

REAL-TIME MUSIC INFORMATION PROCESSING

Elaine Chew* and Alexandre François

University of Southern California Integrated Media Systems Center

{echew, afrancoi}@usc.edu

ABSTRACT

This paper introduces music information processing as an area of research for Industrial Engineers and offers a sample of the types of engineering issues encountered in real-time music information processing. We present the Spiral Array model and the Center of Effect Generator (CEG) algorithm for tonal analysis. We extend the CEG algorithm to real-time in MuSA.RT, a 3D interactive tonal visualization system implemented using the Modular Flow Scheduling Middleware.

Keywords: music information processing, real-time, data stream processing, visualization.

1. INTRODUCTION

The explosion of digital music information has necessitated the developing of engineering decision tools for analyzing, processing and managing music information. This paper introduces music information processing as an area of research for Industrial Engineers and offers a sample of the types of engineering issues encountered in real-time music information processing.

Any computer model for processing music information must begin with a representation that converts content information to numerical data for analysis. In Section 2, we describe the Spiral Array, a mathematical model for tonality (Chew, 2000), and how this geometric model can be used to extract and visualize tonal structure from music data. We then introduce MuSA.RT, a research initiative integrating real-time music processing and content-based graphical rendering in interactive immersive environments. MuSA.RT Op.1 is a system for mapping a live MIDI stream (produced by a keyboard) to the Spiral Array. The integration of live music, real-time computational analysis and interactive graphical representation is facilitated by the Modular Flow Scheduling Middleware (*mfsm.sourceforge.net*).

* USC DJEpstein Department of Industrial and Systems Engineering. 3715 McClintock Avenue GER240 MC:0193, Los Angeles CA 90089-0193. Tel: (213) 8212414. E-mail: *echew@usc.edu*

The growing interest in the building of engineering tools for music information processing is evidenced by the recent surge in the number of conferences devoted to different aspects of this multi-faceted topic such as, the International Conference on Music and Artificial Intelligence (www.music.ed.ac.uk/icmai), the International Symposium on Music Information Retrieval (ismir2002.ircam.fr), the Performance Rendering Piano Concours (Rencon: shouchan.ei.tuat.ac.jp/~rencon) and the International Computer Music Conference (www.icmc2002.org).

The important applications in the domain of music information processing include real-time music transcription and analysis, content-based music categorization and retrieval, pattern recognition and visualization, and computer-generated expressive performance. The human-computer interactions engendered by these systems require engineering decision methodology for extracting pitch (frequency) and time (duration) structures in music (Rowe, 2001). Vertical and horizontal collections of pitches form higher-level pitch structures. In Western tonal music, vertical pitch structures include chords and keys, and horizontal ones include figures, phrases and melodies. Patterns in the duration assigned to pitches relative to the beat (the basic unit for counting) form higher level time structures such as rhythm and meter.

2. The Spiral Array Model

The inter-relations among pitches form a system called *tonality*. The Spiral Array exploits the symmetry and structure in this system to allow for the design of efficient algorithms to analyze, categorize and compare music.

2.1 Model Description

The Spiral Array (Chew, 2000) is a 3D model of tonality that depicts the inter-relations among pitches, chords and keys (see Figure 1). In the Spiral Array, pitch representations are positioned at every quarter turn of a Spiral; adjacent pitch classes are an interval of a Perfect Fifth apart (representing frequency ratios of 3:2), and vertical neighbors are an interval of a Major Third apart. The model is configured so that closely related musical entities are positioned in compact clusters. Each pitch is indexed by its number of perfect fifths from an arbitrarily chosen reference pitch, C (set at position [0,1,0]):

Definition: The radius of the cylinder, r , and the height gain per quarter turn, h , uniquely define the position of a pitch representation, which can be described as: $\mathbf{P}(k) = [\sin(k\pi/2), \cos(k\pi/2), kh]$.

In the model, higher level musical entities are represented in the interior of a spiral array as convex combinations of their lower level parts. Pitch collections that form triads (major and minor) map to triangles. Major and minor triads are represented as convex combinations of their component pitch positions, and the weights on the three pitches are constrained to be monotonically non-decreasing according to the importance of that pitch.

