Lab.2 Time Domain Fundamental Frequency Estimation

1. Introduction

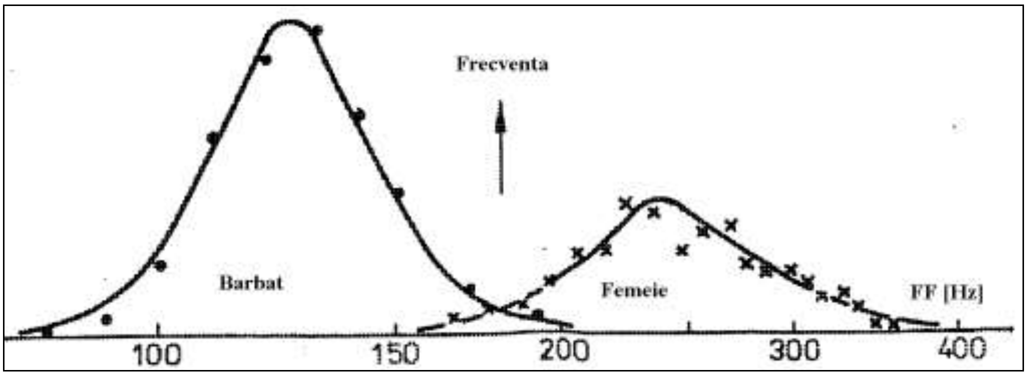
The problem of estimating the fundamental is to take a part of the signal and find the dominant frequency of the repetition. The difficulty lies in the fact that :

- Not all signals are periodic;
- Periodic signals can change the frequency of interest;
- Signals can be contaminated by noise, even with periodic signals of different fundamental frequency:
- Signals are periodic with period T are also periodic with 2T , 3T , etc., so have found the lowest range of periodicitee or highest FF;
- Even constant FF signals can be modified by other means in the area of interest.

There are several methods for determining FF in both the time and frequency domains.

The fundamental frequency (F0) is the measure that is defined for the voiced elements/segments of speech and sound and represents the frequency of oscillation of the vocal cords. The corresponding acustic fundamental frequency is the fundamental period (T0) or pitch. The fundamental frequency of the vocal signal may be missing (e.g. telephone speech). Fundamental period is present even in speech signals with limited bandwidth.

The mean and standard deviation of the fundamental frequency are: 125Hz and 20.5 Hz , respectively, for the male and female voice expresses these values are twice as high. The variation in fundamental frequency during speech was also found to be small, less than 10 Hz .



Fundamental frequency distribution by gender[Fur2018]

Fundamental frequency is often represented on a logarithmic scale to adapt the results to the resolution of the human auditory system.

For voiced parts of speech, the fundamental frequency is between 50Hz and 500Hz, respectively for unvoiced parts, it is considered, by convention, that F0 = 0 (the excitation/glotic wave has a noise-like appearance).

Given the complexity of speech, the fundamental frequency cannot be determined with a simple algorithm. During voiced sounds, air is pushed through the entire vocal tract, including the vocal cords, throat, and mouth. The articulatory parts (tongue, throat, and lips) filter the signal generated by the vocal cords. This filtering may conceal information about the frequency by subtracting the fundamental frequency energy and the increase in energy at other frequencies.

A speech signal does not have a fundamental period throughout its duration. So, the fundamental period is missing during pauses in speech or nonvoiced sounds. Among unvoiced sounds, fricatives are of particular interest as they are generated by restricting the air flow through the vocal cords over a relatively long time. To determine whether a speech segment has a fundamental period (T_0), usual the averaged energy can be used. For the other/fricative frames, the average energy is much lower than for voiced sounds.

There are several ways of calculating the fundamental frequency: time-domain, frequency-domain, and time-frequency domain methods.

In the time domain, most methods are based on the autocorrelation function or derivative.

