METHODS FOR AUDIO MELODY EXTRACTION IN MIREX 2012

Tzu-Chun Yeh
Computer Science Department, National Tsing Hua University, Hsinchu, Taiwan
kenshin.yeh@mirlab.org

Jyh-Shing Roger Jang
Computer Science and Information Engineering Department, National Taiwan University, Taiwan
roger.jang@gmail.com

I-Bin Liao
Multimedia Applications Lab. Telecommunication Laboratories Chunghwa Telecom Co., Ltd.
snet@cht.com.tw

ABSTRACT

In this extended abstract, we introduce 2 methods for audio melody extraction from polyphonic audio music. For the first submission, a hidden-Markov-model-based method[1] is applied to find the main melodies from the polyphonic audio music. For a better result and to smooth the extracted pitch, a combined method, which is the combination of HMM-based method and trend-estimation-based method[2], is applied for the second submission. The result shows that our methods achieve great degrees on the raw-pitch and raw-chroma accuracy.

1. METHOD FOR SUBMISSION TJL3

As mentioned in the abstract, we introduce a HMM-based method for the audio melody extraction. We use the maximum 2 values of the NSHS (normalized sub-harmonics summation) map and their locations (frequencies) as the features for HMM. Fig. 4(a) illustrates the scatter plot of the first two dimensions of the features, including the maximum value of a frame's NSHS and the corresponding index. Fig. 1(b) shows the corresponding NSHS curves for (a). As can be observed in the Fig. 4, the max values and the corresponding indexes are mostly likely to be around the state bin (indicated as the middle bar in both Fig. 1(a) and Fig. 1(b)). For the further information, please refer to [1] to get detailed description of the whole procedure.

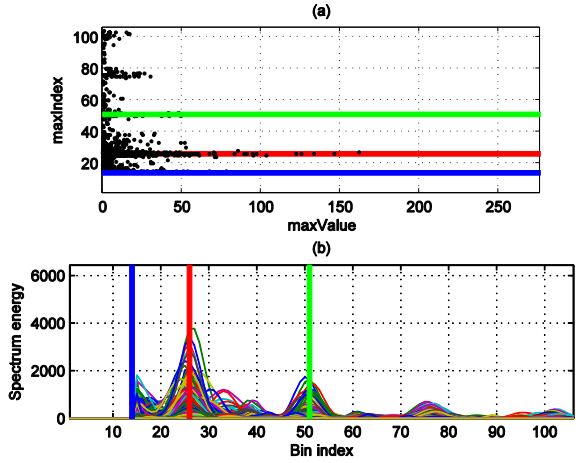

Fig. 1. Illustration of HMM's features of frequency bin 26. (a) Scatter plot of the first two dimensions of the features, including the maximum value of a frame's NSHS and the correspoding index. (b) The corresponding NSHS curves for (a).

2. METHOD FOR SUBMISSION TJL4

For further improvement, we combined 2 methods, which contain trend-estimation-based method proposed by Hsu et al.[2] and HMM-based method, into a more smoothing pitch. Fig. 2 shows the basic blocks of the combined method.

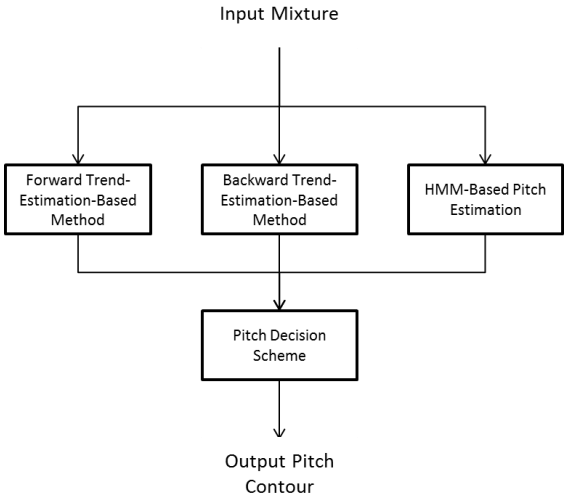

Fig. 2. The basic blocks of the submission TJL4.

We notice that pitch errors are likely to happen at the beginning of a music phrase. If we reverse the input signals in time axis and send it for pitch tracking, the pitch errors occur elsewhere. As a result, forward and backward (in time) signals are likely to generate complementary results. This observation motivates us to combines pitch contours obtained from forward and backward signals to achieve a better accuracy.

For the further information of both methods, please refer to [1][2] for the details.

3. RESULTS ON MIREX 2012

There are several different dataset for the competition of MIREX 2012 including ADC2004, MIREX05, INDIAN08, and MIREX09. The results for different submissions on these dataset are illustrated in figures below:

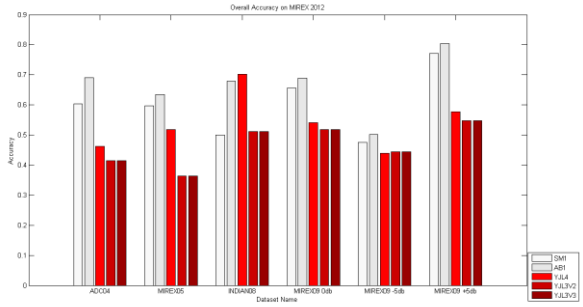

Fig. 3(a). Overall accuracy for different submissions on different dataset.

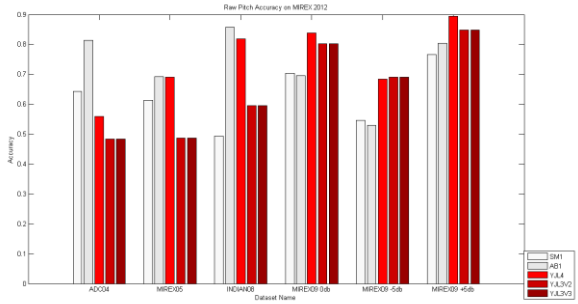

Fig. 3(b). Raw-pitch accuracy for different submissions on differt dataset.

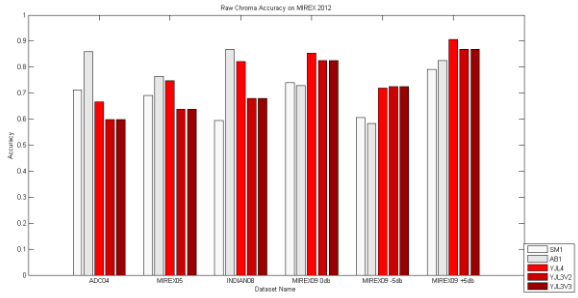

Fig. 3(c). Raw-chroma accuracy for different submissions on differt dataset.

Since our methods focus on the extraction of raw-pitch more than doing voice detection, our methods achieve good results in raw-pitch and raw-chroma accuracy. Especially on MIREX09, we have better results in 3 different Signal-to-Accompaniment Ratio (SAR) levels.

4. REFERENCES

[1] T. C. Yeh, M. j. Wu, J. S. Jang, W. L. Chang, and I. B. Liao, "A Hybrid Approach to Singing Pitch Extraction Based on Trend Estimation and Hidden Markov Models," IEEE ICASSP, pp.457-460, 2012

[2] C. L. Hsu, D. L. Wang, and J. S. Jang, "A Trend Estimation Algorithm for Singing Pitch Detection in Musical Recordings," IEEE ICASSP, pp.393-396, 2011