Definition: The representation for a major triad is generated by the convex combination of its root, fifth and third pitch positions: $C_M(k) = w_1P(k) + w_2P(k+1) + w_3P(k+4)$, where $w_1 > w_2 > w_3$ and $w_1 + w_2 + w_3 = 1$. The minor triad is generated by a similar combination, $C_m(k) = u_1P(k) + u_2P(k+1) + u_3P(k-3)$, where $u_1 > u_2 > u_3$ and $u_1 + u_2 + u_3 = 1$.

Major and minor key representations are generated as convex combinations of their defining triads (see Figure 1). Again, the weights are constrained to be monotonically non-decreasing according to the importance of the triad. The additional parameters α and β for the minor key indicate the relative importance of the major vs. minor versions of the triads.

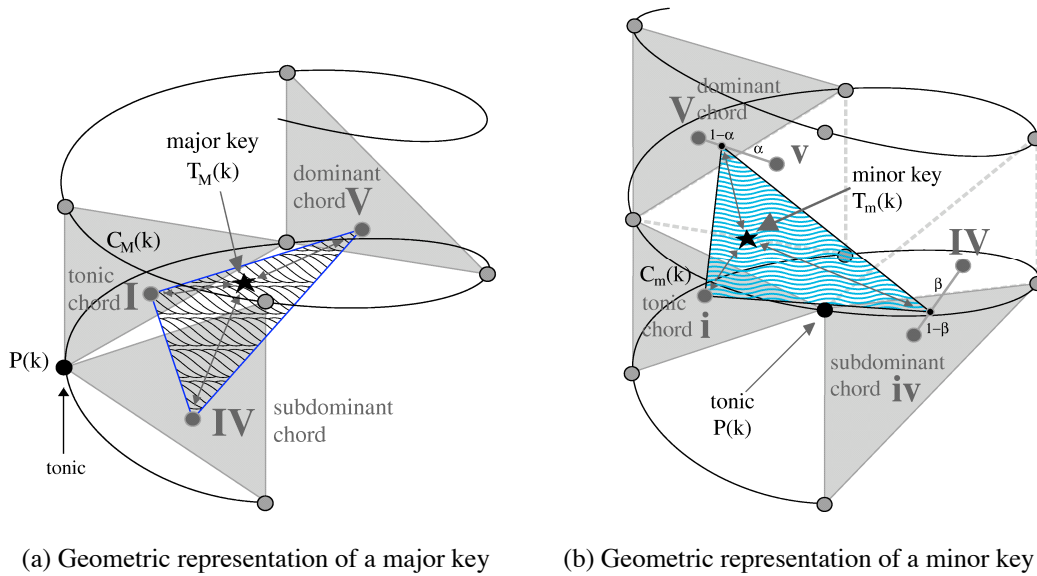


Figure 1: The Spiral Array Model

Definition: A major key is represented by a convex combination of its tonic, dominant and subdominant chords. The weights are restricted to be monotonically decreasing. $T_M(k) = \alpha C_M(k) + \beta C_m(k) + \gamma P(k)$

$\alpha_2 C_M(k+1) + \alpha_3 C_M(k-1)$, where $\alpha_1 > \alpha_2 > \alpha_3$. The minor key representation is generated by a convex combination of its tonic, dominant (major or minor) and subdominant (major or minor) chords as follows:

$$T_m(k) = \alpha_1 C_m(k) + \alpha_2 [\alpha C_M(k+1) + (1-\alpha)C_m(k+1)] + \alpha_3 [\alpha C_m(k-1) + (1-\alpha)C_m(k-1)],$$

where $\alpha_1 > \alpha_2 > \alpha_3$, $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and $0 \leq \alpha \leq 1$, $0 \leq \alpha \leq 1$.

The weights and parameters are selected so that spatial distances in the model mirror the perceived relation between the represented entities. Finding feasible solutions to all the possible constraints is still an open problem [5].

2.1 The Center of Effect

In the Spiral Array model, any collection of notes (also known as pitch events) generates a *center of effect* (c.e.). For any music time series data, the c.e. is a point in the interior of the Spiral Array that is the convex combination of the pitch positions weighted by their respective durations.

