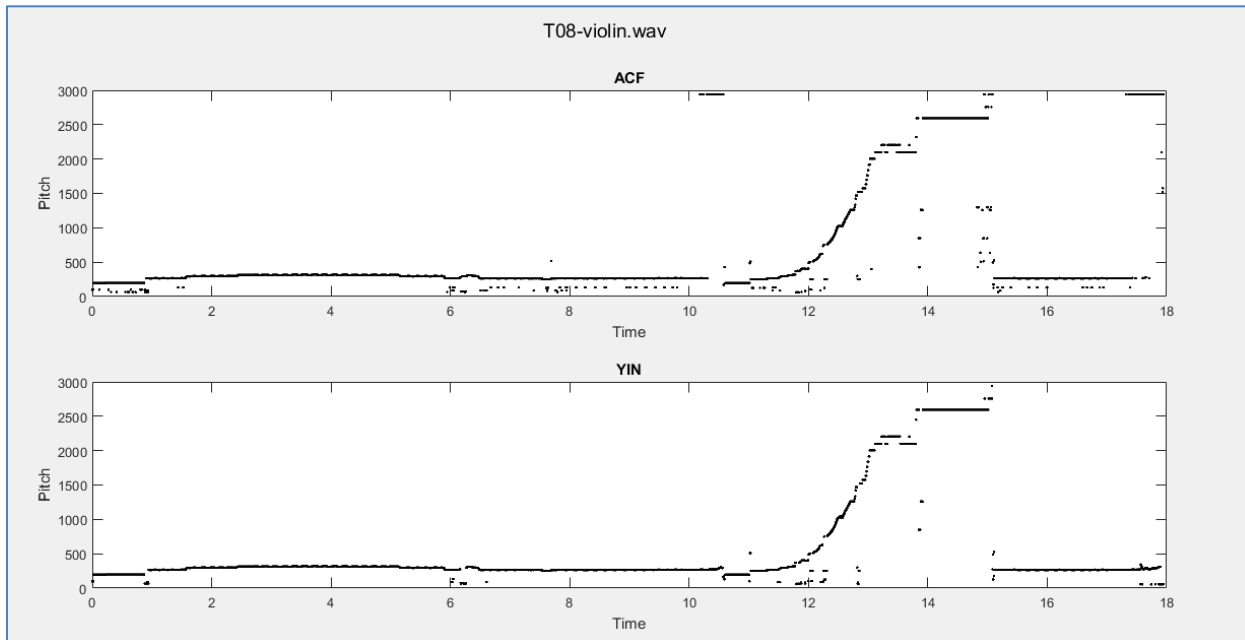


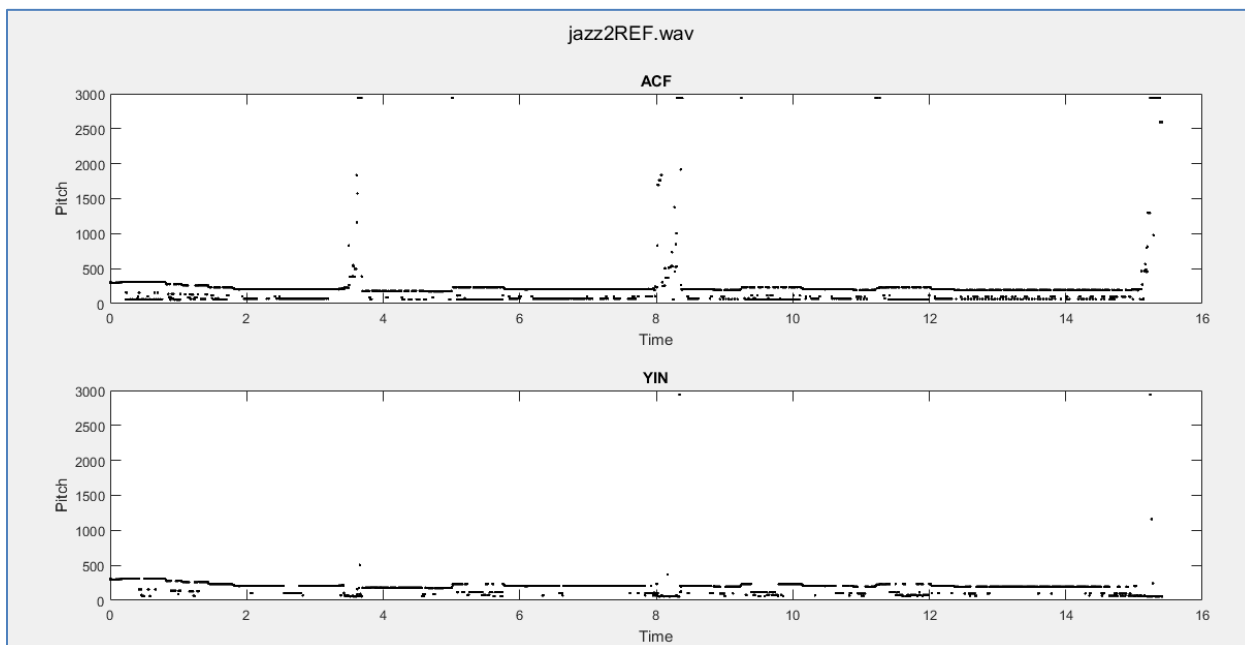
Assignment 4

Q1.

Pitch detection of T08-violin.wav file:

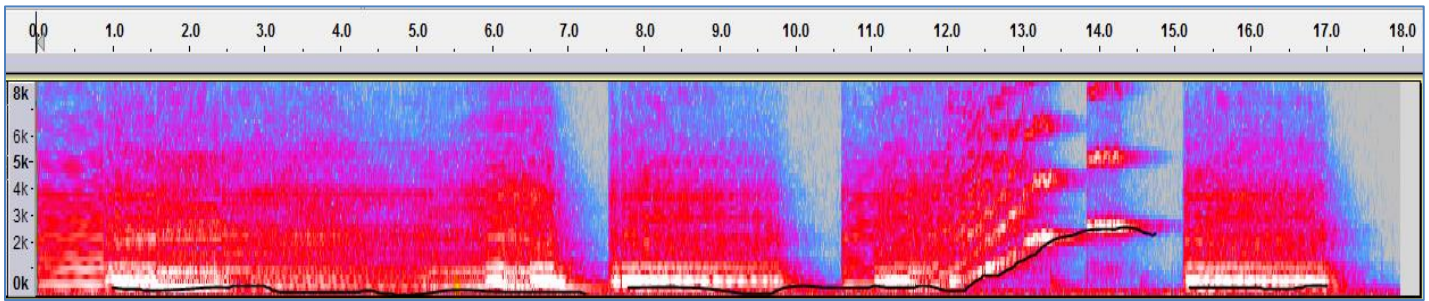


Pitch detection of jazz2REF.wav file:

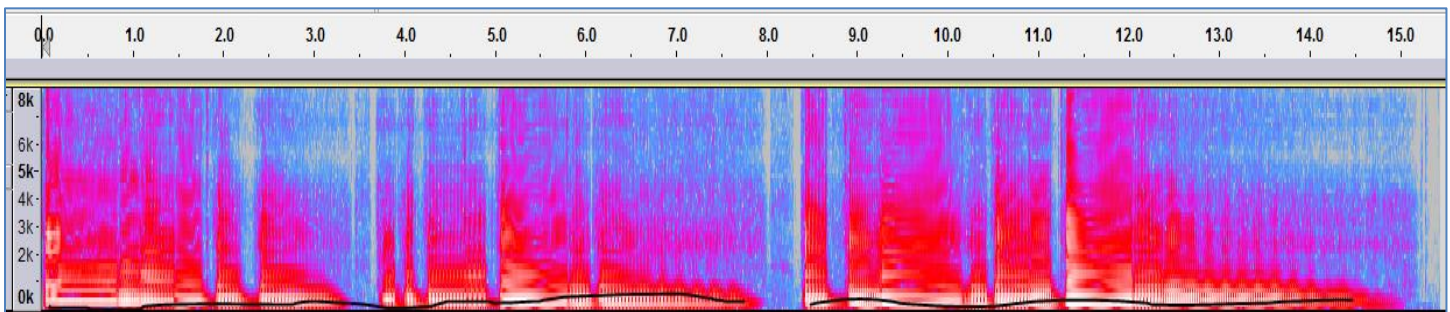


As pitch defines the periodicity – fundamental frequency of the sound thus, through the spectrogram of both the wave file we can approximately say the plots defines pitch of the sound:

Violin:



Jazz:



We can see in violin as the sound fundamental frequency gets higher the pitch detection also increases.