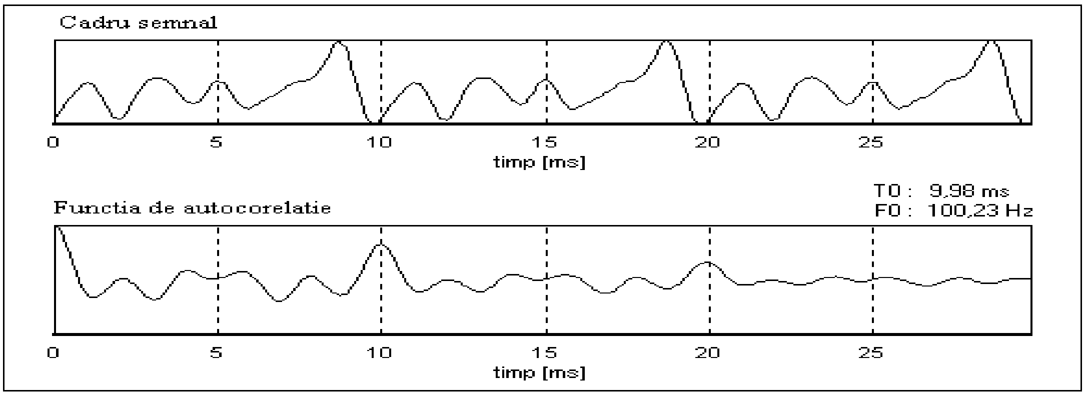
2. Time domain fundamental frequency estimation methods

Study Lecture 3 and the bibliography.

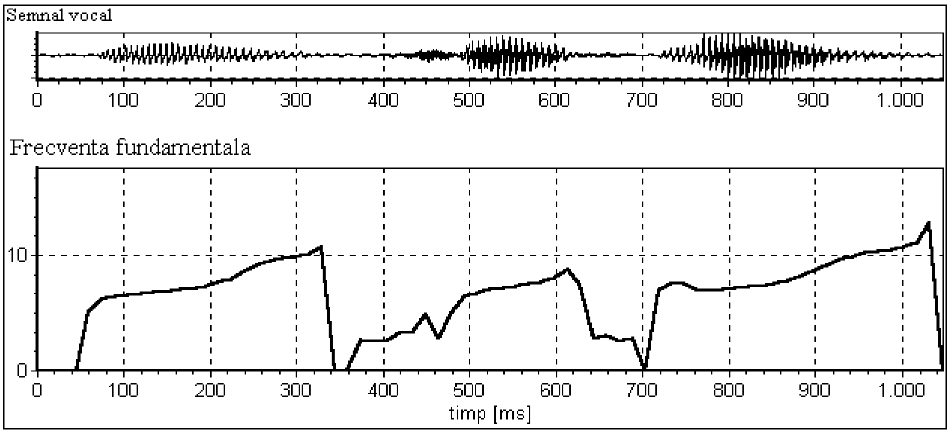
2.1. Autocorrelation method

The auto-correlation function is defined as:

$$R_n(k) = \sum_{m=0}^{N-1-k} [x(m+n)w(m)][x(n+m+k)w(m+k)]$$



Voiced segment and associated autocorrelation function



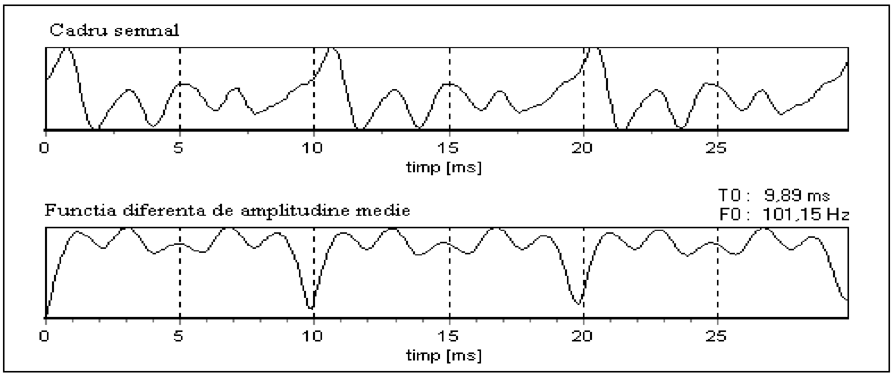
Speech signal and the fundamental frequency contours