Definition: If p_i represents the position of the i -th pitch class in the Spiral Array, and d_i represents its duration, the sequence of notes in a melody up to the t -th pitch event can be written as $(p_1, d_1); (p_2, d_2); \dots; (p_t, d_t)$. The c.e. of this pitch collection is defined as: $c_t = \frac{\sum_{i=1}^t d_i p_i}{D_t}$, where $D_t = \sum_{i=1}^t d_i$.

The moving c.e. tracks the evolving chord and key patterns through spatial proximity to respective musical entities (Chew, 2002, 2003). The c.e. has been shown to be capable of tracking accurately key patterns in the Center of Effect Generator (CEG) algorithm (Chew, 2001). The most likely key is given by the following expression: $\arg \min_{k \in \mathcal{K}} \|c_t - k\|$ where $\mathcal{K} = \{T_m(k) \mid k \in \mathcal{K}\} \cup \{T_m(k) \mid k \in \mathcal{K}\}$.

3. MuSA.RT: an interactive music visualization system

MuSA.RT Op.1 offers the first real-time implementation of the CEG algorithm. The tonal patterns are rendered in real-time on the Spiral Array, using OpenGL so as to take full advantage of available hardware acceleration. A user can manipulate the virtual camera used for rendering via a gaming device. The real-time interaction in MuSA.RT is made possible by the Modular Flow Scheduling Middleware, a software architecture optimized for efficient data stream processing and system integration.

In MuSA.RT, a memory parameter is introduced that makes the most recent event the most salient element in the calculation of the c.e. Depending on the value of the memory parameter, the c.e. can track

the key or the chord patterns. Currently, the parameter has been set to best capture chord patterns, indicated by triangles in the spiral.

Research on content-based visualization of musical structures includes Dixon et.al.'s (2002) work on 3-D visualization of temporal structures and Foote's (2002) 2-D grid for visualizing repeated patterns. MuSA.RT Op.1 is the first real-time 3D interactive tonal visualization system.

3.1 The Modular Flow Scheduling Middleware

The Modular Flow Scheduling Middleware (François, 2000, 2001) is an open source software architecture for media systems integration that made the seamless integration of music and graphical information in a possible. Multithreading support in MFSM enables real-time performance for media stream processing using current state-of-the art single or multi-processor machines (François, 2001, 2002).

The modular framework defines and implements a parallel asynchronous model for generic data streams, and an extensible data model (François, 2000). Applications are implemented as dataflow graphs composed of software component instances. Figure 2 is the application graph for MuSA.RT Op.1. The graph is color-coded: [1] active streams (blue) carry volatile data, such as MIDI messages or rendered frames, through cells (in green) where it is processed; and, [2] sources (red) hold persistent data, such as the Spiral Array model or rendering parameters, on passive streams (red).

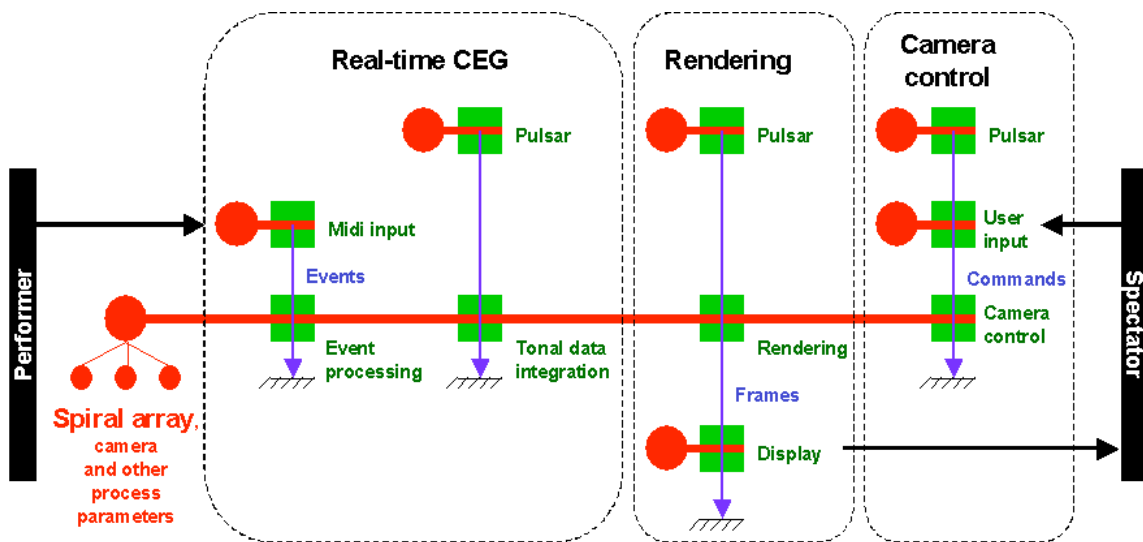


Figure 2: Application graph for MuSA.RT Op.1.

