2016

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*, sjames@ecu.edu.au

https://ro.ecu.edu.au/ecuworkspost2013/2034
In this implementation we use an autocorrelation pitch follower implemented in SuperColliderPitch: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. [7]. 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 realization. To accomplish this the bach library [8] in MaxMSP environment is employed. Using OpenSoundControl the midi note value of the detected pitch is sent from SuperCollider to the MaxMSP environment where the bach.transcribe object is utilized to format the incoming information and present it via a 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 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 bach library enables the quantization of the bach.roll into bach.score object. Having both the raw spacial 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 musicxml file which is brought into 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 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 rapidly 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 for each stream, allowing the blocks of musical material to slide, grate and collide with one another like tectonic plates that crush and reform each other’s placement.

Tectonic: a Networked, Generative and Interactive, Conducing Environment for iPad

This paper describes the concepts, implementation and context of Tectonic: Rodinia, for four realtime composer-conductors and ensemble. In this work, an addition to the repertoire of the Decibel Scoreplayer, iPads are networked together using the bonjour protocol to manage connectivity over the network. Unlike previous Scoreplayer works, Rodinia combines "conductor view" control interfaces, "performer view" notation interfaces and an "audience view" overview interface, separately identified by manual connection and yet mutually interactive. Notation is communicated to an ensemble via scores independently generated in realtime in each "performer view" and amalgamated schematically in the "audience view" interface. Interaction in the work is enacted through a collision avoidance algorithm that modifies the choices of each conductor by deflecting the streams of notation according to evaluation of their "Mass" and proximity to other streams, reflecting the concept of shifting Tectonic plates that crush and reform each other's placement.

4. CONCLUSIONS

The creation of cmetq was motivated by the authors’ interest in the idea of generating music that is both evocative and technologically driven; to explore how these two domains can coexist and offer new possibilities. The experiment aims to demonstrate the potential of using everyday objects to create new musical forms, while also highlighting the importance of collaboration and improvisation within such contexts. The project seeks to challenge traditional notions of musical notation by incorporating new technologies and exploring innovative ways of combining them with established techniques. This approach opens up new possibilities for composers and performers alike, offering a unique opportunity to create music that engages both the auditory and technological senses.

Acknowledgments

A significant and heroic effort was put forward by Alejandro Casazi in realizing the visual aspects of this work. cmetq is able to function on stage as a result of his brilliant work. We also indebted to the Grand Women’s School and the University of Wisconsin Oshkosh Music Department for their support of this work.

5. BIBLIOGRAPHY


ABSTRACT

Rodinia employs generative scores for each of the four streams directed by the composer-conductors. Unlike previous generative notation works by Vickery such as Lyrebird [2] and The Semantics of Reduction [3] Rodinia does not use the analysis of a pre-existing audio artifact to generate notation.

2. IMPLEMENTATION

Rodinia four composer/conductors control separate streams of graphical notation and audio (comprising live instruments reading the notation and their processed audio components) that interact through the algorithmically evaluated Mass and proximity of each stream. The work is performed using the Decibel Scoreplayer on multiple iPads via a manually connected network allowing for each participant conductor or performer to identify independently on the network [1]. The manually connected network was first used in Laura Lowthers’ work for the Decibel ensemble, Loaded [2015]. Previous scores had prioritized synchronization between multiple iPads in order to present uniform representation of fixed scores for all performers. It is made possible by the adoption of the bonjour protocol to manage connectivity over the network. The use of the bonjour protocol also allows connectivity via OSC to stream data to other devices. In Rodinia this is used to stream generative data to a dedicated computer using Wave Terrain synthesis to process and spatialise the audio from the ensemble.

1. INTRODUCTION

TECTONIC: Rodinia is a work for four realtime composer-conductors and ensemble. In geology Rodinia is the name of a supercontinent that contained most of Earth’s landmass between 1.1 billion and 750 million years ago. Tectonic can mean both ‘the study of the earth’s structural features’ and ‘the art of construction’ and this works reflects both aspects of the word’s meaning. The concept of slowly shifting plates that crush and reform each other’s placement is the central paradigm of the work.

Rodinia is the second in a series that began with Tectonic: Vaalbara [2008]. In Vaalbara five instrumental streams are performed independently, using computer generated metronome pulses to manipulate the tempo of each stream, allowing the blocks of musical material to slide, grate and collide with one another like tectonic plates.

Copyright: © 2016 Lindsey Vickery et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
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/Hold – 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);
- 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;
- 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 a 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. 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.

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 jaunp 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 alters 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/slower 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 - a 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” on-coming notation and for the conductors to evaluate strategies to avoid (or seek) collision with the other 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 performers, the audience see it at the moment that it is heard.