2.2 The Average Magnitude Difference Function (AMDF) Method

This method of determining the fundamental frequency is based on the Average Magnitude Difference Function (AMDF), which has the expression:

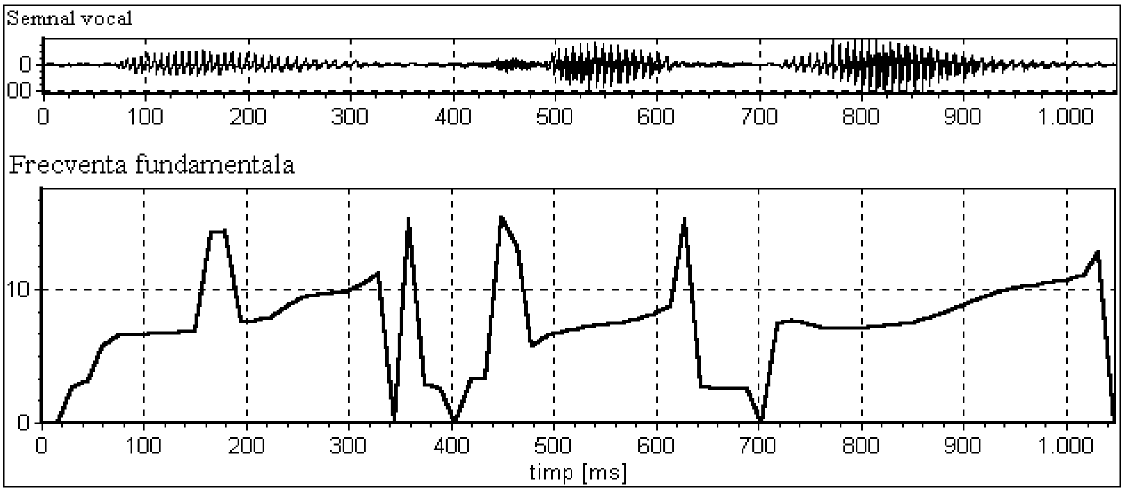
$$D(n) = \frac{1}{N}\sum_{0}^{N-1}|s_k - s_{k-n}| \qquad ,0\leq\ n \leq N\text{-}1.$$

The function D(n) exhibits minima at time intervals of length 1/F0, which generally decrease in amplitude.



Average Magnitude Difference Function (AMDF)

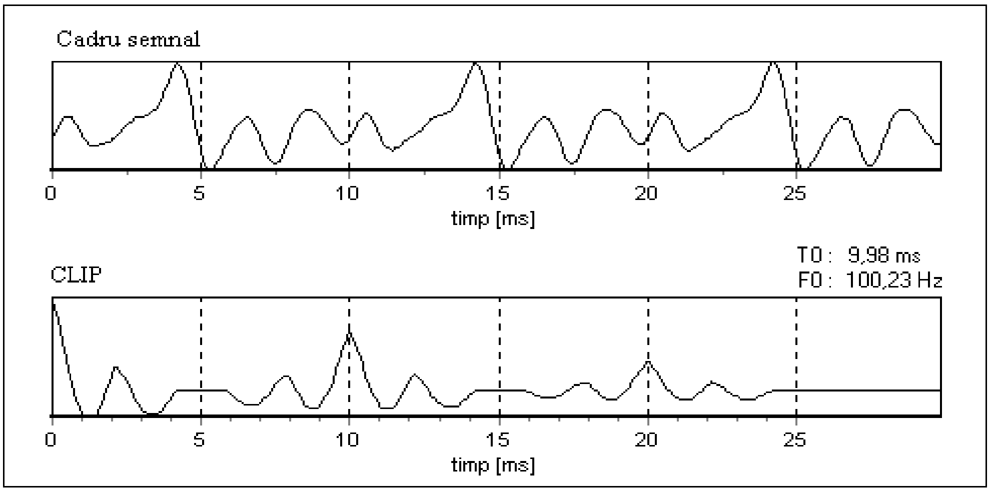
This method also gives erroneous results in the crossover areas at the beginning and end of a voiced segment. The contour for the fundamental frequency is shown in the fig. below.



