Multiple Fundamental Frequency Estimation Based on Harmonicity and Non-negative Matrix Factorization

Gehang Li
University of Rochester
gli28@ur.rochester.edu

ABSTRACT

Multiple fundamental frequency estimation is an important but also challenging task. Since the sounds of different instruments are mixed together, it is very hard to detect different fundamental frequencies in a mono soundtrack audio. In this paper, a method for estimating multiple fundamental frequencies is proposed. The estimating process is divided into two steps. For the first step, the fundamental frequency is detected by using harmonicity. By multiply the original audio signal and the down sampled signal, the result will show a peak at the fundamental frequency. For the second step, a non-zero matrix factorization (NMF) method is applied to get a more accurate result. By using the combination of these two algorithms, the error rate for four instrument music is 34.2%.

1. INTRODUCTION

Fundamental frequency estimation is an important part of understanding sounds. In music that has only a single instrument, there are several computational methods to get the fundamental frequency, for example, the Yin algorithm[6], and non-negative matrix factorization[7]. However, for music with several instruments playing simultaneously, the task becomes difficult. Since the harmonic structure of each instrument is different, overlaps will occur in the spectra. This makes it hard to tell whether a peak in the spectra is a fundamental frequency or a fake pitch caused by the overlap of different pitches.

One way to detect multiple fundamental frequencies in a single time frame is to use harmonicity. Harmonicity means that the spectrum will show peaks, not only at the fundamental frequency, but also at its integer multiples. That property allows us to detect multiple pitches. By multiply the original signal with the down sampled signal, the peaks in the spectra indicate that there is a fundamental frequency.

Another way is to use NMF. NMF can help us to decompose the original music into different parts with its activation time and corresponding frequency property. The problem in NMF is that the number of pitches should be known before decomposing. By using harmonicity, it allows us to detect the number of pitches in a period of time and then use NMF to get a more accurate result.

2. THE METHOD

The proposed method is divided into two steps. First, harmonicity is used to get the flow of pitches. Then, NMF is used to get a more accurate result.

2.1: Step I

To use harmonicity, we first need to get the spectra of a signal, and then multiply it with its down sampled signal. As shown in figure 1, the result will show that, at each fundamental frequency, a peak will appear. The first figure in Figure 2 shows the spectra of a single instrument. It has peaks at the fundamental frequency, the second harmonic, the third harmonic and extra. When we down sample the original spectra, the second harmonic would move to the place where the fundamental frequency was, while the third harmonic would move to the place of 1.5 times of the fundamental frequency. If we multiply the original signal with the down-sampled signal, the peak at the fundamental frequency would be enhanced and other peaks would be decreased.

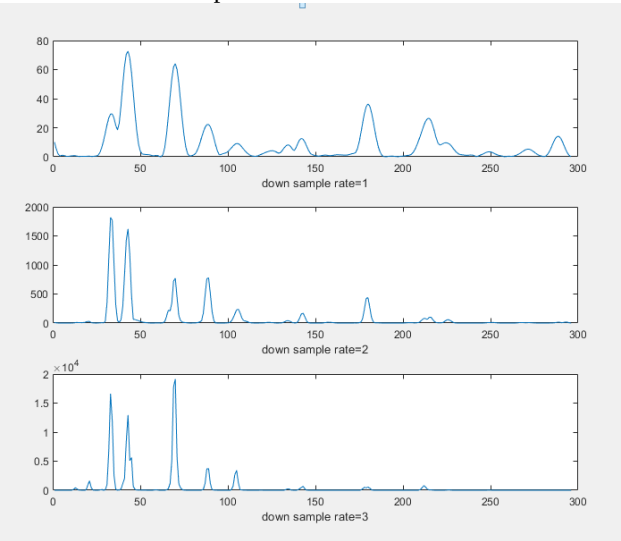

Figure 1: multiplied spectra with different down sample rate. There are 4 fundamental frequencies in total

As in figure 1, the spectra shows peaks at each fundamental frequency. However, it is also clear that for some fundamental frequencies, the peak is very low, since for some instruments the highest peak will appear at the second harmonic instead of the fundamental frequency. To avoid mistakenly detecting the harmonics, we apply a gaussian filter to increase the weight of the low frequency components. Since we already know the instruments in the music, we can increase the weight of the frequencies that are in the instrument's frequency range. Figure 2 shows the result with the designed gaussian filter.

