Network-Enabled Platforms (NEP-2) Program
Progress Report – Feb 28, 2010.

Project NEP54: Open Orchestra
Appendix 5: Preliminary Software Specifications

Introduction

This section consists of two parts, the first pertaining to the score integration, i.e., offline processing of a sheet of musical score and alignment of the score with an audiovisual recording made of the student's performance. The second part describes the online system providing the student with the immersive experience of rehearsing with the conductor and ensemble.

Note that according to discussions during the January User's Group meeting, as well as observations subsequently made by Gordon Foote, we are proceeding with the design objective of a sheet-score-centric user interface, as embodied in a digital musical stand. Necessary features we consider are as follows:

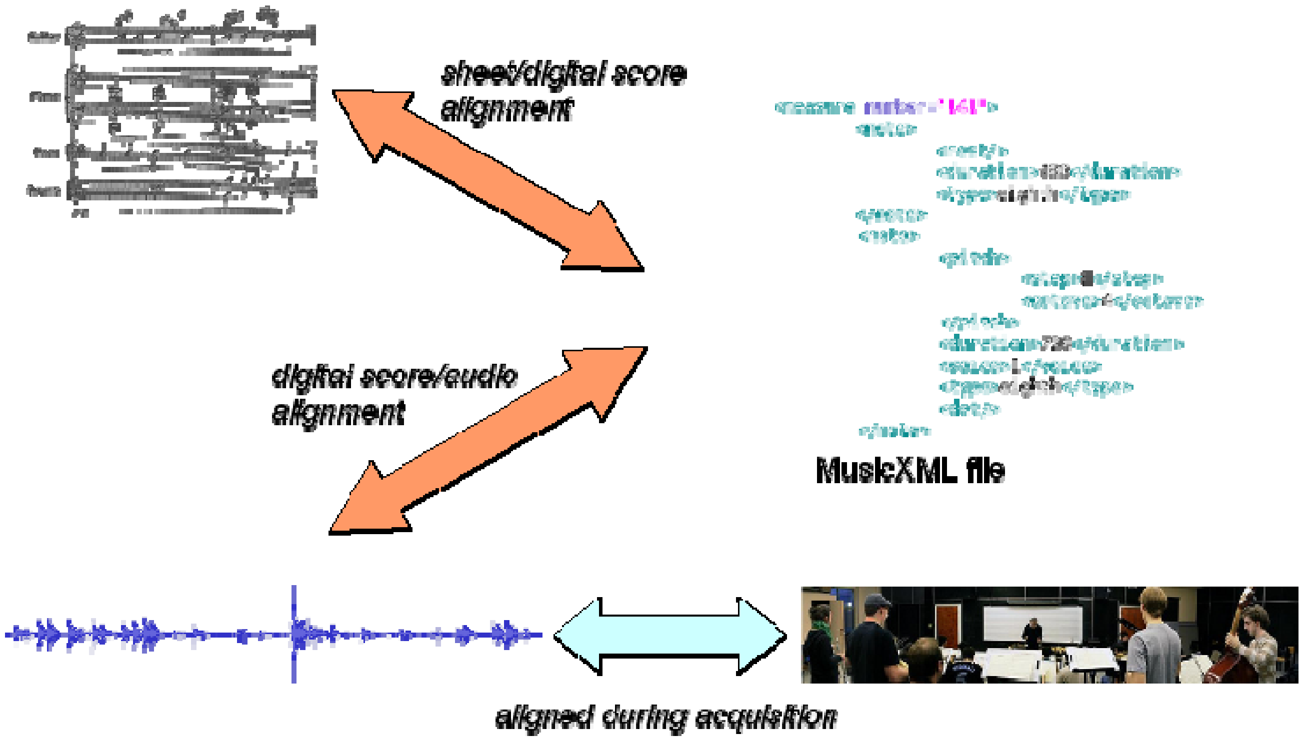
- playback control in a sequencer-like fashion, in which the raw signal waveform is replaced by staffs ; this requires functions of playhead control, loop selection, volume, solo, and mute
- score annotation, preferably by stylus input
- automated feedback visualization

Score Integration

The process of integrating a musical score into our system, for later display on the digital stand, proceeds as follows:

- optical scanning of the scores actually used during rehearsal
- converting the scanned score into a structured description
- aligning data

The following figure summarizes the task of aligning scanned music sheet with audio/video clips. This is divided into two problems: 1) sheet/score alignment: mapping the representation of the score (i.e., which part, staff, measure and note) and its corresponding location on the page, and 2) score/audio alignment: alignment of the representation with the recorded audio stream.



Sheet/digital score alignment. Optical Music Recognition (OMR) converts sheet music to score memory format. Existing commercial tools demonstrate adequate performance for reliable detection of bars, but the issues of recognizing hand-written glyphs remains more challenging. As output format, musicXML (http://www.recordare.com/xml.html), compared to the older MIDI format, offers the advantage greater semantic power, and is more easily integrated into our software architecture.

