Polyphonic Fundamental Frequency Estimation using Score-Informed Source Separation and Probabilistic YIN Algorithm

HAJIM SCHOOL OF ENGINEERING & APPLIED SCIENCES, UNIVERSITY of ROCHESTER

Haoyu Li, Jiyuan Tian
University of Rochester

- hli73@ur.rochester.edu
- jtian7@ur.rochester.edu

UNIVERSITY of ROCHESTER

Dec 14, 2017

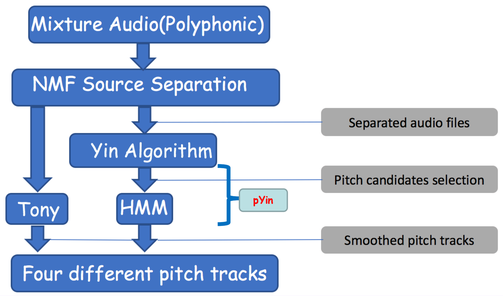
Abstract

We illustrate an innovative approach for fundamental frequency estimation on polyphonic signals. Because pYin algorithm can only operate with monophonic audios, we cannot get the fundamental frequency of polyphonic sound directly using pYin algorithm. The approach is composed of two main stages. First, we do the source separation using non- negative matrix factorization and extract the four monophonic sounds of different instruments. Then we modify Yin to output multiple pitch candidates with associated probabilities (pYin stage 1). We use these probabilities as observations in a hidden Markov model, which is Viterbi-decoded to produce an improved pitch track (pYin stage 2). The polyphonic estimation results calculated using the proposed new approach are compared with the computer-aided melody note transcription software entitled Tony to do the polyphonic pitch estimation. The results show that it has an improvement on both recall and precision rates.

1. Dataset

We use the Bach10 dataset to perform multi-pitch analysis. Bach10 dataset is a polyphonic music dataset that can be applied in versatile research projects, for example, the multi-pitch estimation and tracking, the auto-score alignment and source separations. This dataset consists of 10 pieces of four-part J.S. Bach chorales, as well as the decomposition of the four instruments of each piece. The MIDI scores , the ground-truth alignment between the audio and the score, the ground-truth fundamental frequencies of each part and their corresponding assembly for all pieces are also provided. The audio recordings of the four parts(Soprano, Alto, Tensor and Bass) were performed by the four instruments violin, clarinet, saxophone and bassoon respectively. Each of the music piece contained in this dataset has the four-part playing simultaneously, therefore, at each time instant, several pitches can be determined and tracked, which considered to be extraordinarily good for the scope of this project.

2. Overview



3. Methodology

3.1 NMF Based Source Separation

Non-negative matrix factorization(NMF) algorithm

$$V \approx W \times H$$

- Each data V is approximated by a linear combination of the columns of W weighted by the components of H.

Kullback-Leibler(KL) divergence

$$D(V||WH) = \sum_{i,j}\left(V(ij)\ln\frac{V(ij)}{(WH)_{ij}} - V(ij) + WH(ij)\right)$$

$$W_{ia} \leftarrow W_{ia}\frac{\sum_\mu H_{a\mu}V_{i\mu}/(WH)_{i\mu}}{\sum_v H_{av}}$$

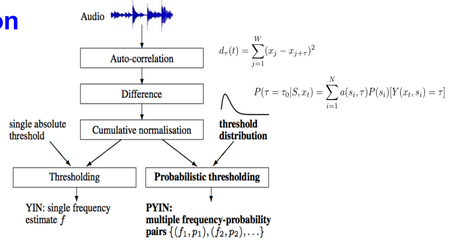
$$H_{a\mu} \leftarrow H_{a\mu}\frac{\sum_i W_{ia}V_{i\mu}/(WH)_{i\mu}}{\sum_k W_{ka}}$$

- The multiplicative update rules have great balance between speed and ease of implementation. We use the KL divergence to examine whether or not the result converge to some local minimum.
- Once the matrices get updated, normalizing W to make each column sum to 1. Scale H accordingly to eliminate the non-uniqueness issue. We have to ensure that W and H do not have zero elements in initialization.

