Recognition of Monophonic Musical Notes Using Short-time Autocorrelation Estimate

L. IHSAN YAZICI, MURSEL ONDER and AYDIN AKAN
Department of Electrical and Electronics Engineering
Istanbul University
Avcilar, 34320, Istanbul
TURKEY

Abstract: - In this paper, we develop a method to obtain pitch frequency of monophonic musical signals. The auto-correlation function is used as the main feature to discriminate the notes. Acoustical signals are recorded from an electronic piano, digitized and stored on a computer. Feature extraction, i.e., the short-time autocorrelation computation, is performed and then notes are recognized by using the proposed peak search method. Examples are presented to illustrate the performance of our method.

Key-Words: - Monophonic music transcription, Musical note recognition, Acoustic signal processing.

1 Introduction

In this paper it is trained to obtain pitch values from monophonic musical signals. Monophonic music means that the performer is playing one note at a time. More than one instrument can be played, but their sounds must not overlap. In this the sound is characterized by only one pitch [1]. This is one of the important process of musical transcription.

Musical transcription of audio is the process of taking a sequence of digital data corresponding to the sound waveform and extracting from it the symbolic information related to the high level musical structures that might be seen on a score [2]. In a very simplistic way, all the sounds employed in the music to be analysed may be described by four physical parameters, which have corresponding physiological correlates [3]:

1) Repetition rate or fundamental frequency of the sound wave, correlating with pitch.
2) Sound wave amplitude, correlating with loudness
3) Sound wave shape, correlating with timbre
4) Sound source location with respect to the listener correlating with the listener's perception

The latter is not considered determinant for music transcription. The other three generate the difference between the parts that can be defined in a musical track [4]: the orchestra and the score.The orchestra is the sound of the instrument itself, the specific characteristics of the instruments (timbre, envelope), which make it sound unique; the score consists of the general control parameters (pitch, onsets, etc), which define the music played by the instrument. In an academic music representation, just the latter can be described*, i.e. which notes to play and when to play them.* In this work only "pitch detection" is studied.

2 Acoustic Features of Music

Since the musical sounds are formed from a sequence of selected frequencies, it can be said that each note played is periodic in stationary parts (attack and decay intervals are not stationary). Time domain methods are based on this *periodicity* information and aim to detect the fundamental frequency value since 1/*period* gives the desired value. Since the periodicity is an important feature for signal processing, autocorrelation of a signal can be used easily and efficiently to obtain the period information. There are several reasons why autocorrelation methods for pitch detection have generally met with good success. The autocorrelation computation is made directly on the waveform and is a fairly straightforward computation. The autocorrelation computation is largely phase insensitive and is simply amenable to digital hardware implementation [5] also.

2.1 Autocorrelation of Musical Signals

Given a discrete time signal x(n), defined for all n, the autocorrelation function is generally defined as

$$R(T) = \lim_{M \to \infty} \frac{1}{2M+1} \sum_{n=-M}^{M} x(n)x(n+T) \qquad (1)$$

The autocorrelation function of a signal is basically a transformation of the signal, which is useful for displaying structure in the waveform. Thus, for pitch detection, if it assumed that *x(n)* is exactly periodic with period P, i.e., *x(n)=x(n+P)* for all n, then it is easily shown that

$$R(T) = R(T + P), \quad (2)$$

The autocorrelation is also periodic with the same period. Conversely, the periodicity in the autocorrelation function indicates the periodicity in signal as can be observed in Figures (1-a) and (1-b). The piece is played with flute and recorded with a microphone.

For a signal that is not stationary, the concept of a long-time autocorrelation measurement as given in Eq. (1) is not really meaningful. The music signals are almost periodic in stationary parts of the played notes and can be called quasi-periodic. Thus, it is reasonable to define a short time autocorrelation function

$$R(T) = \frac{1}{N}\sum_{n=0}^{N-1} s(n)s(n+T) \quad (3)$$

where, *s(n)* is a windowed frame of the signal of length N. So, by using the short-time autocorrelation, the lag between the peaks gives the period of the signal, which is the inverse of the fundamental frequency. [6]

