Tectonic: a networked, generative and interactive, conducting environment for iPad

Lindsay R. Vickery Dr
*Edith Cowan University, lvickery@ecu.edu.au*

Stuart James
*Edith Cowan University, s.james@ecu.edu.au*
In this implementation we use an autocorrelation pitch follower implemented in SuperCollider\texttt{Pitch\_Ugen}. As noted by Roads, autocorrelation is most efficient at mid and low frequencies. Thus it has been popular in speech recognition applications where the pitch range is limited.\cite{Roads2009}

Working with vocal material of a relatively short length, autocorrelation was able to resolve the pitch content of the singer.

The first means of providing immediate feedback is the generation of a real-time transcriptions. To accomplish this the \texttt{bach library}\cite{Roads2008} in \texttt{MaxMSP} environment is employed. Using \texttt{OpenSoundControl} the midi note value of the detected pitch is sent from \texttt{SuperCollider} to the \texttt{MaxMSP} environment where the \texttt{bach\_transcribe} object is utilized to format the incoming information and present it via a \texttt{bach\_roll}. This immediate presentation enables the composer to quickly judge the accuracy and usefulness of the translation and if need be, alter the parameters of \texttt{Pitch\_Ugen}. To further judge the effectiveness, the transcription can be played back with a simple midi instrument while simultaneously playing the audio source. If the translation is judged suitable the \texttt{bach library} enables the quantization of the \texttt{bach\_roll} into \texttt{bach\_score} object. Having both the raw spatial notation and a quantized version side by side for both visual and auditory review means the optimal translation can be quickly determined with a few alterations of quantization settings. Once quantized the information is output as a \texttt{musicxml} file which is brought into \texttt{Finale} and the text is set.

The other control that was utilized in the rendering translations was the dynamic control of the rate at which the pitch analysis is performed in \texttt{SuperCollider}. The analysis routine utilizes a trigger for the rate at which pitches are reported. In previous versions of the translation process it was optimal to set the trigger to repeatedly report notes. This not only renders all of the slight variations in pitch but also helps to show more precisely where a change in pitch occurs. The downside to this approach is that there is an excess of information that the composer must reduce. The addition of dynamic control means that through focused listening and several rehearsals, the composer can control the reporting rate to approximate the ideal rate per each section of the sound file.

3.2.2 Text Setting
Once these translations were completed, they were sent to the performer. The performer took the original text and set it, in our mind, to melody, rhythm, and text as needed. In some cases, rhythms were adjusted for purpose of syllable stress and syllabification. In other cases, certain words in the phrase were extended to become melismatic, which supported the original integrity of the translation. These reworked melodies were recorded and sent to the composer.

The recorded melodies serve as a sonic point of departure for the composer in creating the final works. The melodies are set with a fixed electronic accommodation. The goal of these settings is to create a series of unique songs that explore ideas that the collaborators discussed with each text. Each melodic phrase was approached differently, often using excess material from the initial translations and aims to create songs which can stand on their own and work in the larger piece. The final compositional stage was the creation of connective sonic material between the successive songs. This material took the form of brief fixed electronic works.

4. CONCLUSIONS
The creation of \texttt{cmetaq} was motivated by the authors’ interest in etiquette and its relationship to technology. It is based to two simultaneous conversations. First, the discussion around the development of the text which explores etiquette and communication technology, while using various social media channels to maintain that conversation. Second is the development of the sonic material with the composer and performer communicating through recordings. Each of the conversations is supplemented with conversations via telephone and video chat. The formal design and workflow of \texttt{cmetaq} were directly influenced by both conversations and results in a unique performance piece. It is through conversation, in multiple modalities that we discovered the optimal form of the piece and how to ideally articulate ideas in sound.

Acknowledgements
A significant and heroic effort was put forward by Alejandro Casazi in realizing the visual aspects of this work. \texttt{cmetaq} is able to function on stage as a result of his brilliant work. We also indebted to the Grant Westoff for his support and the University of Wisconsin Oshkosh Music Department for their support of this work.