Speech signal and associated fundamental frequency obtained by the AMDF method

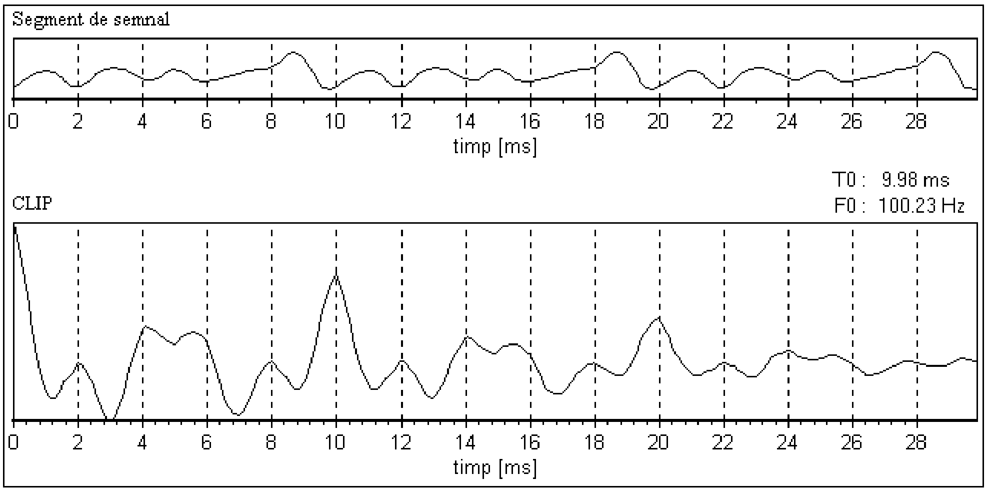
2.3. Center Clipping Pitch Detector (CLIP) Method

This method uses a CL threshold and retains only those elements in the speech signal whose absolute value exceeds the |CL| threshold. For the remaining values, the procedure is as follows: add the CL value to the positive ones, respectively, subtract the CL value from the negative ones, or assign the maximum positive value respectively the maximum negative value (infinite limiting). Next, the autocorrelation function is computed. It presents maxima at time intervals directly correlated with the period associated with the fundamental frequency.

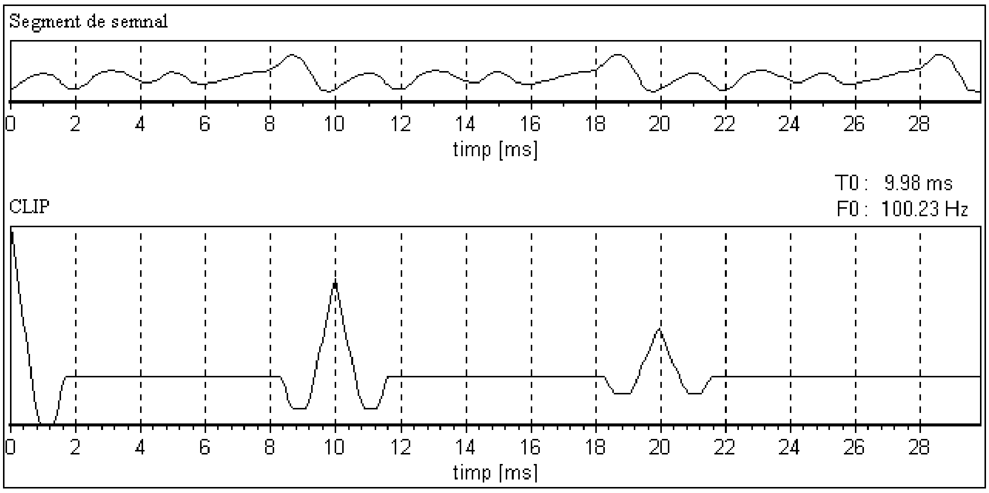


F0 estimation with the CLIP method

The method is sensitive to the threshold value used. The average or maximum value of the signal on the considered segment, possibly multiplied by a subunit coefficient, can be used for CL. Decreasing the threshold brings the function shape closer to the usual autocorrelation while increasing it simplifies the final form of the function.