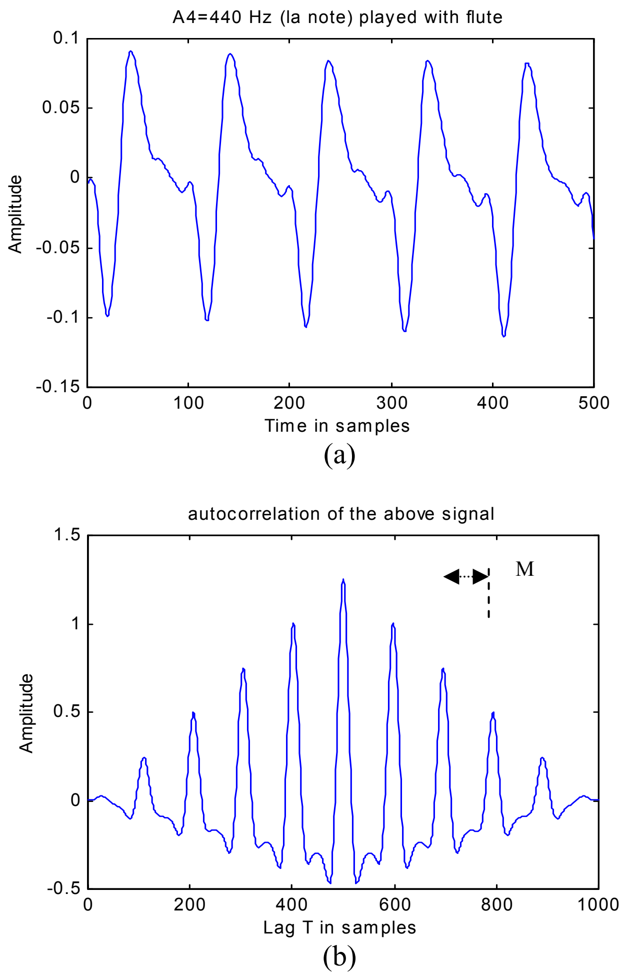


Figure 1. a) The original signal b) The autocorrelation function of the signal, the interval M between the peaks gives the period of the signal

2.2. Window Length Selection

The musical signals, which are transcribed, are sampled at F_s = 44100Hz. The lowest note **C1** has a fundamental frequency value near to 32.75 Hz. So the largest period to be detected is:

1/32.75=0.0305344 s =30.5344 ms

Since at least two periods of signal are required to use the time domain methods (conventional autocorrelation, narrowed autocorrelation, AMDF), the shortest window length can be:

2*30.5344ms=61.0688 ms

36.4ms*44100=2693.13 samples

So the shortest window length, N, can be at least 2694. Most professional musicians can play 16^{th} notes at speeds up to 120bpm (beats per minute). Going one step beyond this, to 32^{nd} notes, at 120bpm corresponds to 960 32^{nd} notes in one minute, roughly one 32^{nd} note, every 16^{th} of a second, which is much faster than most people can play.

1/16=62.5 ms.

62.5*44100=2756.25 samples

As can be seen, if the shortest window size is chosen as N=2750 (higher than 2 period of lowest frequency in the note range, and can evaluate the shortest note duration) the time/frequency resolution requirement is satisfied by the time domain methods. [6]

3 Musical Signal Pitch Tracking Using Autocorrelation

In our study, musical samples are played with electrical piano which has 5 octave and recorded by using windows 'sound recorder' at 44.1 kHz sampling , 16 byte encoding rate. Because of microphone usage (no direct input from instrument to sound card), some noise are included. Recorded samples are filtered with low pass 'elliptical' filter (4^{th} order 0.1, 40, 4 kHz cut off).

The autocorrelation of musical signals is computed with zero lag. Window length is choosen 3000 samples for computational and training simplicity, i.e. sometimes it is necessary to select a window for analysing its content onset, offset,and autocorrelation features characteristics. It can be decreased to 2750 samples as recommended, without any suspection.

The autocorrelation results are normalized to 100 maximum value. This will be explained later. As explained before , autocorrelation with zero lag gives a symmetric result, and it has maximum at self compared point (zero slided point). Peaks are expected after or before this maximum, with respect to periodicity of waveform (Analysing only one side is sufficient). According to harmonic content of musical signals different variations of peaks are observed. These can be categorized as:

1) Only peaks having desired periods and no hormonic related peaks, such as in Fig. 2.

2) Peaks having desired period and harmonic related peaks between these desired peaks. Different variations may be observed with respect to played note range and instrument timbre (i.e. more than one harmonic related peaks, greater peaks than desired peaks due to stronger 1^{st} or 2^{nd} harmonics). Examples are given in Figures 3, and 4.