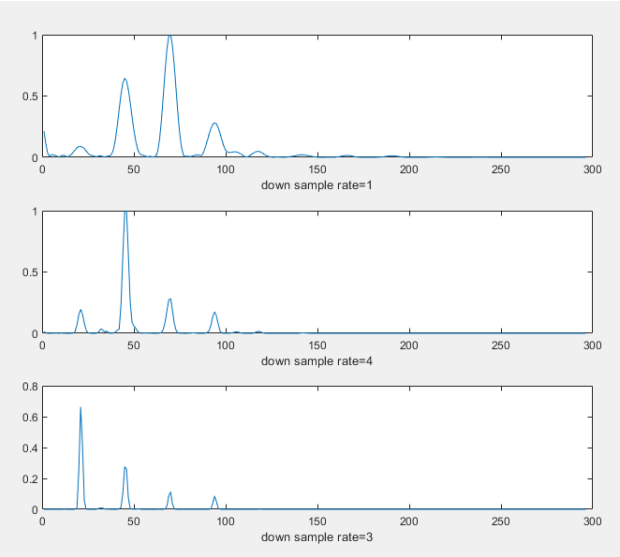

Figure 2: Spectra of a single instrument, result with designed gaussian filter

As we see, the result shows a significant increase in the height of the fundamental frequencies. By applying this gaussian filter, our result shows a decrease of error rate which indicates that it is an effective method.

2.2: Step 2

In step 1, we get an estimation of pitches in every time step. However, this estimation is not accurate enough. Since the spectra of the music has too many overlaps and the energy of each instrument is very different, our result will still have a lot of mistakes. To get a more accurate result, we will use NMF.

In NMF, a segment of music is decomposed into different frequency components with their activation. If we want to use NMF, we need to get the segmentation first. The result we get from step 1 has a lot of fluctuation, and we need to firstly smooth it.

In music, each score will last a fixed time period, which means there will be a minimum lasting time. In our dataset, this minimum would be 70ms. To smooth the spectra, we first filtered the pitches that only last for one frame and its before and next pitch are similar. These kinds of pitches can be considered as mistakenly detected pitches. Then we smooth it again by assume the segments that are shorter than 70ms are wrongly detected and attach it to its before or after segment. The result is shown in Figure 3.

From Figure 3 we can see that this smoothing strategy is a valid method. The fourth graph is the ground truth data. The first graph is the original non-smoothed data. It is messed all together. Then, by smoothing the single pitches that are different from their neighboring pitches, we get the second graph. It looks a little like the ground truth, but still has too many fluctuations. The third graph is the result of short segment smoothing. We can see that after this smoothing method, it is more similar to the ground truth, with all the transforming points (where one note is stopped and another one is played) detected, though many of them are wrongly detected. But these wrongly detected points wouldn't affect the accuracy of NMF. For the NMF method, we firstly detect every transforming point from all 4 tracks and use these points to segment the spectrogram of the music. Then, the wrongly detected points wouldn't affect the accuracy of the result but will only affect how many times we need to do the NMF computation.

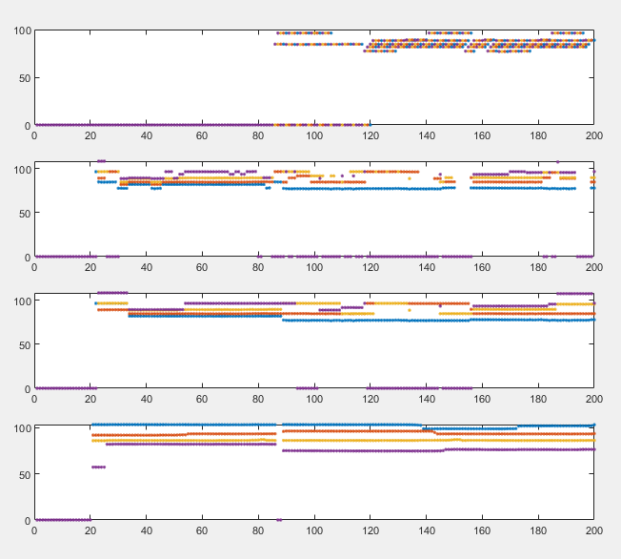

Figure 3: The result of the smoothing method. There are four instruments playing together. The first figure is the non-smoothed figure, the second figure is the result of single pitch smoothing, the third figure is the result of short segment smoothing and the last figure is the ground truth.

In the NMF process, we will initialize the W and H matrix randomly. After the decompose process, if the result is similar to the original input, then we will use it as the pitch in this frame, otherwise we will keep the original input as the pitch in this frame.

