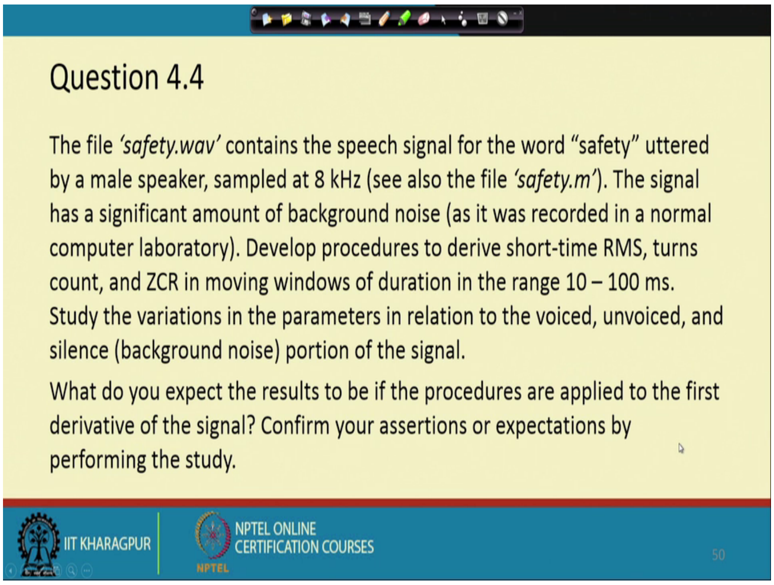
Biomedical Signal Processing
Prof. Sudipta Mukhopadhyay
Department of Electrical and Electronics Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 60
Tutorial - IV (Contd.)

(Refer Slide Time: 00:16)



So, now, we will go for the question 4 of tutorial 4. We have a speech signal which have the utterance of the word safety and sampled at 8 kilohertz. The signal has a significant amount of background noise ok; that means, if we are trying to find out that what part is speech, what part is silenced it becomes difficult because of this background noise. And we need to develop some procedure to find out short time RMS turns count and zero crossing rate.

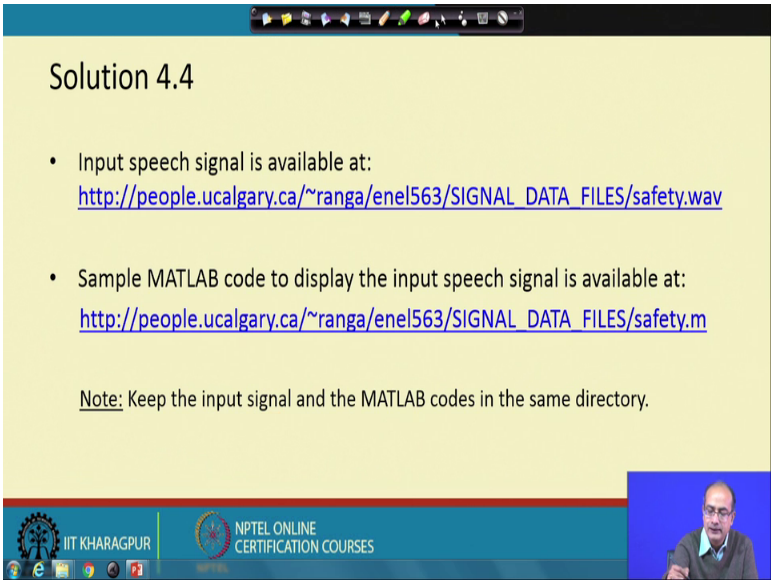
So, already we have seen that short time RMS and turns count. Now, the new thing is zero crossing rate, ZCR, in a moving window of duration 10 to 100 milliseconds. So, moving window that size is changing.

Study the variation of the parameter in relation to voiced, unvoiced and silent that is background noise portion of the signal. So, what is the change in the activity of RMS value turns count or ZCR in all these cases for different types of signal voiced, unvoiced or silence that means, when only the background noise is present. And we are asked to look at that if we apply that the same on the first derivative of the signal what would be the result, ok. Confirm your assertions or explanations by performing the standing we need to do the study to give that.

Now, here you would like to tell that why the first derivative is told. Say if we take a signal in that zero crossing rate depends on where the it is crossing the 0, if it is going through the middle we get the right rate. But say if it is biased if it is going more towards positive we are coming here then sometimes we may miss some of the turns, ok.