5. BIBLIOGRAPHY
\begin{enumerate}
\item N. Carr, \textit{The Shallows: What the Internet is Doing to Our Brains}, New York, NY (2010).
\end{enumerate}
Each composer/conductor in Rodinia uses an iPad interface, the “Conductor View”, to generate notation for their group (Fig. 1). The controller interface is operated by two hands (the iPad permits 11 simultaneous multi-touch points) [4] allowing parameters to be specified simultaneously by the left hand (play/hold, articulation, duration type) and right hand (duration, pitch, dynamic, rate and compass). The variables Conductor View interface are:

- Players – defines the number of performers in each stream and generates a part of varied shade for each performer;
- State – saves a particular configuration of parameters that can be accessed at a later point;
- Play/Stop - stops and starts the generation of new notation;
- Articulation type – defines the graphical shape of the notation events;
- Duration type – generates a morphology of the notation events (line, curve up/down and tremolo);
- Duration – generates events of statistically longer or shorter duration;
- Pitch – designates the central pitch of the notation;
- Dynamic – generates larger/louder or smaller/softer notation events; and
- Compass – designates the statistical range that notation events fall within.

These parameters define the boundaries of stochastically generated graphical events which are distributed to the all of the iPads belonging to the same stream on the network. Like many works for the Decibel Scoreplayer, the notation for the performers is scrolled right to left across the iPad screen; in Rodinia this is designated the “Performer View” (Fig. 2). The scroll time, the duration between the notation’s appearance on the right of the screen and its arrival at the “playhead”, is 12 seconds. The playhead is black line of the left of the screen at which the performer’s execute the notation [5]. This produces a scroll-rate of between 1.1 and 1.8 cm/s depending on the iPad model, falling below the maximal eye-hand span of the average sight-reader (less than 1.9 cm) [6][7]. Therefore, the musicians do not perform the notational event until it arrives – 12 seconds after specification by the conductor. This allows for the performers to comfortably “look ahead” at on-coming notation and for the conductors to evaluate strategies to avoid (or seek) collision with the other 3 streams.

Rodinia also amalgamates the notation from each stream into a single score, the “Audience View”, to be shown on a large screen behind the performers for both the audience and the conductors. Unlike the performer view, audience view shows the streams of notation approaching from four directions (left, right, top and bottom) (Fig. 3). The notation “wraps” around each time it completes the crossing from one side of the score to the other. As notation does not appear until the moment at which it is executed by the performer, the audience see it at the moment that it is heard.

The notation draws on conventions established in works by Cage and his colleagues Earle Brown and Christian Wolff [13], chiefly proportional notation in which the vertical height of the notational event signifies relative pitch (relative to the range of the instrument), horizontal length its (absolute) duration and thickness its dynamic. Unlike Decibel’s scores for Variations I and II, in Rodinia timbre is indicated by the shape of the notational event rather than the shade. Performers are expected to match the qualities of timbral notational types (such as “normal” tone (rich harmonic sounds), “ghost” tone (harmonically poor sounds) and “noise” tone (inharmonic dense sounds)) within each stream. Each conductor controls a group of instruments of similar range so that register choices by the conductors are mirrored in the ensemble. Each stream, and individual parts within a stream are differentiated using shades of four principal colours orange, red, green and blue. Green-Armytage claims that 26 colours should “be regarded as a provisional limit – the largest number of different colours that can be used before colour coding breaks down” [14]. Rodinia is conceived for an ensemble of 16 performers (4 per stream) falling within the limits that of colour differentiation.

3. NOTATIONAL CONVENTIONS

The notational paradigm, semantic spatial notation, employed by Rodinia has been developed over a number of projects by composers working with the Decibel Scoreplayer – in particular the approach to preventing notational events used in the generation of scores from John Cage’s Variations I and II by Decibel [12] Fig. 7.

The notation draws on conventions established in works by Cage and his colleagues Earle Brown and Christian Wolff [13], chiefly proportional notation in which the vertical height of the notational event signifies relative pitch (relative to the range of the instrument), horizontal length its (absolute) duration and thickness its dynamic. Unlike Decibel’s scores for Variations I and II, in Rodinia timbre is indicated by the shape of the notational event rather than the shade. Performers are expected to match the qualities of timbral notational types (such as “normal” tone (rich harmonic sounds), “ghost” tone (harmonically poor sounds) and “noise” tone (inharmonic dense sounds)) within each stream. Each conductor controls a group of instruments of similar range so that register choices by the conductors are mirrored in the ensemble. Each stream, and individual parts within a stream are differentiated using shades of four principal colours orange, red, green and blue. Green-Armytage claims that 26 colours should “be regarded as a provisional limit – the largest number of different colours that can be used before colour coding breaks down” [14]. Rodinia is conceived for an ensemble of 16 performers (4 per stream) falling within the limits that of colour differentiation.

