Pitch Detection

P. Sai Teja

Pitch detection refers to determining the pitch or fundamental frequency of a signal, usually a digital recording of speech or musical note, which is especially important for many speech processing applications. Pitch detection can be classified into two types.

1. Time domain detection
2. Frequency domain detection

In this report we will be discussing one method from each domain to detect the pitch of the signal.

1. Autocorrelation method (Time domain detection)
2. Cepstral method (Frequency domain detection)

**Speech detection:**

Before proceeding to pitch detection, we must identify unvoiced or silence signals and apply the pitch detection algorithm to voice signals.

**Algorithm:**

1. Pass the voice signal through a lowpass filter [1].
2. Segment the entire signal into blocks of X milliseconds (Number of samples=fs\*X) and calculate the mean of absolute values of each segment.
3. While applying pitch detection on a segment check if the calculated mean of the segment is less than threshold for it to be a voice segment and proceed for pitch detection.

**Parameters used:**

Threshold= (5% of the max mean of a segment)

Lowpass cut-off frequency=900Hz [1]

X=10 milli seconds

**Autocorrelation method:**

Autocorrelation method works on the principle that if the given sequence is periodic then its corresponding autocorrelation function is also periodic. But for non-stationary signals like speech signals autocorrelation upon short-time speech signals is more meaningful than upon long-time [1].

To make the periodicity more prominent non-linear methods are useful. In this report I have implemented a non-linear method (compressed centre clipper). These non-linearities intend to provide spectral flattening which make the signal resemble a pure tone [1].

**Center clipping algorithm:**

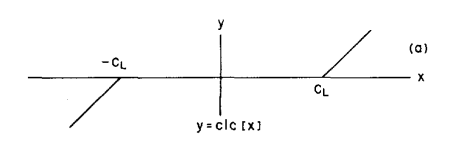


Figure : clc application each segment [1]

* If the signal value is above threshold, deduct threshold value.
* If the signal value is below -threshold, add threshold value.
* If the signal value is in between -threshold and threshold, make it zero.

**Algorithm:**

From the above filtered segments for speech detection.

1. Check whether the segment contains voice, if not assign it’s pitch as 0.
2. Apply the non-linearity on the segment after calculating the clipping threshold (30% of the maximum absolute value of the signal).
3. Apply autocorrelation on the clip and compressed signal.
4. Find the shift difference at which the autocorrelation is maximum from the center of the autocorrelation which is the period of the segment. Since there are many outliers present near to center peak, to avoid erroneous pitch values due to them, max autocorrelation lag is considered only after min expected period which corresponds to (fs/fmax) from the center.
5. Convert period into pitch(fs/(period(in discrete signal)) and plot its comparison with the know results.
6. The pitch results are refined using 1D median filter.

Where fs is sampling frequency, fmax is max pitch possible for voice signal.

