Adriana Olmos, Nicolas Bouillot, Trevor Knight, Nordhal Mabire, Josh Redel, and Jeremy R. Cooperstock
McGill University
Department of Electrical and Computer Engineering
3480 University Street
Montreal, QC H3A 2A7, Canada
{aolmos, nicolas, jer}@cim.mcgill.ca
{trevorknight, nordhal.mabire}@gmail.com
joshredel@mac.com

A High-Fidelity Orchestra Simulator for Individual Musicians' Practice

Abstract: We developed the Open Orchestra system to provide individual musicians with a high-fidelity experience of ensemble rehearsal or performance, combined with the convenience and flexibility of solo study. This builds on the theme of an immersive orchestral simulator that also supports the pedagogical objective of instructor feedback on individual recordings, as needed to improve one's performance. Unlike previous systems intended for musical rehearsal, Open Orchestra attempts to offer both an auditory and a visual representation of the rest of the orchestra, spatially rendered from the perspective of the practicing musician. We review the objectives and architecture of our system, describe the functions of our digital music stand, discuss the challenges of generating the media content needed for this system, and describe provisions for offering feedback during rehearsal.

Ensemble study allows musicians to practice within the context of a group and refine their listening and interaction with other instrumentalists. This requires ample time, availability of the other performers, and large spaces. Individual study, on the other hand, allows the freedom to practice on any schedule, focusing effort on the specific areas of importance to the performer. Ensemble performance requires skills beyond those that can be fully trained in solo practice, however. As part of a whole, a musician must constantly "fit" with the ensemble, as defined through the conductor's instruction or vision; this fit is not fully specified in musical scores (Foote 1979; Progler 1995). Broadly speaking, these factors can affect the sound and temporal characteristics of the musician's performance. The former can be described in terms of inflection, intonation, and balance, and the latter by articulation and accuracy of fingering, as examples.

Several technological solutions have been proposed to overcome the constraints of ensemble study. For instance, the Performance Learning System (Goldmark 1976) and Music Minus One (musicminusone.com) consist of prepared recordings of a musical program from which an instrument or voice is missing. A musician may practice and learn the omitted performance while being accompanied by a recording of the ensemble, much like karaoke systems. Existing technologies of this form suffer from an absence of visual cues, navigation difficulties, and limited sound control of the ensemble.

This article describes Open Orchestra, a system that combines the high-fidelity experience of the ensemble rehearsal or performance, with the convenience and flexibility of solo study, overcoming the limitations noted earlier. Open Orchestra builds on the theme of an immersive orchestral simulator, and also supports the pedagogical objective of instructor feedback on individual recordings, as needed to improve musicians' performance.

The remainder of this article is organized as follows. First, we review relevant literature in the field, then introduce the objectives of our system, followed by a summary of use cases and a description of the environment in which the musicians use the system. Next, we discuss the requirements and needed technologies to construct the repertoire, as well as the overall system architecture. Finally, we conclude with some comments on implementation challenges and future work. (A video of the Open Orchestra system in use is available from openorchestra.cim.mcgill.ca.)

Computer Music Journal, 36:2, pp. 55–73, Summer 2012

Related Work

Many musical practice systems focus on improvement of instrument skills and are thus best suited for amateur players. For instance, the digital violin tutor (Yin, Wang, and Hsu 2005) uses audio analysis for providing feedback to the student. This can detect fingering mistakes and then display the correct position using an avatar within a controllable 3D environment. SmartMusic (www.smartmusic.com) provides similar feedback, but does so through a pop-up display of fingering positions. Similarly, the piano tutor project (Dannenberg et al. 2003) uses score-following software to analyze the student's performance, and then provide feedback where mistakes are made. The system then emulates a teacher by finding the most appropriate lessons according to the student's mistakes. Such systems are generally targeted to amateur musicians wanting to practice their instrument and enhance their basic-playing skills. Experienced musicians, however, develop exceptional listening abilities (Koivunen 2002), and are, therefore, likely to detect most of their own playing mistakes.

IMUTUS (Schoonderwaldt, Hansen, and Askenfeld 2004) was designed to be a complete, autonomous tutor, used without human teachers and providing feedback to the student after each performance. The authors, however, suggest that its benefit would be maximized if used in conjunction with an actual teacher. The system defines a priority of mistakes based on discussions with over 40 teachers. For example, mistakes in articulation are less important for beginning players than control of air flow. The system only informs students of a few mistakes at a time in order to avoid overwhelming them. Students may also request hints in the form of additional annotations made by teachers. This work has been extended and specialized to wind instruments in the VEMUS project (Fober, Letz, and Orlarey 2007).

The i-Maestro (Ng 2008) project targets string instruments, and provides interactive environments for self-learning, private lessons, and classroom use for pedagogy of both theory and performance. The project specifically addresses training support for string instruments with a particular interest in linking music practice and theory training. The resulting i-Maestro framework for technology-enhanced music learning is designed to support the creation of flexible and customizable e-learning courses, including gesture following and exercise generation.

Other systems offer the experience of performing with an orchestra, supporting real-time accompaniment (Dannenberg 1984; Vercoe 1984; Raphael 2006) by synthesis according to anticipation of the performer's trajectory through a non-improvised musical piece. Although potentially useful to obtain a more responsive accompaniment than Music Minus One or its derivatives, such accompaniment systems are not necessarily intended for instructional purposes, and do not provide the critical feedback to help improve one's performance. In their review of various musical tutoring systems, Percival, Wang, and Tzanetakis (2007) discuss recommendations regarding the design of such systems for young and amateur musicians. They conclude that visualization techniques may be useful to confirm or correct a student's judgment, but caution that such feedback should not be used to replace self-assessment by the student.

Virtual environments for orchestral conducting (Borchers, Lee, and Mühlhäuser 2004) have also been developed, allowing the user to control tempo and dynamics. In response to natural conducting gestures made with an infrared baton, the system plays back a real-time audiovisual recording of an actual orchestra. In contrast, the virtual conductor by Reidsma, Nijholt, and Bos (2008) provides visual feedback in the form of a conductor avatar, driven by real-time audio analysis to help communicate tempo mistakes to the musician. A similar system, UBS Virtual Maestro (Nakra et al. 2009) reads the accelerometer data from a hand-held Wii Remote, which the user moves like a conducting baton to affect the playback tempo and dynamics of an audiovisual orchestral recording.

Networked musical performance approached the development of a virtual orchestral environment by allowing musicians to collaborate in real time at a distance (Woszczyk et al. 2005; Shober 2006; Carôt, Rebelo, and Renaud 2007; Zimmermann et al. 2008). These systems focus mainly on small ensembles, however, and the quality often suffers as a result of network latency. Moreover, the ensemble performance experience remains dependent on the simultaneous availability of all performers, in addition to an instructor for feedback and critique.

Open Orchestra Objectives

To address the shortcomings of previous systems, the Open Orchestra project aims to create a high-fidelity, immersive, network-enabled platform that provides individual musicians a rich audiovisual simulation of ensemble performance.

The system is intended to support the full range of training activities for collaboration between a musician and a conductor within the context of a rehearsal scenario. As such, our design was informed by the use cases arising from a series of field studies carried out with a large jazz orchestra. Each of our objectives, resulting from this exercise, is described briefly herein.

Preparatory Activities

At the beginning of a rehearsal, musicians prepare by warming up their lips and fingers, tuning their instruments, and comparing notes with their peers, while awaiting a starting cue from the conductor. During this time, the conductor might also review the score and briefly compare notes with certain members of the orchestra. To support some of these aspects of real-world rehearsal, Open Orchestra keeps track of the instrument(s) the musician plays, personalizing the options presented, such as selections of musical parts and sound settings from previous sessions. At login, Open Orchestra uses the introductory period to display any pertinent text messages to the student. For instance, an instructor might have commented to the saxophonist on the need to apply better intonation to particular notes in the music part. In the case of a first-time user, the system uses this period to recommend default sound settings according to the student's instrument(s) of choice, and encourages the musician to record practice sessions for sharing with one of the listed instructors or mentors.