4. AUDIO PROCESSING APPROACH

The audio of the live instrumentalists is captured and processed digitally in Max/MSP on a standalone computer that is also networked via the bonjour protocol with the
Each composer/conductor in Rodinia uses an iPad interface, the “Conductor View”, to generate notation for their group (Fig 1). The controller interface is operated by two hands (the iPad permits 11 simultaneous multi-touch points) [4] allowing parameters to be specified simultaneously by the Left hand (play/hold, articulation, duration type) and Right hand (duration, pitch, dynamic, rate and compass). The variables Conductor View interface are:

- Players – defines the number of performers in each stream and generates a part of varied shade for each performer.
- State – saves a particular configuration of parameters that can be accessed at a later point.
- Play/Stop – stops and starts the generation of new notation.
- Articulation type – defines the graphical shape of the notation events;
- Duration type – generates the morphology of the notation events (line, curve up/down and tremolo);
- Pitch – designates the central pitch of the notation;
- Dynamic – generates larger/louder or smaller/softer notation events; and
- Compass – designates the statistical range that notation events fall within.

These parameters define the boundaries of stochastically generated graphical events which are distributed to all of the iPads belonging to the same stream on the network. Like many works for the Decibel Scoreplayer, the notation for the performers is scrolled right to left across the iPad screen; in Rodinia this is designated the “Performer View” (Fig 2). The scroll time, the duration between the notation’s appearance on the right of the screen and its arrival at the “playhead”, is 12 seconds. The playhead is - a black line of the left of the screen at which the performer’s execute the notation [5]. This produces a scroll-a black line of the left of the screen at which the performers arrive at the “playhead”, is 12 seconds. The playhead is - a black line of the left of the screen at which the performer’s execute the notation [5]. This produces a scroll after specification by the conductor. Unlike the performer view, audience view shows the streams of notation approaching from four directions (left, right, top and bottom) (Fig 3). The notation “wraps” around each time it completes the crossing from one side of the score to the other. As notation does not appear until the moment at which it is executed by the performers, the audience see it at the moment that it is heard.

The use of an audience view was first employed for the Decibel Scoreplayer in Vickerby’s work with Jon Rose Uhahn c. 1985: the Rosenberg Variations [2012]. For this and other rhizomatic works [8] the projected Audience View provides an overview of the current position of each player and graphically illuminates the choices taken in each stream.

In Rodinia, a mass is calculated for each stream, \( M \), based on its cumulative density: that is, based on the positions of the right-hand parameter sliders selected in the conductor view. This is based on both horizontal and vertical density as pictured in the score view. The deflection angle of each stream, \( \theta \), is based both on the current mass of each stream calculated individually, as well as the total mass. If the distance between the leading point of each stream is below 175px, the deflection angle ranges from 0° to 90° exponentially in inverse of the proximity, as the proximity approaches 0px, such that:

\[
\theta = \frac{M_0}{M} \left(1 - \exp\left(-\frac{d}{175px}\right)\right)
\]

where \( \theta \) is the new angle calculated individually for each stream, \( M_0 \) is the mass of the same stream, and \( M \) is the total mass.

### 3. NOTATIONAL CONVENTIONS

The notational paradigm, semantic spatial notation, employed by Rodinia has been developed over a number of projects by composers working with the Decibel Scoreplayer - in particular the approach to presenting notational events used in the generation of scores from John Cage’s Variations I and II by Decibel [12] Fig. 7.

The notation draws on conventions established in works by Cage and his colleagues Earle Brown and Christian Wolff [13]. Chiefly proportional notation in which the vertical height of the notational event signifies relative pitch (relative to the range of the instrument), horizontal length its (absolute) duration and thickness its dynamic. Unlike Decibel’s scores for Variations I and II, in Rodinia timbre is indicated by the shape of the notational event rather than the shade. Performers are expected to match the qualities of timbral notational types (such as “normal” tone (rich harmonic sounds), “ghost” tone (harmonically poor sounds) and “noise” tone (inharmonic dense sounds)) within each stream. Each conductor controls a group of instruments of similar range so that register choices by the conductors are mirrored in the ensemble. The streams, and individual parts within a stream are differentiated using shades of four principal colours orange, red, green and blue. Green-Armytage claims that 26 colours should “be regarded as a provisional limit – the largest number of different colours that can be used before colour coding breaks down” [14]. Rodinia is conceived for an ensemble of 16 performers (4 per stream) falling within the limits that of colour differentiation.

