# A LOGARITHMIC SPECTRAL COMB METHOD
# FOR FUNDAMENTAL FREQUENCY DETECTION

PHILIPPE MARTIN

Experimental Phonetics Laboratory
300 Huron Str., Toronto, Ontario,
CANADA M5S 2X6

## ABSTRACT

The spectral comb method is a fundamental frequency detection algorithm based on the cross-correlation of a modified power spectrum of the speech signal and a spectral comb function with teeth of decreasing amplitudes and variable intervals. In order to reduce the overall computational complexity and obtain a constant frequency resolution, a modified approach is proposed to compute the cross-correlation function, using a logarithmic scale for both the amplitude and the frequency. The cross-correlation is obtained by iteratively summing each spectral peak shifted on the frequency scale by factors of 1, 1/2,...,1/n, and on the amplitude scale by factors of 1, 2,...,n dB according to the comb tooth order n.

## INTRODUCTION

Fundamental frequency detection plays an important role in phonetic research, as well as in many aspects of speech analysis, such as speech recognition and synthesis. Although many experimental devices and algorithms have been proposed to date [7], none provide error free results, specially in the case of noise or telephonic recordings. The choice of a specific pitch detector will depend on the application, as its structure will define an implicit articulatory model for Fo detection. Any discrepancy between this model and the real conditions of analysis will lead to errors in the fundamental frequency detected.

Among the numerous methods available for pitch detection, those based on short term spectral analysis of the input signal offer usually a good resistance to noise and provide adequate results even if the fundamental component is absent from the speech input. Despite some drawbacks, essentially due to lower time and frequency resolution which precludes their use for medical applications, spectral pitch detection appears to be quite attractive for phonetic and linguistic research.

Most methods of pitch detection based on the short-time spectrum aim to detect some periodicity of the fundamental frequency harmonics. The popular cepstrum approach [1], for instance, computes the Fourier transform of the logarithm of the power spectrum. Other methods are based on a more direct and computationally efficient direct search for periodicity in the spectrum. Schroeder [2] uses an histogram of subharmonics derived from the spectral peaks, and Fo is taken as the smallest common multiple of the periods of its harmonic components. Harris and Weiss [3] use a high resolution Fourier spectrum and retain the most numerous equal spacing of adjacent peaks as fundamental frequency. Sreenivas and Rao [4] use only high quality peaks (well above the noise level), and compute their approximate highest common factor to obtain the pitch value. Sluyter, Kotmans and Leuwaarden [5], in order to reduce the influence of phase distortion in the peak frequency measurement, utilize a minimum distance criterion to recognize harmonic pattern and the resulting fundamental frequency.

## THE SPECTRAL COMB METHOD

By contrast with other pitch detection schemes such as the sieve algorithm, the spectral comb method [6] is based on a direct search of an harmonic structure in the spectrum integrating both the harmonic frequency and amplitude informations. This ensures that a correct value of Fo will be obtained even if no periodicity in the spectrum is found. (Interestingly enough, most frequency domain pitch detection methods such as the cepstrum will fail if the signal has no harmonics, as for a pure tone).

To evaluate Fo, the short-time spectrum |F(w)| is first "groomed" by replacing spectral peaks meeting an appropriate selection criterion by narrow parabola, and by zeroing the remaining of the spectrum. This ensures that non-harmonic related values, usually with low energy, will not interfere in the overall computation.

The groomed spectrum is then crosscorrelated with a spectral comb function C(wp,w) with teeth of decreasing amplitude and variable intervals wp.

$$C(wp,p) = \sum_{n}^{1} An \quad (nwp-w)$$

The maximum of the crosscorrelation function I(wp) is reached when a large number of the comb's teeth coincide with the harmonic peaks of the spectrum. When this value exceeds a voicing treshold, the corresponding tooth interval is taken as 1/Fo.

$$I(wp) = \sum n \exp -1/8 \, |F(wp)|$$

If |F(w)| is represented by m samples, n*(m*m) sums and products are necessary to evaluate n values of the crosscorrelation function I(w). In a 1000 Hz frequency range, with a 4 Hz resolution, this corresponds to
250 * (250*250)= 15,625,000 sums and products.

## A FASTER METHOD

Due to the nature of the groomed spectrum and of the spectral comb, many of the operations involved in this computation involve a zero factor. A much more efficient algorithm can be obtained if only non-zero values were to be taken into account. This can be done if I(wp) is evaluated from the peaks of the spectrum only, whose amplitude and position on the frequency axis are the only information retained. The cross-correlation function is obtained by iteratively adding, for each spectral peak of |F(w)|, parabola

- shifted in frequency according to the order of the comb tooth n;
- shifted in amplitude by an appropriate factor proportional to the comb tooth order n.

## A LOGARITHMIC SPECTRAL COMB

The use of a linear frequency scale ensures the possibility of using the FFT to evaluate the short time spectrum |F(w)|. On the other hand, since all computations are performed on sampled values, a linear frequency scale creates an uneven frequency resolution in the contribution of low and high harmonic components. Using a logarithmic frequency scale, all harmonic components will have a similar impact on the final cross-correlation result. Furthermore, the operations will only involve additions and substractions.

