a more directional mic.  By me using headphones, the echo was eliminated coming from my end and we had things working well enough to proceed with the lesson.”  In some instances, problems were easily solved by checking increasing volume level on the audio mixer.  In other cases, the internet connection appeared to be at fault as reported by the oboe teacher: “Only two issues to report today. That being that a message came up stating that ‘bandwidth is low’ at the [regional school] end. This meant that, especially in full screen mode, the picture was a little bit fuzzy and also movements would be unclear etc. A second issue was with the sound. Initially this wasn't working (possibly due to bandwidth issues), and even when it did arrive, it was very soft- almost like it had been muted somehow. The speakers at my end were up and to the best of my knowledge the microphones were functioning at the Ballarat end.”

## Latency testing

The literature cited concentrated on the issue of latency in the Internet connection between teacher and student.  While this is important, our observations were that the other components in the system were much more likely to be a cause of significant latency.  A typical videoconferencing endpoint consists of a digital video camera, one or more monitors, a microphone, other input devices such as keyboard and mouse all connected to an Internet enabled computer. In order to reduce the Internet bandwidth use the video and audio signals are compressed prior to transmission and then decompressed at the receiver to output on the monitor or speaker.  Each of these components and operations has latency involved.  We worked with several hardware developers and AV system vendors who had high expectations of the performance of their products, but when put together into real end-to-end systems the total signal delay between endpoints was disappointing with delays of 500 to 1000 ms being common.  This was also poorer performance than the best consumer level videoconferencing system we used (*Zoom*, https://zoom.us/).  Our conclusion from this work is that for systems that are going to be widely deployed rather than using specialist studios, latencies of greater than 60 ms (the threshold of human perception) or 150 ms (the design standard for telephony) are likely to be a fact of life and that adjusting the teaching approach will be the best way forward.

## Music score delivery

In a face-to-face class a student may typically have a workbook and a music score that both the student and teacher can work from.  The teacher may also have other printed music scores they can quickly refer to in particular situations.  Furthermore, the teacher can write on the student’s score or workbook to emphasise certain points or to give homework instructions.  This aspect did not receive much attention in the literature, except in an asynchronous system where scores, annotations and text comments.  Being asynchronous, live interaction between the student and teacher was not practiced.  For synchronous lessons various options were trialled.  Using the video signal stream to send an image of a music score and annotations is one possibility, however it is quite expensive in resources to consider this as an operational means.  It also does not allow the student and teacher to both write on the same item.  All the videoconferencing software used had a screen sharing option which was successfully used.  The teacher could share a screen that has a PDF file of the music score and make annotations on it.  This method was reasonably well received. Teachers also emailed music scores and notes to the students, which was very successful because the technology was well understood and easily used, although any interactivity was lost.

As part of the overall project development an information technology student project investigated other options for sharing music scores (Vadakkeveettil 2014). He developed a prototype 'chat room' style software application that relied on HTML5 protocols to use a browser for student teacher interaction. A mark-up language called MusicXML was used to render a music score in a browser window that both the student and teacher could observe. Using a central server to control that window, both the student and the teacher could annotate and highlight objects and regions on the score. Since only XML instructions and vector instructions for annotations were being shared between the teacher and student the bandwidth required was low and latency determined by network latency only. The other advantage of the system is that the instructions for annotations could be stored and retrieved at some future date, for example the next lesson or when the student was doing their homework.

### Teacher reflections and recommendations for future action

During the final weeks of online tuition and after the program had concluded, all participants were invited to summarise their experiences. Despite the technical and interactional difficulties experienced by all participants, the following comments reflect the general satisfaction with and worthwhile nature of both their experiences of and the future potential of online instrumental music tuition.

"I have thoroughly enjoyed being a part of this program. Whether I am making a difference to [my student] as an oboist I cannot be certain as yet, though I'd like to think that she has taken some different approaches to playing the oboe as a result of these lessons. Perhaps more importantly, I hope that the work being done by everyone involved with this project is work that can bear fruit for the University in the coming years, as we expand upon these new possibilities in the teaching of music." (online teacher of oboe).

