4. Pitch detection (F0) through the auto-correlation method (Documentation – see 01)

4. 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.

Unlike the cepstral method valid for the domain of frequencies, the auto-correlation method is valid in the domain of time (working directly with the values of the samples in the vocal signal). The method is frequently met in line literature, being considered a robust method that leads to a good detection of fundamental frequency.

The correlation function applied to two signals x and y provides information concerning the similarities between them. This comparison method is used to detect some regularities (regarding signal periodicity when the signals may be considered periodical or quasi-periodical). The calculus formula in continuum is:

$$R_{XY}(t) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t) \cdot y(t+t)dt; \;\; R_{YX}(t) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t) \cdot x(t+t)dt;$$

In discrete the determination of correlation coefficients reduces to sum computation:

$$R_{XY}[n] = \sum_{k=0}^{N-n} x[k] \cdot y[k+n]; \;\; R_{XY}[n] = \sum_{k=0}^{N-n} y[k] \cdot x[k+n]; \;\; n = \overline{0,N}$$

The auto-correlation is the correlation of a signal with itself (y is replaced by x in formulae). The values of local maximum on the auto-correlation signal determined by the previous calculus method are closely related to the signal periodicity properties. The maximum value found in Rxy[0] must not be taken into account as it represents the signal energy/ power in an analysis window N of a given dimension. The next local maximum is associated with the fundamental period and it can be looked for in the value interval given on the band on which the fundamental frequency [89Hz-500Hz] is looked for. Thus, for a sampling frequency of Fs = 16000Hz, one should seek in the provided auto-correlation signal among the values:

Rxx[Fs/fmax, ... ,Fs/fmin], i.e. Rxx[16000/500, ... , 16000/80] Rxx[32, ... , 200]

It is recommended that the analysis window should be of at least 4-5 times the maximum fundamental period (5* Fs/fmin) in order to get a good determination of the pitch. If the sampling frequency is of 16000Hz, the analysis window should have at least 800-1000 samples. Considering the fact that most often it is the comparison between the results due to various pitch detection methods that is intended, there must exist the same parameters (dimension of the analysis window, **displacement pitch**) when running the program. For the cepstral method, there is the condition to run the program with a dimension of the analysis window of minimum 1024 samples (power of 2) and this value can be accepted as an input parameter for N.

As it is intended for the pitch to be found on vocal segments alone, one must check if for an analysis window the signal energy exceeds in time the minimum energy taken by 10 or the $^1/_{20}$ part of the maximum energy (Emax, Emin are found for all the analysis windows of dimension N), or if the spectral energy in the frequency band [80Hz-2500Hz] exceeds 70% from spectral energy.

4.2. Way of operating

The tool was made under the form of an executable called *F0_corel.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 the 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 Auto-Correlation 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