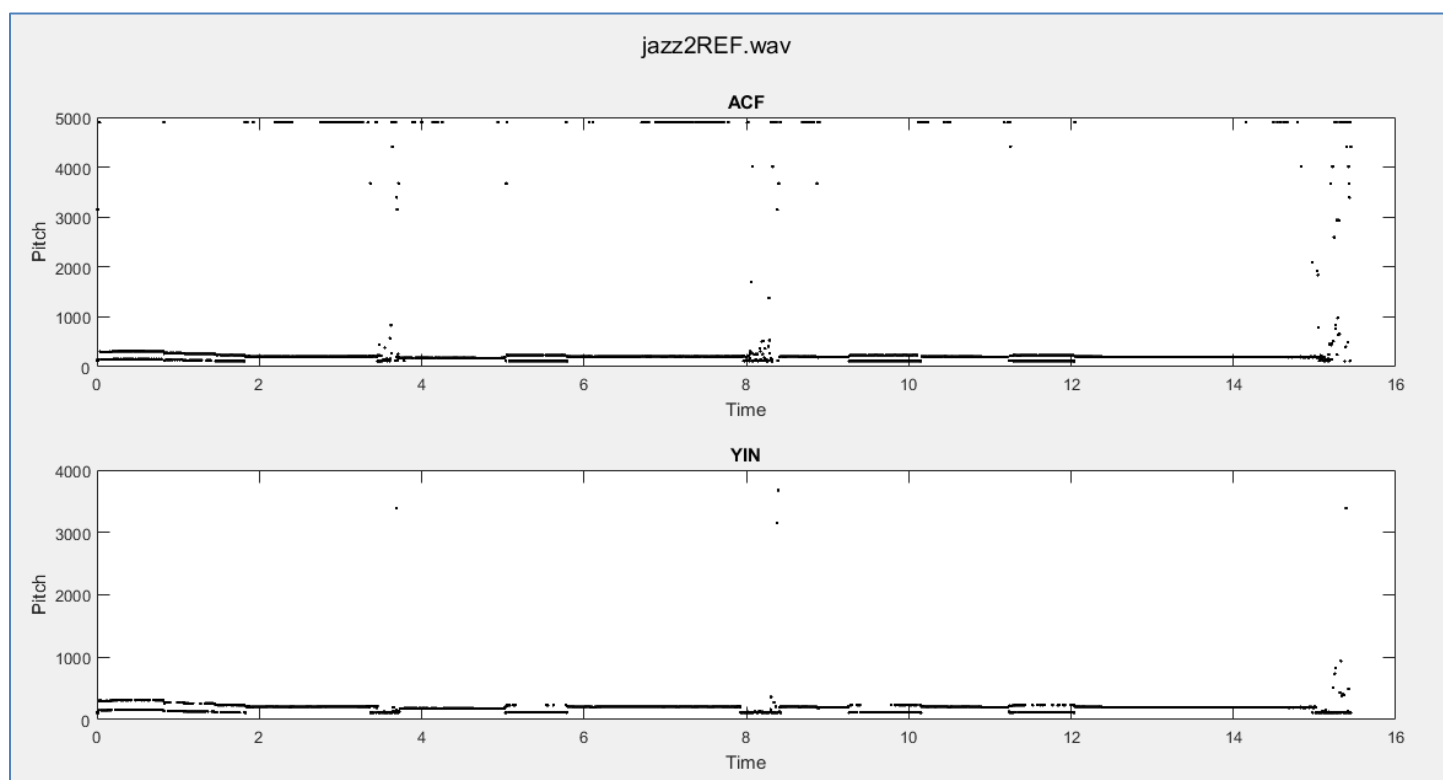
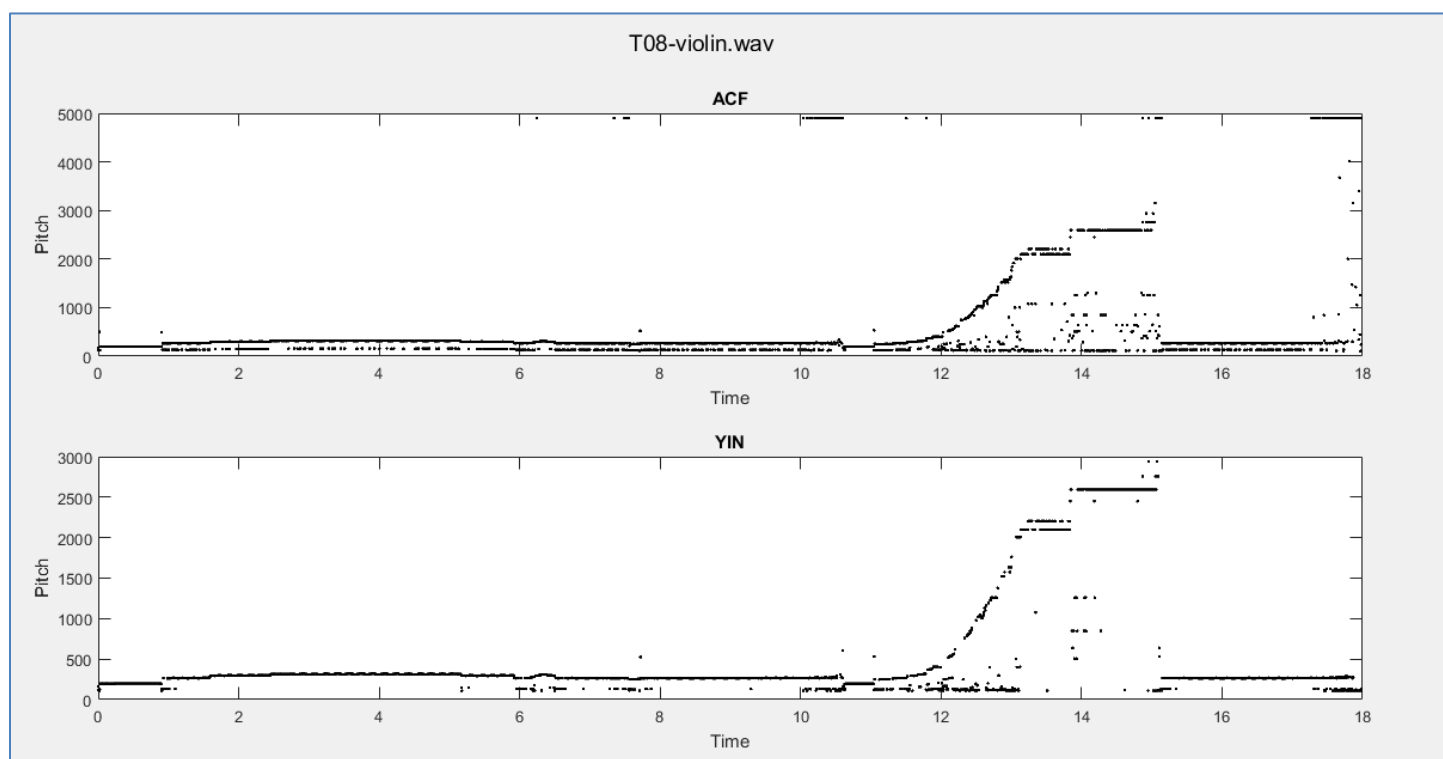
Q2.

Comparing the 2 detections, we can see that where we have no sound, ACF compute it as a high pitch where as YIN compute it as a low pitch. Comparing YIN vs ACF it can be concluded - ACF produce distortion where as YIN produces a confident output.