Tonal analysis, rendering/display and camera control are three independent streams potentially operating at different rates. For MuSA.RT Op.1, analysis, rendering/display and camera control typically operate at 30 Hz, with 640x480 rendering or better.

3.2 The Tracking Algorithm

In MuSA.RT Op.1, a memory factor, r , is introduced and the c.e. is sampled at every Δt time interval. If n_t is the number of active pitches at the time slice $(t-\Delta t, t)$, the modified c.e. is given by:

$$c_t = (1 - r) \cdot c_{t-\Delta t} + r \cdot \sum_{i=1}^{n_t} p_i / n_t .$$

At every point in time, the triangle representing the triad closest to the current c.e. lights up, blue if it is a minor triad, and pink if it is a major triad. The closest triad is indexed by: $\arg \min_{\Delta} \|c_t - \Delta\|$ where $\Delta = \{C_m(k) \Delta k\} \cup \{C_M(k) \Delta k\}$.

Figure 3 shows our current prototype, which uses a preliminary (simple) rendering routine to create a sketch of the Spiral Array model and the pitch structures. The final presentation will feature high resolution rendering of the analytical structures that preserve their elegant geometric simplicity while providing an informative and visually pleasing environment.



Figure 3: François demonstrating an early prototype of the MuSA.RT Op.1 system.

4. Summary

We have described the Spiral Array model and the CEG algorithm for tonal analysis. We extended this original algorithm to real-time in MuSA.RT. The MuSA.RT platform allows the immediate visualization and analysis of interactions between the pitch and time structures. The demonstrated system is an extensible platform that forms the basis for future research in real-time music analysis.

Acknowledgements

This project was supported in part by funding from the Integrated Media Systems Center (an NSF Engineering Research Center), Cooperative Agreement No. EEC-9529152, and an NSF Major Research Instrumentation Program (MRI-2001) Grant No. EIA-0116573.

References

- Chew, E. & Chen, Y.-C. (2003). Mapping MIDI to the Spiral Array: Disambiguating Pitch Spellings. To appear in Proceedings of the 8th INFORMS Computing Society Conference.
- Chew, E., 2002. The Spiral Array: An Algorithm for Determining Key Boundaries. In Proceedings of the 2nd International Conference on Music and Artificial Intelligence, pp. 18-31.
- Chew, E., 2001. Modeling Tonality: Applications to Music Cognition. In Proceedings of CogSci2001.
- Chew, E., 2000. Towards a Mathematical Model of Tonality. Ph.D. Dissertation. Massachusetts Institute of Technology, Cambridge, MA.
- Dixon, S., Werner, G. & Widmer, G., 2002. Real Time Tracking and Visualisation of Musical Expression. In Proceedings of the 2nd International Conference on Music and Artificial Intelligence, pp. 58-68.
- Foote, J., Cooper, M. & Nam U., 2002. Audio Retrieval by Rhythmic Similarity. In Proceedings of the 3rd International Symposium on Musical Information Retrieval (ISMIR), September 2002, Paris, pp. 81-85.
- François, A.R.J., 2002. Components for Immersion. In Proceedings of the IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, August 2002.
- François, A.R.J., 2001. A Modular Software Architecture for Real-Time Video Processing. In Proceedings of the International Workshop on Computer Vision Systems, Vancouver, B.C., Canada, July 2001, pp. 35-49.
- François, A.R.J., 2000. A Modular Middleware Flow Scheduling Framework. In Proceedings of ACM Multimedia 2000, Los Angeles, CA, November 2000, pp. 371-374.
- Rowe, R., 2001. Machine Musicianship. MIT Press: Cambridge, MA.