MULTIPLE F0 ESTIMATION FOR MIREX 2007

Chunghsin YEH
IRCAM/CNRS-STMS
Analysis Synthesis team

ABSTRACT

This extended abstract describes the system proposed for MIREX (Music Information Retrieval Evaluation eXchange) 2007 in the **Multiple Fundamental Frequency Estimation and Tracking** contest. This system estimates the concurrent F0s at each single analysis frame of polyphonic signals without information from neighboring frames. Therefore, the proposed system is submitted for single frame evaluation.

1 INTRODUCTION

The proposed F0 estimation system is based on a score function which evaluates the plausibility of a set of F0 hypotheses [1]. It evaluates all possible combinations among F0 hypotheses for the concurrent source number from 1 to the maximum N_{max} determined by a threshold on score improvements [2]. Then, the best set of F0s is selected consecutively by means of two criteria on residual energy and envelope smoothness.

2 SYSTEM DESCRIPTION

The system is composed of four stages. At first, the adaptive noise level estimation [3] distinguishes the sinusoidal components. F0 candidates are then iteratively extracted until no sinusoidal components are left. The score function joint evaluates all the combinations of F0 candidates and the best set is selected together with the number of F0s.

2.1 Noise level estimation

Under the assumption that noise is nearly white within a narrow frequency band, we model narrow band noise by means of Rayleigh distribution. The process starts by subtracting spectral components classified as sinusoids [4]. The noise level is defined as the low-quefrency liftered spectrum [5]. Spectral peak reclassification (according to the estimated noise level) and noise peak distribution fit (to Rayleigh distribution) iteratively approximate the underlying noise level.

2.2 F0 candidate extraction

This stage aims at reducing the F0 hypotheses in the search range to improve the efficiency of the joint evaluation at the next stage. It is a simple iterative estimation/subtraction process with harmonic F0 extraction. The predominant F0 at each iteration is estimated by the score function and the components not explained by the harmonic model are used for the next estimation. The iteration stops when the harmonic-to-noise ratio (see definition in [5]) is below 3dB. Since the subtraction remove components around all the harmonics of an extracted F0, the extracted F0 in fact represents a harmonic group of F0s. Therefore, we propose to further detect harmonically related F0s within each F0 group by means of detecting partials disturbing the envelope smoothness.

2.3 Joint evaluation of F0 hypotheses

To construct and evaluate hypothetical sources, we follow the three physical principles for nearly-harmonic sounds:

1. Spectral match with low inharmonicity
2. Spectral smoothness
3. Synchronous amplitude evolution within a single source

These principles are formulated as four criteria: harmonicity , mean bandwidth and centroid of Hypothetical Partial Sequences, and the standard deviation of mean time of hypothetical partials. The linear combination of the four criteria forms the score function which evaluates the plausibility of each combination of F0 hypotheses.

2.4 Estimation of the number of F0s

An iterative search to infer the plausible hypothetical number of F0s has been proposed [2]. The true number of F0s is denoted as N, while the inferred hypothesis is denoted as S_M. Starting with S_1, the system iteratively evaluates the score improvements of all possible hypotheses $\{S_1, \ldots, S_M, S_{M+1}\}$, where S_{M+1} is the last hypothesis. S_{M+1} provides a score improvement (w.r.t. the score of S_M) under a threshold, which leads to the termination of the joint evaluation process. Then, the hypothesis S_M is considered as the most plausible number of F0s in the current frame. However, it is found that the threshold for the score improvement is difficult to set when a variety

of instruments are mixed in the observed signal. Therefore, we propose two criteria to find the best combination within the top-ranked combinations.

We start by ranking F0 hypotheses with their individual probabilities defined by the score criteria [2]. Starting from the best-ranked F0, we gradually add another F0 from the candidate rank list if (1) the added F0 and previously selected F0s together appears in the top-ranked combinations, and (2) the added F0 either explains more energy or improves the smoothness of hypothetical envelopes. There are two criteria involved: the additional energy explained by the added F0 and the smoothness improvements.

If the added F0 does not correspond to higher harmonics of the selected F0s, it should explain more energy. To evaluate if the added F0 helps to explain more energy, we make use of the signal-to-noise ratio of the residual components. if the decrease of this ratio is below 3dB, the added F0 is discarded.

If the added F0 corresponds to higher harmonics of the selected F0s, it should improve the envelope smoothness of the F0s corresponding to its subharmonics. The threshold for the smoothness improvement is trained on the decrease of mean bandwidth criterion of music instrument sound samples. If the added F0 helps to decrease the average mean bandwidth less than the amount that the envelope of monophonic sound may decrease while smoothing out the largest partial, the added F0 overly smooths out the hypothetical envelopes and is regarded as superharmonics

3 DISCUSSIONS

The proposed method does not make use of instrument models and thus the submitted system is tuned to disfavoring harmonically related F0s in order not to report spurious F0s. Therefore, higher F0s buried in the partials of lower F0s are expected to be missing in the estimation result. Instrument models could help to extract harmonically related F0s but model selection in complex polyphonic signals remains a challenge.

4 REFERENCES

[1] Yeh, C., Roebel, A. and Rodet, X. “Multiple fundamental frequency estimation of polyphonic music signals”, *Proc. IEEE, International Conference on Acoustics, Speech and Signal Processing(ICASSP’05)*, Philadelphia, USA, 2005.

[2] Yeh, C., Roebel, A. and Rodet, X. “Multiple F0 tracking in solo recordings of monodic instruments”, *120th AES Convention*, Paris, France, May 20-23, 2006.

[3] Yeh, C. and Roebel, A. “Adaptive noise level estimation”, *Proc. of the 9th Int. Conf. on Digital Audio Effects (DAFx’06)*, Montreal, Canada, 2006.

[4] Roebel, A. “A new approach to spectral peak classification”, *Proc. of the 12th European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, 2004.

[5] Qi, Y, and Hillman, Rebert E. “ Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals”, *Journal of Acoustical Society of America*, 102(1), pp. 537-543, 1997.