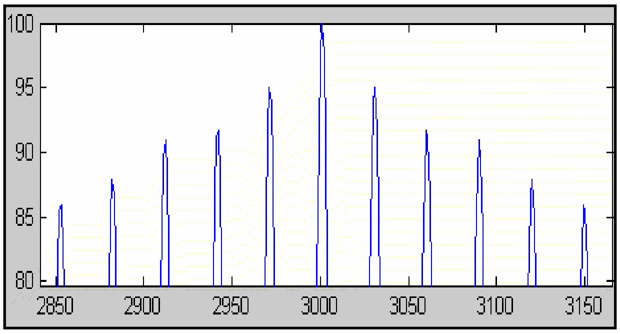


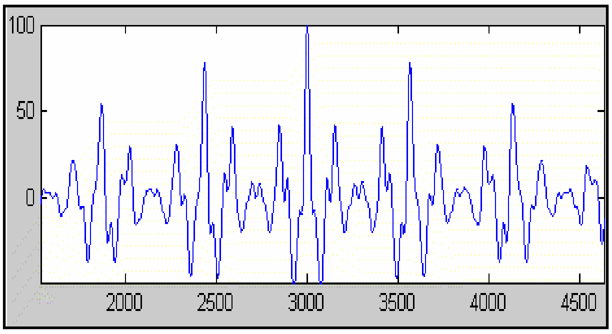
Figure 2. Autocorrelation output containing peaks at desired periods.

In our study, a peak search algorithm is used to estimate the fundamental or pitch frequency of the played notes from the autocorrelation sequence. The peaks mentioned in category 1 are observed at 3^{th} octave of 5 or more octave electrical piano. It is easy to analyse: the distance between maximum and first peak gives period of played note (T) and (1/T) is fundemental frequency of it.

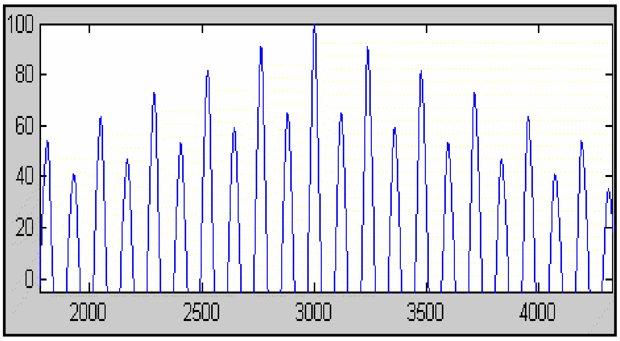
The peaks mentioned in category 2 are observed at 2^{nd} octave of 5 octave electrical piano and lower one.
At this stage it is necessary and important to mention an experimental knowledge about peaks amplitude level comparised to maximum peak. Fig. 4 shows the differences between maximum and first peaks amplitude after maximum.

Starting from 95% of maximum of the peak searching level is decreased . Two peaks are searched for every 1% decrementation. The peaks mentioned at category 1 are tracked over 85. Over 85% of maximum, the length between maximum and first peaks and first and second peaks are compared. If the difference between these two lengths are small or equal to one sampling period (1/44100) it is declared that the length between maximum anf the first peaks is the period of played note (T). Then the fundamental frequency is 1/T.

It can also be declared that there is no problem breaking peak search under 85% level when first peak is tracked.



(a)



(b)

Figure 3. Harmonic related peaks between desired ones

The difference between this peak and maximum gives the fundamental period. The decrementation is choosen as 1% to avoid harmonic related peaks. If they are tracked as two peaks, length comparison is applied. If maximum and first and first and second length difference is gretaer than one sampling period (1/44100), it is declared that first peak is resulted from harmonics. The difference between the maximum and the second is taken as fundamental period T. It is permitted to track two peak for each level (between 1% amount gap), but if only one peak is detected this is preferred and desired for avoiding and overcoming harmonic reaulted peaks.