Performance Guidance

When an ensemble was in the early stages of learning a new piece of music, we observed the conductor counting the beat with his fingers and providing verbal guidance as to what was expected in certain sections of the music. Similarly, Open Orchestra provides the musician with an optional digital metronome for tempo guidance and access to the conductor's guidelines, recorded as audiovisual content, for how he or she would like certain sections to be played. Because this is most likely relevant only in the early stages of rehearsal, the student can deselect playback of these general comments at any time. During regular (non-virtual) rehearsal sessions, we observed that musicians take advantage of small breaks or pauses to practice a specific bar that might be hard to play, just as they might repeatedly practice a specific complicated section at home. To support this functionality in Open Orchestra, the musician can directly set the starting and the ending points of a section using the written music, and then play that section repeatedly in a loop. This offers greater flexibility than is available in a traditional rehearsal. As one student commented, during the rehearsal he "cannot pause the conductor" to repeat a music section with the orchestra, but with our system, "here I can."

Collaborative Tool

Ensemble rehearsal is more than refining one's ability to play a piece of music on a given instrument. Equally important is that it entails performing as a group, in coordination with the full orchestra. To support training of these aspects, the system includes recording and playback features so that the musicians can review their own performance. These features are augmented by a synchronized display of the music part and optional performance evaluation tools that provide feedback concerning the quality of the student's recording, based on signal-analysis metrics and comparison with professional recordings. The analysis is integrated into the display of the written music part for visualization purposes, and can be accessed by the instructor as well, if requested. The musician can write comments or questions on the recording, and share it with an instructor who, in turn, can also provide feedback, either through written comments or through recorded audio. These interactions occur asynchronously. When either the musician or the conductor logs into the system, all of their messages are displayed, allowing them to continue the conversation or interaction at their convenience.

Feedback and Student–Instructor Interaction

During a rehearsal, the conductor uses gestures and singing to describe the temporal variation of note dynamics, inflection, and intonation, and also clapping to indicate tempo, among other forms of feedback. These expressions can be conveyed by the audio and video of the conductor within Open Orchestra. In addition, to help students better understand the subtleties of performance, Open Orchestra attempts to offer feedback through a visual representation of performance parameters, including articulation, timing, intonation, inflection, and dynamic contrast, in relation to similar parameters from a reference track of a professional musician. Although feedback is typically the prerogative of the conductor, students often paraphrase the conductor's comments, or ask questions related to the interpretation of a certain section of the music part. This capability is supported in Open Orchestra through a feature for posting questions directly (but asynchronously) to the conductor. Musicians can also record their performance and include this with their question(s) to the conductor. This allows the conductor to "zoom" in to the performance, listening either just to the instrument on its own, or in the context of the full orchestra. Because of time constraints, this capability is normally not feasible in an actual orchestral rehearsal setting.

High-Definition Immersive Experience

Open Orchestra provides the musician with the experience of sitting within the ensemble; but, unlike previous systems, it does so with both high-definition video and high-resolution audio, rendered from the perspective of the instrumentalist. From previous studies in virtual reality (Psotka 1995), we believe that this perspective is critical to achieving a compelling sense of immersion in the simulated environment. To increase the quality of this experience, the musicians see the conductor and the relevant part of the orchestra on a panoramic video display, and hear the rest of the orchestra with their own part removed. In addition, the system displays the written music part and system controls on a digital music stand that can be operated by touch. Among other capabilities, this offers the musicians access to comments made by the conductor at certain bars, emphasizing certain aspects of the piece to which the orchestra should be paying attention.

Network-Enabled Platform

The potential storage requirements for the collection of professional audio and video recordings of multiple selections of ensemble performance, recorded or rendered from a variety of instrumentalist perspectives, could easily overwhelm the capacity of typical servers. To ensure content delivery in the highest-quality format possible, it is more effective to adopt a network-enabled platform in which one high-capacity server is dedicated for streaming the high-definition content, on demand, over a high-speed network. Subsequent interactions, in the form of recordings, comments, and messages, are then stored in the cloud. This architecture serves the need to communicate self-recorded samples of the student's performance to a remote instructor (or conductor) for asynchronous review and comment. Further details are provided under the System Architecture section.

System Overview

We now provide an overview of the system components necessary to achieve our objectives, including the audiovisual display, the various functions of the digital music stand, and the feedback capabilities of Open Orchestra. At the time of writing, we have

Figure 1. Early prototype of the Open Orchestra hardware.

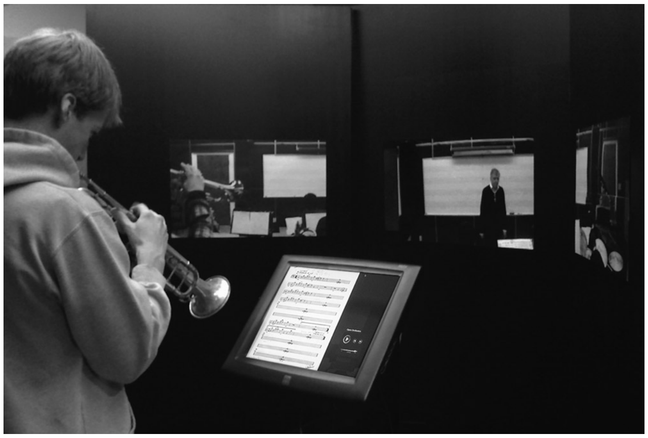


Figure 2. The Open Orchestra system as currently deployed. The touch screen, seen in the middle, serves as the music stand for display of the music part and controls.

completed the design and implementation of a preliminary prototype (see Figure 1) that supports most of the functionality described subsequently, apart from visualization of the performance feedback. Currently, the workstation is being tested at five music schools across Canada.

Video and Audio Display

The panoramic view of the orchestra is delivered through three 32-in NEC MultiSync V321 screens (see Figure 2); the audio is delivered through Bose QC3 noise-canceling stereo headphones. The decision to use headphones instead of loudspeakers was motivated by the demand for an immersive audio experience that could be deployed in a typical university space, without the need to construct a dedicated room with special acoustic properties. This provides the further benefit of isolating the performer's own audio from the accompanying orchestral tracks when the student captures recordings for instructor feedback.

Control over audio rendering is necessary to select between an idealized audio playback and a more realistic rendering of the performance from the instrumentalist's position within the orchestra, with the sound of distant instruments attenuated. As one of the drum players in our studies commented, "When you are playing in a big band as a drummer you are so far back that all the sound is going out and you have nothing." From this student's perspective, the normal orchestral listening conditions make it very difficult to play. Similarly, a trombonist commented, "All you can hear is the first trumpet in your ear and whoever is beside you, everybody else is a big wash; I never hear any piano or anything, at least from where I am sitting." Although studies are required to demonstrate the pedagogical utility, we are interested in supporting initial learning from a more idealized rendering of the ensemble sound, and then gradually moving to more realistic listening conditions. This prompted our studies of audio rendering perspective, summarized under the Preliminary Evaluation section.

Digital Music Stand

Written musical scores can be regarded as formal symbolic objects that provide methods and procedures to communicate common processes and goals to various entities within an orchestra. Scores also support teamwork and collaboration within an orchestra, with annotations marking the location where clarification, augmentation, or modification of the written instructions is needed (Winget 2008). Given the importance of annotation functionality, and the desire to integrate it in a manner that supports its storage and communication, we incorporated an electronic tablet as a digital music stand for display of the score and the user-interface elements. The stand serves as the student's interface to the system, providing a display of the music part and performance feedback, control of playback and audio levels through finger-touch input, and music annotations through stylus input. Given its centrality to operation of the Open Orchestra system, the stand must be easily adjustable to accommodate different player heights as well as the player's preference of sitting or standing positions. This is accomplished by mounting the tablet on an adjustable arm that allows the stand to lift, tilt, pan, and rotate (see Figure 2).

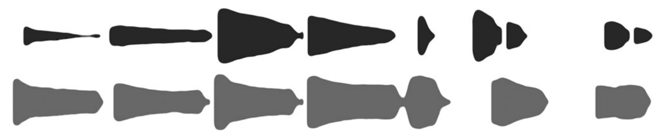
Actions that musicians perform while interacting with their music part include selection of music pieces, play/pause control, flipping through pages, and specifying start points from which to rehearse. Given their high frequency of expected use, we want to ensure that these operations do not become cumbersome. This motivates the use of a touch-sensitive display, allowing efficient control by finger touch or gesture, rather than requiring interaction with a dedicated input device, whether mouse or stylus. To support annotation in a natural manner during rehearsal (for example, marking up the music with feedback provided by the conductor), however, stylus input is desirable, to ensure a similar level of writing efficiency and control as with a pen. Moreover, finger-touch sensing alone is likely inadequate to indicate with accuracy a specific note, which in most cases occupies only a small number of pixels on the display. These factors motivated our adoption of a dual-input technology, supporting both finger and stylus input, with high-accuracy position detection.

