PITCH DETECTION BASED MONOPHONIC PIANO TRANSCRIPTION

Qiaozhan Gao
University of Rochester
qgao5@ur.rochester.edu

ABSTRACT

In consideration of making the creation of music score easier, music transcription walks into the engineering field since technology becomes more advanced than the old ages. Music transcription is designed to generate a music score from a simple audio recording. Although transcribing music pieces into music score is not hard for professional musicians, the process of music transcription can still save a lot of work.

The goal of this paper is to design a project to complete the entire transformation, which includes in multiple processing steps. For an integrated process, there needs to be a pitch detection method, onset detection and the final step of generating transcription. Although the topic is limited to monophonic music recordings, it is just a first trial of all kinds of music transcription.

Key Words: pitch detection, YIN, onset detection, music transcription

1. INTRODUCTION

As can be found in documentaries and historical records, musicians usually used pens to make notes of a music piece while playing along with the instrument. Decades went by, science and technology today are far more advanced than the old ages. The appearance of new technology like computers makes music transcription possible.

Generally, there are two kinds of music data: audio recordings, such as those found on compact discs, and MIDI which is widely used as symbolic music representations for computers. Audio recordings are right now more easily approachable to everyone because of the popularity of smart devices. This is also the reason of building the entire detection system on those daily recordings.

Unlike string instruments produce sounds by plucking and bowing the string, the piano is heard by the hit on the string, which is why piano is considered to be the instrument with less change on the frequency of the pitch. Therefore, piano is chosen to be the instrument for music transcription because of its stability.

The problem of music transcription can be viewed as identifying the notes played in a certain period, which means that the system need to consists the resolutions of pitch detection, onset and offset detection and automatic transcription.

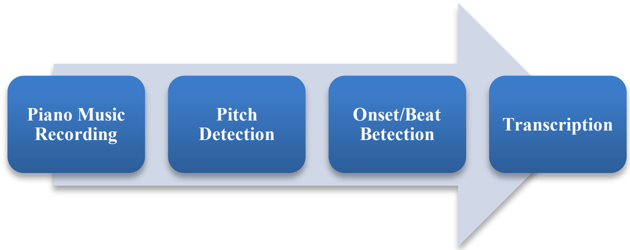

Figure 1.1 Structure of a music transcription system

Figure 1.1 illustrates the structure of this paper. The paper gives a brief overview of the pitch detection algorithm used in Section 2, and Section 3 and 4 provides an introduction of onset and beat detection separately. Section 5 clarifies the process of how to create a music score by computer. Experimental results for data sets, discussion of results and future works are provided in Section 6.

2. PITCH DETECTION: YIN

For pitch detection, the algorithm used is designed to estimate the pitch or the fundamental frequency of a signal, usually from a digital recording of speech or music.

YIN algorithm is designed for the estimation of the fundamental frequency (F0) of sounds. It utilizes the autocorrelation method with a number of modifications that combine to prevent errors. For pitch detection, taking the detection of fundamental frequency of music recording as the target, YIN is perfectly suitable for the work.

The YIN algorithm can be simplified into six steps [1]:

Step 1: The autocorrelation method

The autocorrelation function can be defined as

$$r_t(\tau) = \sum_{j=t+1}^{t+W-\tau} x_j x_{j+\tau}$$

where $r_t(\tau)$ is the autocorrelation function of lag τ calculated at time index t, and W is the integration window size. This step separates the entire wave data into multiple windows and calculates autocorrelation value of each frame.

Step 2: Difference function

The difference function of an unknown period is

$$d_t(\tau) = \sum_{j=1}^{W} (x_j - x_{j+\tau})^2$$

and searching for the values of τ for which the function is zero. There is an infinite set of such values, all multiples of the period.

Step 3: Cumulative mean normalized difference function

Replacing the difference function by the "cumulative mean normalized difference function":

$$d'_t(\tau) = \begin{cases} 1 & \tau = 0 \\ \dfrac{d_t(\tau)}{\left(\frac{1}{\tau}\right)\sum_{j=1}^{\tau} d_t(j)} & \text{otherwise} \end{cases}$$

This new function is obtained by dividing each value of the old by its average over shorter-lag values.

Step 4: Absolute threshold

Setting an absolute threshold and choose the smallest value of τ that gives a minimum of d' deeper than that threshold.