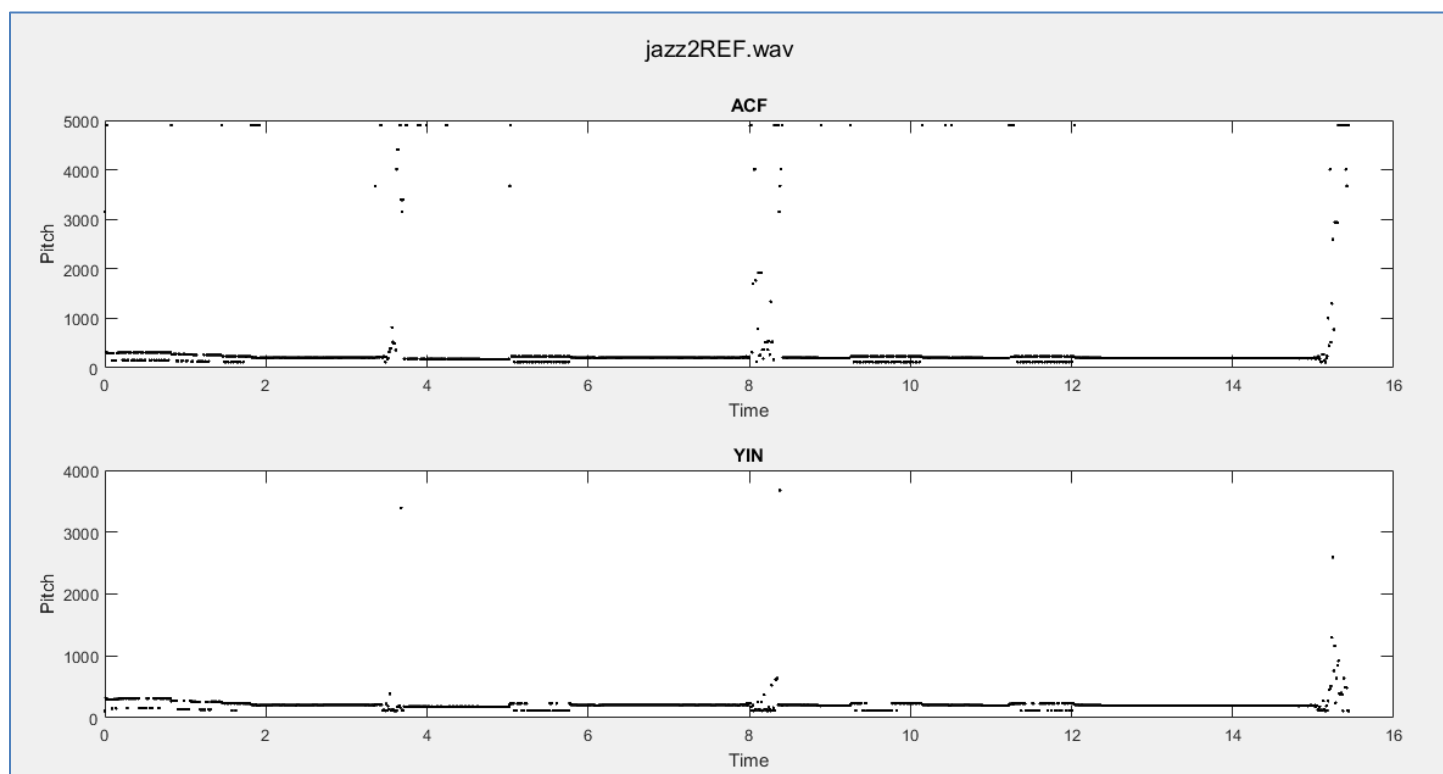
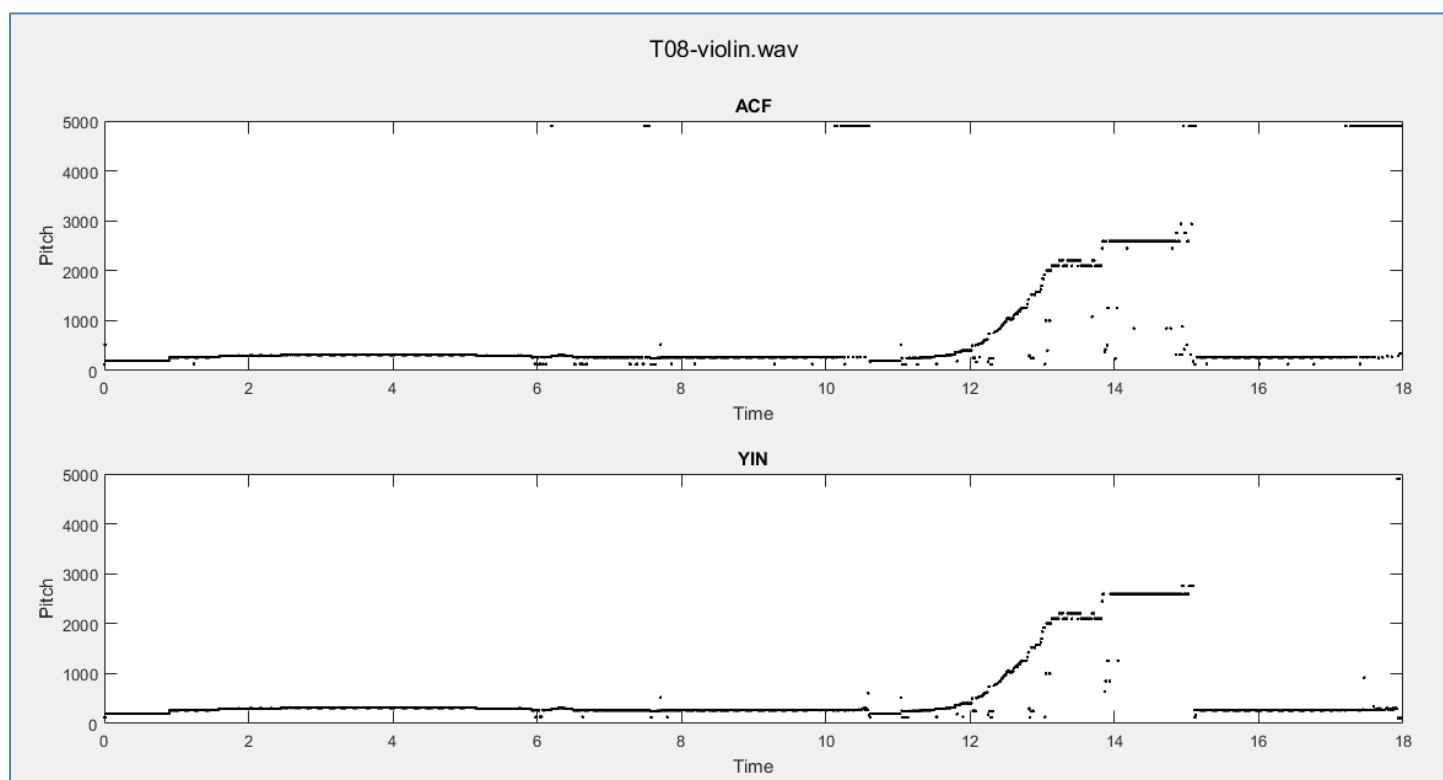
Q3.

min_lag and max_lag values can be defined on the bases of hearing range whereas, hopsize must be less than windowlength/2. Thus, varying the window length, we will define lags for the range – 100Hz-5kHz(9-400lags) and hop as $\frac{1}{2}$ of the window length.