Automatic Performance Feedback

Machine-based performance evaluation can offer musicians an objective assessment of their performance, independent of an instructor, and can be used by the instructor as one measure of a student's progress. The latter aspect is important when considering scalability to large groups, as this not only offloads some of the responsibilities of the instructor, but helps identify sections of the music or aspects of the performance that are proving most difficult for the ensemble as a whole.

Analysis of the audio signal is used to extract fundamental frequency, harmonic amplitudes, and root-mean-square (RMS) amplitude, from which a number of musical features, including pitch and intonation, timing and rhythm, articulation, timbre, and dynamics can be inferred, similar to the approach used in previous literature (Geslin and Lefevre 2004; Schoonderwaldt, Askenfelt, and Hansen 2005; Daudin et al. 2007). These features can then be compared with those from a reference audio track, through an appropriate visualization technique for articulation and dynamics.

Figure 3. Visualization of articulation with the reference (top) and student (bottom) performance.

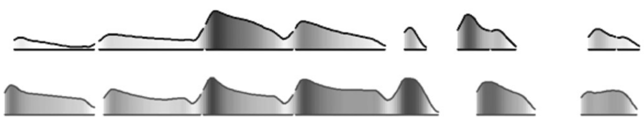


Visualization of a student's performance requires a model against which comparisons can be made to pinpoint areas in need of improvement. Unfortunately, an objective model, such as one generated from symbolic data (e.g., the music score), lacks the subtlety and nuance of interpretation naturally contained in an audio recording. For this reason, an audio recording is preferred as source data for the visualizations, representing the changes of both articulation and dynamics over time. We performed feature extraction of the RMS values of the recordings with Sonic Visualiser (Cannam et al. 2006) and used Processing (processing.org) to generate the visualizations.

Figure 3 shows an example articulation visualization of student and reference musicians playing the same nine-note phrase, with the reference performance (top) and the student's performance (bottom). In the reference performance, the musician has articulated the last two pairs of notes more clearly, whereas the student musician slurs them together. Timing differences among the performances can also be seen on the corresponding two pairs, with the student playing late on the first and early on the second.

Figure 4 shows the same phrase with a different visualization, this time mapping amplitude to both height and color, from yellow to red (shown as light to dark gray in this publication). The data have also been smoothed with a moving average to focus on the overall contour rather than the fine-grained

Figure 4. Visualization of dynamics with the reference (top) and student (bottom) performances.



detail. In this view, the crescendo-decrescendo contour of the first five notes of the reference phrase (top) contrasts with the flatter student performance (bottom).

Unlike most music tutoring systems, the evaluation provided by Open Orchestra is intended to be oriented toward characteristics of ensemble performance, rather than independent, technical proficiency of the musician. Although it may prove valuable to include the latter as part of the feedback provided by our system, we are consulting with several conductors to ensure that we retain an initial focus on the most appropriate metrics of ensemble integration.

Asynchronous Expert Feedback

To maximize the benefit to the student, it will usually be necessary to complement the automatic performance feedback, described earlier, with expert feedback and technical instruction that only a professional music teacher or conductor can provide. Open Orchestra supports this capability by making selected student recordings accessible to an instructor or conductor, who may review these at any time and offer feedback as appropriate. The system thus establishes a collaborative platform where musicians, conductors, and instructors can interchange ideas regarding the performance and interpretation of a specific musical piece.

Building the Repertoire

For music schools and students, access to a wide musical repertoire through our platform represents a significant opportunity, not only for ensemble practice, but also for discovery and/or consideration of new pieces. The process of building such a repertoire requires considerable effort for each piece, however. In addition, the content must be correlated with a digital representation of each of the music parts, allowing for temporal and content association between the practicing musician and the full performance.

Audiovisual Content Generation

Two options exist to obtain perspective-specific audiovisual rendering with independent volume control for the multiple channels, the first involving simulation of the entire performance, the second based on live content acquisition.

Simulated Content Acquisition

The synthetic-reproduction approach offers the advantage of being able to generate a large number of audio pieces without the need for an expensive and complex recording process, nor even the presence of an instrumentalist. There is, no doubt, a significant continuum of quality that can be obtained from a synthetic reproduction, ranging from a direct, "naive" Musical Instrument Digital Interface output, to one that attempts to reproduce the various annotations and descriptions provided by the composer, as well as the deliberate shaping by performers of expressive parameters such as tempo, dynamics, and articulation (Widmer and Goebl 2004). In this manner, highly complex scores can be "performed" without the substantial costs of hiring professional musicians. Various computational approaches have been investigated, for example, using rule-based (Friberg 2006), structure-level (McAngus Todd 1992), and mathematical (Mazzola and Göller 2002) models to formulate hypotheses of expressive music performance. Widmer and Goebl (2004) provide a review of such methods, including their own combination of note-level rules with structure-level expressive patterns induced by machine learning (Widmer 2002).

A significant benefit of these models is the automation of the generative process that would otherwise require a highly detailed description of each part, including precise intonation, inflection, timing, and other features. Such models, however, focus on common principles of performance and,

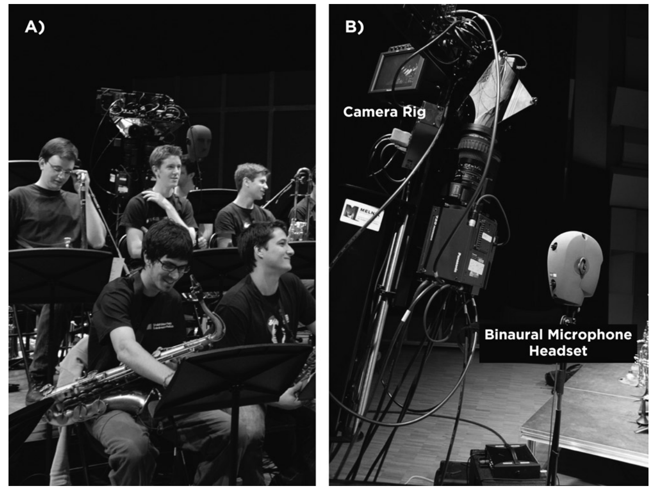


Figure 5. The three-camera recording apparatus inserted into the position of the lead trumpet player. The cameras can be seen in close-up.

accordingly, permit limited control over personal artistic aspects of performance style (Widmer and Goebl 2004). For those implementations that support some degree of expressive control over the generated music, the complexity of such control may require expertise beyond the musical skills of the target users.

Another important aspect of synthetic musical reproduction is the synthesizer itself; most do not yet model a natural-sounding transition between notes (Lindemann 2007). This results, for example, in synthesized violin and trumpet sounds that are perceived more as a concatenation of separate sounds than as a continuous musical phrase. Synful (synful.com) attempts to preserve the realism and complexity of natural sound, employing reconstructive phrase modeling, which is an analysis-synthesis system related to both additive and concatenative synthesis, and includes a database of rapidly varying components derived from original recordings of human performers.

Although it is perhaps only a matter of time before the subtleties of human performance can be modeled and controlled adequately, current systems rely on information provided in the score, for which the details of performance are typically not specified completely. For example, as frequently observed in jazz scores, a piano staff may be left as a sequence of chord names, or a drum staff might only designate particular instants to emphasize. Nevertheless, we are encouraged to consider the possibilities of synthetic reproduction for Open Orchestra, at least for specific repertoires, under the assumption that advances in this regard are only a matter of time.