### 4. AUDIO PROCESSING APPROACH

The audio of the live instrumentalists is captured and processed digitally in Max/MSP on a standalone computer that is also networked via the bonjour protocol with the...
iPad scores. This processing is informed by the move- ments of the four user controlled streams in order to gen- erate and gradually deform a two-dimensional terrain map [15].

The terrain is initially generated by a method of perlin noise functions and undergoes both spatial deformation using a 2D spatial lookup process and 2D modulation amplitude. The 2D spatial lookup process involves translating four separate planes from a point of origin $(x_1, y_1)$, $(x_2, y_2)$, $(x_3, y_3)$, and $(x_4, y_4)$ translated by the movement of the four separate streams $(x_1, y_1)$, $(x_2, y_2)$, $(x_3, y_3)$, and $(x_4, y_4)$. The surface is also modulated by the relative direction and interactions of these four streams. A 2D surface and is generated as follows using the relative direct- ion and distances between the four streams. Equation 3 describes this process for just two different streams $(x_1, y_1)$ and $(x_2, y_2)$ if the change in direction between these streams brings them closer together, an additive function is applied:

$$f(x, y) = f(x, y) + f(x_1, y_1) + f(x_2, y_2)$$

where $f(x, y)$ is the new 2D function, and $f(x_1, y_1)$ is the previous 2D function. The iterative process is also ap- plied subtractively for streams that are moving away from each other.

The terrain surface that is generated is then used to con- trol the audio processing by using Wave Terrain Synthe- sis to control complex sound synthesis [16]. Similar tech- niques have been explored using Wave Terrain Synthesis as a framework for controlling timber spatialisation in the frequency domain [17]. However, in this project, this ap- proach is used for controlling both granular synthesis and spectral spatialisation [18].

![Figure 8. A trajectory of white noise reading values off the terrain after 1 second. A trajectory of white noise reading values off the terrain after 10 seconds. The audio-rate trajectory that is used to read information from the terrain is a random 2D signal (white noise, as shown in Fig. 8), a curve that is considered to have effec- tive space-filling properties. This means that details of the contour can be mapped to spatial details of the pro- cessing with great precision and resolution. The control information generated, in the way of 8192 individual pa- rameters, those being 352,800 parameters generated per second, are used to control the relative distribution of grains and spectra across 8 loudspeakers.

Controlling granular synthesis via such an interface may take grain time or grain size into consideration. In order to control 1000 simultaneous grains, parameters would be updated at 44.1Hz. Depending on the implementation of the synthesis model, parameter assignments are multilar- ous. For example 2D data could determine the grain pan and grain length of individual grains.

Swarm-based spatialisation is also used where 2D data is mapped to the spatial position of individual grains. In this case the space-filling properties of the 2D trajectory sig- nal will also correlate with the level of immersion of the resulting sound spatialisation.

Spectral spatialisation is also explored in Rodinia. Each spectral bin is assigned an independent spatial trajectory. 1024 simultaneous frequency bands are updated at lower-dimensional audio rates, that is, at approximately 431Hz. This is used to create complex immersive effects that would otherwise be more cumbersome if using standard control-rate methods.

5. CONTEXT

Prestly defines generative music as in indeterminate music played through interaction be- tween one or more persons and a more or less prede- termined system, such that the players control some but not all — performance parameters, and relin- quish choices within a selected range to the system. Tectonic: Rodinia conforms to this broadest definition of generative art work, through its use of algorithmically determined system, such that the players control some but not all — performance parameters, and relin- quish choices within a selected range to the system. Tectonic: Rodinia conforms to this broadest definition of generative art work, through its use of algorithmically determined modification of the intentions of human con- ductors. The term most specifically refers here, however, to the use of generative "emergent: non-repeatable" [20] music notation, a category of the emerging genre of ani- mated notation [21].

