Real-Time Pitch Tracking Algorithms Overview

François Thibault
Music Technology Group
McGill University
Montreal, Qc, Canada
thibault@music.mcgill.ca

ABSTRACT

In this short text, we wish to define clearly the requirements for real-time pitch tracking algorithms in musical applications, briefly introduce the *major* developments in the field over the last 40 years and provide some insight on what techniques might be more suitable for certain applications.

1. PITCH TRACKING REQUIREMENTS

Historically, a lot of algorithms that are used in music processing come from the research and developments operated in the speech telecommunication industry. When used in musical contexts however, a sometimes radically different set of requirements is defined, which may or may not compromise the usability of the given algorithm. Critical to the evaluation of the number of pitch tracking algorithms portrayed in the literature is thus the definition of general requirements for fundamental frequency detection in musical contexts.

First, the algorithm must be able to run in real-time using only a fraction of the computing power available, to leave some resources for actual sound transformations and processing. In many cases where synthesis is dependent on the detected fundamental frequency, the algorithm must also run with a minimal output latency to avoid a feedback effect that could confuse the performer. Note here that for an algorithm which output discrete pitch instead of just the fundamental frequency tracking there is an additional delay required to ensure a valid discrete guess (e.g. vibrato might slow down pitch recognition).

Another requirement is accuracy in the presence of noise. In offline processed algorithms, one can always come back and fix detection errors, but in a typical musical concert situation where the pitch tracker operate in real-time, mistakes in the detection process will cause an undesired synthesized result which cannot be further corrected. The musical environment setup often using loudspeakers for the audience creates important feedback problems that the pitch tracker must be able to handle correctly for proper operation.

One more constraint on pitch tracking algorithms is the required frequency resolution demands which may be very high depending on the application. Tracking pitch microvariations such as vibrato in many string instrument and human voice might require resolution to much less than one percent of the fundamental frequency for example.

Ideally, the pitch tracker should be able to operate on a wide range of input signals (e.g. different instruments or room characteristics) to be useful in most musical settings. The pitch and amplitude ranges that the algorithm is designed to handle is particularly important for musical applications.

2. TIME-DOMAIN METHODS

2.1 Autocorrelation method

The autocorrelation method is based on the fact that a periodic signal will correlate strongly with itself when offset by the fundamental period. The autocorrelation function is just a measure to which extent a signal correlates with a time-shifted version of itself. The time shift corresponding to the maximum in the autocorrelation function of the signal indicates a likely candidate for fundamental period of the signal.

One advantage of the autocorrelation function [Rabiner 1977] is that it is largely phase insensitive and is thus well suited for application where the signal might have been degraded with respects to its time evolution. Disadvantages of the techniques however are a possible confusion of the peaks because of spectral formants and poor resolution at high frequencies. Note that pre-processing (often spectral flattening) is required to achieve better results with this technique.

2.2 AMDF

The Amplitude Magnitude Difference Function is again based on the idea that a periodic signal will be similar to itself shifted by a fundamental period. The AMDF is therefore similar in concept to the ACF but rather looks at the *difference* with the time shifted version of itself. The time shifts which yields minima in the AMDF function are considered likely period estimates of the signal.

The AMDF suffers from the the same drawbacks as the ACF, but is less computationally expensive due to the use of substractions rather than multiplications and additions. [Kobayashi and Shimamura 1995] achieved more robust results by combining the ACF and AMDF functions together, than those obtained by each function used separately.

2.3 Fundamental Period Measurement

In the Fundamental Period Measurement method [Kuhn 1990], the signal is first ran through a bank of half-octave bandpass filters. If the chosen filters are sharp enough, the output of one of the filters should display the input waveform fundamental frequency freed from its upper partials (a nearly sinusoidal signal). It is then up to a decision algorithm to decide which of the filter outputs corresponds to the current fundamental frequency of the signal. Once the decision is made, it is then trivial to obtain the fundamental period by performing an inexpensive zero-crossings analysis.