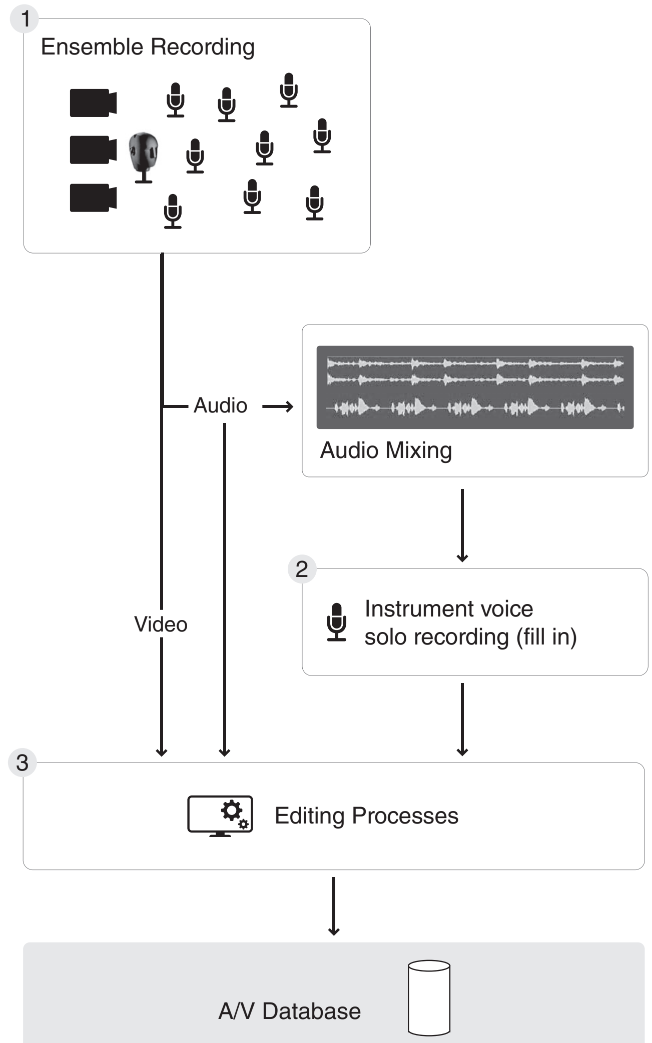
Similar synthesis of a video representation of the performance is, perhaps, as challenging. This might be achieved using computer-generated video avatars of the ensemble members in place of actual musicians, such as is done in the Harmonix games series (e.g., Guitar Hero and Rock Band; www.harmonixmusic.com/). In these games, audio is acquired from a live band and motion capture is used to assist the animation process of the characters (Lord 2009). Other music-driven avatar motion production includes motion and rhythm synchronization for dance (Kim, Park, and Shin 2003) and facial emotion animation generated from music (DiPaola and Arya 2006). These techniques still involve acquisition of human gesture, however, which would, for our purposes, require the involvement of actual musicians.

High-Fidelity Content Acquisition

The second approach, which we have recently used, involves using multiple audiovisual recordings to capture the performance from each spatial perspective associated with the instrument location intended for the simulator. This allows for rendering of an audiovisual experience for the practicing musicians that is reasonably consistent with their perception in a real, physical ensemble. To satisfy the requirements of our three-screen panoramic display, an assembly comprising three Panasonic AK-HC900P cameras, as shown in Figure 5, was positioned appropriately at each seat from which the video perspective was to be captured, with the corresponding musician removed from the recording. For multiple recording positions, this requires several iterations, with the remaining performers repeating the performance for each capture.

All video was acquired at 1280×720 resolution at 60 frames per second (HD 720p format), and recorded at AVC-Intra 100 using three Panasonic AJ-HPM110 recorders, then edited with Final Cut Pro and re-encoded in H.264, at various bitrates,

Figure 6. The recording process for live content acquisition: (1) The ensemble performance is recorded removing one instrument position at a time. (2) Individual musicians or sections of the orchestra re-record their audio in a studio, filling in the position removed from the recordings obtained in the previous step. (3) Performer-specific listening mixes are generated, using spatialized rendering of the individual tracks from step 2 to recreate the appropriate experience from each perspective. The binaural recording from the position of the player is used as a reference in this process.



as required to accommodate the capabilities of heterogeneous clients.

We had initially considered the possibility of recording the audio of each musician (or a set of same instruments) independently to ensure complete isolation of the associated tracks, and then combining these to create the desired mix during playback. This approach would have required that the musicians play along in precise synchronization with one of the previous recordings. Early trials demonstrated the difficulty of this approach, however, resulting in imperfect alignment, which would manifest most visibly in asynchrony between the audio and video tracks. For instance, the visual recording of a percussionist hitting the drum might lead or lag the recorded audio on a number of beats. Such misalignments would likely lead to students deliberately ignoring the video display, thus undermining the value of our technology and depriving students of the intended immersive experience. Instead, we decided to carry out the acquisition of audio and video tracks in parallel, as illustrated in Figure 6.

An Avid Pro Tools HD system was used as the main recording platform, capturing a total of 28 audio tracks at 24-bit, 96-kHz resolution in Broadcast Wave Format. A synchronization signal was provided by a Brainstorm DCD-8 Distripalyzer, which also generated tri-level sync signals for the three HD cameras. Single- or multi-directional microphones were placed for each musical instrument at the most appropriate positions for the instrument's specific musical acoustics. To ensure the highest quality of audio tracks, Grace Design microphone pre-amplifiers and Prism Sound analog-to-digital converters were used for the recording session. In parallel, a binaural recording was made from the position of each target musician (see Figure 5b) to provide a reference of the auditory balance from that perspective. The audio editing and mixing were completed in the Critical Listening room at the Centre for Interdisciplinary Research in Music Media and Technology of McGill University. This room was equipped with a Pro Tools HD system and B&W 802D speakers powered by a Classé Audio power amplifier.

At a later stage, solo audio recordings (of the removed musicians) can be carried out, with each musician listening through headphones to the previous recording of the rest of the orchestra. We found that the musicians were quite capable of playing along in this manner, much as the student musicians will be doing when they use the resulting Open Orchestra system for rehearsal. Acquisition of the solo recordings ensures a pristine audio track, with the signal quality necessary for subsequent analysis and comparison with the student's performance, but not one that is used for rendering the actual playback mix.

Enabling Score-Centric Control

In rehearsal, the music part serves as a dominant focus of attention, motivating its use, in digital format, as the control interface. At present, scores and music parts are mostly available in paper; these must be converted into a digital format suitable for use by our system. For representation of the score and music parts, we chose MusicXML (www.recordare.com/xml.html), a de facto standard for musical score interchange that is semantically more powerful than the older MIDI format.

Figure 7 presents an overview of the process of converting acquired content, including both the music part itself and the performance recording of professional musicians, into a format suitable for the simulation environment of the Open Orchestra system. This involves translation of the sheet into a digital representation, followed by alignment of the music part and its associated audio.

Sheet/Digital Score Translation

The first step is done through optical musical recognition (OMR) software (Choudhury et al. 2000), which scans the paper sheets to extract a machine-readable representation that can be used by our system. We initially used SharpEye (www.visiv.co.uk; Fremerey et al. 2008), a MusicXML-compliant software application, to extract both the notes themselves and their spatial information on the page. In theory, this would allow us to determine what note the students select based solely on the position they "touch" on a digital display of the scanned music part. Although the typical performance of commercial OMR tools allows for reliable detection of bars, however, OMR tools in general remain substantially error-prone (despite significant progress in the last few years), resulting in systematic errors that require subsequent (manual) correction (Byrd and Schindele 2006; Fremerey et al. 2009). Unfortunately, manual correction upsets the spatial relationship between the original music part and its digital representation. As a result, our software must reconstruct the display of the music part in order to maintain a correspondence between on-screen locations and their associated musical elements.

