Sound Processing

Task 1 Report

Fundamental frequency analysis

T2 F1

Maciej Badura 194401
Panos Lennon 201622

1. Introduction

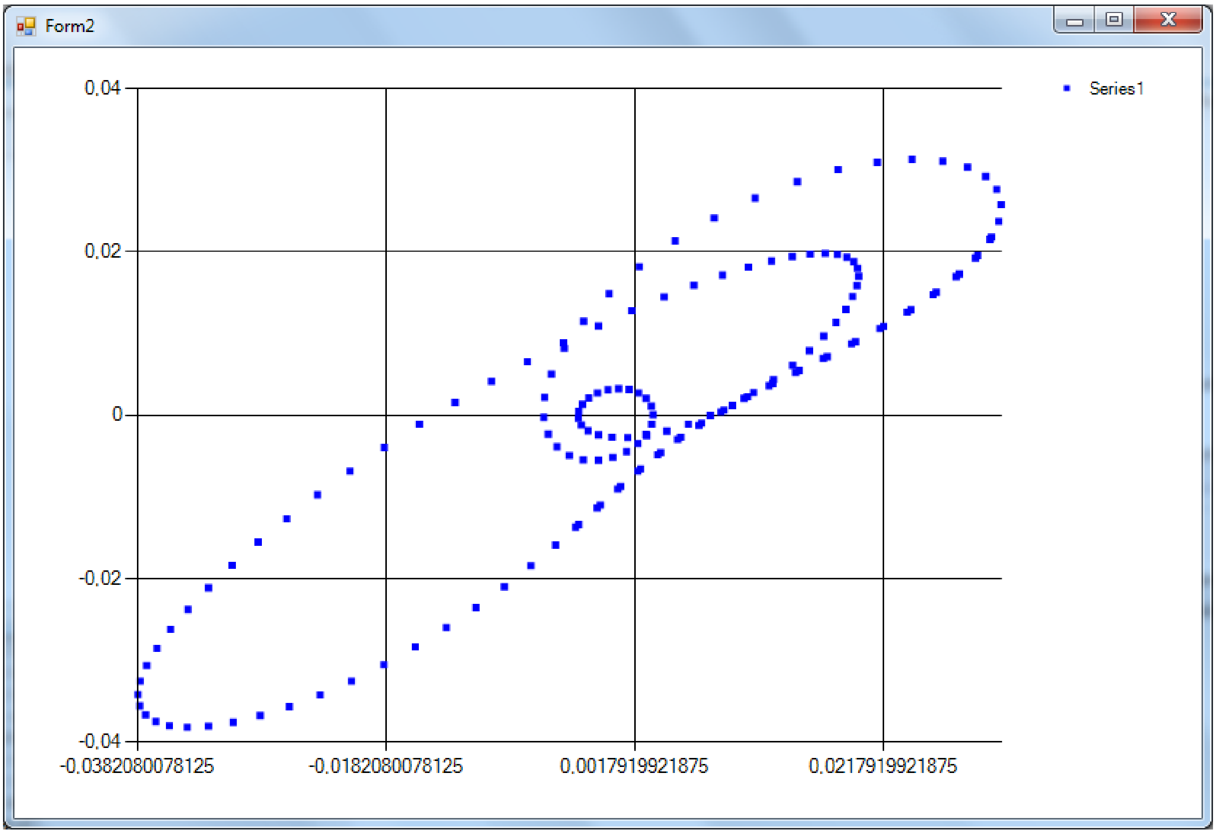
The aim of our task was to analyse sound signal in time and frequency domain. For this purpose we created an application that reads a *.wav file, generates a plot for its phase space and detects its fundamental frequency.

2. Phase space analysis

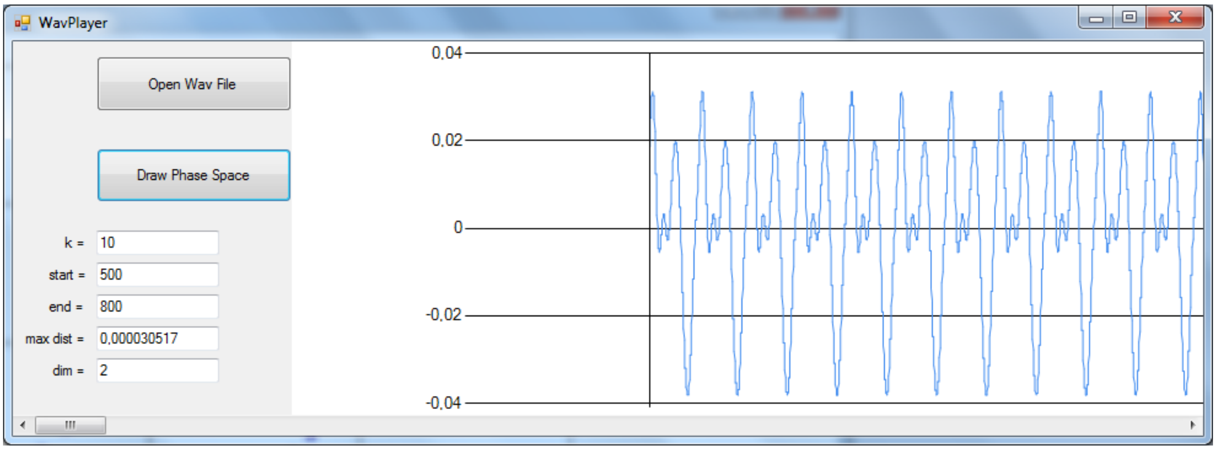
One of our task was to analyse phase space of a provided *.wav file. The signal is presented in two dimensional space. The x axis represents the original values of the signal and y axis represents values of the original signal shifted by index value k which is an integer. The formula for x and y coordinates of the points in two dimensional phase space can be described as:

$$(x(i),y(i)) = (f(i),f(i-k))$$

Where i is the index of the sample in the original set of samples, k is the index of the shift. Algorithm stops when it comes back to the beginning of the phase so when the distance from the starting point and last point is close to zero. Below is an example of the plot that we obtained:



The problem that may appear is when we have a 'luck' and start from the point that is a crossing point of phase but is not the end of it the algorithm will stop. To get rid of this issue we started to experiment with different factors. Firstly we started playing with k index, and minimum distance factor which should be close to 0 but not exactly zero. Then we added also more dimensions to the plot to increase distance between crossing points. Below are presented parameters chosen for the plot obtained in the previous image.