**Cepstral method:**

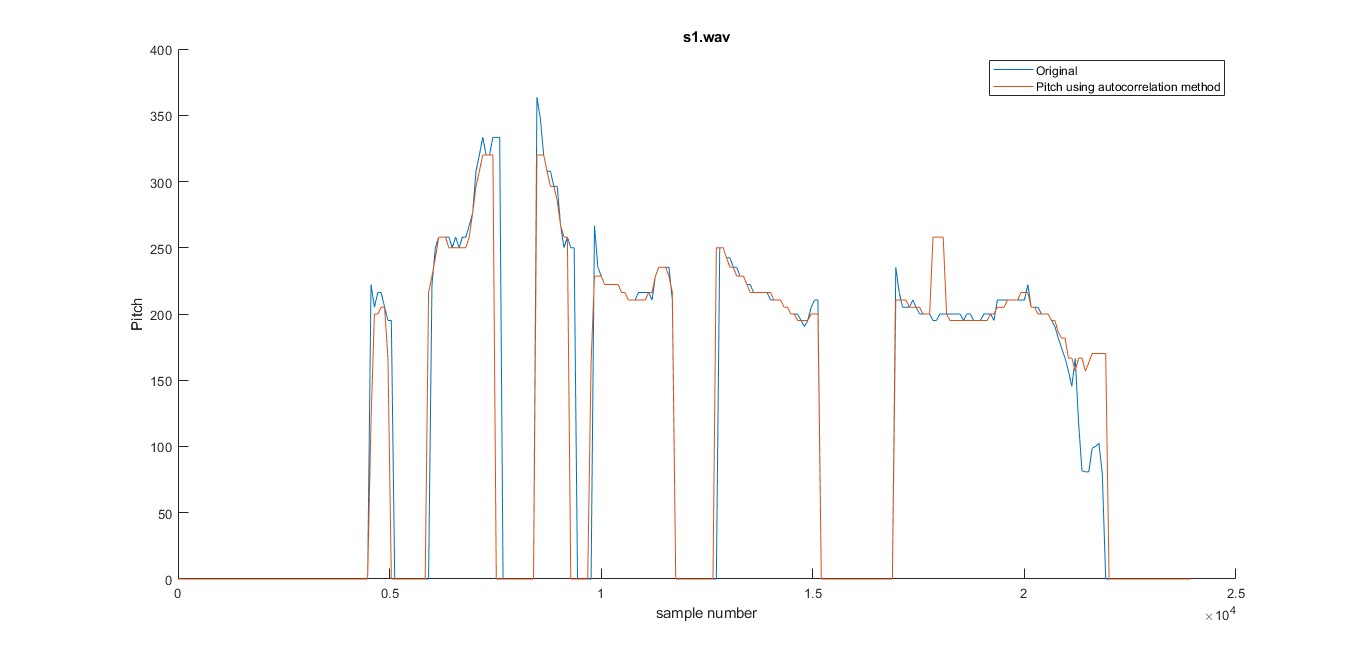
From the above filtered segments for speech detection.

1. Check whether the segment contains voice, if not assign it’s pitch as 0.
2. Apply FFT upon the segment.
3. Apply log upon the absolute value of obtained FFT output.
4. Apply IFFT upon this signal to get a cepstrum signal (cepstrum indicates inverse of spectrum).
5. Find the quefrency (inverse of frequency, which is this method’s way of representing time) peaks at a distance greater than threshold points, find the maximum successive difference between quefrencies obtained. This is the period of the segment.
6. Convert period into pitch (fs/ (period found in discrete signal) and plot its comparison with the know results.
7. The pitch results are smoothened using 1D median filter.