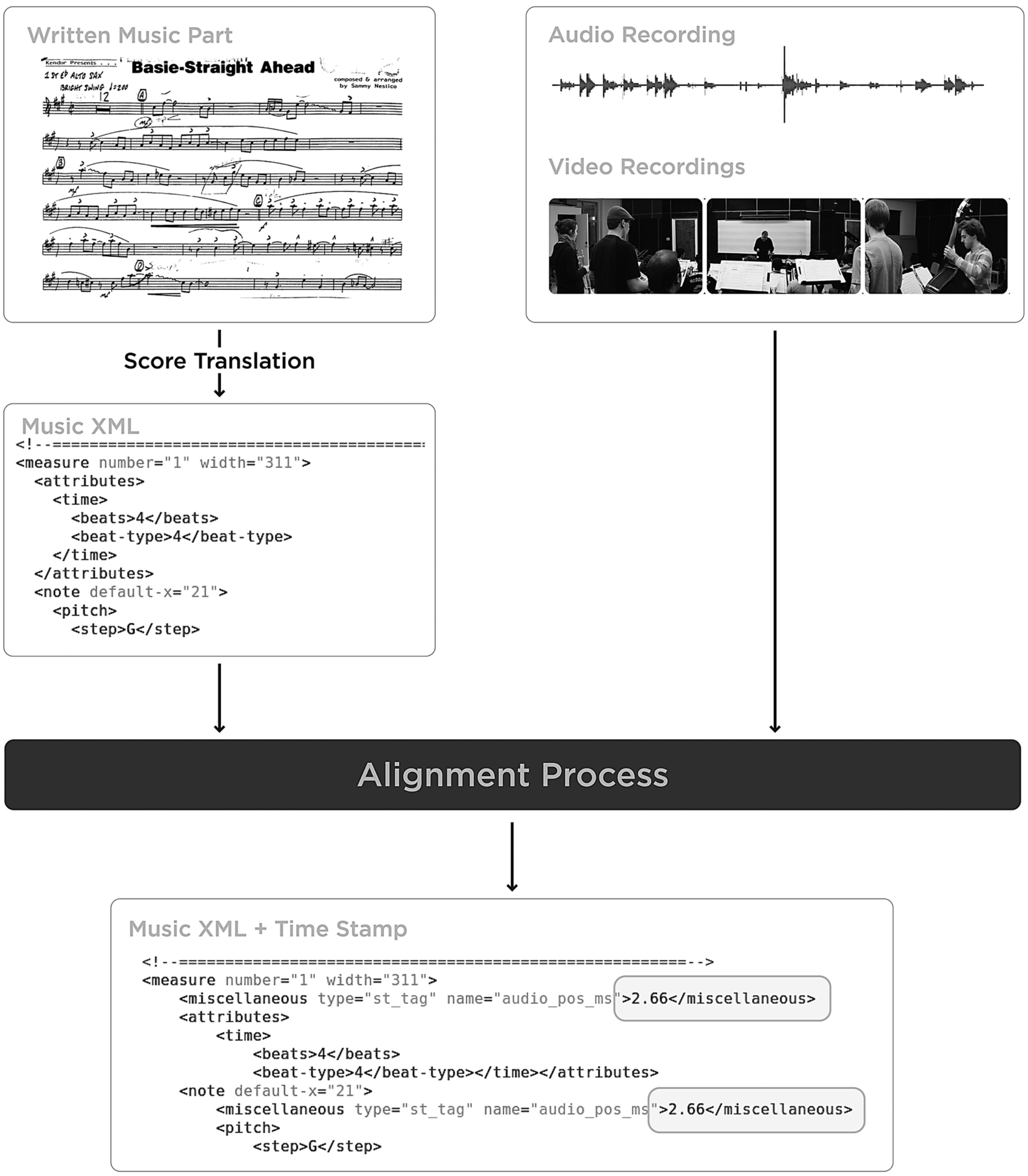
Digital Score/Audio Alignment

During the acquisition of the high-fidelity audio and video that serve as our reference media for the Open Orchestra system, the ensemble plays the music parts, but interpretation leads to possible variations of tempo and/or note placement among different performances. This precludes a straightforward association between the musical elements of the score and their associated timing in the actual recording. Short of recording the piece with a metronome guiding the ensemble (a possibility that was quickly discarded because of its negative impact on interpretation), a computational approach to temporal alignment is required.

As an immediate solution for our initial deployment, we used a manual tagging system to obtain the timestamp for each bar. As a longer-term, automated solution, however, we considered two different approaches. The first, score following (Dannenberg 1984; Vercoe 1984), is used as an online algorithm for real-time alignment, which enables the computer to follow a particular musician, often for accompaniment purposes (Dannenberg and Raphael 2006). Probability models and a training step are commonly used in order to improve the performance of the algorithm, although more-recent systems (Cont 2010) adaptively update their internal state during performance, and thus avoid the need for offline training or parameter tweaking.

The second approach computes the alignment between an audio recording and its score representation after the fact. This has been investigated for several applications, including query-by-humming music retrieval systems and media player synchronization of animations with recorded audio (Kurth et al. 2004). One of the most popular strategies, dynamic time warping (DTW), is based on non-linear warping of two data sequences to find similarities between them. These sequences are composed either of audio samples or extracted features, such as onset, pitch, or chroma (Hu, Dannenberg, and Tzanetakis 2003), as appropriate for the given instrument (Devaney, Mandel, and Ellis 2009). (Chroma corresponds to

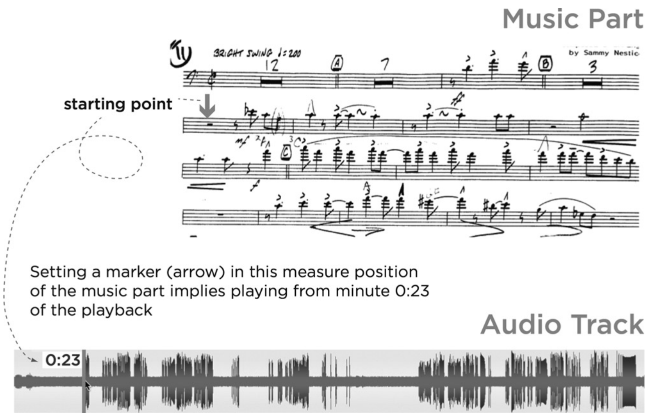
Figure 7. Alignment of the music part with audio (and video) data identifies the associated time instants between media, allowing for simulator playback control from a score-centric user interface.



the musical notes of the equal-tempered scale, and does not consider the register [octave] of a pitch [Shepard 1964].) For instance, wind instruments are better analyzed by chroma, whereas percussion and instruments with percussive attacks having a well-defined energy increase can be aligned using note-onset features (Müller 2007).

When alignment has to be performed on ensemble recordings, however, the challenges posed by the presence of multiple instruments (i.e., polyphony) can render these features irrelevant. In this case, it is preferable to apply DTW to a version of the score that is synthesized into a sequence of audio samples. For the alignment to succeed, the synthesized

Figure 8. *In this example, a trombone player could select a given measure and set a marker (represented by an arrow) to indicate that the desired playback position in the part should start from 0:23.*



audio interpretation must be sufficiently realistic, respecting accurate timing parameters of note onsets and offsets. Notation software, such as Finale and Sibelius, or composition tools, such as OpenMusic (Agon, Stroppa, and Assayag 2000), can be used for this work. Some systems, such as OpenMusic and MaxScore (Didkovsky and Hajdu 2008), support this process in real time, but require manual steps in their use, which thus precludes automatization of the alignment process.

Our implementation of these steps is based on the Vamp plug-in architecture (www.vamp-plugins.org). This software takes advantage of MATCH (Dixon and Widmer 2005), an existing audio-audio alignment tool, to find the optimal alignment between two audio sequences. The algorithm applies a spectral analysis and matches similar increases in energy along frequency bins. Once audio frames are aligned, our system selects interesting instants, computed during initialization, to extract related position in the audio performance file. As output, the software produces a table that translates logical score time, expressed in units of a configurable quarter-note division, into its related timestamp in the audio file, expressed in milliseconds.

Current practical use of the DTW with audio sequences provides reasonably accurate results, but variation occurs according to parameter settings (window and step size) for the spectral analysis. Currently, the synthesized audio is generated using external software, but we hope to automate this step in the near future. This requires further investigation into the tradeoff between the quality of the generated audio file (e.g., with respect to its sound bank or the tempo) and the resulting alignment accuracy. Our initial tests indicate that alignment suffers from an under-specified part in the score—for instance, in improvised solo sections, and also in some rhythm-only sections where piano chords are cued without specification of rhythm or chord inversion. The latter results in simplistic audio synthesis, which causes significant degradation in alignment accuracy. We are also considering the integration of other strategies to cope with these problematic situations, thereby providing a more accurate alignment throughout the score. Although this alignment process results in an average temporal accuracy of one quarter note, this is sufficient to ensure that a selected position on the music part always corresponds to the desired measure. This technique can thus be used in the future to align the music piece with the actual performance.

