AUDIO MELODY EXTRACTION – LATE BREAKING AT ISMIR 2010

Karin Dressler
Fraunhofer IDMT, Ilmenau, Germany
kadressler@gmail.com

ABSTRACT

In this document, we outline our current approach to the automatic extraction of the melody from polyphonic music recordings. The presented algorithm is an advanced version of our successful submission to the audio melody extraction task at MIREX'09 [1]. During the late breaking demo session at ISMIR 2010, we will discuss the impact of the further developments and present some preliminary results.

1. METHOD

1.1 Spectral Analysis and Pitch Estimation

A multi resolution spectrogram representation is obtained from the audio signal. Then, the instantaneous frequency (IF) is computed. The weighted magnitude and the IF of the spectral peaks are evaluated in order to identify the pitch of the strongest periodic sounds. The pitch estimation algorithm performs two tasks: 1. the spectral peaks are added as harmonics to existing tones, 2. the identification of salient pitches. To find relevant starting points for new tone objects, it is important to give more weight to peaks that are not covered by the spectral envelope of existing tones. Hence, all peaks that are well explained by the spectral envelope of tones are virtually inhibited.

The estimation of the salient pitches is based on the pair-wise analysis of spectral peaks. The main idea lies in the identification of partials with a successive harmonic number. The identified peak pairs are evaluated according to a perceptually motivated rating scheme. The resulting pitch strengths are then added to a pitch spectrogram.

1.2 Tones and Midi Notes

A new tone is started, if the most salient pitch passes the magnitude threshold. The added partials are used to establish a timbre representation. The spectral envelope of the tone, determines how much each partial of the current frame will influence pitch and magnitude of the tone. This way the impact of noise and other sound sources can be decreased noticeably. After pitch and magnitude estimation, the tone height is predicted for the next analysis frame.

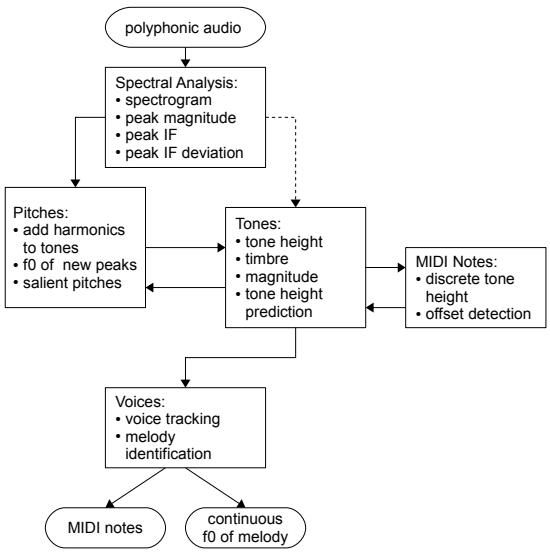

Figure 1. Algorithm Overview

Moreover, the estimation of the discrete tone height (e.g. MIDI note) and offset detection is performed. This information is used to provide MIDI note output.

1.3 Voices and Melody Identification

The tone objects are processed to build voices. In the scope of the melody extraction algorithm, a voice is an object that is defined by its magnitude, its frequency and the frequency range. A rating is calculated for each tone depending on loudness, frequency dynamics, tone salience and tone to voice distance. A tone is assigned to a voice, if it passes the magnitude threshold and if it lies within the voice's frequency range. In competitive situations, the tone with the maximum rating is elected. Finally, the melody voice must be chosen. In general the most salient auditory stream is identified as the melody. If no clear decision can be taken, the voice magnitudes are weighted according to their frequency.

2. REFERENCES

[1] K. Dressler, "Audio Melody Extraction for MIREX 2009," in *5th Music Information Retrieval Evaluation eXchange (MIREX)*, 2009, available online http://www.music-ir.org/mirex/2009/results/abs/kd.pdf