Step 5: Parabolic interpolation

Each local minimum of $d'_t(\tau)$ and its immediate neighbors is fit by a parabola. The abscissa of the selected minimum then serves as a period estimate.

Step 6: Best local estimate

Repeat detecting around the vicinity of each analysis points for a better estimate.

The six steps above are the entire process of YIN pitch detection algorithm. However, the first five steps are enough to build a pitch detection algorithm which is also used as a part of our transcription system.

3. ONSET DETECTION

The start of an acoustic recording or other sound is marked by the onset, in which the amplitude of the sound rises from zero to an initial peak. Note onset detection and localization in analyzes of musical signals is undeniably important. Even a short period of a single note contains a number of changes in the signal spectral content. Locating the position of the onset is one of the most essential part in music segmentation and music transcription. Thus, it forms the basis of high level music retrieval tasks.

Unlike other music information retrieval studies which focus more on beat and tempo detection via the analysis of periodicities, an onset detector faces the challenge of studying single events. Onset detection determines the physical starting time of the note or other musical events as they appear in a piece of music recording.

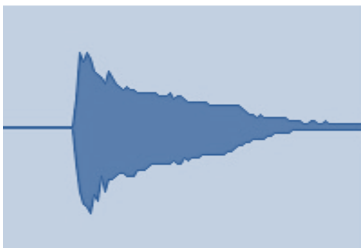

Figure 3.1 Waveform of a single note played by piano

In the waveform of a single note played by piano, there is always a sudden increase of energy at the beginning of the note, as Figure 3.1 shows. Because of the sudden increase in waveform, the onset detection is never an easy job.

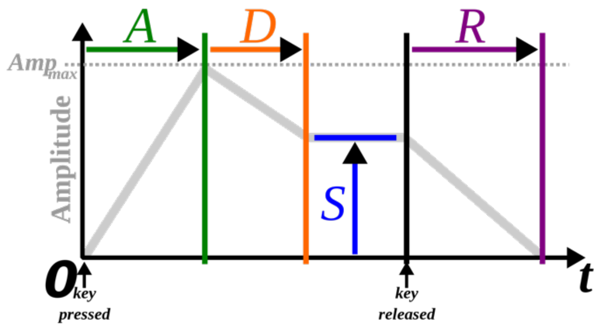

Figure 3.2 ADSR Envelope

To know about the structure of a single note, the ADSR (Attack Decay Sustain Release) Envelope should be demonstrated in the first place, which as shown in Figure 3.2. When an acoustic musical instrument produces sound, the loudness and spectral content of the sound change over time.

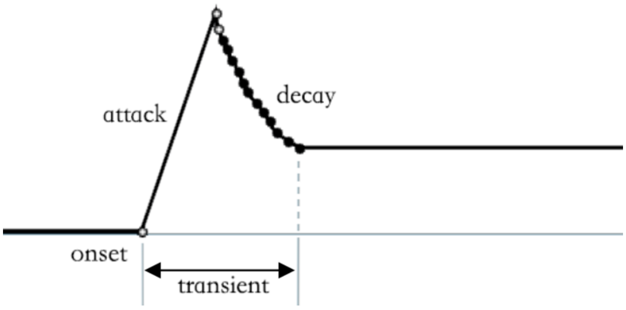

Figure 3.3 Definition of different parts in a note

Figure 3.3 shows the general structure of a single note. The attack of a note refers to the phase where the sound builds up, which typically goes along with a sharply increasing amplitude envelope. A transient can be described like part of noise where sound component is a short duration with high amplitude occurring at the beginning of a tone. Different from attack and transient, the onset of a note refers to single instant or the earliest time point at which the transient can be detected first.

An onset detector usually includes in four detection aspects: energy increases in spectral diagram, changes in spectral energy distribution or phase, changes in detected pitches and spectral patterns recognizable by machine learning techniques such as neural networks.

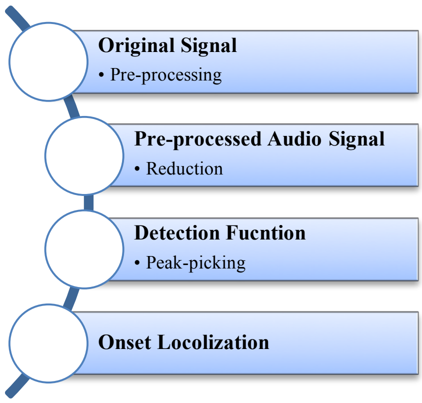

Figure 3.4 Structure of a standard onset detection algorithm

Figure3.4 shows the standard procedure for an onset detection algorithm. The original audio signal can be pre-processed to enhance the performance. The detection function can be used to derive the pre-processed audio signal at a lower sampling rate, while the peak-picking is applied to locate the series of onsets.