Major advantages of the technique are its easy implementation and efficiency of computation. Some obvious disadvantage being its strong dependance on thresholds levels crucial for the decision algorithm. Fortunately, an automatic threshold setting algorithm was also implemented.

3. SPECTRAL-DOMAIN METHODS

3.1 Cepstrum

The cepstrum is defined as the power spectrum of the logarithm of the power spectrum of the signal and is a tool often used in speech processing. One property of the cepstrum is that it clearly separates the contribution of spectral formants and the periodicity characteristics of the signal. This is useful to solve some of the problems notably encountered in the autocorrelation method, to avoid misidentification of fundamental frequency due to the presence of strong formants in the spectrum. The strongest peak can be looked for in the excitation part of the cepstrum (high cepstral region) and the given quefrency can be related to the fundamental frequency of the input signal.

Cepstral pitch determination [Noll 1967] is a proven method that is especially suitable for signals which can be efficiently represented by a source-filter model (e.g. human voice). One disadvantage however is its relatively high computational cost due to the two cascaded Fourier transformations.

3.2 Harmonic Product Spectrum

The Harmonic Product Spectrum (HPS) approach [Schroeder 1968] measures the maximum coincidence of harmonics by simply looking at the bin amplitudes from a Fourier transform for each spectral frame (they are multiplied together). The resulting periodic correlation array is searched for a maximum which should correspond to the fundamental frequency.

Since the algorithm is highly sensitive to octave errors, it must be followed by a correction algorithm looking at the power at the octave below to ensure the correct octave of the fundamental frequency is detected. The algorithm is very simple to implement, and does well under a wide variety of conditions. The poor low frequency resolution of this technique is often compensated by zero padding the signal before taking the Fourier transform to obtain better interpolation of the low frequencies. This however augments the computational requirements of the algorithm substantially.

3.3 Constant-Q transform method

It is a well known fact that the linear scale outputs of the Fourier transformation does not match well the way we perceive audio frequencies. To better approximate this phenomena, the constant-Q transform turns the Fourier transformation into a logarithmic scale representation of frequency. A periodic signal analyzed through this procedure will yield a constant distance pattern between partials in the transform representation. [Brown 1992] used the cross-correlation of this result with a perfect comb pattern (corresponding to ideal partial positions for a given fundamental frequency) to determine the periodicity of the signal. The result is peak picked and the strongest peak corresponds to the fundamental frequency of the signal.

The complexity of the constant-Q transform was significantly reduced in [Brown and Puckette 1992] and allows a real-time implementation to be viable. The algorithm however is sensitive to octave errors and other peaks could be considered as pitch candidates.

3.4 Least-Squares fitting

All the spectral algorithms seen so far rely heavily on Fourier transformation. [Choi 1997] however wanted to overcome the delay introduced by Fourier analysis and does spectral analysis using a least-square approach. The idea is to fit sinusoids to the current signal segment and minimize the error signal to determine where are located the sinusoidal components. Strong sinusoidal components are identified as sharp valleys in the least-squares error signal. To achieve the detection of the first couple of partials, few evaluation of the error function are required and this fact keeps the computational complexity of the algorithm reasonable.

The fundamental frequency is obtained by averaging the first few partial frequencies over their partial number to obtain a better resolution of the fundamental frequency. The major advantage of the algorithm is that it can use rectangular windowing (as opposed to methods using DFT which must use more dynamic window functions), and can thus operate on shorter window lengths, providing an appreciable gain in responsiveness of the algorithm. Least-squares method is by far out performing spectral method with respect to the output latency of the technique, and is thus particularly recommended for applications with minimum latency requirements.

3.5 Maximum Likelihood

Maximum likelihood algorithms search through a set of possible ideal spectra and chooses the one which matches more closely the sound analyzed by a Fourier transformation. This algorithm was adapted to sinusoidal modelling theory [McAulay and Quatieri 1990], by finding the best possible fit for given sets of harmonic partials to the set of partials measured. With this technique, it is possible to further enhance the discrimination of pitch by neglecting small amplitude valued partials (possibly belonging to reverberation or else).