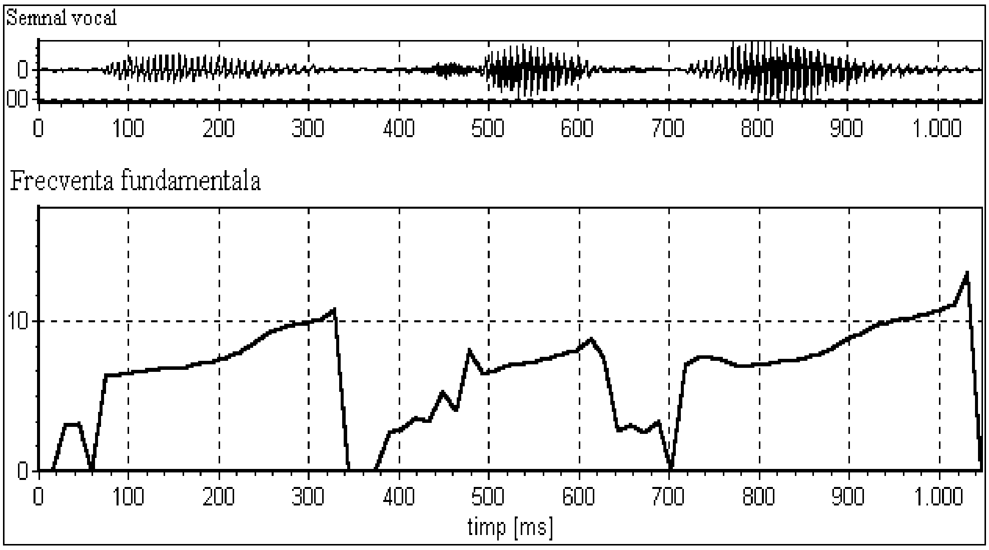


F0 determination by CLIP procedure, threshold 20%.

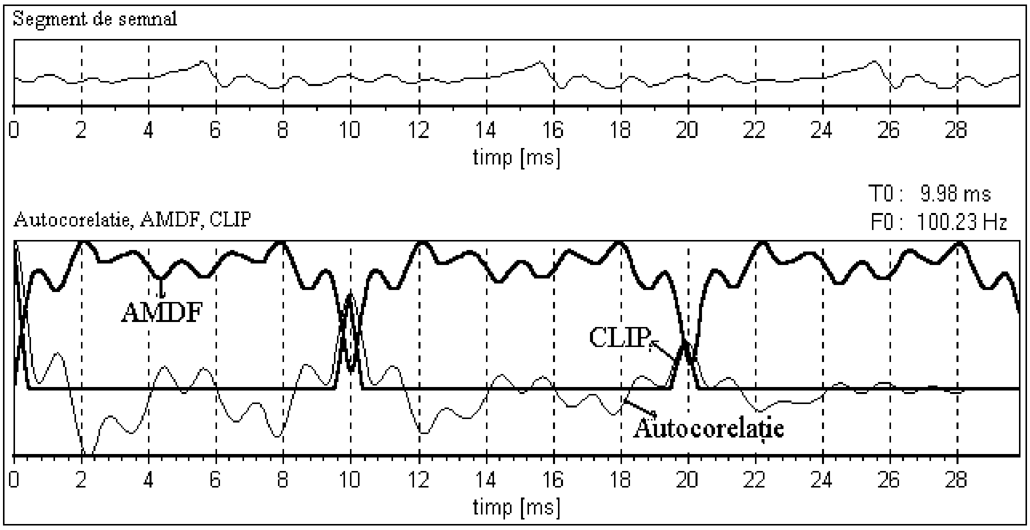


F0 determination by CLIP procedure, threshold 60%.

The fundamental frequency contour is shown in Fig. below.



