Jonathan Mooser
Review of
A Model for Tonal Context Time Course Calculation from Accoustical Input.
By
Izmirli & Bilgen

In *A Model for Tonal Context Time Course Calculation from Accoustical Input.* Izmirli and Bilgen propose a two stage algorithm for extracting key context from an acoustical signal. The two stages are described and analyzed separately. First, raw audio data is translated in to a set of notes categorized by pitch class. Then the resulting sets of notes are processed through a "leaky integrator" to give an estimation of tonal context as a function of time. Results are analyzed for two sample inputs and compared to the results of a similar system proposed by Leman.

As with Emilia Gonzales' paper, one of the most useful features of Izmirli's proposed method is the fact that it takes raw audio data as input. This means that virtually any piece of recorded music can be used, as opposed to only those for which MIDI representations exist.

Moreover, he analyzes each of the two stages separately, allowing them to be isolated and used independently. This separability is potentially extremely useful. If one only needs to finds the notes in a piece, or only needs to find the key context from a set of know notes, it is easy to implement only the relevant sections.

The first stage uses a Q transform on 7 ms intervals to find the frequency spectrum as a function of time. Then starting from the low end of the spectrum, peaks are found and identified as notes. For found notes, overtones (integer multiples of the note's frequency) are eliminated from considerations. The author readily points out that this will leave some notes unfound, but that those hidden notes are unlikely to contribute to the tonal context analyzed in stage 2.

Having found which notes are present in each time slice, stage 2 uses a leaky integrator estimate tonal context. The leaky integrator basically iterates though the piece, keeping a score for each possible key at each time. At any given time, the key context with the highest score is taken to be the best estimate and the result is a map of musical key as a function of time. If the piece has a modulation, then, the system should find that and correctly identify the change in key.

Izmirli usesw Debussy's Arabesque #1 as input, applying each stage separately and together. It is worth noting that when human-generated sets of notes were used as input to stage 2, the results were fairly similar to when the input was generated by stage 1. That is to say that even though stage 1 does not correctly find all notes in the raw audio data, its output is good enough to effective find tonal context.

Results are then compared to a system proposed by Leman. This part of the analysis is somewhat lacking in that the author never defines any sort of ground truth. Presumably there is a "correct" view of the tonal context at each moment in the piece that most human listeners could agree on. We don't know however, which system most closely approximates the correct results because that data is not given.

Overall the method proposed by Izmirli is simple and based on intuitive notions of tonality. It appears easy to implement and could be readily incorporated into a music information retrieval system.