Synthesis of the Digital Score

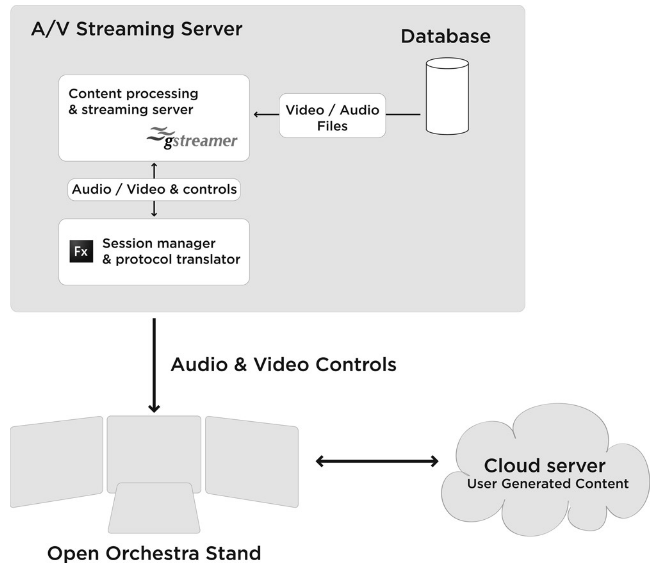
The digital score can then be synthesized, building on the extracted musical elements from the OMR step and the temporal associations computed with the recorded audio. The corresponding graphics contents of the score are rendered from the MusicXML file, and then arranged based on a set of heuristics encoded in the software. As intended, this allows for playback control using a given music part as an interface, as shown in Figure 8. For example, the spatial position within the music part serves as an index into the recorded audio file to select a section for looped playback.

Software Architecture

The Open Orchestra system deals with two main types of data, each best suited to a different approach for storage and serving:

1. Large, persistent data: high-definition audio and video files that are provided to the user for immersion within the simulated orchestra.
2. Dynamic data: users' own recordings and annotations, comments, and feedback provided by instructors.

Figure 9. Audiovisual server architecture allowing for real-time control of audio parameters



We designed our architecture around the premise of heterogeneous client machines, especially anticipating potential access not only from music schools, but also from students' home machines, albeit with a (likely) less-immersive rendering capability. This mandated that client storage and processing requirements be kept minimal, and in turn, any demanding computation, for example, audio spatialization, be performed on the server side, streaming the requested content to clients.

For the large, immutable data of audio and video recordings of the orchestra, a dedicated physical server appears to be the most cost-effective solution. This machine must provide processing capacity for audio perspective rendering, connected to clients over a high-speed data network for real-time media streaming. For dynamic, user-generated content, however, including student recordings and instructor feedback, the requirements will vary significantly depending on student activity; storage duration is shorter term; and real-time serving capability is unnecessary. In addition, this data only needs to be shared between the student and a few individuals, such as instructors and peers. As such, cloud storage and processing may be more appropriate, as this route offers more flexible scaling options as required for an expanded client base.

Integration and synchronization of user-generated content stored in the cloud is being investigated. Approaches may be client-oriented, in which the ensemble rendering will be provided by the server, with user tracks overlaid locally, or server-driven, in which the server simply includes the user-generated content as a regular audio channel when synthesizing the playback (e.g., for the instructor to evaluate). In either case, synchronization between the orchestral recordings and the user-generated ones requires attention to any delays induced by the local recording process.

Accessing the Orchestral Simulator

Operation of the system is illustrated in Figure 9. The user specifies parameters of the requested streams through the score-centric interface. A table lookup obtains the corresponding timestamp that indexes into the recorded audio for the requested instruments. User commands are then conveyed to a server-side audio/video sequencer, which selects the appropriate media contents and performs multichannel volume control and mixing in real time. The server, which hosts the pre-recorded media, consists of a custom streaming engine and the session managers, described subsequently. Because the musician hears the rest of the band through headphones, results are transmitted live to the client as a stereo audio stream synchronized with a single video channel (Bouillot, Tomiyoshi, and Cooperstock 2011), typically a "stitched together," triple-width video that is split at the client for display over three screens. It is equally possible to render the experience through a surround-sound system, but this risks audio feedback during recording.

Our custom streaming engine is dedicated to content processing and stream serving. Its features include play/pause/seek controls in addition to volume control over individual channels of audio. Internally, it handles simultaneous reading of multiple files, mixing of the various audio channels into a stereo stream, synchronization of audio with video, AAC encoding of audio at 320 kbps, stream marshaling, and transmission. The engine is implemented as a pipeline of components, building on the GStreamer open-source multimedia framework (Kost 2010). To maintain synchronization, the system clock is used to slave all elements in the pipeline, including file reading and marshaling of network protocols. GStreamer derives several times from the system clock and the playback state (e.g., the running time of the pipeline and the current position in the media). This "stream time" is handled by the data payloader that determines header timestamps, allowing the marshaling element to specify synchronization information between the audio and video streams. The protocol translator receives the streams, decodes synchronization information, and converts it into the format handled by clients (e.g., Adobe RTMP), maintaining inter-channel synchronization for rendering. Interaction between these server components is described in detail in a related publication (Bouillot, Tomiyoshi, and Cooperstock 2011).

The session manager is responsible for translating streaming protocols and maintaining sessions with multiple, possibly simultaneous, clients. Accordingly, each client is assigned a dedicated session, allowing for content serving and control forwarding to the streaming engine without affecting other client sessions. Our Java implementation of this component extends the default functionality of the Wowza Media Server 2 (www.wowzamedia.com).

We have not yet evaluated control response time, a critical factor to ensure an effective experience with the system, especially for continuous control such as volume adjustment. Fortunately, the latency characteristics of each system component are generally controllable, but tuning of buffer and audio frame sizes is anticipated to balance between reliability and delay.

Automated Performance Feedback

The automated performance tools are intended to highlight relevant elements of ensemble performance where the student needs to improve, hopefully enhancing the student's understanding of their part in the orchestra. This might be achieved through a combination of comparative audio playback and data visualization approaches; the latter can exaggerate nuances as well as tell the student what to listen for and where.

The display is based on automated comparison between the recording of the student musician's performance and a reference "expert" audio track (see Figures 3 and 4), which may be a recorded example of the same part or of other instruments from the same section or ensemble. The former has the advantage of being the exact same musical content as what the student is intended to play, and serves well as a reference to evaluate the student's fit with the rest of the ensemble. There are benefits as well, however, to visualizing performance in relation to different instruments, as, for example, in seeing one's entrance in relation to the timing of other instruments.

An important question for investigation is the choice of which features to analyze. Section- and ensemble-relative dynamic levels, timing, and intonation are likely candidates because they all indicate whether the instrument is playing cleanly within the ensemble. In contrast, other features, such as instrumental tuning, fingering, placement, and duration of notes relative to the score, articulation, and timbre, are more indicators of technical instrumental mastery. A ranking of the importance for feedback purposes of these and other possible features can be obtained through an experiment or survey of musical professionals, such as was done with the IMUTUS system (Schoonderwaldt, Askenfelt, and Hansen 2005). Similarly, the algorithms used to extract musical features, find dissimilarities with the reference track, and determine the importance of the different musical features will need to be validated by comparison to the values assigned by an independent human expert.

The analysis and comparison of musical features will no doubt generate considerable data, for which both judicious selection and due attention to the limitations of information visualization will be critical for usability. Categorizing and selectively displaying feedback will prevent overwhelming the user. Because a written music part is a familiar representation of both the time progression and content of the piece of music for a given instrument, we intend to base as much of the feedback visualization and navigation around this representation as can be supported effectively. Such feedback will likely take the form of highlighting, annotations, or glyphs. The rich literature on visualization techniques provides ample guidelines (Hiraga and Matsuda 2004; Daudin et al. 2007; Robine, Percival, and Lagrange 2007; Aggelos, Kostas, and Stefanos 2008) for display of parametric data, but given the variety of representational choices available, the actual appearance of the visualizations will have to be thoroughly investigated. Early testing with musicians has indicated that the use of standard notation, when possible, helps convey clear and contextualized meaning. For example, an easy contrast can be made between the dynamic markings on the music part and dynamic markings generated from the students' playing. Some features, however, such as timbre, lack such easy notation methods. Similarly, describing subtle variations in timing may require a different method from standard notation. Moreover, different features will require visualization at different time-scales: from the note- or measure-level to the macro-level, showing the entire piece.