**Note:** MATLAB does the above steps (2-4) using a function named **rceps** if the data is real valued.

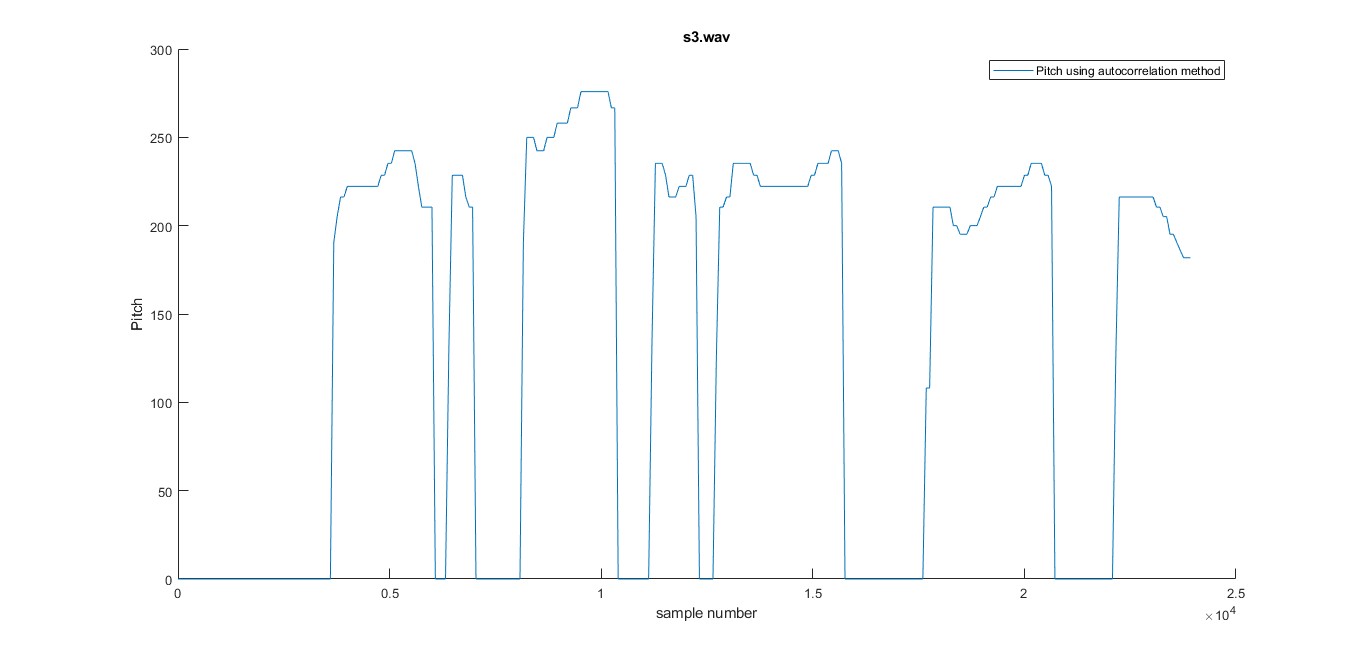
**Results:**

**Autocorrelation method:**



Diagram, histogram

Description automatically generated with medium confidence



Chart, histogram

Description automatically generated

**Parameters used:**

* Median filter(1D) of order five has been used.
* For s1 the shift required to remove the outliers is 20(fs/fmax=8k/400) from center. fmax is taken 400Hz as it is a female voice including some tolerance to frequency. Similarly, for s3 also.
* For s2 the shift required to remove the outliers is 32 from center. fmax is taken 250Hz as it is a male voice including some tolerance to frequency. Similarly, for s4 also.

**Observations:**

For pitch below 150Hz in s2.wav it is observed that there is a 20-50Hz difference in the calculated pitch.

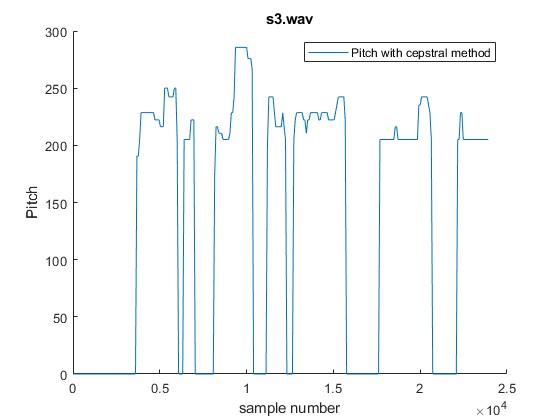
**Cepstral method:**

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated



Chart, histogram

Description automatically generated

**Observation:**

* When there is not much difference between successive samples the calculated pitch is tending to be flat.
* Even though there is a shift to above or below the original pitch, the calculated pitch is following the slope like that of original pitch.

**Comparison between both methods:**

Diagram, histogram

Description automatically generated with medium confidence

Chart, histogram

Description automatically generated

* Both the methods seem to give same kind of performance, but autocorrelation method is giving a better fit than the cepstral method.
* Without the non-linearity (clc) cepstral method is performing better than autocorrelation method.
* Cepstral method is adapting better than autocorrelation method when the steep of pitch variation is more.
* When there isn’t much difference between successive samples the calculated pitch is tending to be a flat for cepstral method.
* Even though the fit isn’t perfect for cepstral method, the max error is comparatively less.
* In cepstral method, even though there is a shift to above or below the original pitch, the calculated pitch is following the slope like that of original pitch in comparison to autocorrelation method.
* Cepstral method took 0.1187 secs for execution whereas autocorrelation method took 0.2338 secs (twice that of cepstral method and seems to be same even with increase in number of segments).

**References:**

[1] S. S. Upadhya, "Pitch detection in time and frequency domain," 2012 International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, India, 2012, pp. 1-5, doi: 10.1109/ICCICT.2012.6398150.

[2] A.M. Noll, Cepstrum pitch determination. J. Acoust. Soc. Am. (1967). <https://doi.org/10.1121/1.1910339>