Some worthwhile suggestions for future action were also made by the online teachers including the desirability of a dedicated studio for online teaching that is set up with cameras and microphones tuned to the space, the desirability of have quick access to technical assistance if problems are encountered with the complex set up of computers and peripherals, and the desirability of having an initial face-to-face lesson to help 'calibrate' the quality of interaction in subsequent online lessons.

## Conclusions

This research investigated various aspects of using videoconferencing technology to assist learning of a psychomotor intensive activity, playing a musical instrument, by students in remote locations. Significant advances were made on a number of fronts, however, latency in the videoconferencing environment remains an issue that is unlikely to be solved due to technical limitations. Our research found that by subtle adjustments of teaching methods, and a willingness on the part of the teachers to accept that they were not working in a face-to-face environment, significant progress could be made by the learners. If possible occasional face-to-face classes should be held, but of the two alternatives of no tuition by a specialist teacher and tuition by videoconferencing, the latter is clearly advantageous. In designing teaching and learning activities using videoconferencing, whether it be for music tuition or engineering education, as much attention needs to be paid to the associated aspects of the activity such as document exchange and annotation and teaching or collaboration style as is paid to the technical aspects of video and audio transmission. Even with broadband networks, if it is important that the video and audio signals are synchronised then noticeable latency will be inherent in the system and therefore the activity must be designed around that delay. There are many consumer level video conferencing solutions available but the performance is not uniform and careful testing is advised to find one that meets the need. Professional level video conferencing solutions are not necessarily any better because they are largely optimised around giving priority to audio transmission which may not suit specific applications.

## References

Allochio, Claudio, Buso, Nicola, Drioli, C. (2012). LOLA Presentation. *LOLA Workshop*.

Cooperstock, J. R. (2011). Supporting demanding collaborative human activities ]. *IEEE Signal Processing Magazine*.

Duffy, S., Williams, D., Stevens, T., Kegel, I., Jansen, J., Cesar, P., & Healey, P. G. T. (2012). Remote Music Tuition. In *Proceedings of 9th Sound and Music Computing Conference* (pp. 333–338).

Kolb, D. A. (1984). *Experiential learning : experience as the source of learning and development*. Englewood Cliffs, N.J.: Prentice-Hall.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1973). *Taxonomy of Educational Objectives, the Classification of Educational Goals. Handbook II: Affective Domain.* . New York: David McKay Co., Inc.

Lancaster, H. (2007). Music From Another Room : Real-time Delivery of Instrumental Teaching. In *NACTMUS National Conference, Music in Australian Tertiary Institutions–Issues for the 21st Century* (pp. 1–20). Retrieved from http://espace.library.uq.edu.au/view/UQ:136601

Lester, D., & Thronson, H. (2011). Human space exploration and human spaceflight: Latency and the cognitive scale of the universe. *Space Policy*, *27*(2), 89–93. http://doi.org/10.1016/j.spacepol.2011.02.002

Ruippo, M. (2003). Music Education Online. *Sibelius Academy*, (2), 1–8.

Tait, H. a, & Blaiklock, B. a. (2005). Outreach, Piano Pedagogy and the Future with Technology. In *Australasian Piano Pedagogy Conference* (pp. 1–5).

Vadakkeveettil, S. T. (2014). Multi Camera Video Conferencing for Online Music Learning :Approaches to Real Time Sharing of Music Sheet and Annotations. Melbourne School of Engineering, University of Melbourne. Master of Information Technology: 24.

## Acknowledgements

We acknowledge the staff and students of Ballarat Clarendon College who participated in this research. Staff of the Melbourne Networked Society Institute (MNSI) , Julien Ridoux and Ken Clarke are acknowledged for their contribution to work, as are research assistant John Baratta and technician David Collins. The work was partially funded by the University of Melbourne through the MNSI.

### Copyright