It is an interactive form of generation that has game-like aspects to the conductors’ interactions with the algorithmic modifications: a dynamic obstacle game. In this sense it resembles "4-ways"-concept structures in which "four agents traveling in four opposing directions, meeting at nearly the same time [22] or (form the individual conductors perspective) a "Frogger"-like structure in which "one agent encounters many perpen- dicular crossing agents" [23].

The game analogy is perhaps amplified by the inclusion of an Audience View, allowing the audience both to hear and view the interactions of the streams, and the conduc- tors’ attempts to maintain control under conditions in which their choices are undermined and their ability to utilise the algorithmic modifications to subvert the con- trol of the other conductors.

Musically, the work is something of a concert for con- ductors themselves that are silent but create sound through their gestures. The Rodinia environment gives significant freedom of choice to the conductors, which is curtailed only by the interactions between their choices.

6. CONCLUSIONS

Tectonic: Rodinia adds a series of new capabilities to the Decibel ScorePlayer. Many of these advances have been dependent upon the adoption of the Bonjour network protocol and the subsequent ability to stream data be- tween a variety of devices.

There is arguably some value in engaging the audience with a visual representation of the sound they are hearing, but the requirements of the performer are quite different to those of the listener and displaying the performer’s score to the audience and allowing them to “see what is coming” may reduce the effectiveness the musical dis- course when it is actually heard. Delaying the audience score until the moment of its execution by the performers goes some way to alleviating the issue.

Rodinia is somewhat unusual in its combination of gener- ative and interactive qualities in the context of notated music for live instrumentalists. Although the “tectonic” concept is distinct, the implementation of this work pro- vides a framework capable of accommodating a wide range of generative and interactive/generative works em- ploying varied conceptual approaches.

Acknowledgments

The XCode programming for Tectonic: Rodinia was de- veloped by Aaron Wyatt. Many Thanks!

Partial funding for this project was provided by an Early Career Researcher Grant from Edith Cowan University.

7. REFERENCES


function is applied: translating four separate planes from a point of origin modulation. The 2D spatial lookup process involves noise functions and undergoes both spatial deformation. The terrain is initially generated by a method of perlin noise functions and updated at 44.1Hz. Depending on the implementation of the synthesis model, parameter assignments are multifarious. For example 2D data could determine the grain pan and grain length of individual grains. Swarm-based spatialisation is also used where 2D data is mapped to the spatial position of individual grains. In this case the space-filling properties of the 2D trajectory signal will also correlate with the level of immersion of the resulting sound spatialisation. Spectral spatialisation is also explored in Rodinia. Each spectral bin is assigned an independent spatial trajectory. 1024 simultaneous frequency bands are updated at gener-dimensional audio rates, that is, at approximately 44.1kHz. This is used to create complex immersive effects that would otherwise be more cumbersome if using standard control-rate methods.

5. CONTEXT

Preistly defines generative music as indeterminate music played through interaction between one or more persons and a more or less pre-determined system, such that the players control some but not all — performance parameters, and relinquish choices within a selected range to the system. Tectonic: Rodinia conforms to this broadest definition of generative art work, through its use of algorithmically determined modifications of the intentions of human conductors. The term most specifically refers here, however, to the use of generative “emergent: non-repeatable” music notation, a category of the emerging genre of algorithmic compositions/S13/final_projects/benedetti_minto.pdf, 2013.

6. CONCLUSIONS

Tectonic: Rodinia adds a series of new capabilities to the Decibel ScorePlayer. Many of these advances have been dependent upon the adoption of the Bonjour network protocol and the subsequent ability to stream data between a variety of devices. There is arguably some value in engaging the audience with a visual representation of the sound they are hearing, but the requirements of the performer are quite different to those of the listener and displaying the performer’s score to the audience and allowing them to “see what is coming” may reduce the effectiveness of the musical discourse when it is actually heard. Delaying the audience score until the moment of its execution by the performers goes some way to alleviating the issue.

Rodinia is somewhat unusual in its combination of generative and interactive qualities in the context of notated music for live instrumentalists. Although the “tectonic” concept is distinct, the implementation of this work provides a framework capable of accommodating a wide range of generative and interactive/generative works employing varied conceptual approaches.

Acknowledgments

The XCode programming for Tectonic: Rodinia was developed by Aaron Wyatt. Many Thanks!

