Monophonic Fundamental Frequency Extraction

Bertrand Scherrer

March 8, 2007

1 Introduction

Fundamental frequency estimation of monophonic acoustic signals is a very rich and active field of research. Similarly to a lot of signal processing techniques, fundamental estimation was first developped for, and applied to speech signals. One can literally find hundreds of methods in [Hess83]. It is also interesting to note that [Rabiner76], and [deCheveigné01] made an attempt to compare different commonly used fundamental frequency estimation techniques.

This quick review, will not present all the different methods. Instead, we will give a general overview of the topic based on the model-level typology of algortihms proposed in [Klapuri04].

2 Spectral Location Algorithms

2.1 Time-Domain Periodicity Analysis

The most famous algorithm of this type is the **autocorrelation function** (ACF) defined as :

$$r(\tau) = \lim_{N \to \infty} \frac{1}{2.N + 1} \sum_{n=-N}^{N} x(n).x(n + \tau) \tag{1}$$

A thorough presentation of this algorithm is presented in [Rabiner77]. In simple terms, the autocorrelation measures the degree of similarity between samples of a signal. This is done by multiplying[1] the signal by a time shifted version of itself, for different lags τ. In the case of periodic signals or quasi-periodic ones (like speech or music) $r(\tau)$ exhibits local maxima for lags τ corresponding to integer multiples of the period of the signal. A computationally efficient implementation of the ACF is done using the Fast Fourier Transform (FFT) algorithm where the ACF is computed as :

$$r(\tau) = FFT(|FFT(x[n])|^2) \tag{3}$$

Raising the magnitude spectrum to the second power emphasizes spectral peaks in relation to noise. On the other hand, this operation makes this method very sensitive to ‘spectral peculiarities of the target sound’ [Klapuri04] such as formants in speech signals.

2.2 Cepstrum-Based Analysis

By applying the logarithm function instead of the power of two to the magnitude spectrum in eq.3, we get the expression of the **cepstrum** [Noll67]:

$$r(\tau) = FFT(\log (|FFT(x[n])|)) \tag{4}$$

[1] the exact mathematical operation is called a convolution of the signal with a conjugate version of itself that has been time shifted and for which the time has been reversed :

$$\int_{-\infty}^{+\infty} x(t).\check{x}(\tau - t).dt \tag{2}$$

In a source-filter modeling context, the main interest of this method is that it separates the contribution of the filter from that of the source . Especially, the voice sound spectrum is the product of the glottis excitation (the source) and the vocal tract (formant filter) spectra. By taking the log of the spectrum of the sound, this product becomes a sum and the contributions of the excitation and of the filter appear clearly separated. In the case where the source is considered periodic, we have a clear peak corresponding to the period of the excitation.

There is a continuum of techniques in between but the common aspect of all them resides in the fact that they emphasize frequency partials at the harmonic locations in the spectrum.

2.3 Harmonic Pattern Matching in the Frequency Domain

Brown [Brown92] uses a filter-bank approach (using 1/24th octave filters) to yield logarithmically spaced spectral components. This means that there is a constant spacing between partials of a harmonic sound regardless of the fundamental frequency. Thus, the fundamental estimation is done by correlating the filterbank output with spectral patterns corresponding to different ideal fundamental frequencies. The mask yielding the highest correlation value indicates the value of the fundamental frequency.

A maximum likelihood spectral pattern matching fundamental estimator was presented in [Doval91, Doval93]. It consists in maximizing the likelihood of a fundamental frequency candidate given the observation of the sound partials. This technique gives a list of most likely fundamental frequency candidates along with a certain score. The presence of such a score might be useful in teh case of simple polyphony. This particular aspect will be the object of my final project for this class.

In the two-way mismatch method [Maher94] the fundamental is chosen to minimize the discrepancies between observed frequency components and harmonic frequencies generated by trial F0 values. For each of the trial values, the first mismatch is calculated as the average of the frequency differences between each observed partial and its nearest neighbour among the predicted harmonic frequencies. The second mismatch measure is obtained by averaging the frequency differences between each predicted harmonic frequency and its nearest neighbour among the observed partials. The chosen fundamental is identified as the one which minimizes the weigthed sum of the two mismatches.

The main problem with the spectral location estimators is that they do not perform well when there is inharmonicity. The spectral interval algorithms are more adapted to that kind of problem.

3 Spectral Interval Algorithms

This category of techniques is based on the fact that periodic signals have periodic spectrum of period $F0$. Their goal is to measure this periodicity in the frequency domain.

A first example is the spectrum autocorrelation method written as follows, over the positive frequencies of a K-length magnitude spectrum :

$$\tilde{r}(m) = \frac{2}{K} \sum_{k=0}^{\frac{K}{2}-m-1} |X(k)||X(k+m)| \tag{5}$$

The big difference between eq.1 and this expression is the type of data the ACF is taken on: here we consider *frequency intervals* m and not time lags. Such an approach can be found in [Lahat87] where the fundamental frequency is estimated after preprocessing of the data to flatten the spectrum (used to minimize the effect of formants) and the fundamental frequency is estimated from the computation of the ACFs on all the channels, using a decision method preventing lower octave errors.

In general, spectral interval estimators behave well with inharmonicity because, even though the interval do not remain constant, they vary less than the position of the partials. Moreover, they are more robust to high level of additive noise (unlike the cepstrum-based F0 estimation family).

