Overview of Homophonic Pitch Detection algorithms

Alexandre Savard
Schulich School of Music - McGill University
555 Sherbrooke St. West
Montreal, QC Canada H3A 1E3

***Abstract*— In this paper are described the most commonly encountered algorithm for pitch detection. We restrict ourselves exclusively to monophonic sound source. Both frequency and time domain algorithm are explored. The techniques presented here are zero-crossing, autocorrelation function, average magnitude difference function, weighted autocorrelation function, harmonic product spectrum, cepstrum and the teager operator function.**

***Index Terms*— pitch-tracking, pitch detection.**

I. Introduction

TO provide computers the ability to be sensitive to pitch has been a concern in music technology since the very beginning. Many efforts have been made in order to acheive an efficient algorithm for pitch detection. There exist about a hundred of different methods based on variation of some basic general ideas. However, in spite of all those attempts, none of these algorithms have been as succesful as the reference in that domain, and I said the human ear, especially in realtime context. Peoples who claim to have the absolute pitch have the ability to evaluate the frequency almost instantaneously.

We will first make a distinction between the two concepts of pitch and of fundamental frequency. Those two abstractions appear easily to be completely indistinguishable. However, they come from two contrasting research field. The pitch is a psychoacoustic concept and refers to the perceptual appreciation of the highness or lowness of a sound. The frequency is a physical attribute of signals of any nature that describe the amount of times that a repeated event occur per unit of time. Since we are mostly dealing with complex signals, the fundamental frequency is defined as the lowest component of the signal. The sensation of pitch is mainly related to the notion of periodicity.

Musical automatic transcription is probably the most obvious exemple of application for automatic pitch detection. The idea is simple, the computer has to be able to transcribe into common musical notation the pitch information in an audio content. In the same kind of idea, recent development in digital signal processing make possible to realize signal treatments in realtime. This bring the need to the computer to accomplish score following. Otherwise, the possible applications for such ability are musical queries by humming or singing, acoustic features for human-computer interaction and pitch-dependent sound editing process such as time-scaling and pitch-shifting with PSOLA algorithm.

Many pitch detection techniques present some similarities between each other or are directly derived from the same fundamental ideas. Most of the techniques can be sorted using this non-exclusive classification: voice (speech or singing), instrumental, monophonic, polyphonic, time-based or spectral-based algoritm and other alternative technique. For this paper, only algorithm for monophonic sound will be considered.

The most important criteria for realtime consideration is obviously a minimal output delay. Refering to the definition of periodicity, latency problems should be more accentuated in low frequency due to the fact that most of the algorithm needs at least more than a single iteration of the period. Fastness of the process is not only important for realtime consideration but also because several sounds present a very short sustained part. Additional requiremnents are robusteness to noise and reverberation as well as efficiency against musical expressiveness factors as we think about glissando, vibrato or thrill.

II. Time Domain Algorithm

A. Zero-Crossing Detection

A direct application of the definition of periodicity leads to the zero-crossing technique. The idea is to count the number of number of time that the signal crosses a reference level. The fundamental frequency should be, in theory, half this number. However, in practice, this technique some weakness for noisy signals or for signals with energy in high frequency.

Zero-crossing detection can be improved using an adaptive filter [Tadokoro, Matsumoto, Yamaguchi 2002]. First, the fundamental frequency is grossly estimated. The signal is then band-pass filtered by a filter centered on that frequency. Repeating this process improve considerably the algorithm..

B. Autocorrelation Functon

The autocorrelation function is described as below [Cheveigne, Kawahara 2002]:

$$\phi(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n+\tau)$$

It provides a measurement of the amount of similarity of the signal with a shifted version of itself in respect with an offset τ. Because of the the properties of periodicity, there should be a maximum of similarity at an offset value which coresponds to the period of the signal. Autocorrelation becomes less efficient in very high frequency, frequency above the range of human voice, since the period of the signal becomes shorter while the errors in frequency approximation becomes greater.

C. Average Magnitude Difference Function

The average magnitude difference function is an alternate to the autocorrelation function. Instead of loocking at the product of the signal with a time-shifted version of itself, it will compute the difference between the two signals:

$$\psi(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} |x(n) - x(n+\tau)|.$$

While auttocorelation have peaks at maximum similarity, there will be valleys in the average magnitude difference function.

D. Weighed Autocorrelation Function

The autocorrelation function and the average magnitude difference function can be combined together [Cuadra, Master, Sapp 2001]. We then obtain a more noise-robust estimation of the pitch especially in high frequency. We can define this function as following:

$$f(\tau) = \frac{\phi(\tau)}{\psi(\tau) + 1}.$$

III. Spectral Domain Algorithm

A. Harmonic Product Spectrum

The basic idea related to this technique is that partial of harmonic sounds are easily superposable by compressing the spectrum by a factor coresponding an the harmonic number [Cuadra, Master, Sapp 2001]. When multiplying those resulting rescaled spectrum we should obtain in theory a unique pick which corespond to the fundamental frequency.

$$Y(\omega) = \Pi_{r=1}^{R} |X(\omega r)|$$

In practice, octave error can occur from that algorithm due to the presence of two distinctive peaks in the result of the multiplication. Another weakness of this method comes from the fact that it is useless for non-harmonic sounds.

B. Cepstrum

One of the most popular technique for pitch tracking, the cepstrum is calculated by taking the DFT of the log of the magnitude spectrum of a signal [Cuadra, Master, Sapp 2001]. It detects the periodicity of the spectrum which means the distance between harmonic partials. This distance coresponds to the fundamental frequency and will appear to be a peak at this precise value in the cepstrum. The logarithim operation flatten the spectra so that it gives more robustness to formants. However, this same operation rises the noise level.

IV. Alternative Technique

A. Teager Operator

This technique comes from research in in signal processing for speech. In the source filter-model, voiced speech s(t) can be considered as the convolution of a pulse train p(t) with the impulse respond of the vocal tract h(t). The pulse train is produced by the successive opening and closure of the glottis. The opening/closure of the glottis result in an increase of the energy in the signal.

The Teager energy function is a non-linear operator that defines the instantaneous energy in the signal as:

$$E_{Teager}(n) = s^2 n(n) - s(n-1)s(n+1)$$

where s(n) represents a sample of the signal. It is derived from the total energy of an oscillatory spring-mass system. For each opening/closure of the glottis, there should be a peak in the teager energy function. From those results, we can estimate the fundamental frequency. It works well for synthesized voice but improvement needs to be done before using it for real voice.

V. Conclusion

As we can see, there exist a large number of different algorithms for pitch detection. Most of them are derivated from research in signal processing for speech. Each one presents some weakness and strength. It is now possible to combine them taking the result from the one we know that is the more approptiate for the analysed situation.

References

[1] Curtis Roads, The Computer Music Tutorial, p.497-533, Boston, The MIT Press, 1996.

[2] Cheveigne A. and Kawahara H., "YIN, a fundamental frequency estimator for speech and music ", The Journal of the Acoustical Society of America, Volume 111, Issue 4, pp. 1917-30, 2002.

[3] Cuadra P., Master A. and Sapp C., "Efficient Pitch Detection Techniques for Interactive Music" Proceedings of the International Computer Music Conference, 2001.

[4] Tadokoro Y., Matsumoto W. and Yamaguchi M., "Pitch detection of musical sounds using adaptive comb filters controlled by time delay", Proceedings of the international Conference on Multimedia and Expo, p.109-12, 2002.