This algorithm inherits the high computational cost of sinusoidal modeling but might still be acceptable depending on the application. It provides very robust estimation of the fundamental frequency even when a significant number of partials are missing (including the fundamental).

4. OTHER METHODS

It is impossible to summarize all attempts to fundamental frequency tracking in such a small text. Other undiscussed approaches are here briefly mentioned but we are still far from having covered all major implementations.

The recent re-emergence of interest for adaptive learning systems has also been witnessed in fundamental frequency tracking with approaches using Neural Networks and Hidden Markov Models [Doval and Rodet 1991]. [Marchand 2001] came up with a promising approach using Fourier of Fourier transformation. Sub-harmonic to harmonic ratio was also recently proposed by [Sun 2002] for effective detection in noisy environments. It is also important to note that there is also an increasing number of implementations following the steps of [Gold and rabiner 1969] towards using several algorithms in parallel to increase the validity of the detection.

5. CONCLUSIONS

Although real-time pitch detection is often considered as a solved problem, there is still quite a bit of research performed on this topic to try to achieve more robust and faster implementations. It was noted that despite the abundant literature on the topic since the 1950's, complete and objective performance evaluation reviews are still missing.

6. REFERENCES

Brown, J. C. (1992). Musical fundamental frequency tracking using a pattern recognition method. *The Journal of the Acoustical Society of America 92*, 1394–1402.

Brown, J. C. and M. S. Puckette (1992). An efficient algorithm for the calculation of a constant q transform. *The Journal of the Acoustical Society of America 92*, 2698–2701.

Choi, A. (1997). Real-time fundamental frequency estimation by least-square fitting. *IEEE Transactions on Acoustics, Speech, and Signal Processing 5*, 201–205.

Cirotteau, D., D. Fober, S. Letz, and Y. Orlarey (2001). Un pitchtracker monophonique. Journes Informatique Musicales.

Dannenberg, R. B. (1984). An on-line algorithm for real-time accompaniement. In *Proceedings of the International Computer Music Conference*. International Computer Music Association.

Doval, B. and X. Rodet (1991). Fundamental frequency estimation using a new harmonic matching method. In *Proceedings of the International Computer Music Conference*, pp. 555–558. International Computer Music Association.

Gold, B. and L. rabiner (1969). Parallel processing techniques for estimating pitch periods of speech in the time domain. *The Journal of the Acoustical Society of America 46*, 442–448.

Kobayashi, H. and T. Shimamura (1995). A weighted autocorrelation method for pitch extraction of noisy speech. pp. 343–344. Acoustical Society Japan.

Kuhn, W. B. (1990). A real-time pitch recognition algorithm for music applications. *Computer Music Journal 14*, 60–71.

Marchand, S. (2001). An efficient pitch-tracking algorithm using a combination of fourier transforms. In *Proceedings of the COST G-6 Conference on Digital Audio Effects*.

McAulay, R. and T. Quatieri (1990). Pitch estimation and voicing detection based on a sinusoidal speech model. *IEEE Transactions on Acoustics, Speech, and Signal Processing 1*, 249–252.

Noll, A. M. (1967). Cepstrum pitch determination. *The Journal of the Acoustical Society of America 41*, 293–309.

Puckette, M., M. Apel, and D. Zicarelli (1998). Real-time audio analysis tools for pd and msp. In *Proceedings of the International Computer Music Conference*. International Computer Music Association.

Rabiner, L. (1977). On the use of autocorrelation analysis for pitch detection. *IEEE Transactions on Acoustics, Speech, and Signal Processing 25*, 24–33.

Schroeder, M. R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurements. *The Journal of the Acoustical Society of America 43*, 829–834.

Sun, X. (2002). Pitch determination and voice quality analysis using subuarmonic-to-harmonic ratio. *IEEE Transactions on Acoustics, Speech, and Signal Processing 1*, 333–336.