4 Auditory Model Algorithms

Algorithms trying to model the human auditory system were developed from the observation that any signal with more than one frequency component exhibits periodic fluctuations, beating, in its time-domain amplitude envelope. The fundamental period can then clearly be identified as the duration between the two highest beatings.

Let us note that the time-domain amplitude envelope is obtained by operating what is called a half-wave rectification on the signal (what is negative is taken to be 0) [Klapuri04]. As a result, the spectrum of the half-wave rectified signal contains partials at the locations of the original partials as well as frequencies corresponding to the intervals between the partials. Hence, we can see that these methods allow for a compromise between spectral location and spectral interval method.

The root of these approach is the unitary auditory model proposed in [Meddis91] that can be decomposed in four main steps :

- the signal is passed through a bank of bandpass filters (40 to 128) representing the frequency selectivity of the inner ear (the channels are almost regularly spaced on a logarithmic frequency scale).
- the signal in each channel is compressed, half-wave rectified and low pass filtered to reporoduce the behaviour of the signal measured going into the auditory nerve.
- a periodicity estimation is carried out in each channel using short-time ACF.
- the autocorrelations are summed over all the channels, at each frame t :

  $$s_t(\tau) = \sum_c r_c(\tau) \tag{6}$$

  The value of the lag corresponding to the highest value of $s_t(\tau)$ is taken to be the fundamentla frequency on frame t.

The first two steps are widely accepted as general by the research community. The third and fourth part however are still subject to debate and research: indeed, they model the process that take place in the central nervous system to infer periodicity out of the auditory nerve ... and, so far, no measures of that phenomenon have been performed.

The global processing chain, however, has been very successful in reproducing phenomena in human hearing and this is why this model is still a reference today.

5 Conclusion

As a conclusion that your choice of fundamental frequency to use can/should be dictated by your a priori knowledge on the studied signal. If you are working with harmonic signals or inharmonic signals will guide you in the use of one or another family of technique. It is interesting to note that mallet percussion instruments where pitch is perceived but which spectra do not exhibit the same kind of structure as harmonic sound are typically not included by the single-f0 estimation methods.

Also, it is important to keep in mind the different obstacles to f0 estimation. As pointed out in [Rabiner76] this task is made difficult by the fact that we are looking for something that does not exist : a strictly periodic signal. Also, we often have to process sounds with interfering information in it (other instrument can be heard, or the reverberation of the room).

Finally, we should note that the present typology is not completely exhaustive and that there exist exceptions that can not be classified in its categories. An example of that is a method presented in [Abu-Shikhah99] using the Teager Energy Function measuring the instantaneous energy of a signal. Also methods based on instantaneous frequency estimation such that presented in [Abe95] do not appear to fit into these categories.

References

[Abe95] Abe, T., Kobayashi, T., and Imai, S. [1995]. Harmonics tracking and pitch extraction based on instantaneous frequency. In *Proceedings of the 1995 IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1. pp. 756–9.

[Abu-Shikhah99] Abu-Shikhah, N. and Deriche, M. [1999]. A novel pitch estimation technique using the Teager energy function. In *Proceedings of the IEEE Fifth International Symposium on Signal Processing and Its Applications*.

[Brown92] Brown, J. C. [1992]. Musical fundamental frequency tracking using a pattern recognition method. *Journal of the Acoustical Society of America*, vol. 92 (3): pp. 1394–1402.

[deCheveigné01] de Cheveigné, A. and Kawahara, H. [2001]. Comparative evaluation of f0 estimation algorithms. In *Proceedings of Eurospeech*.

[Doval91] Doval, B. and Rodet, X. [1991]. Estimation of fundamental frequency of musical sound signals. In *Proceedings of the 1991 IEEE International Conference on Acoustics, Speech and Signal Processing*.

[Doval93] —— [1993]. Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMM's. In *Proceedings of the 1993 IEEE International Conference on Acoustics, Speech and Signal Processing*.

[Hess83] Hess, W. [1983]. *Pitch Determination of Speech Signals*. Springer-Verlag.

[Klapuri04] Klapuri, A. [2004]. *Signal Processing Methods for the Automatic Transcription of Music*. Ph.D. thesis, Tampere University of Technology.

[Lahat87] Lahat, M., Niederjohn, R. J., and Krubsack, D. A. [1987]. A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35 (6): pp. 741–50.

[Maher94] Maher, R. C. and Beauchamp, J. W. [1994]. Fundamental frequency estimation of musical signals using a two-way mismatch procedure. *Journal of the Acoustical Society of America*, vol. 95 (1): pp. 2254–63.

[Meddis91] Meddis and Hewitt, M. [1991]. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America*, vol. 89 (6): pp. 2866–82.

[Noll67] Noll, A. M. [1967]. Cepstrum pitch detection. *Journal of the Acoustical Society of America*, vol. 41 (2): pp. 293–309.

[Rabiner76] Rabiner, L., Cheng, M., Rosenberg, A., and McGonegal, C. [1976]. A comparative performance study of several pitch detection algorithms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24 (5): pp. 399–418.

[Rabiner77] Rabiner, L. [1977]. On the use of autocorrelation anaysis for pitch detection. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25 (1): pp. 24–33.