3.2 Polyphonic Pitch Track Estimation

pYin algorithm

- These equations show a signal correlates strongly with itself when offset by the period and multiple periods.
- When the possibility value is above 0, any τ inside can produce a fundamental frequency candidate f = 1/ τ.
- By using pYin algorithm, the first stage of the output is a set of fundamental frequency candidates with their corresponding probabilities.
- After stage one is performed, a HMM-based pitch tracking is used to choose one pitch candidate per frame by uniformly divide the pitch space.



4. Experiment

4.1 A Case Study Implementing Proposed Method Using Christ- ederdubistTagundLicht Audio Piece

NMF-based Source Separation

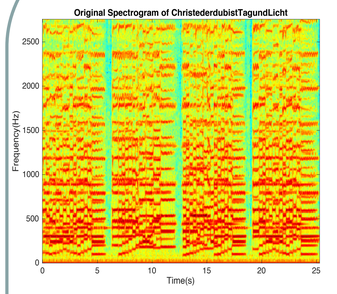


Fig1. Original Spectrogram of the Polyphonic Audio Piece

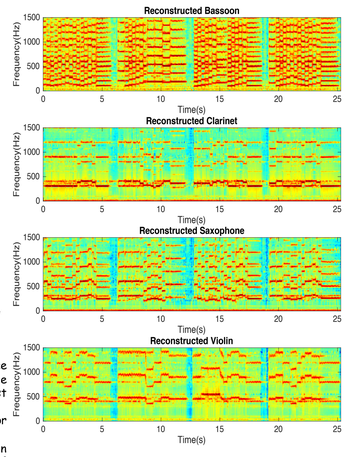


Fig2. Reconstructed Spectrogram for 4 Separated Sources

- The frame size and hop size we used in NMF source separation are 4096 and 1024 respectively. These parameters are chosen based on our comparison test results.
- The training data we used are different for different pieces of music.
- The KL divergence converges really quick, around n =30. We did several tests and used n=200 for both of our training and separation processes.
- Only the low frequency components are shown in Fig.1 and Fig.2.

Fundamental Frequency Estimations

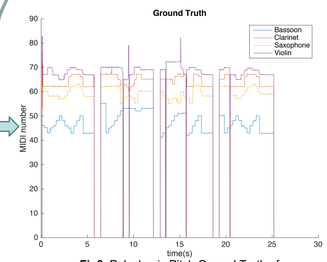


Fig3. Polyphonic Pitch Ground Truth of ChristederdubistTagundLicht

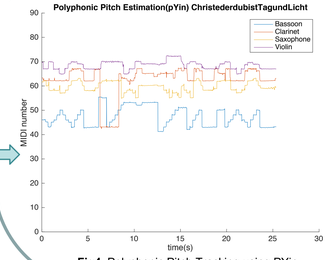


Fig4. Polyphonic Pitch Tracking using PYin Estimation

- The four monophonic audio files are individually used to perform the pYin fundamental frequency estimations .
- The critical parameters that we used to perform the pYin algorithm are $\beta = 0.15$ for the Yin threshold distribution, the onset sensitivity of 0.7, and the duration pruning threshold of 0.1.
- Tony is a melody transcription software that enables us to do the multi-pitch analysis task. It implements several melody estimation methods including the fully automatic pitch estimation and note tracking based on pYin.
- Tony automatically resamples the input source file to 44,100Hz. It uses a frame size of 2048 samples and hop size of 256 samples.
- Our results show a very promising polyphonic pitch estimation using our proposed algorithm.