4. BEAT DETECTION

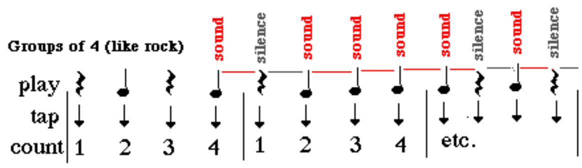

Figure 4.1 Beat definition

Beat is a sequence of instants derived from audio signal, with which the listener may tap his foot. Figure 4.1 demonstrates what beat is by tap. The selected instants generally correspond to moments in the audio signal where a beat is indicated.

In acoustic signal analysis, beat detection uses the computer to detect the beat of a music piece. Looking through previous work, the methods available for beat detection are always a tradeoff between the accuracy and the speed. Beat detectors are pretty common in music visualization such as media player plugins. The beat detection algorithms may utilize some simple statistical models whose theories based on sound energy, or they may involve in sophisticated comb filter networks.

After the pitch detection process, we can acknowledge how many frames are in this piece of recording. Within the sampling rate and the presupposed BPM (Beats Per Minute), it is not hard to figure out the relationship between frames and beats. Discarding the silence before the first note, after calculating the numbers of frames of each note, we can set the minimum number of frames to be the quarter note, and twice of its size to be a half note. If there are other types of notes, they can be done in the same manner. The detection is based on the local results among the minimum number of frames.

5. IMPLENMENTATION

5.1 Pitch Detection

Apply the YIN algorithm to the piano recording wave file. There are two simple music pieces were played for whole testing which are Twinkle Twinkle Litter Star and Yankee Doodle.

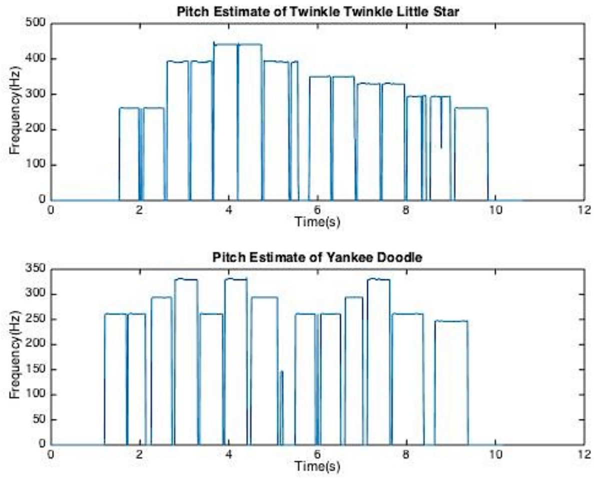

Figure 5.1 Pitch estimation result of YIN algorithm. (a) is the music piece Twinkle Twinkle Litter Star and (b) is Yankee Doodle.

These two music pieces are chosen because they are easy, simple and monophonic. Since YIN algorithm is a complicated processing method, a shorter music piece is better for repeated tests.

As shown above in Figure 5.1, although the results of estimation aren't perfectly stable, we can see that the YIN algorithm can retrieve all notes in the music piece. However, there are also some imperfections. In the pitch estimate figure of Twinkle Twinkle Litter Star, there are some little gaps inside one note, possibly because of the noise in the environment. In the figure of Yankee Doodle, there are some little peaks after one note, which can be the noise caused by hitting on the keyboard.

5.2 Transcription

In music analyzes, transcription means notating a piece of music or sound which was not previously notated. When a musician wants to create a sheet music from a recording, they write down the music notes that make up the song in notation. Automatic music transcription (AMT) is the process of converting a music signal into a musical score.

For transcription on computer, there are several different ways. First, which is also widely used, it is to generate a MIDI format file and transfer that into a music score software, such as Sibelius or Muse Score. MIDI (short for Musical Instrument Digital Interface) is a digital technical standard, which is a file format that provides a standardized method for sequences to be saved. With MIDI, the transportation and edition in other systems become possible, which also makes instruments, computers and other related devices can be connected and communicated with each other. The other way of transcription is to create the sheet music by Matlab toolbox. In this experiement, we imitate the sheet music in Matlab by drawing plots with solid and hollow dots.