While OMR is commercially available in several products, the inverse operation, i.e., mapping the digital representation of the notes, bars and staffs to their corresponding positions in the printed score, is less frequent. This function will likely be highly desirable for the analysis purposes of our system. To the best of our knowledge, SharpEye (http://www.visiv.co.uk/) is the only OMR software providing such spatial information.

Digital score/audio alignment. Accurately identifying a particular note in the score from its audio signal is an active research problem in the Music Information Retrieval (MIR) community.

The first algorithm family is based on converting the score representation to audio and aligning the result with the actual recorded audio using the similarity matrix strategy. Dixon's Musical Alignment Tool Chest (MATCH) (http://www.elec.qmul.ac.uk/people/simond/match/) is freely available software that provides such audio/audio alignment, either as a vamp plugin or standalone. However, this family of algorithms suffers a known limitation of alignment accuracy, which may deviate from the level of individual notes of the piece is played quickly.

The second family is based on comparison of extracted onsets and pitch information from the audio with directly available equivalent information from the score. However, this strategy is not accurate if the piece exhibits variations in pitch and rhythm. No available software is yet known for this family.

Should the MATCH software prove insufficient in accuracy to satisfy our requirements, future development may include the design and implementation of a new or existing strategy.

Online System for Student Experience

The following material describes the components of the online system for student experience. A high-level architectural diagram appears below.

Score Tagging and Online Alignment Management

Alignment information will be stored as tags in the musicXML file, allowing direct access to specific portions of data as required by the student. The score-tagging feature may also be used for recording user-specific data as annotations, although further investigation is required this possibility. The “scoretag” software, currently under development, allows for writing and reading tag(s) during navigation in a score memory representation. The navigation allows for reading at variable speed, as required for alignment with audio recording, as well as arbitrary steps through the score.

Spatial Audio Rendering

In order to consider various strategies for audio rendering, we will investigate both offline processing and real-time rendering. While the former may offer the maximum simulation accuracy, real-time rendering provides greater flexibility and adapts well to user-specific configurations. Headset listening is highly recommended to ensure a well-defined auditory environment and adequate sound isolation, although speakers likely provide a greater level comfort to the student.

A first prototype has been developed using the Pure Data (PD) environment with the nSlam library. This prototype provides real-time audio spatialization of multiple sound sources, limited only by processor capacity, and allows for source positioning (either manually or through external software control) and multiple output formats (stereo/HRTF, quad-channel, 5.1, 8.1, etc.). Future developments will include source-position presets and possibly echo concealment.

Analysis

A fundamental aspect of signal processing for the Open Orchestra project is to perform automated analysis of the students' performance. These tools will be integral for self-directed student learning as well as instructor-led lessons. For the individual student, the system should be capable of identifying potential trouble spots in one performance but also identifying chronic problems and trends over time. Instructors will also be able to view individual student reports. However, the ability of the system to collate and summarize an analysis of an entire class or ensemble will be crucial to the scalability and usability of the system for an instructor.

Work in this area has focused on identifying parameters of the students' performance that are musically meaningful and helpful for learning as well as methods and tools. As outlined above, the system will maintain an XML representation of the score in addition to the raw audio of the students' performance, as well as a reference audio file for the part the student is supposed to be playing.

Depending on the different parameters to be measured, the students' tracks may be compared against one or the other. A preliminary list of musical features for investigation includes those related to:

