

Haojun Wang  
haojunwa@usc.edu

## Review on Tonal Description of Polyphonic Audio for Music Content Processing

In this paper, the author presents an approach to extract a description of the tonal context of music from polyphonic audio signals. Two levels of abstraction are used to differentiating the low-level signal descriptors and high-level textual labels. Temporal scales for description are also addressed; defining some features as being attached to instants of certain times.

Western tonal music uses keys as relationships between a series of pitches. The dominant degree is defined as the fifth degree of the scale. The subdominant degree is defined as the fourth degree of the scale. There are 24 keys since both a major and a minor mode manages by each tonic. This paper only consider aspects of the harmonic content related to the combination of notes into chords, and its relation to the tonality of the piece.

The model presented in the paper uses different levels of abstraction and temporal scales to describe the tonal aspects of a music excerpt. There are two main categories of temporal scales defined: Instantaneous and Global features. Also two levels of abstraction are defined as Low-level signal descriptor, which is related with the audio signal, and High-level descriptor, which represents textual labels related to the musical content of signals. These two descriptors are used to represent the key or the tonal content and later to be used for content retrieval and navigations across digital music collections.

The computational approach in this model consists of three steps: the pre-processing of audio signals, the instantaneous descriptor computation of the Harmonic Pitch Class Profile (HPCP) vector, and the Global descriptor computation. In the pre-processing stage, a transient detection algorithm to eliminate regions where areas located within 50 ms of transients is regarded as noises. In the stage of HPCP computation, several steps are required: first, the FFT is used in spectral analysis of the input signal. Second, the author defines the local maxima of the magnitude spectrum. Next, the Pitch Class Profile (PCP) is used in HPCP computation to measure the intensity of each of the twelve semitones of the diatonic scale, by mapping each frequency bin of the spectrum to a given pitch class. Finally, the normalization is applied to HPCP values. In the stage of global descriptor computation, it first computes statistics of the HPCP vector over an audio excerpt, and then estimation of the tonality is computed by using a related key model.

Three music excerpts are used in experiments verifying the accuracy of this model. Also PCP and HPCP are performed respectively to compare with the model. The results show at least 20 % improvement in terms of the correctness comparing with PCP and HPCP. This is a very promising. Whether only main chords or all chords considered in the experiments seems have little effect on the model output.

Seems in this paper, the author uses a more complicated two-level structure to capture the tonal content. Compare to another paper I read this week by O. Izmirli. Seems the HPCP is more accurate to capture the tonal context than simply applying a peaky integrator. However, the Q transform is used in the latter one because they claim it is more

Haojun Wang  
haojunwa@usc.edu

appropriate to process music signals than FFT transform, which is used in this paper. Also, the global descriptor in this model helps to capture the related tonal context in wider rangel, which is not considered in the other model. I would like to see more verification of the model presented in this paper with other related works, as the author described in the future work section.