MULTIPLE-F0 ESTIMATION FOR MIREX 2010

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) 2010 in the **Multiple Fundamental Frequency Estimation and Tracking** contest. The submitted system is based on the one submitted in 2009 with some modifications.

1. INTRODUCTION

The system is based on a frame-by-frame analysis with a tracking mechanism as post-processing. Two parameter settings for the noise level estimation and an optional note correction are submitted. There are four results generated for the first sub-task **frame-by-frame evaluation**: two parameter settings with or without tracking, and four results generated for the second sub-task **note contour evaluation**: two parameter settings with or without note correction.

2. NOISE LEVEL ESTIMATION

Noise level estimation serves to distinguish the sinusoids from noise in the observed spectrum. Different from the algorithm in [1], the distribution fit of the residual spectrum is based on the magnitude distribution of spectral bins. The combined skewness [2] 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 more coherently derived, too. Two settings for the number of sinusoids to remove at each iteration is submitted: remove 1 or 20 sinusoids at a time.

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). 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 propose an iterative combination method. The polyphony inference algorithm begins with the most likely F0 hypothesis and then progressively combined with other hypotheses. The scoring of hypothetical combinations is used to select the most plausible one and the contribution of an added F0 is verified by the individual score criteria. An F0 hypothesis is considered *valid* if it either improves the envelope smoothness of the hypothetical sources that have partials overlapping with its partials, or explains a significant amount of salient peaks. The iteration stops when the most plausible combination does not improves either of the two criteria. Different from an usual iterative approach, care has been taken to replace a subharmonic-like F0 with a more plausible one during the iteration.

4. SOURCE STREAM TRACKING

Source stream tracking aims at tracking the estimated F0s into note contours. 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 optionally applied to further segment the trajectories with respect to the MIDI note numbers.

5. REFERENCES

[1] Yeh, C., "Multiple fundamental frequency estimation of polyphonic recordings" *Ph.D. thesis, Université Paris VI*, 2008.

[2] Raiford, T. E., "Skewness of combined distributions", *Journal of the American Statistical Association*, vol. 37, no. 219, pp. 391393, September 1942.

[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.