Preliminary Evaluation

The Open Orchestra system has so far been evaluated with respect to audio perspective and the performance visualization. For the former, we conducted a preliminary evaluation of the respective benefits of different audio rendering perspectives, one from the musician's perspective, and another from that of the audience, the latter being similar to Music Minus One. This involved an experiment with eight jazz musicians, taking on their respective orchestral positions of lead trumpet, lead alto sax, lead trombone, and drums. Although the results indicated a slight preference for the audio experience rendered from the musician's perspective, this difference was significant only for the trumpet, suggesting that the value of a dedicated audio image is instrument-dependent. Our ongoing work is examining the customization of these audio parameters for a given musician based on recommendations from a mentor or conductor. Further details of this study are available in a related publication (Olmos et al. 2011). The ability of Open Orchestra to tailor the learning environment in this manner may prove valuable to allow musicians to practice and improve their skills in a pedagogically optimal orchestral context.

Our study of automatic performance visualization (Knight, Bouillot, and Cooperstock 2012) considered support for musical interpretation. This involved subjective judgment of a student's performance compared to reference "expert" performance for particular aspects of musical performance: articulation and dynamics. The experiment presented the two samples by either audio only, visualization only, or both together. Assessment of the effectiveness of the feedback condition was based on the consistency of judgments made by the participants (all music students) regarding how well the student musician matches the reference musician, the time taken to evaluate each pair of samples, and the subjective opinion about the feedback's perceived utility.

For articulation, differences in the mean scores assigned by the participants to the reference-versus-student performance were not statistically significant for each modality. This suggests that, whereas the visualization strategy did not offer any advantage over presentation of the samples by audio playback alone, visualization nevertheless provided sufficient information to make similar ratings. For dynamics, four of our six participants categorized the visualizations as helpful. The means of their ratings for the visualization-only and both-together conditions were not statistically different from each other, but were statistically different from the audio-only treatment, indicating a dominance of the visualizations when presented together with audio. Moreover, the ratings of dynamics under the visualization-only condition were significantly more consistent than under the other conditions.

Conclusions

Open Orchestra was motivated by the desire to provide an immersive, computer-assisted learning environment that enables an individual musician to practice with a simulated ensemble. In approaching the task of its design, many existing technologies were investigated, in particular, computer-assisted pedagogical musical systems that focus on improvement of instrument skills. These systems do not address the needs of ensemble listening or interaction with other instrumentalists, however. Conversely, present-day ensemble rehearsal systems of the Music Minus One format suffer from critical drawbacks, including an absence of visual cues, limited navigation possibilities, and lack of ensemble sound control. The development of our high-fidelity orchestral simulator required us to solve many challenges that had not been addressed previously.

Access to both the dynamically rendered Open Orchestra repertoire and to user-generated data requires a network-enabled platform. Our approach is based on the integration of several existing networking technologies, including cloud computing and several widely used streaming protocols such as RTMP and RTP. Our resulting design is centered around a digital music-stand interface that controls, among other features, the playback of stereo audio and panoramic video from the musician's position within the orchestra. This interface also allows volume adjustment of each instrument section, and annotation of the score. The initial repertoire made available through the Open Orchestra system is being acquired from high-fidelity audio and video sequences of semi-professional musicians, to ensure that nuances of interpretation are preserved. The process of extracting a digital representation of the score, aligning this with the audio and video recordings, and rendering the results in a manner that reproduces the ensemble experience has been described, along with the associated challenges for each of these steps.

Open Orchestra enables rehearsal in a simulated ensemble with the freedom to control and tune the playback. One of the most important aspects of the learning experience relates to feedback, however. This is provided in two forms. First, students gain direct access to automated analysis tools that visualize several audio features. The selection of appropriate features to analyze is currently under investigation, with our emphasis on those that relate to ensemble integration rather than instrument playing skills. Second, students can request external feedback from the conductor. Supporting the latter requires that Open Orchestra serve as a collaborative tool, achieved by offering students the ability to record their performances and annotate the written music part for later review. Despite the unusual asynchronous nature of the resulting interaction between musicians and conductor, this offers several benefits. Notably, the ability to obtain the conductor's comments directly on recordings of a student's rehearsals offers the possibility of individualized, private discussions of specific relevance to each musician. Furthermore, the Open Orchestra system may facilitate the task of the conductor, allowing for a rapid focus on specific elements that require attention, since the student's recording can be played back either in isolation, or in conjunction with any or all of the master tracks from the professional recording.

Although the current prototype is functional, we are, at present, investigating issues of control latency of distributed real-time media delivery, obviously a critical factor to the user experience. Considerable work remains to evaluate the prototype system, from both the students' and the conductor's perspectives, and make refinements to its interface, music-analysis strategies, and visualization techniques. In addition, our system presently has no means of coping with improvised parts, as are commonly found in jazz performance; these, naturally, will not align precisely with reference samples from previous recordings. This raises obvious challenges for future development.

Overall, simplicity and system reliability remain our principal goals, as the distraction of technical sophistication should not, and cannot, impede musical performance, which requires a great deal of attention and concentration on the part of the performer. We look forward to evaluating the success of the Open Orchestra system in achieving these objectives in actual testing with music students.

Acknowledgments

The authors would like to thank Canada's Advanced Research and Innovation Network (CANARIE) for funding of this work, conducted under a Network Enabled Platforms (NEP-2) program research contract. Many of the design decisions described here were informed by our highly valuable sessions with Gordon Foote, Director of McGill University's Jazz Band I, and his students. The overall project described here is developed in collaboration with colleagues at the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) at McGill University. Specifically, the authors are most indebted to the efforts of John Roston, Wieslaw Woszczyk, Doyuen Ko, and Jake Woszczyk, as well as the CIRMMT technical staff.

References

Aggelos, G., P. Kostas, and N. Stefanos. 2008. "Real-Time Detection and Visualization of Clarinet Bad Sounds." *Proceedings of the 11th International Conference on Digital Audio Effects*, pp. 59–62.

Agon, C., M. Stroppa, and G. Assayag. 2000. "High Level Musical Control of Sound Synthesis in OpenMusic." *Proceedings of the International Computer Music Conference*, pp. 332–335.

Borchers, J., E. Lee, and M. Mühlhäuser. 2004. "Personal Orchestra: A Real-Time Audio/Video System for Interactive Conducting." *Multimedia Systems* 9(5):465–468.

Bouillot, N., M. Tomiyoshi, and J. R. Cooperstock. 2011. "Extended User Control over Multichannel Content Delivered over the Web." *Proceedings of the 44th Conference on Audio Networking* (pages unnumbered). Available online at www.aes.org/e-lib/browse.cfm?elib=16148.

Byrd, D., and M. Schindele. 2006. "Prospects for Improving Optical Music Recognition with Multiple Recognizers." *Proceedings of the 7th International Conference on Music Information Retrieval*, pp. 41–46.

Cannam, C., et al. 2006. "The Sonic Visualiser: A Visualisation Platform for Semantic Descriptors from Musical Signals." *Proceedings of the 7th International Conference on Music Information Retrieval*, pp. 324–327.

Carôt, A., P. Rebelo, and A. Renaud. 2007. "Networked Music Performance: State of the Art." *Proceedings of the 30th International Conference of the Audio Engineering Society* (pages unnumbered). Available online at www.aes.org/e-lib/browse.cfm?elib=13914.

Choudhury, G. S., et al. 2000. "Optical Music Recognition System within a Large-Scale Digitization Project." *Proceedings of the International Society for Music Information Retrieval* (pages unnumbered). Available online at ciir.cs.umass.edu/music2000/papers/choudhury_paper.pdf.

Cont, A. 2010. "A Coupled Duration-Focused Architecture for Realtime Music to Score Alignment." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32-6:974–987.

Dannenberg, R. B. 1984. "An On-Line Algorithm for Real-Time Accompaniment." *Proceedings of the International Computer Music Conference*, pp. 193–198.

Dannenberg, R. B., et al. 2003. "Results from the Piano Tutor Project." *Proceedings of the Fourth Biennial Arts and Technology Symposium*, pp. 143–150.

Dannenberg, R. B., and C. Raphael. 2006. "Music Score Alignment and Computer Accompaniment." *Communications of the ACM* 49(8):38–43.