3. Fourier spectrum analysis

The main aim of this task was to use a spectral method for approaching the fundamental frequency detection. Using discrete Fourier transform to transform sample of sounds to complex number. After the transformation, a creation of a 2d graph was needed for further analysis.

$$X(k) = \mathrm{DFT}_N\{x(n)\} = \sum_{n=0}^{N-1} x(n) W_N^{nk} \text{ , for } k = 0, 1, ..., N-1$$

where

$$W_N^{nk} = e^{-i2\pi nk/N} = \cos\left(\frac{2\pi nk}{N}\right) - i\sin\left(\frac{2\pi nk}{N}\right) \text{ , } i = \sqrt{-1}$$

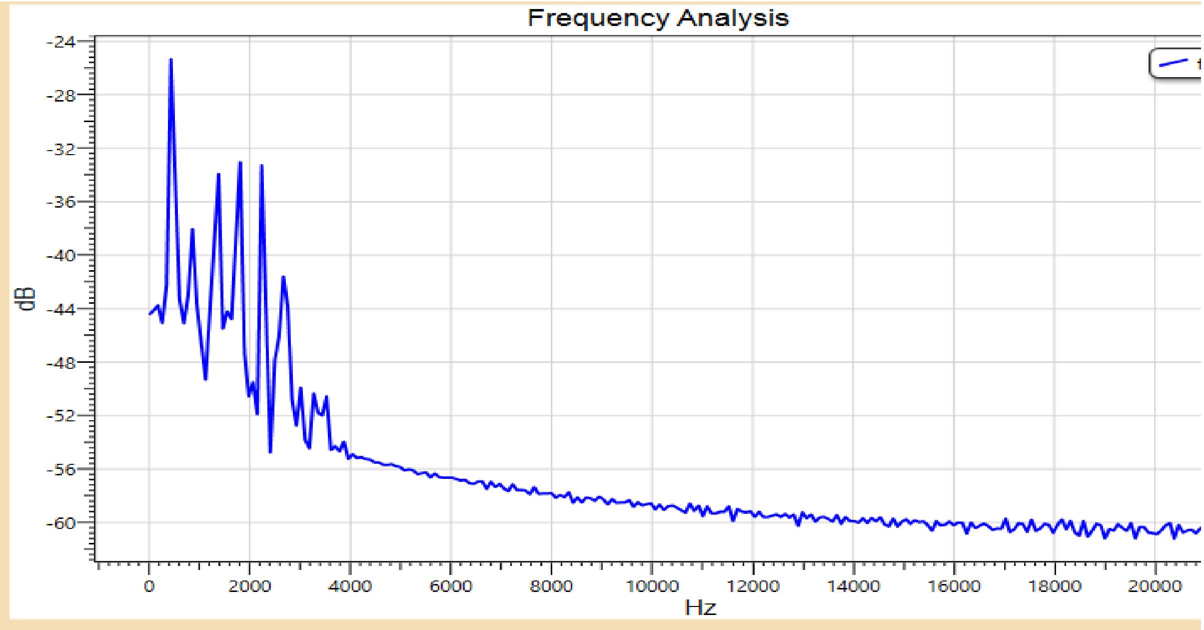
and where $x(n)$, represents the input signal for $n = 0, 1, ..., N - 1$.

Before applying DFT we used a window to multiply our input called Hamming Window This method is used for reducing spectral leakage.

W(n) = 0.53836 – 0.46164 cos (2pn/n-1)

This method is used for reducing spectral leakage

As an example we will take the sound sample given in the artificial wav sounds with the name 455Hz.



After the graphical analysis we had to take in consideration the fundamental frequency and also the harmonics. Choosing a thresholding value was very important at obtaining a more stable results. By using a threshold and trying to collect the amplitude peaks.

After obtaining the peak amplitude we had to extract their frequencies and compare the distance of each pair. The frequency in Hertz of a partial corresponding to X(k) may be computed on the basis of the spectral resolution (f s / N) of the DFT. f s is the sampling frequency in Hertz and N is the number of samples in the analysis window.

In our case sampling frequency was 44100 and number was samples needed was 512. Overall a buffer used for storing the complex numbers of the FFT transformation was 1024 and we actually need half of those due to the Nyquist frequency. The next step of obtain the fundamental frequency was to store the amplitude peaks distances in a list and perform a sorting operation. This was needed to calculate the median of this list which is the fundamental frequency.

Median = x(n-1)/2 if n is odd and ½(xn/2+xn/2+1) if n is even

Depending on the sound we had to adjust the threshold to obtain better results.
In our cases a threshold of -40dB was giving better results. The overall threshold f the graph was -90dB.

4. Test Results using the Fourier spectrum analysis

Sound	Fundamental frequency result	Threshold
Artificial/med/455Hz	430Hz	-40dB
Artificial/med/683Hz	688Hz	-40dB
Artificial/easy/225	215Hz	-40dB
Artificial/diff/405	430Hz	-40dB