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

\[
\theta = \frac{M_{\text{pitch}} - M_{\text{compass}}}{M_{\text{mass}}} \quad (1)
\]

\[
M_{\text{pitch}} = M_x + M_y + M_z + M_{\theta}
\]

\[
\text{where } \theta \text{ is the new angle calculated individually for each stream, } M_{\text{pitch}} \text{ is the mass of the same stream, } M_{\text{pitch}} \text{ is the total mass, } M_{\theta} \text{ is the angle scalar, and } \beta \text{ is positive or negative scalar determining a turn in direction either left or right of the current direction of each stream. The height parameter is used to calculate whether an interaction results in an upward or downward deflection. The total mass, } M_{\text{pitch}} \text{ is the sum of all stream masses such that:}
\]

\[
M_{\text{pitch}} = M_x + M_y + M_z + M_{\theta}
\]

\[
\text{Figure 5. a. example of a point in the plane performing a self-avoiding random walk using Chappell’s model. b. Greenfield’s “avoidance drawing” (2015).}
\]

\[
\text{The key difference in Rodinia is that since music is a time-based medium, it can never “double-back” on itself and therefore in a generative score the deflection can never be greater than 90º.}
\]

\[
\text{Early studies conducted in Jitter, by Vickery for testing collision avoidant lines explored this paradigm, exploring “proximity only” avoidance (all lines were of equal density) to illustrate the kinds of pathways generated by this strategy (Fig 6).}
\]

\[
\text{Figure 6. Vickery “collision avoidant lines” study for Tecktonik: Rodinia (2013): first, second, and twelfth passes.}
\]

\[
\text{In Rodinia, a mass is calculated for each stream, } M_{\text{pitch}}, \text{ based on its cumulative density: that is, based on the positions of the right-hand parameter sliders selected in the conductor view. This is based on both horizontal and vertical density as pictured in the score view.}
\]

\[
\text{The deflection angle of each stream, } \theta \text{, is based both on the current mass of each stream calculated individually, as well as the total mass. If the distance between the leading point of each stream is below } 175\text{px the deflection angle rises from } 0^\circ \text{ to } 90^\circ \text{ exponentially in inverse of the proximity, as the proximity approaches } 0\text{px, such that:}
\]

\[
\text{Figure 7. Decibel’s scrolling, proportionally notated screen-score for Variations I and II.}
\]
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 amplitude modulation. The 2D spatial lookup process involves translating four separate planes from a point of origin \( (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4) \) by the movement of four separate streams \( (x_1', y_1'), (x_2', y_2'), (x_3', y_3'), (x_4', y_4') \).

The surface is also modulated by the relative direction and interactions of these four streams. A 2D surface terrain 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 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 predetermined system, such that the players control some but not all — performance parameters, and relinquish choices within a selected range to the system

\[
 f(x,y) = (f(x,y) + f(x,y)) + \left( \frac{1}{2} (f(x,y) + f(x,y)) \right) 
\]

where \( f(x,y) \) is the new 2D function, and \( f(x,y) \) is the previous 2D function. The iterative process is also applied subtractively for streams that are moving away from each other.

The terrain surface that is generated is then used to control the audio processing by using Wave Terrain Synthesis to control complex sound synthesis [16]. Similar techniques have been explored using Wave Terrain Synthesis as a framework for controlling timbre spatialisation in the frequency domain [17]. However, in this project, this approach is used for controlling both granular synthesis and spectral spatialisation [18].

\[ \text{Figure 8.} \text{ A trajectory of white noise reading values off the terrain after 1 second.} b. \text{ A trajectory of white noise reading values off the terrain after 10 seconds.} \]

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 effective space-filling properties. This means that details of the contour can be mapped to spatial details of the process with great precision and resolution. The control information generated, in the way of 8192 individual parameters, 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 44.1kHz. 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 lower-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.

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


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 amplification. The 2D spatial -lookup process involves translating four separate planes from a point of origin \((x, y)\), \((x', y')\), \((x'', y'')\), \((x'''', y''')\) translated by the movement of two separate planes \((x, y)\) \& \((x' , y')\). The surface is also modulated by the relative direction and interactions of these four streams. A 2D surface is generated by using the relative direction and distances between the four streams. Equation 3 describes this process for just two different streams \((x, y)\) and \((x', y')\). If the direction in between these streams brings them closer together, an additive function is applied:

\[
f(x, y) = f(x', y') + \left( \frac{x-x'}{2} + \frac{y-y'}{2} \right)
\]

where \((f(x, y))\) is the new 2D function, and \((f(x', y'))\) is the previous 2D function. The iterative process is also applied subtractively for streams that are moving away from each other.

The terrain surface that is generated is then used to control the audio processing by using Wave Terrain Synthesis to control complex sound synthesis [16]. Similar techniques have been explored using Wave Terrain Synthesis as a framework for controlling timbre spatialisation in the frequency domain [17]. However, in this project, this approach is used for controlling both granular synthesis and spectral spatialisation [18].

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. Delving 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 Chew2
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, wind, and brass instrumental playing. We present AVA, an off-line system for automatic vibrato and portamento analysis. The system detects vibrato 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 visualise 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, 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.

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 module for vibrato analysis, 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 portamenti 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 additional detection and remove spurious results.

Copyright: ©2016 Luwei Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.