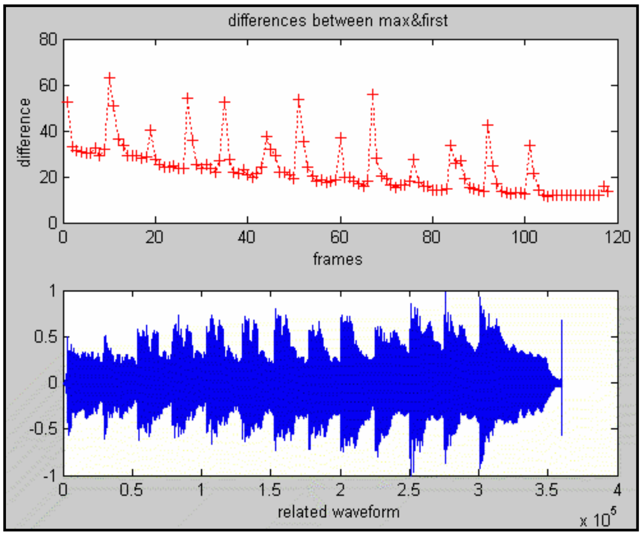


Figure 4. Maximum and first peaks difference for the first octave of elctrical piano.

It may seem as a nice approach to take peaks having GREATEST amplitude. But, waveforms have some tricks; "waving waveform", "node", "stronger harmonics" etc. The windows having these and their combinations may give higher peaks than desired one. Thus estimating the pitch by this approach might be very difficult.

4 Results and Discussion

The pitch detection results are given in this section. First, every note of each octave are played consecutively from low to high. The recording parameters are same as before. Fundamental frequency estimation results are given in the following figures. Fig. 5 and 6 show the fundamental or pitch frequency estimate of all the notes in the first and the fifth octaves.

The results given are obtained without onset usage. The related waveform added to distinguish onset points visually. Sometimes abnormal disagreements exist at starting or end points of notes. At this point it is not aimed to declare which frequency referring to which note. It is recommended that comparing agreements of frequency with respect to waveform changes are sufficient. Similar results are obtained with flute recordings played 'es' between every notes. Onset detection is also studied in our ongoing research. Figure 5 shows the estimated onset times for the fifth octave notes.

5 Conclusion

In this study we coclude that instrument characteristics are strongly important in the estimation of musical notes. Pitch detection algorithms have to be developed considering the instrument characteristics.

The peak "difference" mentioned in this study can be very useful to analyse autocorrelation and detect pitch frequency, especially instruments having onsets like piano.

References:

[1] G.Monti, M.Sandler, "Monophonic Transcription with Autocorrelation," 2000.

[2] Eric D. Scheirer, "Extracting Expressive Performance Information from Recorded Music" Master of Science Thesis, Massachusetts Institute of Technology, August 1995.

[3] R.F. Moore. "Elements of Computer Music". Prentice Hall, Englewood Cliffs, New Jersey, 1990.

[4] Eric D. Scheirer, " The MPEG-4 Structured Audio Standard", IEEE ICASSP Proc., 1998

[5] L.R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection", IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. Assp-25, No.1, February, 1977

[6] Dinc, "Automatic Transcription of Monophonic Music," Master's Thesis, METU, 2002.

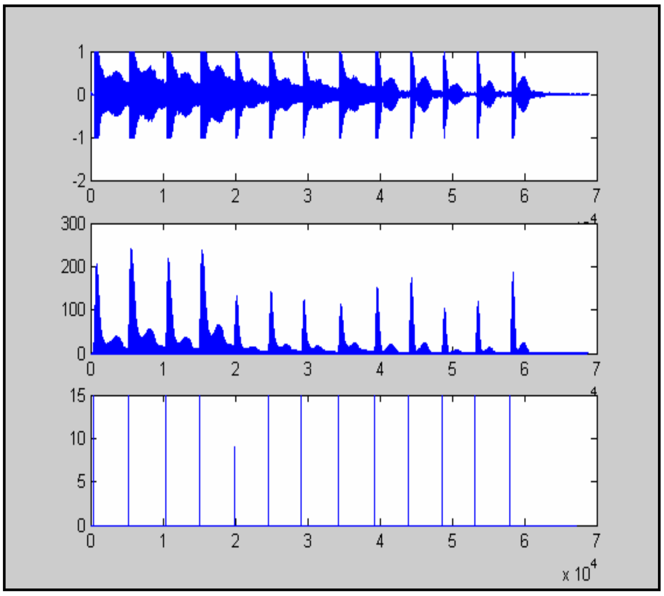


Figure 5. Onset time estimates for the fifth octave.