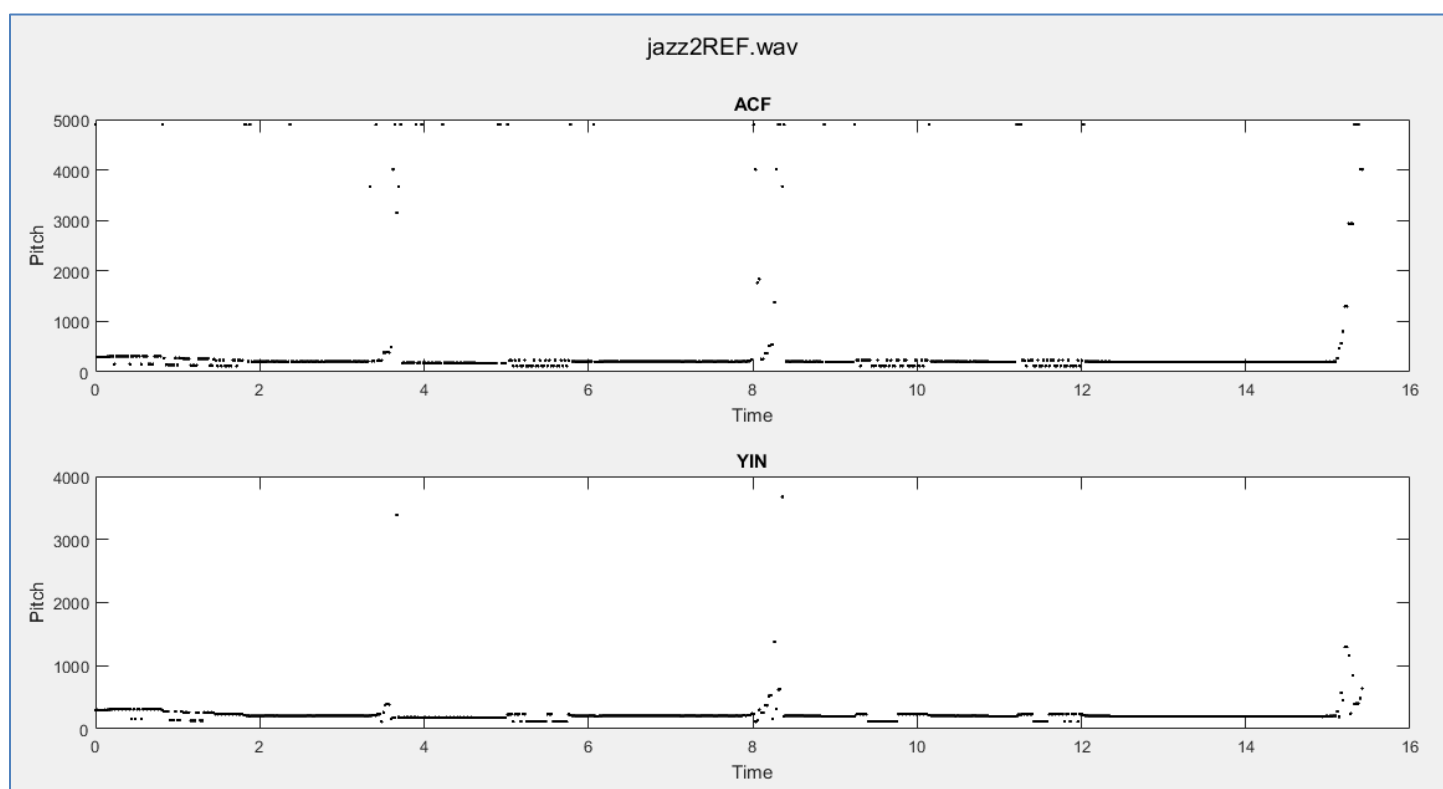
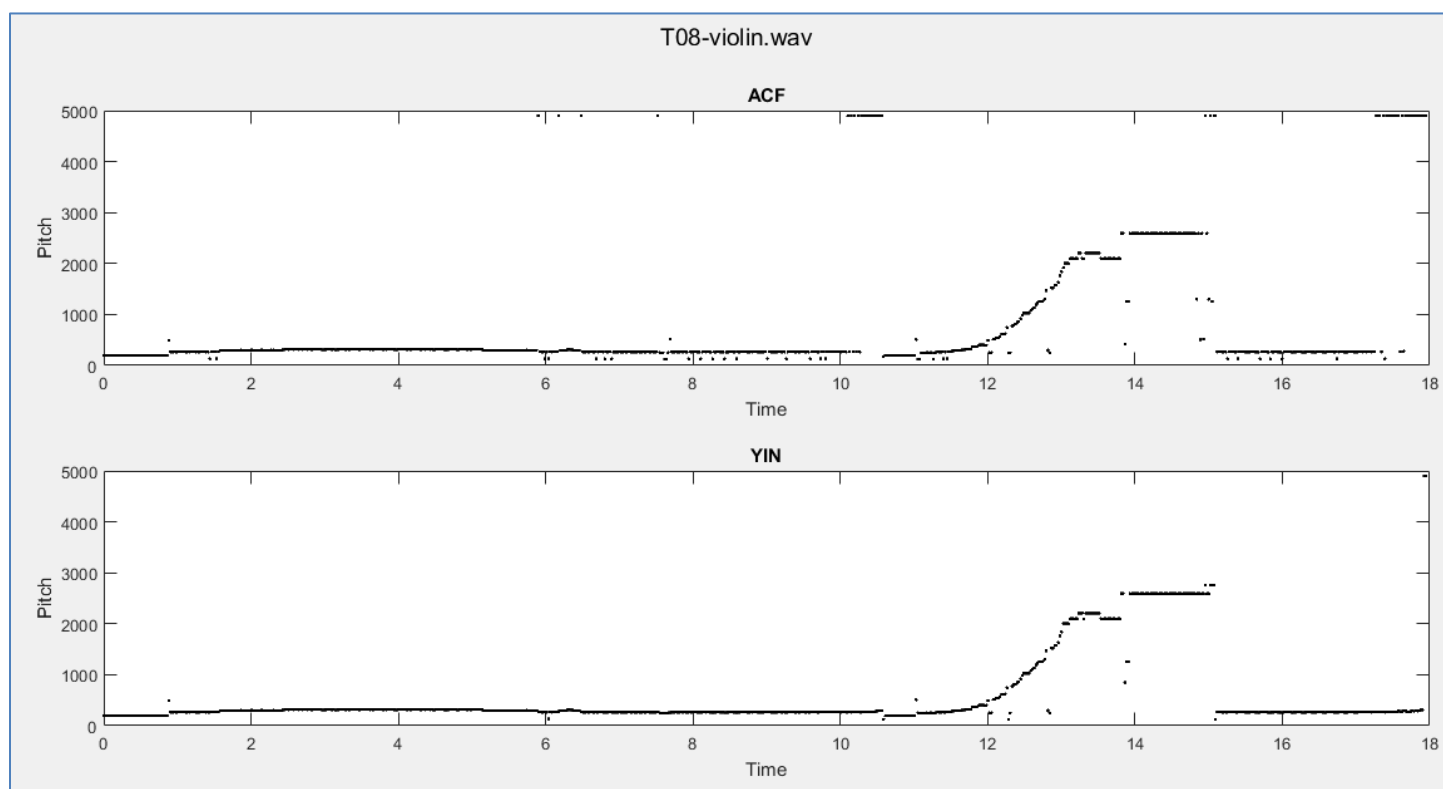
Window length – 512:



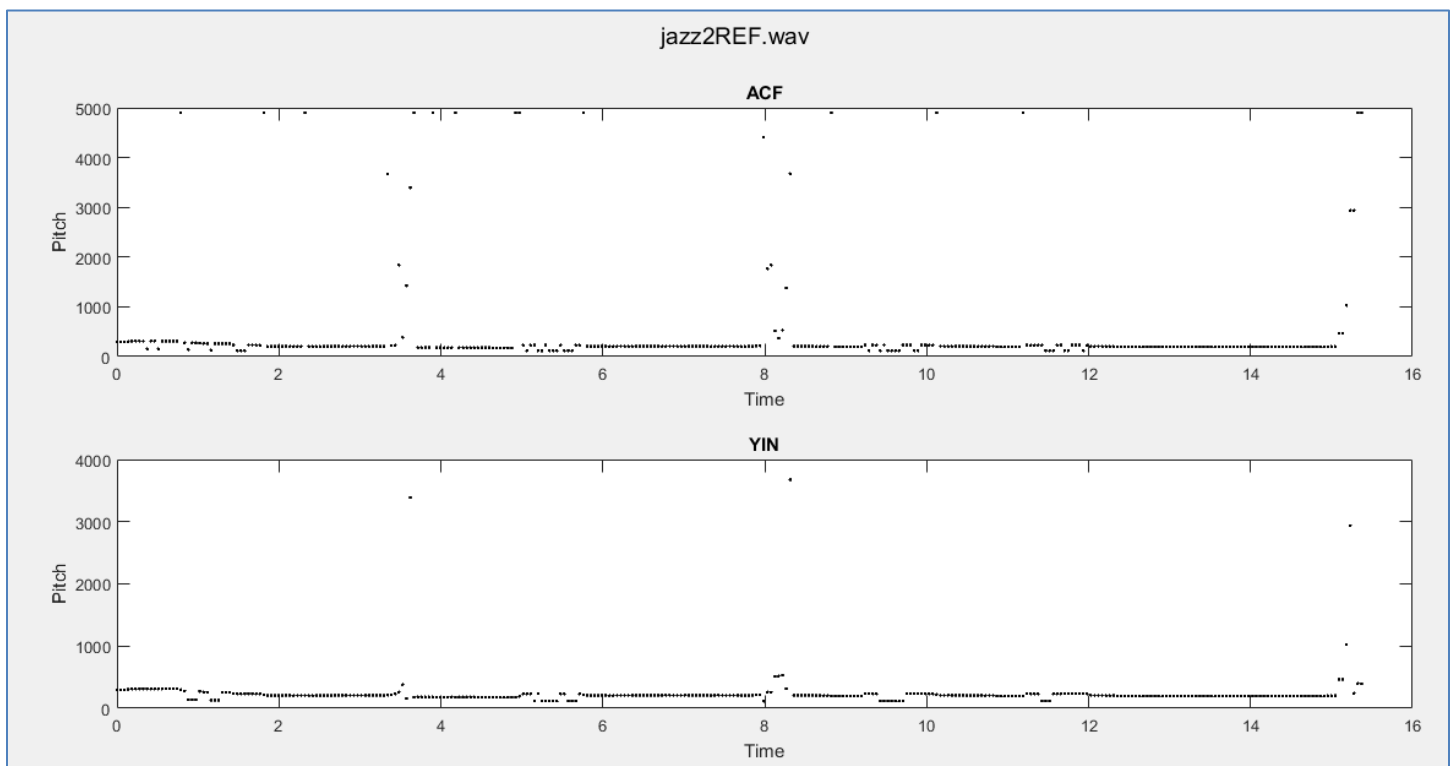
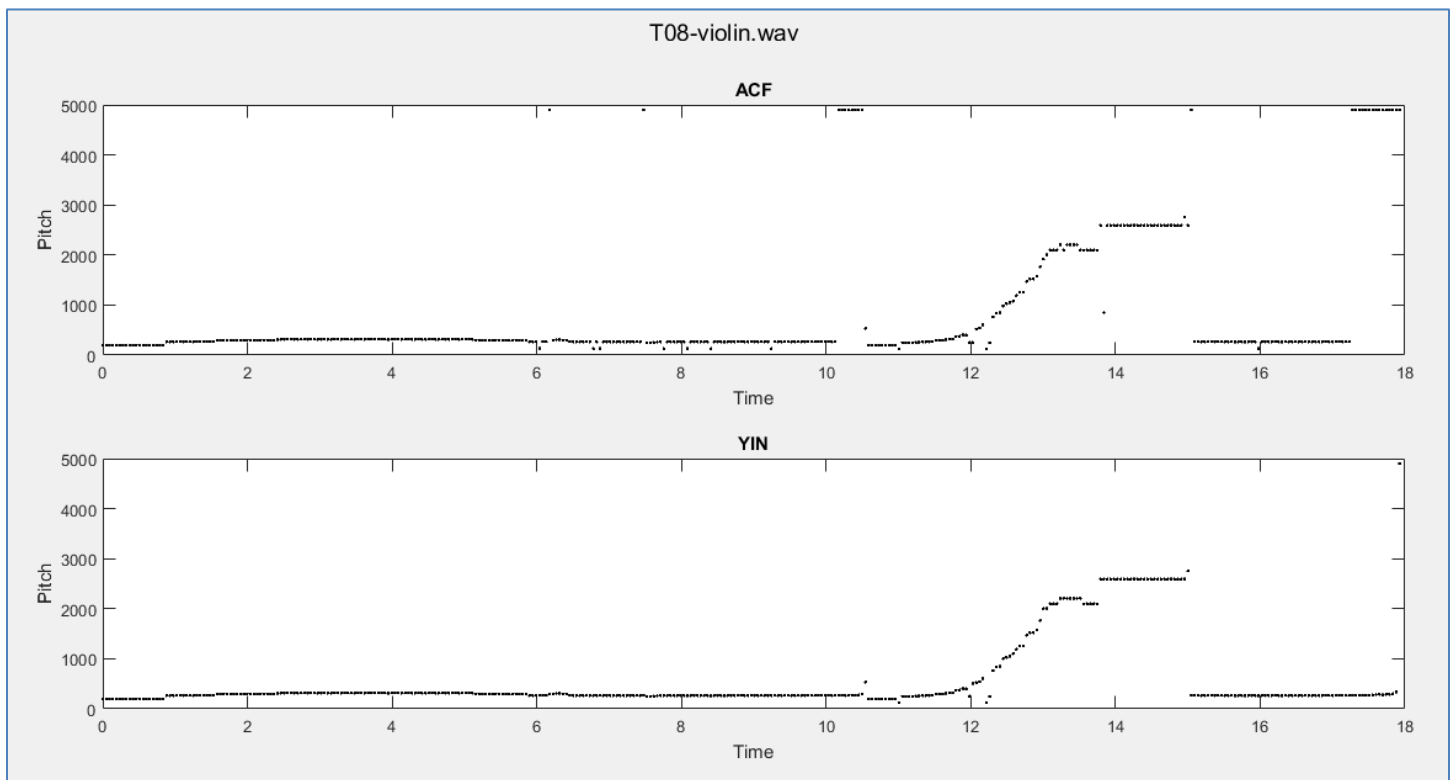
Window length – 1024:



Window length – 2048:



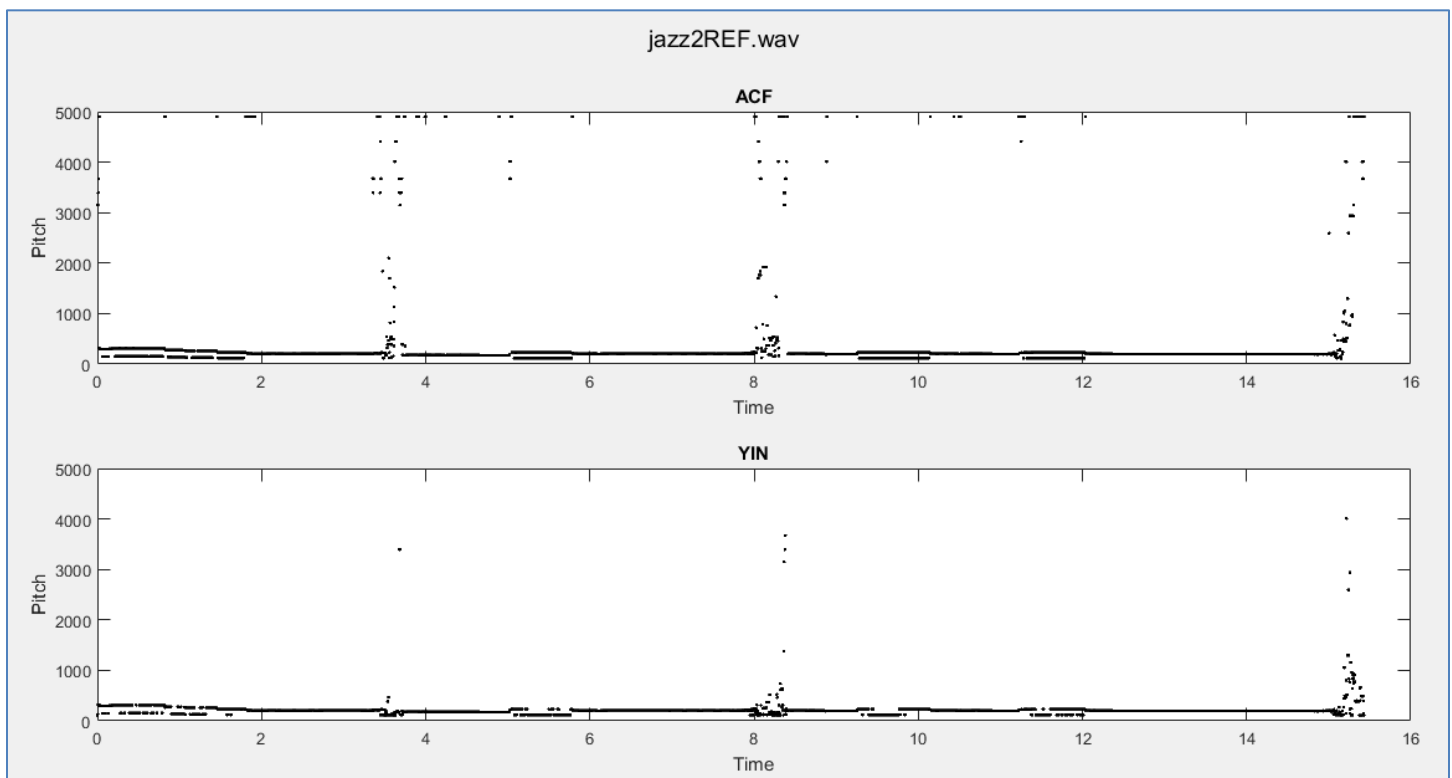
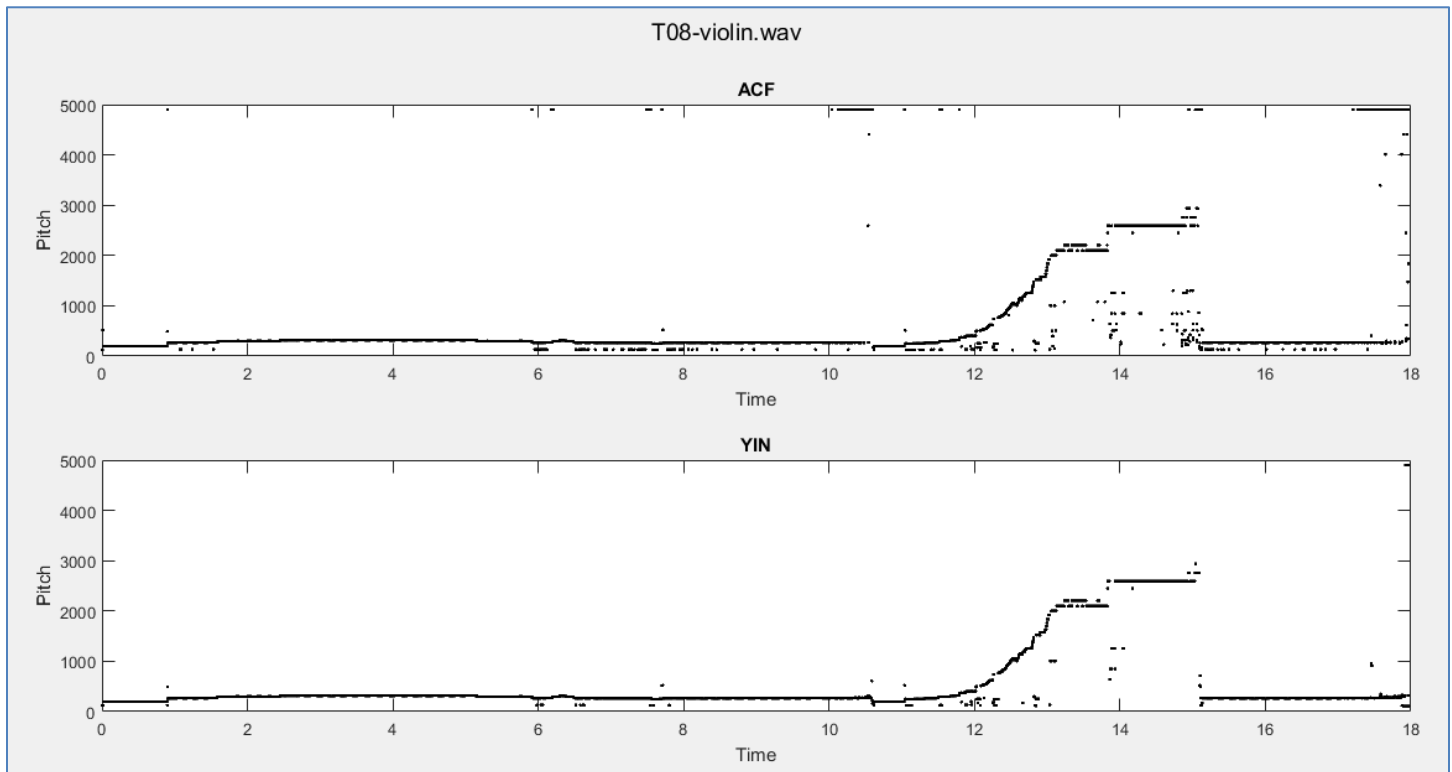
Window length – 4096:



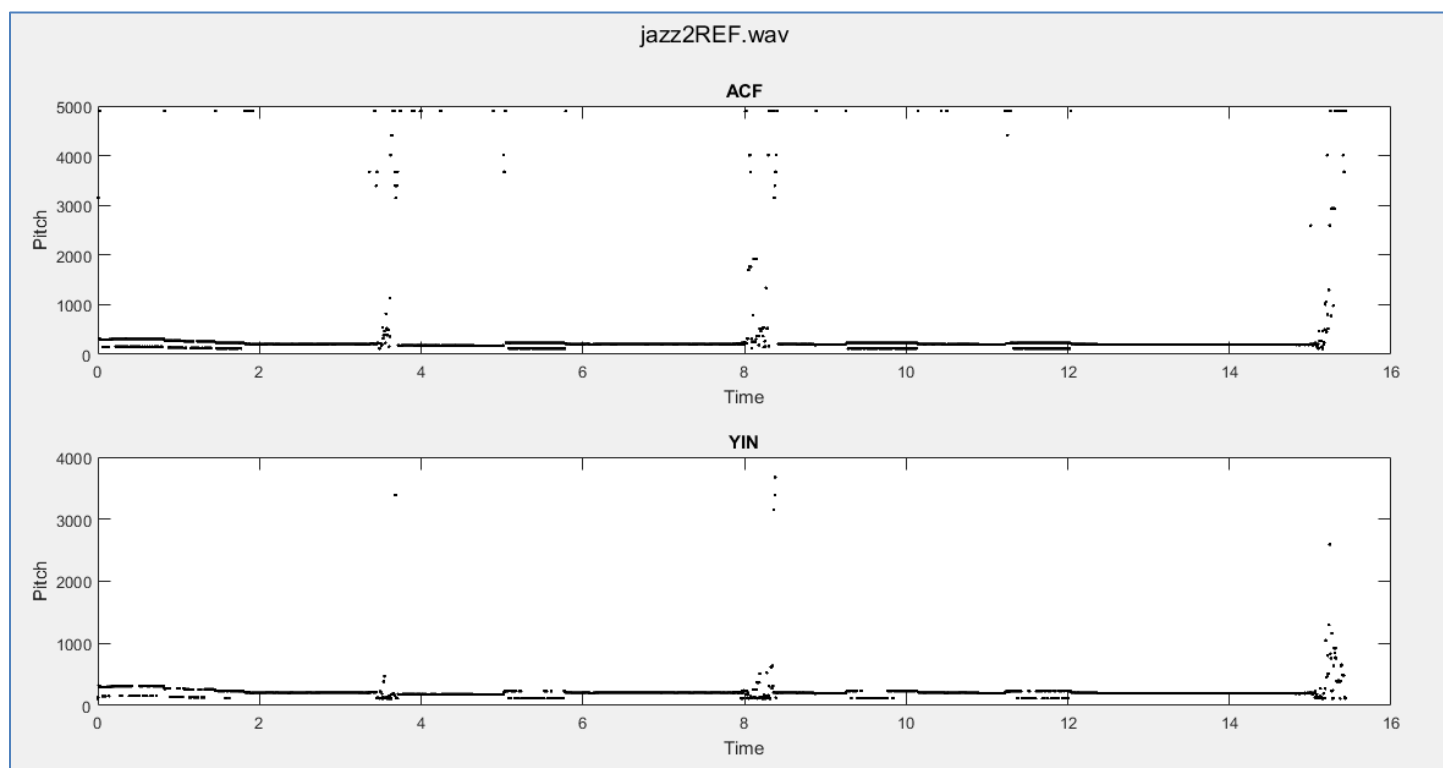
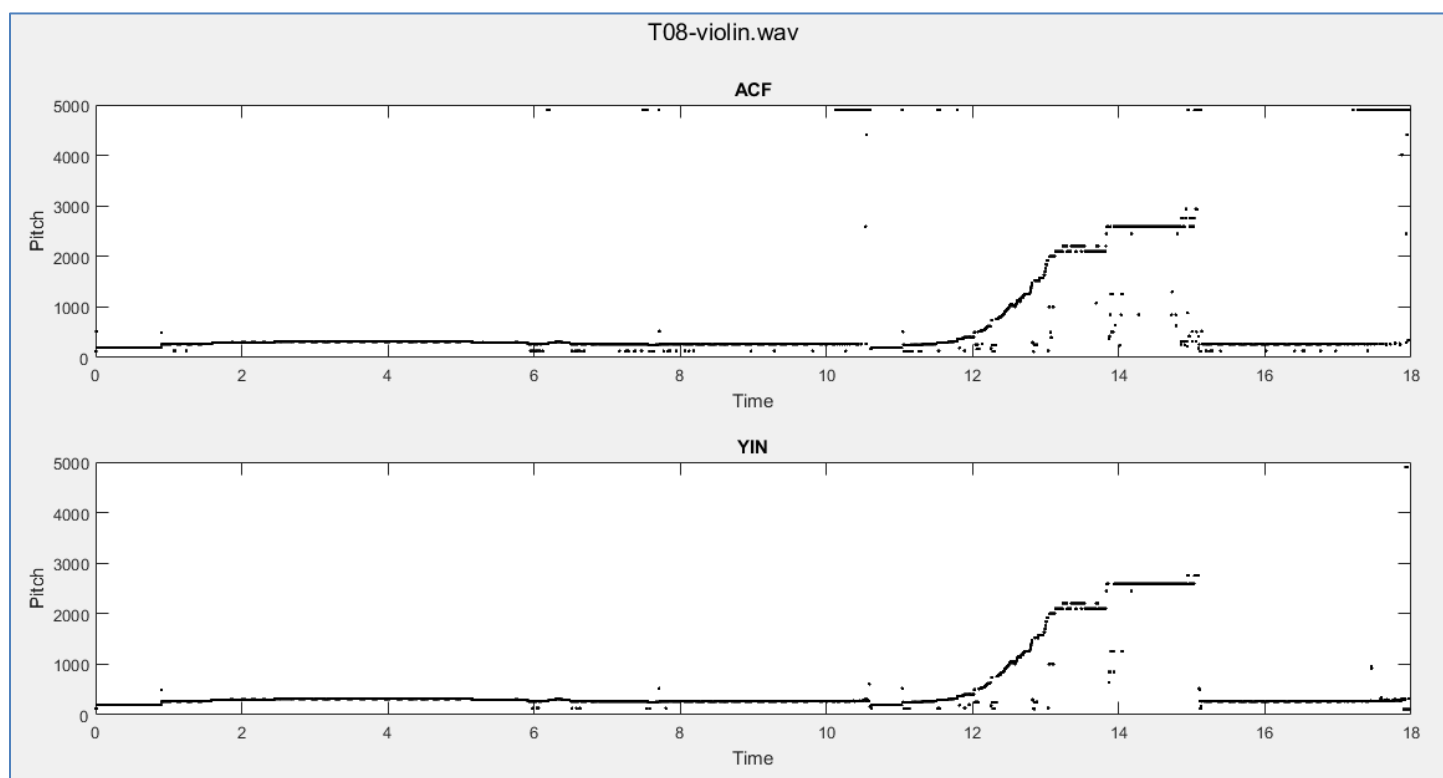
As the window length increases, most of the pitch information for YIN corresponding to the low frequency, whereas low window length ACF is not able to correctly measure pitch and then maps it to high frequency thus we might take window length as 1024,2048.