5.2.1 MIDI

For creating a transcription from MIDI, primarily, there are three parameters needed to build a complete MIDI file, which are note onset, midi note number and lasting beats of the note.

First, transfer the detected pitches into midi number. The transformation needs the equation below:

$$MIDI\ number = 69 + 12log_2(\frac{f}{440Hz})$$

Table 5.1 illustrates the MIDI information results detected by the system. With those information, we can create a MIDI file of recording and then transfer into the sheet music by the music score software.

Twinkle Twinkle Little Star			
Onset time	**MIDI number**	**Note name**	**Beat (duration)**
0.00	60	C4	1
0.50	60	C4	1
1.00	67	G5	1
1.50	67	G5	1
2.00	69	A5	2
3.00	67	G5	1
3.50	67	G5	1
4.00	65	F5	1
4.50	65	F5	1
5.00	64	E5	1
5.50	64	E5	1
6.00	62	D4	1
6.50	62	D4	1
7.00	60	C4	1

Table 5.1(a) MIDI information results of Twinkle Twinkle Litter Star

Yankee Doodle			
Onset time	**MIDI number**	**Note name**	**Beat (duration)**
0.00	60	C4	1
0.50	60	C4	1
1.00	62	D4	1
1.50	64	E5	1
2.00	60	C4	1
2.50	64	E5	1
3.00	62	D4	2
4.00	60	C4	1
4.50	60	C4	1
5.00	62	D4	1
5.50	64	E5	1
6.00	60	C4	2
7.00	62	D4	1
7.50	60	C4	1

Table 5.1(b) MIDI information results of Yankee Doodle

5.2.2 Matlab Score

Use the information got in the former step, we can generate a brief music score from Matlab.

As shown in Figure 5.2, the x-axis represents the beat of the whole piece of music, while the y-axis is the MIDI note number of each beat. The solid dot in the plot means a quarter note while the hollow dot represents a half note. Because of the choose of music piece, quarter and half notes are enough to demonstrate a music score in this situation.

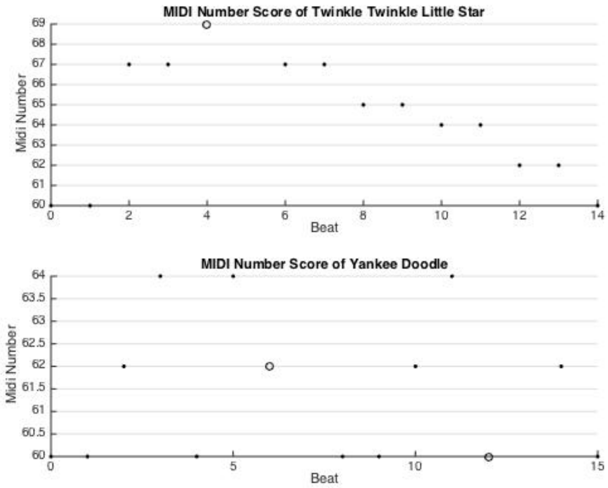

Figure 5.2 MIDI number score generated by MATLAB. (a) is Twinkle Twinkle Litter Star and (b) is Yankee Doodle.

5.2.3 Music Score

Figure 5.3 Music Score comparison between original music piece and the result gained from the transcription system. (a) is Twinkle Twinkle Litter Star and (b) is Yankee Doodle.

With the MIDI information and Matlab Score, we can get the final result of the music sheet. Figure 5.3 shows the comparison between the music sheet of the piece of music recording should be and the actual result got from the experiment. The difference between expected result and experiment result of each trial is shown with the blue-color notes which makes the comparison much clearer. From Figure 5.3, we can see the system generally works well with accurate pitch detection and beat detection, although there are also some small detection faults when the duration of one note changed.

5.3 Result Analysis

Form the MIDI information and music score, we can obviously find out the advantage and shortage of this system.

After multiple tests of different recordings of one same music piece, we can conclude average ratio of results into the table below.

Piano Piece	Note Accuracy	Beat Accuracy	Onset Accuracy	Offset Accuracy
Twinkle	96.88%	90.63%	100%	50%
Yankee	91.67%	95.83%	100%	66.67%

Table 5.1 Average detection accuracy ratio of different music recordings

With results shown in Table 5.1, we can know that the system is satisfied with note and beat detection, as well the great job in onset detection. However, the offset accuracy still needs to be improved.

