

SINGING MELODY EXTRACTION IN POLYPHONIC MUSIC BY HARMONIC TRACKING



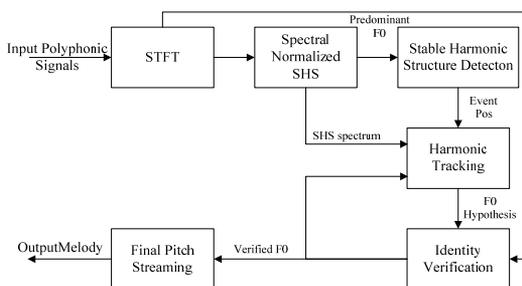
Chuan Cao, Ming Li, Jian Liu and Yonghong Yan
ThinkIT Speech Lab, Institute of Acoustics,
Chinese Academy of Sciences, Beijing



TASK DESCRIPTION

Given a polyphonic music audio, this method extracts the fundamental frequency values of the melodic music instrument (sometimes human singing).

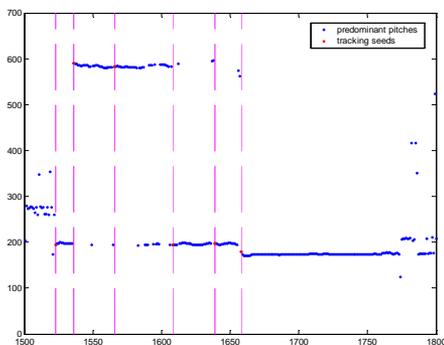
Our method is based on **Sub-Harmonic Summation (SHS) spectrum** and a novel **harmonic structure tracking strategy**.



STABLE HARMONIC STRUCTURE DETECTION

Stable condition:

- dominating the mixture
- lasting for enough long time



HARMONIC TRACKING

We use the **SHS spectrum** to track harmonic structure forward and backward.

- Taking stable harmonic structures as **track seeds**
- Tracking process goes **forward** and **backward**

CANDIDATE SELECTION

Hypothetic F0s that satisfy the following condition are selected into the candidates pool.

$$\{f_i \in F_{Cand} \mid |f_i - f_c| < \theta_c \cdot f_c, H(f_i) \geq H(f_{i+x})\}$$

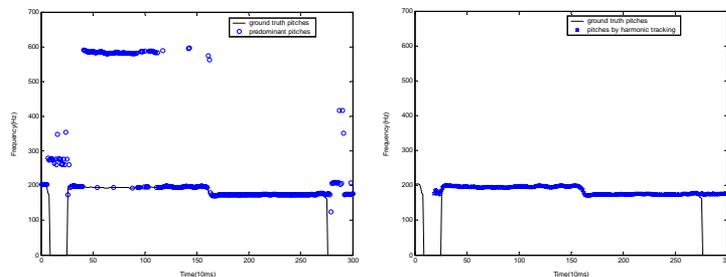
in which, f_c is the **last confirmed pitch**, θ_c is continuity factor, $H(*)$ is the **SHS** value of a certain pitch hypothesis and $x \in [-d_s, +d_s]$, d_s is the check range for **local SHS spectrum maximum**.

IDENTITY VERIFICATION

We try to use **timbre information** and calculate the **cross-correlation** between the **hypothetic harmonic family** and the **confirmed harmonic family**. Hypotheses survive if they satisfy:

$$Corr(F_H, F_C) > \theta_{Thres}$$

then F0 with the biggest saliency is chosen and the tracking process **goes on!**



EXPERIMENTAL RESULTS

Test Sets:

- Set1: **LabROSA** database
- Set2: **vocal part of Set1 + 4 pop songs** in IS-MIR04 test set

Test Set	Acc_p	Acc_f
Set1	78.30%	82.23%
Set2	74.12%	79.39%

Acc_p is the pitch accuracy of the **predominant pitch** and **Acc_f** is the accuracy of the **tracked out pitch**, the tolerance is $\pm 1/4$ tone.

CONCLUSION

The **improvement** upon predominant pitch is **3.87%** on test Set1 and **5.27%** on Set2.