3. RESULT

3.1: The Dataset

In this paper, we use Bach10 v1.1 as our dataset. The Bach10 is a dataset that contains 10 pieces of music, and each piece is played by four instruments: violin, bassoon, saxophone and clarinet. It contains the original sound of the four instruments, the mixture and the ground truth pitch information. I heard the music myself and find that the scores in these music always last longer than 70ms, and the frequency range is 40~80 in MIDI.

3.2: The result

To analyze our method, we will show the accuracy of the non-smoothed result, smoothed result and NMF result. Figure 4 shows these results.

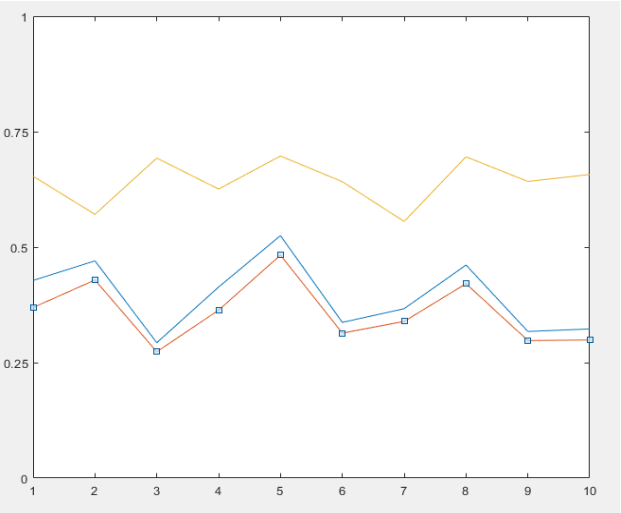

Figure 4 The accuracy of result. The x-axis is the piece number, while the y axis shows the accuracy. The yellow line is the result of NMF, the blue line is the result of non-smoothed and the orange line is the result of smoothed

From the figure, we can see that NMF is an effective way to increase the accuracy. We can also notice that the smoothed result is less accurate compared with the non-smoothed result. Though the transforming points are more accurate after the smoothing process, the value of the filtered point may be changed to an incorrect value. Since these kinds of mistakes don't affect the continuity (the position of transforming points), the NMF result isn't affected by them.

4. CONCLUSION

This paper proposed a method of multiple fundamental frequency estimation by using harmonicity and NMF. We first use harmonicity to get the continuity for future computation. To get the estimated pitches in this step, we employ a spectra filtering method to avoid the incorrect detection of the second harmonic. Then we use a NMF method to get the accurate pitch estimation. We employ a time domain pitch smoothing method to get the continuity of the pitches, and then use this continuity to decompose the spectrogram to get the final result. We achieved an error rate in 34.2% on average with music played by four instruments.

Compared with others work, our result is quite low. It is reported in [1] that the error rate can be reduced to 22%. In our method there's still much that can be improved. In the first step, the accuracy is quite low. This will cause the quality of computed continuity to be pretty low. And in the smoothed part, the smoothing window length is fixed without any musical constraint, this will also cause mistakes in continuity. In the NMF part, the method to detect fundamental frequency from the decomposed spectra components is too simple and may cause many inaccurate detections. As a project of the course, this method is still not perfect.

5. REFERENCES

[1] Abe, T., Kobayashi, T., and Imai, S. (1995). "Harmonics tracking and pitch extraction based on instantaneous frequency," Proc. IEEE-ICASSP, pp. 756–759.A.

[2] Shepp,L. A. & Vardi,Y. Maximum likelihood reconstruction for emission tomography. IEEE Trans. Med. Imaging. 2, 113–122 (1982). X.

[3] E. Vincent, N. Bertin and R. Badeau, "Adaptive Harmonic Spectral Decomposition for Multiple Pitch Estimation," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 528-537

[4] Chunghsin Yeh, A. Robel and X. Rodet, "Multiple fundamental frequency estimation of polyphonic music signals," *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, Philadelphia, PA, 2005, pp. iii/225-iii/228 Vol. 3.

[5] C. Yeh, A. Roebel and X. Rodet, "Multiple Fundamental Frequency Estimation and Polyphony Inference of Polyphonic Music Signals," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1116-1126, Aug. 2010

[6] A. de Cheveign´e and H. Kawahara, "YIN, a fundamental frequencyestimator for

speechandmusic," *J. Acoust.Soc. Am.*, vol. submitted, 2001.

[7] E. Vincent, N. Bertin and R. Badeau, "Harmonic and inharmonic Nonnegative Matrix Factorization for Polyphonic Pitch transcription," *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, 2008, pp. 109-112.