Speech signal and associated fundamental frequency contour obtained by CLIP method, threshold 30%



Convergence of autocorrelation-based methods for fundamental frequency determination

3. Model for implementation

One way to estimate F0 directly from the waveform is by using autocorrelation.

The autocorrelation function for a frame (window) of the signal shows how well the waveform correlates with itself at different delays. We expect a periodic signal to correlate well with itself at small delays and multiple delays of the fundamental period.

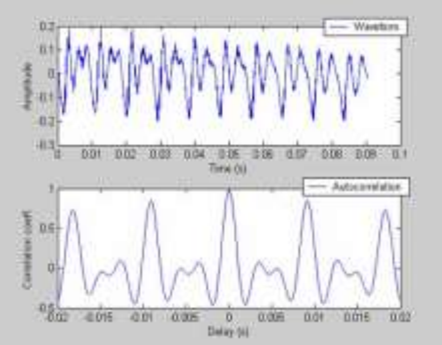
The following code displays the autocorrelation function of a frame of a voiced speech signal.

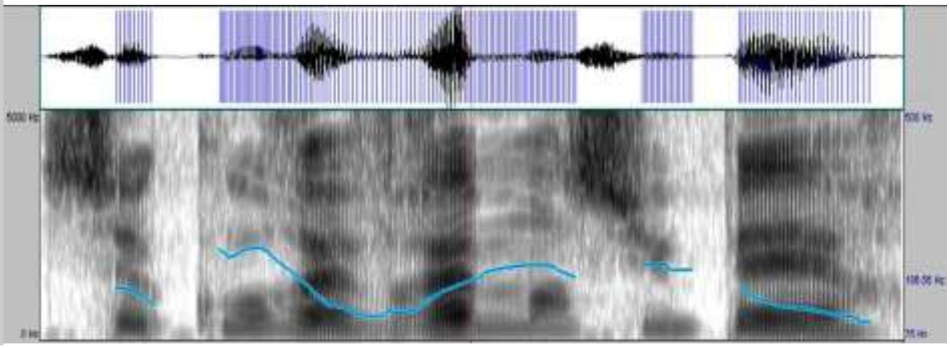
```
%MODEL

 % get a section of vowel

[x,fs]=wavread('six.wav',[24120 25930]);
ms20=fs/50;                  % minimum speech Fx at 50Hz
%
% plot waveform
t=(0:length(x)-1)/fs;        % times of sampling instants
subplot(2,1,1);
plot(t,x);
legend('Waveform');
xlabel('Time (s)');
ylabel('Amplitude');
%
% calculate autocorrelation
r=xcorr(x,ms20,'coeff');
% plot autocorrelation
d=(-ms20:ms20)/fs;           % times of delays
subplot(2,1,2);
plot(d,r);
legend('Autocorrelation');
xlabel('Delay (s)');
ylabel('Correlation coeff.');
%Track the peaks of the autocorrelation function at delay 0, T, 2T...
```

We can estimate FF by the distance between successive peaks 2ms(=500Hz) and 20ms (=50Hz).

For example:

```
ms2=fs/500              % maximum speech Fx at 500Hz

ms20=fs/50              % minimum speech Fx at 50Hz
                        % just look at area corresponding to positive delays
r=r(ms20+1:2*ms20+1)
[rmax,tx]=max(r(ms2:ms20))
fprintf('rmax=%g Fx=%gHz\n',rmax,fs/(ms2+tx-1));
```





4. Progress of work

Download the files wav.zip and colea.zip
- Analyze the COLEA application.
- Cut off a long enough frame of a vowel and record it with a distinctive name.
- Display the F0 values for the selected section in the COLEA application and note the results.
- Create your implementation using either (AMDF/CLIP/AUTOC) or another method
- Test the application on the same file and compare results.
- Choose from the attached files and determine the F0 contour for a phrase after voiced/unvoiced detection. Compare your results with those provided by COLEA. Comment on the results.

SEND REPORT OF THE LABORATORY WORK:
IMPLEMENTATION (personal scripts), results, captures, COMMENTS.