MELODY EXTRACTION USING HARMONIC MATCHING

Vishweshwara Rao
Indian Institute of Technology Bombay
vishu@ee.iitb.ac.in

Preeti Rao
Indian Institute of Technology Bombay
prao@ee.iitb.ac.in

ABSTRACT

This extended abstract describes our submission to the MIREX 2009 evaluation task on Audio Melody Extraction. This system has specifically been designed for vocal F0 extraction in the presence of harmonic interference. It utilizes a spectral harmonic-matching pitch detection algorithm (PDA) followed by a computationally-efficient, optimal-path finding technique that tracks the melody within musically-related melodic smoothness constraints. An independent vocal segment detection system then identifies audio segments in which the melodic line is active/silent by the use of a melodic pitch-based energy feature.

1. INTRODUCTION

The problem of melody extraction from polyphonic audio, involving the detection and extraction of the pitch contour of the lead melodic instrument, has received considerable attention from researchers in the past; as reflected by the large number of entries for audio melody extraction task in the MIREX 2005, 2006 and 2008 evaluations. However melody extraction is still not a solved problem. In order to assess the performance of current systems this task has been proposed for the 2009 MIREX evaluations as well.

Our submission to the 2009 audio melody extraction task was primarily developed for north Indian classical vocal music. However it was also found to perform well on publicly available western music datasets (at http://www.ee.columbia.edu/projects/melody/). A brief description of the modules in our system is presented here.

2. ALGORITHM DESCRIPTION

The algorithm comprises two modules, a melody extraction module, which estimates a melodic pitch in every frame of audio, followed by a voice detection module that uses the estimated melodic pitch to determine whether the melodic voice is actually present or not.

2.1. Melody Extraction

The core pitch detection algorithm (PDA) used by the melody extraction module is based on the Two-Way Mismatch (TWM) method [1]. The TWM PDA falls under the category of harmonic matching PDAs that are based on the frequency domain matching of measured spectrum with an ideal harmonic spectrum.

The inputs to the melody extraction module are the magnitudes and frequencies of detected sinusoidal components. These are detected using a slightly relaxed main-lobe magnitude matching criterion [2] on local maxima detected from the magnitude spectrum, which is computed using a high resolution FFT with a fixed data window length of 30 ms.

Unlike typical harmonic matching algorithms that maximize the energy at the expected ideal harmonic locations, the TWM PDA minimizes a spectral mismatch error that is a particular combination of the energy of the partial and its frequency deviation from the ideal harmonic location. The error is computed by comparing the measured peaks in the signal spectrum with a predicted harmonic spectral pattern for each candidate F0. The TWM PDA has been found to be more robust to sparse, but strong, harmonic interferences as compared to other harmonic matching PDAs [3], which makes it suitable for melody extraction of a harmonically rich voice in the presence of tonal accompaniment. In the interest of computational efficiency, the present implementation of the TWM PDA computes the TWM error only at possible trial candidate F0s that are pre-computed from the list of measured sinusoidal components [4] and fall within the F0 search range (between 100 and 1280 Hz).

The TWM PDA is operated within the framework of dynamic programming-based (DP) smoothing. DP uses a combination of suitably defined local measurement and smoothness costs into a global cost, which is optimized over a continuous voiced segment. Here the measurement cost is the TWM error, normalized to lie in the interval [0, 1]. The smoothness cost is derived from the distribution of inter-frame pitch transition values over a training dataset of clean, sung-voice pitch contours [5].

2.2. Melodic Voice Detection

The TWM-DP PDA produces an estimate of the predominant pitch at every instant irrespective of the underlying signal content. For any useful representation of the melody, it is necessary to find a means to automatically detect frames where the melodic voice is indeed present.

We use a measure of voicing based on the signal energy associated with the predominant pitch estimate, called normalized harmonic energy (NHE) [3]. It is defined as the sum of the energies of individual harmonics corresponding to the predominant pitch. All frames are labeled as having/missing the melodic voice by applying a static threshold to this feature.

The frame level labels are further smoothened over homogenous segments as determined by an automatic boundary detection method [6] based on detecting abrupt but stable changes in the harmonic energy feature. Grouping is done by a process of majority voting, i.e. the segment assumes the label of that class (voiced/unvoiced) into which the majority of the frames in that segment have been classified.

ACKNOWLEDGEMENTS

The authors would like to express their heartfelt thanks to the group at IMIRSEL for their considerable effort to run the independent evaluations. We would also like to thank the NCPA, Mumbai for giving us permission to contribute their Indian classical music data to the evaluation.

REFERENCES

[1] R. Maher and J. Beauchamp, "Fundamental Frequency Estimation of Musical Signals using a Two-Way Mismatch Procedure," *J. Acoustical Soc. America*, vol. 95, no. 4, pp. 2254-2263, 1994.

[2] D. Griffin and J. Lim, "Multiband Excitation Vocoder," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, no. 8, pp. 1223 – 1235, 1994.

[3] V. Rao and P. Rao, "Vocal melody detection in the presence of pitched accompaniment using harmonic matching methods," in *Proc. of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, 2008.

[4] P. Cano, "Fundamental frequency estimation in the SMS analysis," in *Proc. of COST G6 Conf. on Digital Audio Effects 1998*, Barcelona, Spain, 1998.

[5] A. Bapat, V. Rao and P. Rao, "Melodic contour extraction of Indian classical vocal music," in *Proc. Intl. Workshop on Artificial Intelligence and Music (Music-AI '07)*, Hyderabad, India, January 2007.

[6] J. Foote, "Automatic audio segmentation using a measure of audio novelty," in *Proc. IEEE Intl. Conf. Multimedia and Expo (ICME)*, vol. 1, pp. 452-455, 2000.

[7] G. Poliner, et. al., "Melody Transcription from Music Audio: Approaches and Evaluation," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no.4, pp. 1247-1256, May 2007.