Starting from a logaritmic short-time spectrum, the cross-correlation with a logarithmic spectral comb function is then obtained by recursively adding n times

- shifted in frequency by log n
- shifted in amplitude by n dB

Again, a more efficient algorithm will proceed from the spectral peaks added recursively after having been shifted by log n on the frequency scale and by n on the amplitude scale (n=1, 2,..., n).
With this approach, assuming that each peak is represented by p values and that n comb's teeth are considered, the total number of additions is reduced to p*n*h , with h= number of harmonics taken into account. With typical values of p=16, n=8 and h=8, we have thus 1024 additions to perform to obtain the cross-correlation function (Each addition involving an extra address calculation).

The price to pay to implement this logarithmic approach is the spectral analysis of the speech input, which has to be obtained either by a relatively high resolution FFT followed by a logarithmic mapping of the frequency scale, or a direct logarithmic DFT. The latter solution would be more easily implemented in hardware form.

Furthermore, the sides of each peak parabola must be constant on the logarithmic frequency scale. This suggests a possible improvement in the computation of the logarithmic DFT : since the width of each peak is proportional to the duration of the time window used, shorter blocks of sampled speech input can be used for higher frequencies. Starting for example at +10 Hz at -20 db below a spectral peak at 100 Hz, the equivalent logarithmic width at 1000Hz will correspond to a frequency width of +100 Hz. Using for instance a Gaussian window, this would imply a duration of the time window equal approximatively to 20 times the period involved, i.e. 20* 1/100 Hz=200 ms and 100 Hz and 20 ms at 1000 Hz. This variable window length roughly corresponds to the time resolution of the ear for pure tones.

## CONCLUSION

Using a logarithmic scale for both the amplitude and the frequency scale of the short-time power spectrum, the computational effort to evaluate the cross-correlation function in the spectral comb method is dramatically reduced. Typically, only 1024 sums are necessary, compared to more than 15,000,000 sums and products in the direct approach. This method, which requires the Fourier transform of the speech input to be logarithmic, seems suitable for hardware implementation leading to reliable real-time operation.

## REFERENCES

[1] A.M. Noll, Short-time Pitch Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection, JASA, 36: 296-302, 1969.

[2] M. R. Schroeder, Period Histogram and Product Spectrum: New Methods for Fundamental Frequency Measurement, JASA, 43: 829-834, 1968.

[3] C. M. Harris and M. R. Weiss, Pitch Extraction by Computer Processing of High Resolution Fourier Analysis Data, JASA, 35: 339-343, 1963.

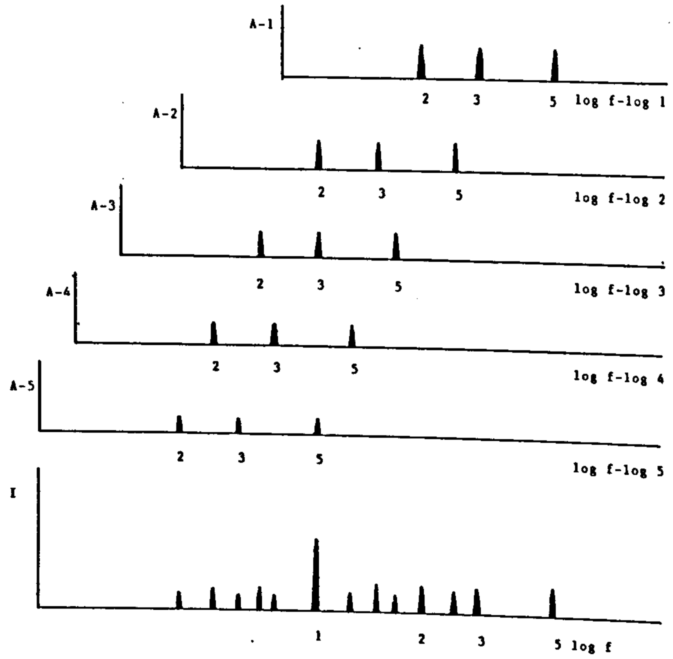
[4] T. V. Sreenivas and P. V. S. Rao, Pitch Extraction from corrupted Harmonics of the Power Spectrum, JASA, 65: 223-228, 1979.

[5] R. J. Sluyter, H. J. Kotmans and A. V. Leuwaarden, A Novel Method for Pitch Extraction from Speech and a Hardware Model Applicable to Vocoder Systems, Proc. of the ICASSP-80, I: 45-48, 1980.

[6] Ph. Martin, Comparison of Pitch Detection by Cepstrum and Spectral Comb Analysis, Proc. of the ICASSP-81, I: 180-183, 1981.

[7] W. Hess, Pitch Determination of Speech Signals, Springer, Berlin, 1983.

[8] Ph. Martin, Spectral Comb Gives Real-Time Pitch Analysis, Speech Technology, Sep-Oct. 1983.



Cross-correlation-function for a 3 components spectrum (200 Hz, 300 Hz and 500 Hz). The maximum is obtained for Fo=100 Hz.