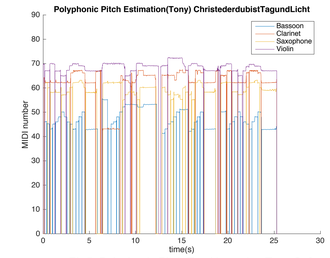


Fig5. Polyphonic Pitch Tracking using Tony Software

4.2 Optimal Parameters Determination

Instrument	Bassoon	Clarinet	Saxophone	Violin
N_{frame}, N_{hop} = 1024,512	22.74%	44.25%	59.03%	74.76%
N_{frame}, N_{hop} = 4096,1024	65.85%	91.76%	90.50%	97.73%

Table 1. Pitch Track Accuracy Comparisons

- The selection of the frame size and hop size will hugely impact on the source separation results.
- Using the optimized N_{frame} , N_{hop} , the reconstructed audio files can be clearly identified by human ears.
- The number of iterations has very minimal impact on the separation accuracies.

Instrument	Bassoon	Clarinet	Saxophone	Violin
Mean predominant f_0 Accuracy	71.80%	81.68%	93.43%	96.50%

Table 2. Mean Predominant Frequency Estimation

- The violin has the highest mean f_0 Accuracy among the four instruments.
- The pYin algorithm with selected parameters outperforms the Tony software in terms of accuracy and recall rates.

5. Statistical Study on Polyphonic Pitch Estimations

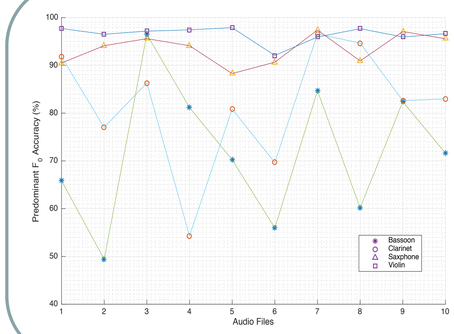


Fig6. Predominant F_0 Accuracy of the Bach10 Dataset

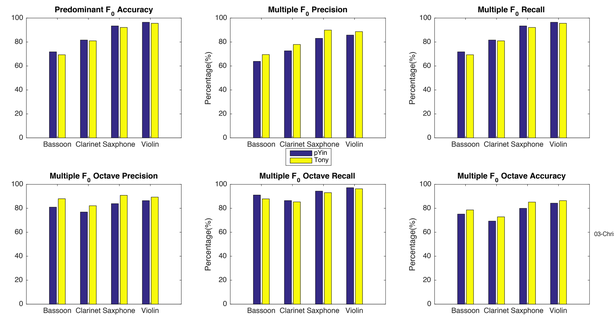


Fig7. Comparisons between pYin and Tony

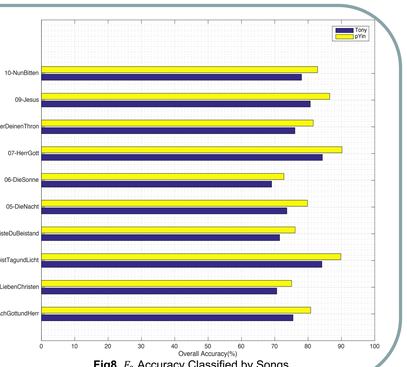


Fig8. F_0 Accuracy Classified by Songs

6. Conclusions and Discussion

- The proposed method successfully resolve the problem of fundamental frequency determination for a polyphonic audio source. It gives us an overall accuracy of 85.85% by running through the entire Bach10 dataset.
- The polyphonic pitches can not be identified correctly sometimes when the notes are rapidly changing.
- The source separation implementing NMF algorithm requires lots of computational time, the parameters could be further optimized to reduce the computational time while retaining the high polyphonic source separation accuracies.
- The low pitch source(i.e. Bassoon) from a polyphonic audio piece has lower accuracy comparing to other sources. Improving that can be the future research direction.

References:

[1] M. Mauch and S. Dixon,*pYIN: A Fundamental Frequency Estimator Using Probabilistic Threshold Distributions*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), 2014.

[2] Ewert, Sebastian, and Meinard Mller. *"Using score-informed constraints for NMF-based source separation."* Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012.

[3] A. P. Klapuri,*Multiple fundamental frequency estimation based on harmonicity and spectral smoothness*, IEEE Transactions on Speech and Audio Processing, vol. 11, no. 6, pp. 804 816, 2003.