Partial funding for this project was provided by an Early Career Researcher Grant from Edith Cowan University.

7. REFERENCES


AVA: A Graphical User Interface for Automatic Vibrato and Portamento Detection and Analysis

Luwei Yang1 Khalid Z. Rajab2 Elaine Chew1
1 Centre for Digital Music, Queen Mary University of London
2 Antennas & Electromagnetics Group, Queen Mary University of London
{l.yang, k.rajab, elaine.chew}@qmul.ac.uk

ABSTRACT

Musicians are able to create different expressive performances of the same piece of music by varying expressive features. It is challenging to mathematically model and represent musical expressivity in a general manner. Vibrato and portamento are two important expressive features in singing, as well as in string, woodwind, and brass instrumental playing. We present AVA, an off-line system for automatic vibrato and portamento analysis. The system detects vibratos and extracts their parameters from audio input using a Filter Diagonalization Method, then detects portamenti using a Hidden Markov Model and presents the parameters of the best fit Logistic Model for each portamento. A graphical user interface (GUI), implemented in MATLAB, allows the user to interact with the system, to visualize and hear the detected vibratos and portamenti and their analysis results, and to identify missing vibratos or portamenti and remove spurious detection results. The GUI provides an intuitive way to see vibratos and portamenti in music audio and their characteristics, and has potential for use as a pedagogical and expression analysis tool.

1. INTRODUCTION

Musicians introduce a high degree of acoustic variations in performance, above and beyond the categorical pitches and durations indicated in the musical score [1]. The sources of these acoustic variations include dynamic shaping, tempo variation, vibrato, portamento, staccato, and legato playing. While some expressions have been notated in the score (e.g. tempo and dynamics), musicians sometimes alter the instructions to create their own expressions [2]. We call these devices expressive features as they are usually not denoted in the composition but adopted in performance. These devices result in unique performance styles that differentiate one musician from another.

We focus on two expressive features: vibrato and portamento. Vibrato is a periodic modulation of frequency, amplitude, and even spectrum [3]. Portamento is the note transition that allows musicians to adjust the pitch continuously from one note to the next [4]. Vibrato and portamento characteristics can be used to reveal differences in performance styles, and performance variation among different musicians [4, 5, 6, 7, 8].

This paper presents an off-line system, AVA, which accepts raw audio and automatically tracks the vibrato and portamento to display their expressive parameters for inspection and further statistical analysis. We employ the Filter Diagonalization Method (FDM) to detect vibrato [9]. The FDM decomposes the local fundamental frequency into sinusoids and returns their frequencies and amplitudes, which the system uses to determine vibrato presence and vibrato parameter values. A fully connected three-state Hidden Markov Model (HMM) is applied to identify portamento. The resulting portamenti are modeled as Logistic Functions which are well suited to displaying the characteristics of a portamento [4]. The AVA system has been implemented in MATLAB and consists of a graphical user interface (GUI) and all relevant functions [1].

The structure of the paper is as follows: Section 2 presents the vibrato and portamento feature detection and analysis modules. Section 3 introduces AVA’s MATLAB interface, and Section 4 presents discussions and conclusions.

2. FEATURE DETECTION AND ANALYSIS

The basic architecture of the AVA system is shown in Figure 1. Taking the audio as input, the pitch curve (fundamental frequency) is extracted using the pYIN method [10], a probabilistic version of the original Yin method [11]. The resulting pitch curve is sent to the vibrato detection module, which identifies vibrato existence using an FDM-based method. The detected vibratos are forwarded to the portamento detection module, which outputs the vibrato statistics. To ensure the best possible portamento detection performance, we flatten the detected vibratos using the built-in MATLAB ‘smooth’ function as the oscillating shape of the vibrato degrades portamento detection. The HMM-based portamento detection module uses this vibrato-free pitch curve to identify potential portamenti. A Logistic Model is fitted to the detected portamentos for quantitative analysis. Moreover, if there are errors in detection, the interface allows the user to indicate missing vibratos or portamenti and remove spurious results.

2.1 Vibrato Detection and Analysis

There exist two kinds of vibrato detection methods: note-wise and frame-wise methods. Note-wise methods require

[9] The beta version of AVA is available at luweiyang.com/research/ava-project.