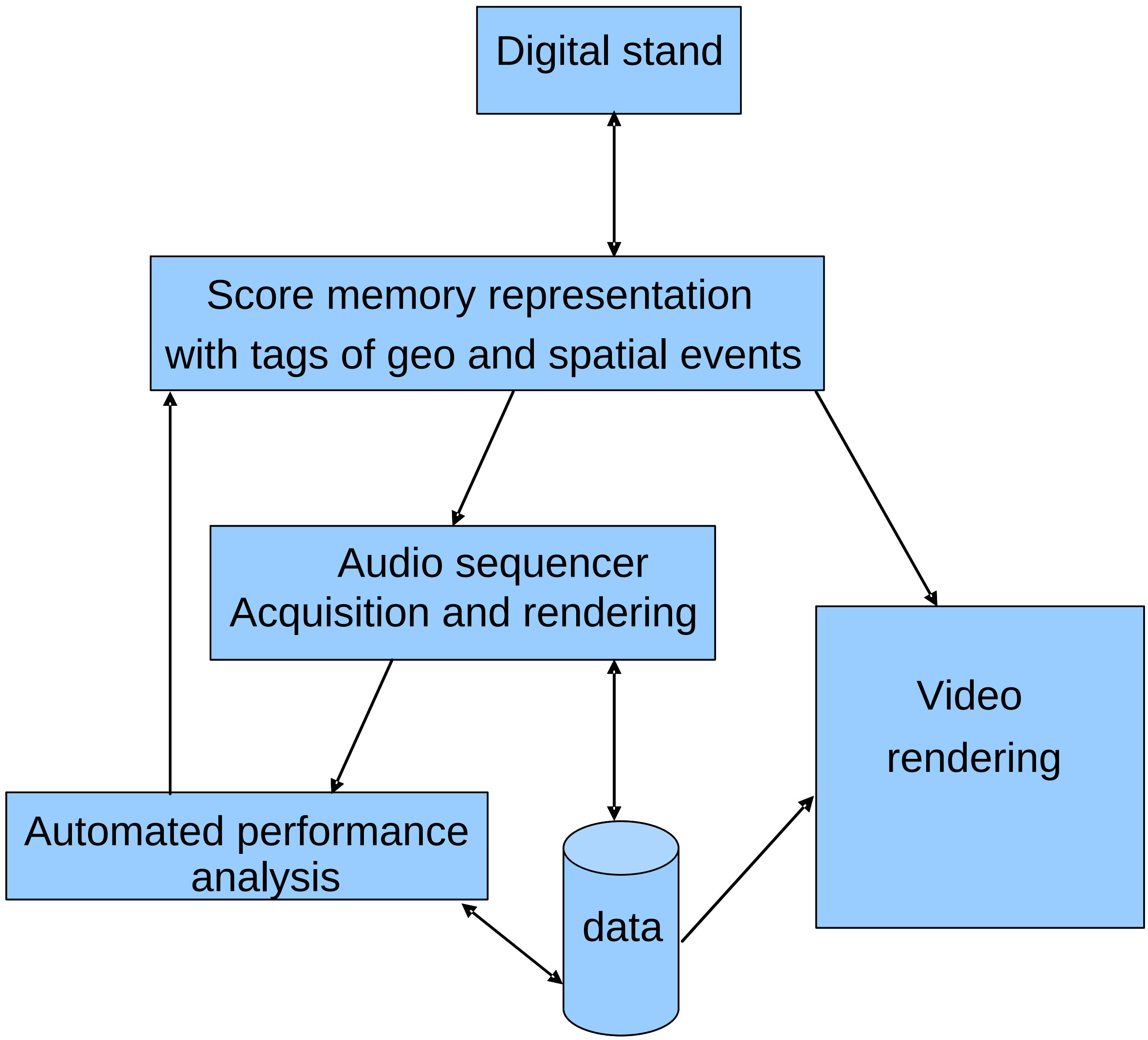
1. timing of events: tempo, rhythm, note length
2. frequency or pitch: tuning, correct note, intonation, further expression such as vibrato
3. volume: dynamics and group balance

The VAMP plugin architecture (http://vamp-plugins.org/) is a well-suited tool to support implementation of the Open Orchestra system. These plugins can be implemented on both Windows and Linux platforms and can be easily called and controlled by the Open Orchestra software in whatever form it takes. Additionally, several institutions have open-source suites of VAMP plugins for audio analysis, which include parameters that will be examined in this project.

Pitch analysis of a single monophonic instrument by frequency content is a straightforward operation and is well integrated into several VAMP plugins. The detection of onsets and offsets of notes is a more complex operation both on the analysis side and making use of the output. The VAMP plugins featuring onset detection combine several techniques -- transient detection, peak detection, pitch changes and silence detection. These work together to form robust onset detection as any one method by itself is insufficient. For example, the second note of a pair of slurred notes requires pitch detection. These tools can be tweaked by adjusting their function, sensitivity, window size, increment and shape. Since the audio from Open Orchestra has relatively narrow parameters -- monophonic with a known instrument, it should be possible to optimize the onset detection for the application and vary the parameters depending on the instrument or instrument family.

The last category of analysis -- balance and dynamics -- is best considered by looking at the envelopes of the waveforms. Comparing points within a student's performance and comparing that range to the reference track determines the dynamic accuracy. The student's balance can be determined by comparison of the waveform envelopes against those of the ensemble. This comparison, however, may be dependent on room acoustics, playback volume and other factors; further investigation is required to determine if such feedback can be considered useful.

The collection of analyses, as described above, necessarily results in a large volume of data. A particular challenge of this part of the project will be to determine the sections or points of interest and visualize their associated feedback in a pedagogically effective manner. Comparing any of the parameters to the score or reference audio provides a numeric value of the accuracy of the student’s playing. Any of the resulting output, however, representing, for example, timing, pitch and dynamic accuracy, will vary continuously over the duration of the piece. While waveforms and chromagrams are common methods used for audio visualization, they require a trained eye to read and analyze. It will thus be a significant visual design challenge to represent relevant content most effectively in order to contrast the student’s performance with that of an “ideal” player.


Digital stand
Score memory representation
with tags of geo and spatial events
Audio sequencer
Acquisition and rendering
Video
rendering
Automated performance
analysis
data