# MULTIPLE-F0 ESTIMATION FOR MIREX 2008

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

**Wei-Chen Chang**
Dep. of Computer Science and Information Engineering
National Cheng-Kung University, Tainan, Taiwan

### ABSTRACT

This extended abstract describes the system proposed for MIREX (Music Information Retrieval Evaluation eXchange) 2008 in the **Multiple Fundamental Frequency Estimation and Tracking** contest. This system is based on a frame-by-frame analysis with a recently developed tracking mechanism. It is submitted for the first two sub-tasks: (1) frame-by-frame evaluation (2) note contour evaluation.

## 1 INTRODUCTION

The proposed multiple-F0 (fundamental frequency) estimation system is composed of two parts: frame-based F0 estimation and source stream tracking. The number of sources, or *polyphony*, and the related F0s are first estimated on a frame-by-frame basis [1]. The tracking mechanism then refines the estimation of the number of source streams in a maximum likelihood manner [2]. Compared with the version submitted in 2007, the frame-based F0 estimation part has been improved, especially in the estimation accuracy of higher polyphony. This year, two versions are submitted for the first sub-task *frame-by-frame evaluation*: (i) frame-based estimation *without* tracking and (ii) frame-based estimation *with* tracking. For the second sub-task *note contour evaluation*, the tracking results are reported.

## 2 FRAME-BASED MULTIPLE-F0 ESTIMATION

The frame-based F0 estimation part is based on a score function which evaluates the plausibility of a set of F0 hypotheses [3]. It evaluates all possible combinations among F0 hypotheses for the concurrent source number from 0 to the maximal polyphony hypothesis. Then, the best set of F0s is selected progressively by means of two criteria related to the residual and the spectral smoothness. It is composed of four stages. At first, the adaptive noise level estimation distinguishes the sinusoidal components. F0 candidates are then iteratively extracted until no significant sinusoidal components are left to explain. The score function joint evaluates all the combinations of F0 candidates and the best set is selected by a polyphony inference algorithm.

### 2.1 Noise level estimation

Under the assumption that the power spectrum of noise is nearly flat within a narrow frequency band, the magnitude distribution of narrow band noise is modeled by means of Rayleigh distribution. Consequently, the noise level is modeled as a succession of Rayleigh distributions, each of which is a function of frequency. An adaptive noise level estimation algorithm has been developed to iteratively approximate the underlying noise level [4]. According to the estimated noise level, the spectral peaks are classified into sinusoids (above the noise level) and noise (below the noise level).

### 2.2 F0 candidate selection

This stage aims at select the F0 candidates in a precise and concise manner such that the number of their combinations is reduced to a reasonable amount [1]. The NHRF0s (non-harmonically related F0s) are first extracted in an iterative estimation/suppression process. Each NHRF0 represents a harmonic group of partials which do not overlap completely with the partials of the other groups. Then, HRF0s (harmonically related F0s) are detected within each harmonic group by means of detecting partials disturbing the envelope smoothness.

### 2.3 Joint evaluation of F0 hypotheses

Given a set of F0 hypotheses, the hypothetical sources are constructed by partial selection and overlap treatment. The related combination is evaluated by a score function composed of four criteria:

1. Harmonicity: harmonic matching
2. Mean Bandwidth: envelope smoothness
3. Spectral centroid: energy concentration in lower partials
4. Synchronicity: synchronous amplitude evolution within a single source

The linear combination of the four criteria forms the score function which evaluates the plausibility of a given combination of F0 hypotheses.

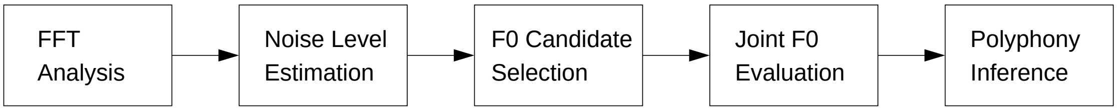

**Figure 1**. Overview of the frame-based multiple-F0 estimation system.

### 2.4 Estimation of the number of sources

The strategy is to progressively increase the polyphony hypothesis and calculate the score of all possible combinations of F0 candidates[1]. The scoring of hypothetical combinations is used to select the most plausible ones, among which the best combination is determined by iteratively verifying the related F0 hypotheses to consolidate the estimates. The estimation of the largest polyphony possible is determined by the *score improvement* [5]. All the top-five combinations of all polyphony hypotheses are retained for the consolidation of the F0 estimates.

The polyphony inference algorithm begins with listing the individual F0 hypotheses in order of their individual salience. Beginning with the most likely F0 hypothesis, each hypothesis is progressively combined with the current F0 estimates and its contribution is verified by the previously calculated score criteria. If an F0 hypothesis (to be added) is higher in frequency than the lowest one previously selected, it 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. On the other hand, if an F0 hypothesis (to be added) is lower in frequency than the lowest one previously selected, it is considered valid provided that it explains a significant amount of salient peaks. Otherwise, it is considered a spurious source that is composed of noise.

## 3 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*. The reason to establish candidate trajectories beforehand is that the connection of the intermediate F0 estimates usually form broken segments of the underlying source streams. Candidate trajectories are more complete, which provides a good initial estimate of the source streams.

A note is described by a hidden Markov model (HMM) having two states: the attack state and the sustain state. Based on a high-order hidden Markov model, the tracking of F0 candidates is carried out in a forward-backward dynamic programming scheme. The propagated weights are first calculated in the forward tracking stage, followed by an iterative tracking of the most likely trajectories in the backward tracking stage. Then, the estimation of the underlying source streams is carried out by means of iteratively pruning the candidate trajectories according to the intermediate F0 estimates in a maximum likelihood manner.

## 4 DISCUSSIONS

In MIREX 2007, some participants have included tracking mechanisms, either a simple post-processing or more sophisticated algorithms, in the submission for the frame-by-frame evaluation. It is not clear to see how the tracking mechanisms, if can be separated from the system, contribute to the overall accuracy. In MIREX 2008, therefore, we submit an improved version of frame-based F0 estimation system with a "plug-in" of tracking, which reports two results: with tracking or without tracking. It is expected to understand more clearly how much improvement has been achieved by each part of the system. In addition, the tracking algorithm allows to evaluate the estimation of note contour. However, the integration of onset detection is left as a possible improvement.

## 5 REFERENCES

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

[2] 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 International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, 2008.

[3] 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.

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

[5] Yeh, C., Roebel, A. and Rodet, X. "Multiple F0 tracking in solo recordings of monodic instruments", *120th AES Convention*, Paris, France, 2006.