MULTIPLE-F0 ESTIMATION FOR MIREX 2011

Chunghsin Yeh and Axel Roebel
IRCAM/CNRS-STMS Paris, France
Analysis-Synthesis team

ABSTRACT

This extended abstract describes the system proposed for MIREX (Music Information Retrieval Evaluation eXchange) 2011 in the **Multiple Fundamental Frequency Estimation and Tracking** contest. The submitted system is based on the one submitted in 2010 with some modifications.

1. INTRODUCTION

The system is based on a frame-by-frame analysis with a tracking mechanism as post-processing. Two versions with slightly different candidate extraction methods are submitted. There are four results generated for the first sub-task **frame-by-frame evaluation**: two candidate extraction methods with or without tracking, and two results generated for the second sub-task **note contour evaluation**: two candidate extraction methods with or without note correction.

2. NOISE LEVEL ESTIMATION

Noise level estimation serves to distinguish the sinusoids from noise in the observed spectrum. The distribution fit of the residual spectrum is based on the magnitude distribution of spectral bins. The combined skewness of noise and sinusoidal bins are used as the statistical measure to test against the skewness of Rayleigh distribution. Thresholds for the skewness test and the residual energy are coherently derived. Based on the evaluation of our submission last year, the number of sinusoids to remove at each iteration does not influence the results much and therefore we set it to 20 for better efficiency.

3. JOINT EVALUATION OF F0 HYPOTHESES AND POLYPHONY INFERENCE

Similar to the candidate extraction method described in [1], the NHRF0s (non-harmonically related F0s) are iteratively extracted and all their harmonics are extracted as HRF0s (harmonically related F0s). We also propose another version of candidate extraction which is slightly different in dealing with repeated selection. Given a set of F0 hypotheses, the hypothetical sources are constructed by partial selection and overlap treatment and the related combination is evaluated by a score function. To infer the number of the sources, we follow the iterative combination method proposed in [2].

4. NOTE CONTOUR TRACKING

Instead of tracking the *intermediate F0 estimates* (frame-based estimation), it is proposed to connect the F0 candidates across the frames to establish *candidate trajectories* [3]. The reason to establish candidate trajectories beforehand is that candidate trajectories are more complete, which provides a good initial estimate of the source streams. However, a candidate trajectory may accidentally connect several notes of the same pitch when, for instance, the notes falling at a harmonic of a NHRF0. Therefore, the connection is constrained by the intermediate F0 estimates. Due to the frequency resolution limit at low frequencies, a simple correction is applied to further segment the trajectories with respect to the MIDI note numbers.

5. REFERENCES

[1] Yeh, C., Roebel, A., Rodet, X. “Multiple fundamental frequency estimation and polyphony inference of polyphonic music signals”, *IEEE Trans. on Audio, Speech and Language Processing*, vol. 18, no. 6, August, 2010.

[2] Yeh, C., “Multiple-F0 Estimation for MIREX 2010”, *the 6th Music Information Retrieval Evaluation eXchange*. 2010.

[3] Chang, W.C., Su, W.Y., Yeh, C., Roebel, A., Rodet, X. “Multiple-F0 tracking based on a high-order HMM model”. *Proc. of the 11th Intl. Conf. on Digital Audio Effects (DAFx-08)*, Espoo, Finland, 2008.