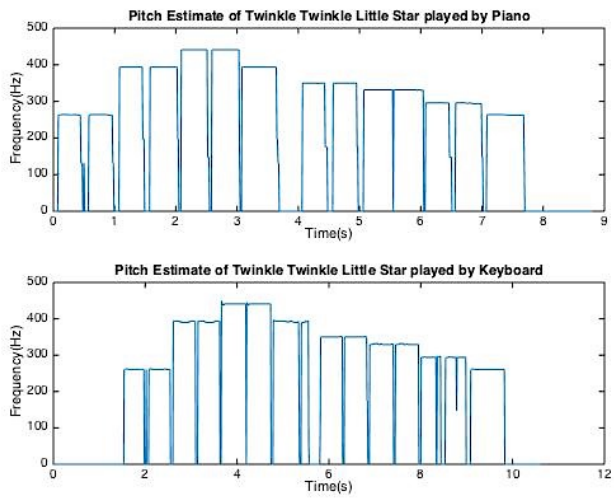

Figure 6.1 Pitch Estimate of same music piece played by different instruments. (a) is played by Original Piano and (b) is played by Electronic Keyboard.

6. DISUCUSSION AND FUTURE WORK

The project is currently just a special case of all kinds of music transcription, which means that it is not practical enough to implement to all kinds of recordings. In consideration of building a better system, in the future, there are more work can be done to make the system more robust.

6.1 Signal Pre-processing

Besides the pre-processing of noise reduction which can be done by computer, there are also some features which are better to be pre-done before the recording.

Unlike violin, guitar those string instruments, piano generates sounds by the hit on the string. When you play the piano, you need to hit the key, and then the hammer connected to the key will strike the string. With string's vibration, the signal was transmitted to soundboard and starts to resonate and provides with a steady and sustained music sound. In consideration the structure of producing sound, it is piano that the instrument with less change on pitch.

Although piano is stable on playing a tune, there are still few influences needed to be think about.

In order to build up a clean analysis of the music piece, there should be no use of piano pedal. Undeniably, the piano pedal can make the music piece sound more charming and beautiful, however, it will also build up a sound effect which is hard to eliminate with Matlab processing.

What is more, it is better to use a piano to play rather than a keyboard. The different results played by different instruments are shown in Figure 6.1. As can be seen, electronic keyboard will generate more noise than the original piano. The reason of this is the material of the key of a keyboard is usually plastic which can make some noise at the moment the finger hits the key.

When making a recording, the environment is also an important factor that needs to be checked. Lots of studies show that we can also operate pitch detection on human speech. That is why we should avoid speaking sound or natural noise when we are recording something. As a result, a more silent and quieter environment is always a better choice to record.

6.2 Polyphonic Music

The pitch detection method from this project is YIN algorithm, which can only be used for monophonic music because the detection of pitches can only analyze one frequency at a time. It will definitely be great if this system can be applied to polyphonic music recordings. To that kind of concern, the pitch detection method of the system needs to be modified to a different one such as HMM or SVM or other multi-pitch detection methods.

7. REFERRENCE

[1] Alain de Cheveigne, Hideki Kawahara: "YIN, a fundamental frequency estimator for speech and music", *Acoustical Society of America*, Vol. 111, No.4, 2002.

[2] Daniel P.W. Ellis: "Beat Tracking by Dynamic Programming", 2007.

[3] Juan P. Bello, Laurent Daudet and Mark B. Sandler: "Automatic Piano Transcription Using Frequency and Time-Domain Information", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol.14, No.6, 2006.

[4] Emmanouil Benetos, Simon Dixon: "Joint Multi-Pitch Detection using Harmonic Envelope Estimation for Polyphonic Music Transcription", *IEEE Journal of Selected Topics in Signal Processing*.

[5] Graham E. Poliner and Daniel P. W. Ellis: "A Discriminative Model for Polyphonic Piano Transcription", *EURASIP Journal on Advances in Signal Processing*.

[6] Graham E. Poliner, Daniel P.W. Ellis, Andreas F. Ehmann, Emilia Gomez, Sebastian Streich, Beesuan Ong, "Melody Transcription from Music Audio: Approaches and Evaluation".