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**

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

\[ \text{Corr}(F_H, F_C) > \theta_{\text{Thres}} \]

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

<table>
<thead>
<tr>
<th>Test Set</th>
<th>Acc_p</th>
<th>Acc_f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set1</td>
<td>78.30%</td>
<td>82.23%</td>
</tr>
<tr>
<td>Set2</td>
<td>74.12%</td>
<td>79.39%</td>
</tr>
</tbody>
</table>

\( \text{Acc}_p \) is the pitch accuracy of the predominant pitch and \( \text{Acc}_f \) is the accuracy of the tracked out pitch, the tolerance is +1/4 tone.

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