Now we will vary the hopsize as $w/8$, $w/4$, $w/3$, $w/2$ for $w = 1024$:

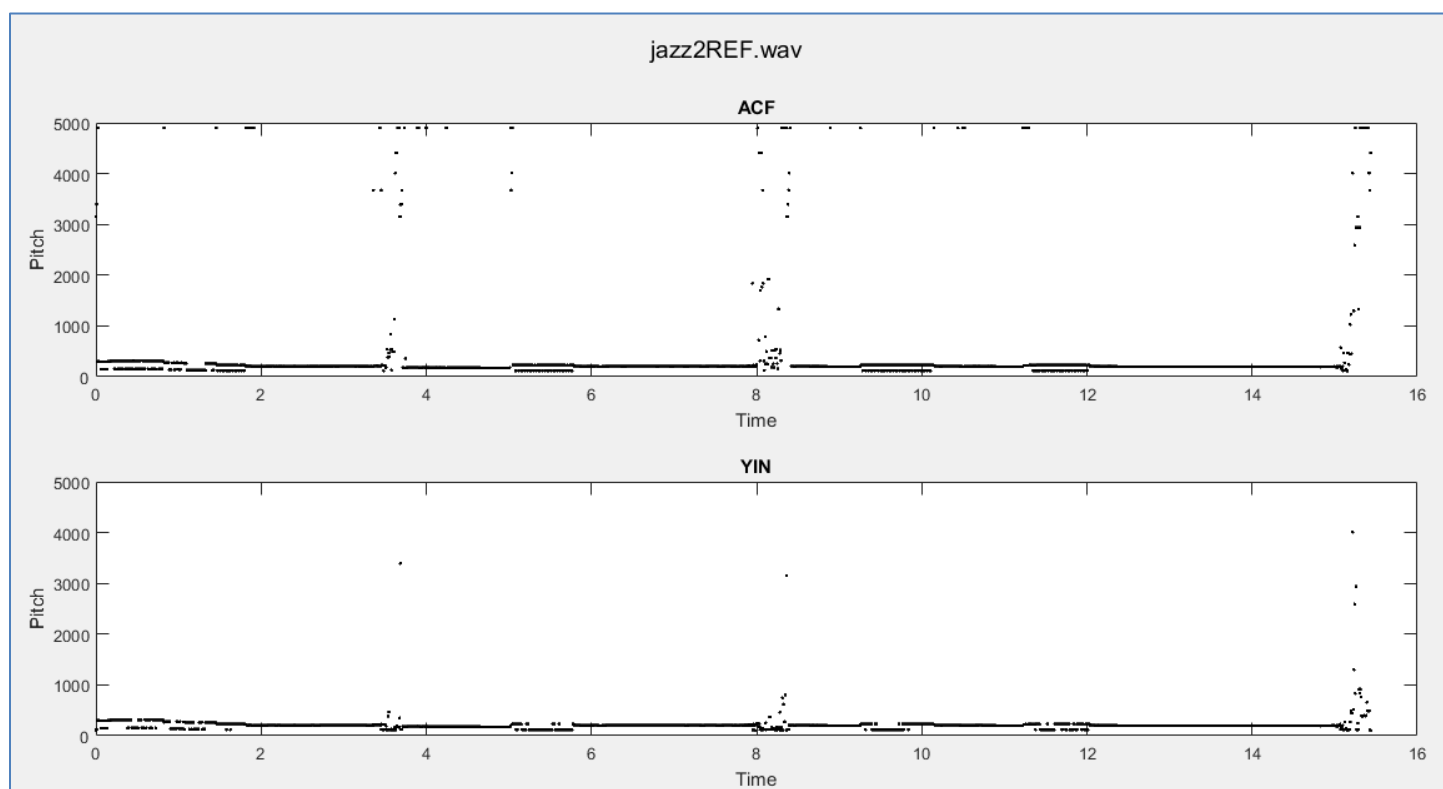
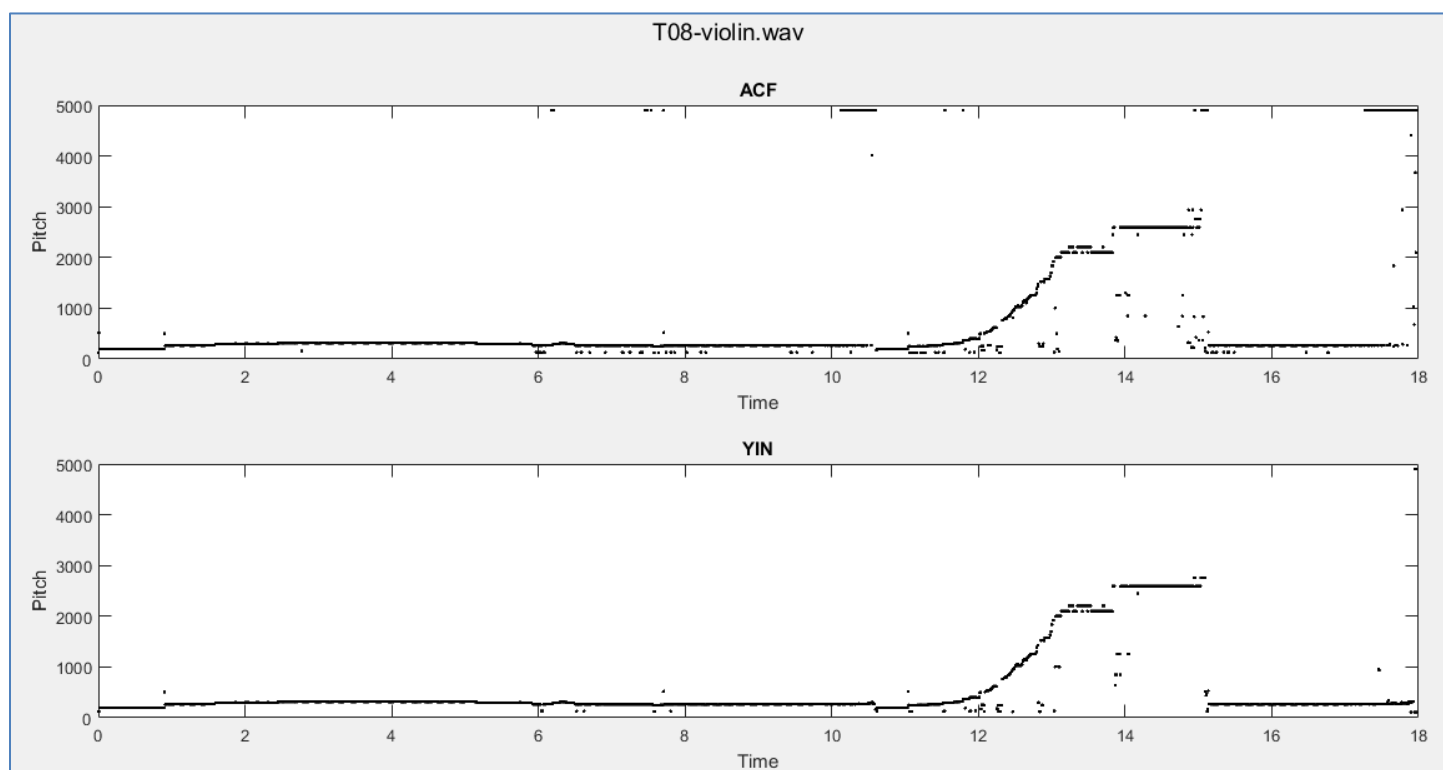
Hop = $w/8$:



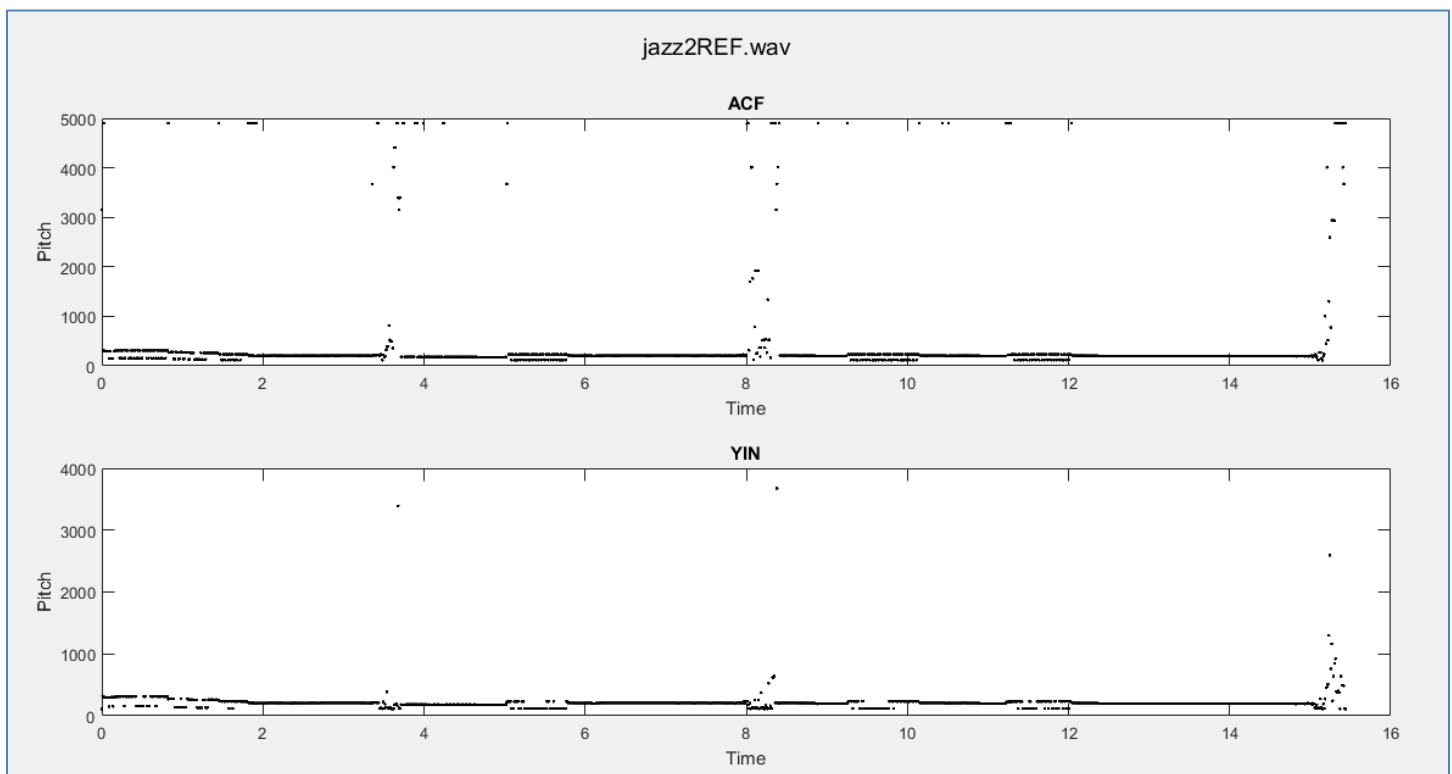
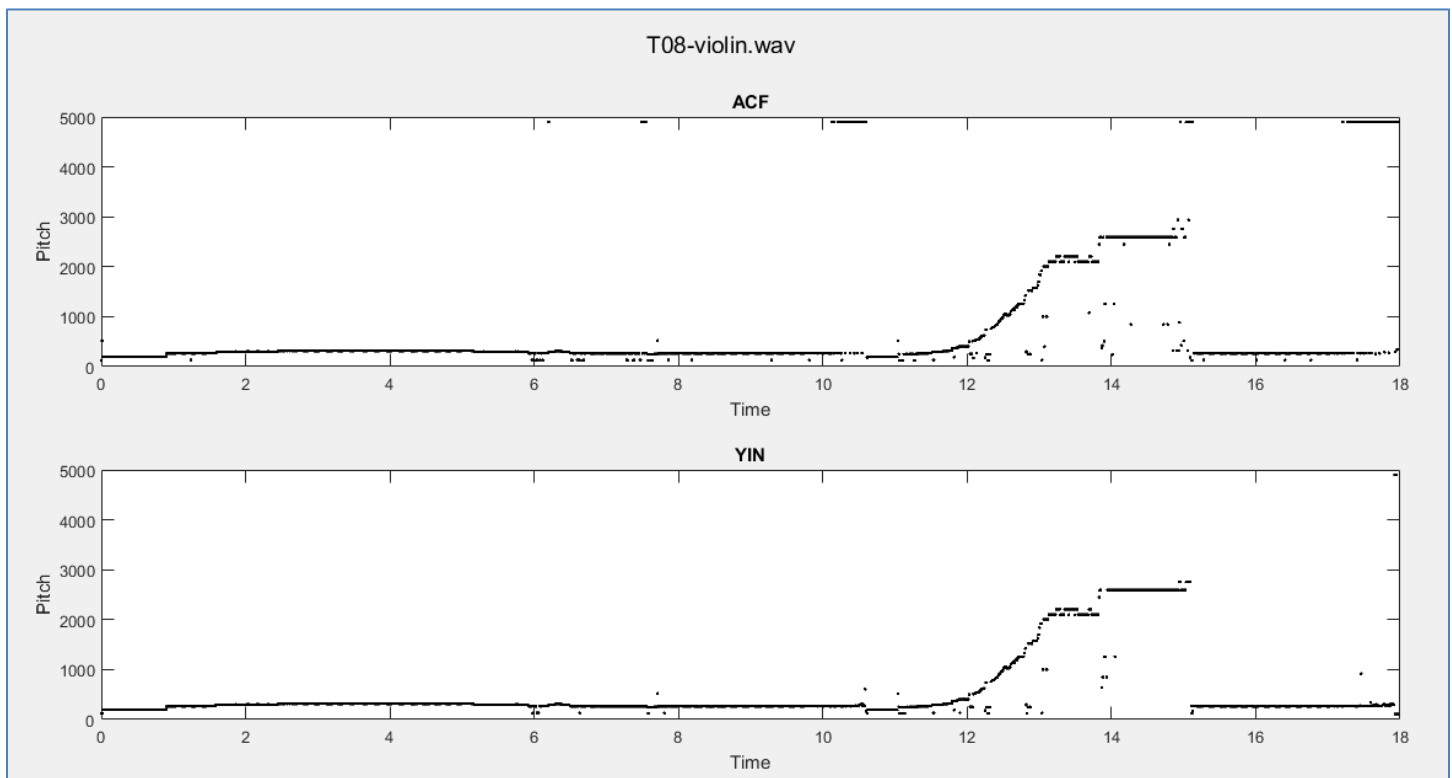
Hop = $w/4$:



Hop = $w/3$:



Hop = $w/2$:



As the hop size decreases, computation time increases whereas, we cannot determine much difference in the pitch plots.

Thus, we can take $\text{hop} = w/4 - w/2$.

Thus, we can take parameters = win_size = 1024, hop = 512, min_lag = 9, max_lag = 400.

Q4.

The pitch detecting function can be improved by improving the peak detection function as:

1. Detect all the local maximas/minimas position in the min_lag and max_lag range with a thresholding function.
2. Compute the difference between all the consecutive maximas/minimas.
3. Take the average of these differences to compute pitch value (pitch value = $1/\text{average_time_diff}$).