Daudin, C., et al. 2007. "Visualisation du Jeu Instrumental." *Actes des Journées d'Informatique Musicale*, pp. 64–72.

Devaney, J., M. I. Mandel, and D. P. W. Ellis. 2009. "Improving MIDI-Audio Alignment with Acoustic Features." *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 45–48.

Didkovsky, N., and G. Hajdu. 2008. "MaxScore Music Notation in Max/MSP." *Proceedings of the International Computer Music Conference* (pages unnumbered). Available online at quod.lib.umich.edu/cgi/p/pod/dod-idx?c=icmc;idno=bbp2372.2008.034.

DiPaola, S., and A. Arya. 2006. "Emotional Remapping of Music to Facial Animation." *SIGGRAPH Sandbox Symposium on Videogames*, pp. 143–149.

Dixon, S., and G. Widmer. 2005. "MATCH: A Music Alignment Tool Chest." *Proceedings of the 6th International Conference on Music Information Retrieval*, pp. 492–497.

Fober, D., S. Letz, and Y. Orlarey. 2007. "VEMUS—Feedback and Groupware Technologies for Music Instrument Learning." *Proceedings of the Sound and Music Computing Conference*, pp. 1–8.

Foote, G. C. 1979. "A Rhythmic Analysis of Swing Music From the Count Basie Library." Master's thesis, School of Music, University of Minnesota.

Fremerey, C., et al. 2008. "Automatic Mapping of Scanned Sheet Music to Audio Recordings." *Proceedings of the 9th International Conference on Music Information Retrieval*, pp. 413–418.

Fremerey, C., et al. 2009. "Sheet Music-Audio Identification." *Proceedings of the International Conference on Music Information Retrieval*, pp. 645–650.

Friberg, A. 2006. "pDM: An Expressive Sequencer with Real-Time Control of the KTH Music-Performance Rules." *Computer Music Journal* 30(1):37–48.

Geslin, Y., and A. Lefevre. 2004. "Sound and Musical Representation: The Acousmographe Software." *Proceedings of the International Computer Music Conference* (pages unnumbered). Available online at quod.lib.umich.edu/cgi/p/pod/dod-idx?c=icmc;idno=bbp2372.2004.138.

Goldmark, P. C. 1976. "Performance Learning System. US Patent 3,955,466." Available online at www.patentlens.net/patentlens/patent/US_3955466/en.

Hiraga, R., and N. Matsuda. 2004. "Visualization of Music Performance as an Aid to Listener's Comprehension." *Proceedings of the Working Conference on Advanced Visual Interfaces*, pp. 103–106.

Hu, N., R. B. Dannenberg, and G. Tzanetakis. 2003. "Polyphonic Audio Matching and Alignment for Music Retrieval." *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 185–188.

Kim, T.-H., S. I. Park, and S. Y. Shin. 2003. "Rhythmic-Motion Synthesis Based on Motion-Beat Analysis." *SIGGRAPH*, pp. 392–401.

Knight, T., N. Bouillot, and J. R. Cooperstock. 2012. "Visualization Feedback for Musical Ensemble Practice: A Case Study on Phrase Articulation and Dynamics." *Proceedings of the Conference on Visualization and Data Analysis* (pages unnumbered).

Koivunen, N. 2002. "Organizational Music, the Role of Listening in Interaction Processes." *Consumption, Markets and Culture* 5(1):99–106.

Kost, S. 2010. "Writing Audio Applications using GStreamer." *Proceedings of the Linux Audio Conference*, pp. 15–20.

Kurth, F., et al. 2004. "Syncplayer: An Advanced System for Multimodal Music Access." *Proceedings of the International Conference on Music Information Retrieval*, pp. 381–388.

Lindemann, E. 2007. "Music Synthesis with Reconstructive Phrase Modeling." *Signal Processing Magazine, IEEE* 24(2):80–91.

Lord, A. 2009. "Harmonix 'Rock Band II'." *SIGGRAPH Computer Animation Festival*, pp. 62–63.

Mazzola, G., and S. Göller. 2002. "Performance and Interpretation." *Journal of New Music Research* 31(3):221–232.

McAngus Todd, N. P. 1992. "The Dynamics of Dynamics: A Model of Musical Expression." *Journal of the Acoustical Society of America* 91(6):3540.

Müller, M. 2007. *Information Retrieval for Music and Motion.* Secaucus, New Jersey: Springer-Verlag.

Nakra, T., et al. 2009. "The UBS Virtual Maestro: An Interactive Conducting System." *Proceedings of the International Conference of New Interfaces for Musical Expression*, pp. 250–255. Available online at www.cs.illinois.edu/~paris/pubs/nakra-nime09.pdf.

Ng, K. 2008. "An Overview of the i-Maestro Project and an Introduction to the 4th i-Maestro Workshop." *Proceedings of the 4th i-Maestro Workshop on Technology Enhanced Music Education* (pages unnumbered). Available online at www.i-maestro.org/workshop.

Olmos, A., et al. 2011. "Where Do You Want Your Ears? Comparing Performance Quality as a Function of Listening Position in a Virtual Jazz Band." *Proceedings of the 8th Sound and Music Computing Conference* (pages unnumbered). Available online at www.padovauniversitypress.it/proceedings-of-the-smc-2011-8th-sound-and-music-computing-conference.

Percival, G., Y. Wang, and G. Tzanetakis. 2007. "Effective Use of Multimedia for Computer-Assisted Musical Instrument Tutoring." *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education*, pp. 67–76.

Progler, J. A. 1995. "Searching for Swing: Participatory Discrepancies in the Jazz Rhythm Section." *Ethnomusicology* 39(1):21–54.

Psotka, J. 1995. "Immersive Training Systems: Virtual Reality and Education and Training." *Instructional Science* 23(5):405–431.

Raphael, C. 2006. "Demonstration of Music Plus One: A Real-Time System for Automatic Orchestral Accompaniment." *Proceedings of the 21st National Conference on Artificial Intelligence*, pp. 1951–1952.

Reidsma, D., A. Nijholt, and P. Bos. 2008. "Temporal Interaction between an Artificial Orchestra Conductor and Human Musicians." *Computers in Entertainment* 6(4):1–22.

Robine, M., G. Percival, and M. Lagrange, 2007. "Analysis of Saxophone Performance for Computer-Assisted Tutoring." *Proceedings of the International Computer Music Conference*, vol. 2, pp. 381–384.

Schoonderwaldt, E., A. Askenfelt, and K. F. Hansen. 2005. "Design and Implementation of Automatic Evaluation of Recorder Performance in IMUTUS." *Proceedings of the International Computer Music Conference*, pp. 97–103.

Schoonderwaldt, E., K. Hansen, and A. Askenfeld. 2004. "IMUTUS—An Interactive System for Learning to Play a Musical Instrument." *Proceedings of the International Conference of Interactive Computer Aided Learning*, pp. 143–150.

Shepard, R. N. 1964. "Circularity in Judgments of Relative Pitch." *Journal of the Acoustical Society of America* 36(12):2346–2353.
Shober, M. F. 2006. "Virtual Environment for Creative Work in Collaborative Music-Making." *Virtual Reality* 10(2):85–94.
Vercoe, B. 1984. "The Synthetic Performer in the Context of Live Performance." *Proceedings of the International Computer Music Conference*, pp. 199–200.
Widmer, G. 2002. "Machine Discoveries: A Few Simple, Robust Local Expression Principles." *Journal of New Music Research* 31(1):37–50.
Widmer, G., and W. Goebl. 2004. "Computational Models of Expressive Music Performance: The State of the Art." *Journal of New Music Research* 33(3):203–216.
Winget, M. A. 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):1878–1897.
Woszczyk, W., et al. 2005. "Shake, Rattle, and Roll: Getting Immersed in Multisensory, Interactive Music via Broadband Networks." *Journal of the Audio Engineering Society* 53(4):336–344.
Yin, J., Y. Wang, and D. Hsu. 2005. "Digital Violin Tutor: An Integrated System for Beginning Violin Learners." *Proceedings of the 13th Annual ACM International Conference on Multimedia*, pp. 976–985.
Zimmermann, R., et al. 2008. "Distributed Musical Performances: Architecture and Stream Management." *ACM Transactions on Multimedia Computing Communications and Applications* 4(2):1–23.