3. Pitch detection (F0) through the cepstral method (Documentation – see 01)

3.1. Presentation

Among the numerous methods to detect the fundamental frequency (F0) met in literature, the pitch detection through both the cepstral and auto-correlation methods has been implemented so far. The results provided by the two methods are compared and in case of major differences between the values, a selection algorithm is applied that is to be presented later on.

The application of the cepstral method in pitch detection is justified by the fact that the spectrum of an uttered sequence is the product of two spectral components – one depending on the way in which the sound is *generated* (resonance frequency of vocal cords, dimension of the generating tube – the larynx) and another depending on the way in which the generated signal is *filtered*.

Given the cepstrum calculus formula, the multiplication together of the spectrum of the exciting signal and that of the transfer function is transformed by applying the logarithm to an addition operation.

$$H(w) = H_g(w) \cdot H_f(w)$$

$$cepstrum = IFFT\left(\log FFT(s)\right)$$

$$cepstrum = = \Im^{-1}\left(\log H_g(w) \cdot H_f(w)\right) = \Im^{-1}\left(\log H_g(w)\right) + \Im^{-1}\left(\log H_f(w)\right)$$

The two spectrums being clearly separable, it is justifiable to look for the cepstral maximum that corresponds to the fundamental frequency among the values corresponding to the frequency band [70Hz-500Hz]. For voices coming from male subjects, the usual values of F0 are about 100-150 Hz; for women and children respectively they amount to 250-350 Hz.

To the recommendation that a weighting window (other than the rectangular one) should be employed in order not to distort the generated spectral results and implicitly the cepstral ones, we also add that of using analysis windows of at least 1024 samples.

For example for a sampling frequency of Fs = 22050 for the frequency band [70Hz-500Hz], the cepstral vector (corresponding to positive frequencies) has to contain at least Fs/f_low = 22050/70 = 315 samples. As half of a signal spectrum is associated with positive frequencies and the other half with negative ones, the minimum dimension of the analysis window is of 2*Fs/f_low = 650 samples. As the algorithm of the Fourier fast transformed needs analysis windows of dimensions equal to powers of 2, 1024 samples will be analysed for each window.

To eliminate the noise induced by network frequency, the signal is filtered by means of a high-path *Butter* filter at the 70Hz frequency.

If for an analysis window the signal energy does not exceed in time the minimum energy taken by 10 or the $^1/_{20}$ part of the maximum energy computed for every analysis windows of dimension *N*, or if the spectral energy in the frequency band [80Hz-2500Hz] does not exceed 70% from the spectral energy, then it is considered that the information contained is not *vocalic* and hence the fundamental frequency F0 should not be determined in that area.

3.2. Way of operating

The tool was conceived under the form of an executable called *F0_cepst.exe*. This has to be enclosed in the same folder with the sound files (wav) to be analysed. The user selects the following parameters:

- Name of file (it may be selected from the wav file list in the folder). After inputting the sound

file name, the information in the header is checked and pieces of information such as sampling frequency, number of channels, number of bits per sample, and the total number of samples are displayed. Only mono-channel sound files are accepted.

- Number of processed samples – dimension of analysis window (for example *1024*)
- Displacement pitch (a non-null positive integer whose value depends on the number of samples in the sound file)
- Select the type of window (implicitly = 0, the rectangular window)

The data – the frequencies corresponding to the pitch are saved under the form of a column, the dimension of the generated data vector being determined based on the displacement pitch and the number of samples processed on an analysis window *N*:

(total_number_of_samples - N)/pitch + 1

Citation and Copyright

The program was written by Marius Zbancioc in collaboration with Horia-Nicolai Teodorescu.

The program may be used free in educational and research applications provided that the following citation is displayed:

Marius Zbancioc, Horia-Nicolai Teodorescu: "Pitch Detection (F0) through Cepstral Method" Application. Tools for the Archive of Romanian Spoken Language – Romanian Sounds http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/ro/instrumente.htm.

In the absence of a proper citation, we reserve the right to apply to justice. In order to use the program in commercial applications, one needs the written, contractual permission of the authors. Contact us on the address below and provide us with the description of your application and the way you intend to use the program. In the absence of a written permission, we reserve the right to apply to justice.

Authors: H.-N. Teodorescu and M. Zbancioc