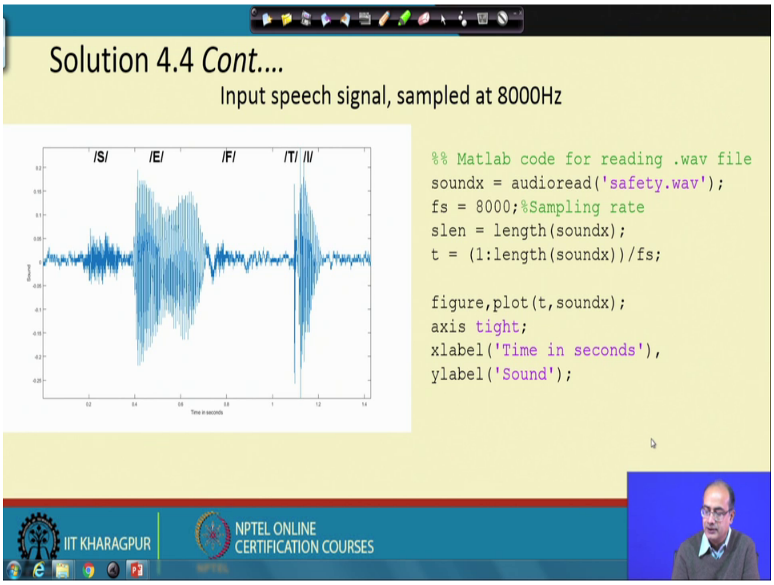
So, zero crossing rate would be biased if the DC value is there, DC bias is there or; that means, baseline wondering is there or there is a pedestal. So, when we take the first derivative actually we can get rid of that, ok, then the zero crossing rate should be better. But then when we are taking the derivative it can also accentuates that accentuate the high frequency noise, ok. So, that is the challenge that is what we are asked to see that how we get that result in this example.

(Refer Slide Time: 03:41)



So, first we start with that data that, safety dot wav we download next we take the safety dot m to read that the signal. And we keep that input data as well as the MATLAB code in the working directory of the MATLAB the same directory and let us proceed with the work.

(Refer Slide Time: 04:08)



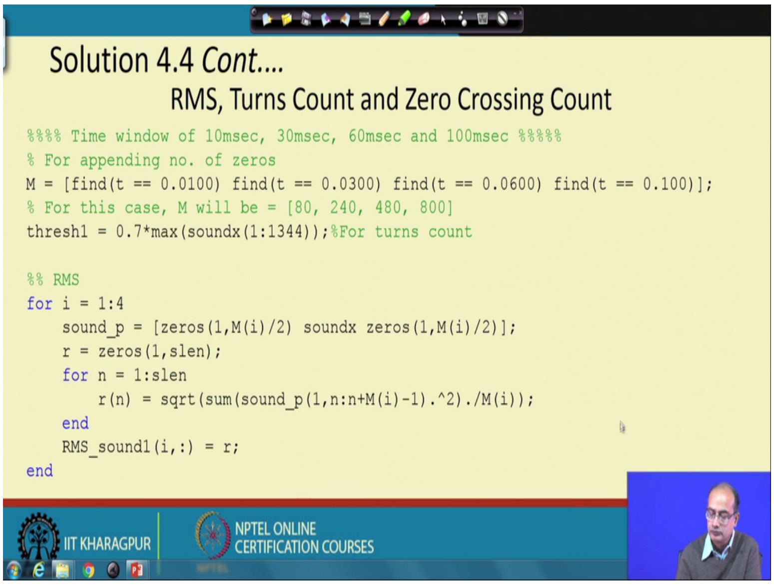
The first thing that the signal what we get we need to load that signal. So, we need to because it is in wav file, not that file. We are using a different command audioread note that load command as we are doing in the previous case, ok. And this is given in the MATLAB code so we need not have to worry much about the same.

Then we know that what is the sampling frequency that is 8 kilohertz. So, we take that number of samples present that that is given by the length of the signal sound signal, and we create the time index out of it, and we plot that axis tight makes it that there would not be any gap, that we would not leave much space above and below and we just give the x and y level here, ok, with that we get the output.

So, here is the plot we get. Here in this we should note one point that here we have also shown that what are the different phonemes they are uttered. First is yes before that this part this is background noise, ok. Then in safety then this is a consonant S is a consonant, then a phoneme E sound this is a voice sound this is after consonant this is voiced, again F is a consonant or we can say unvoiced sound, then T is a plosive then again that safety that I sound this is a voice out, ok.

So, we get the different kind of sounds this is unvoiced, then this is voiced, this is unvoiced we can say, this is voiced sorry this is again unvoiced, this is voiced, and then again after that we are getting noise here. So, all those silent portions in between the that we get they are not actually looking like no signals because of the presence of the that noise ok. So, with that we start our experiment.

(Refer Slide Time: 06:59)



First we have to calculate the that the different things like RMS, turns count and zero crossing and we have to calculate them for varying size of the that window. So, first we pick up that size of the window that is given in samples, we put that in the value of M in a vector corresponding to 10 millisecond, that 30 millisecond, then 60 and 100 millisecond, ok.

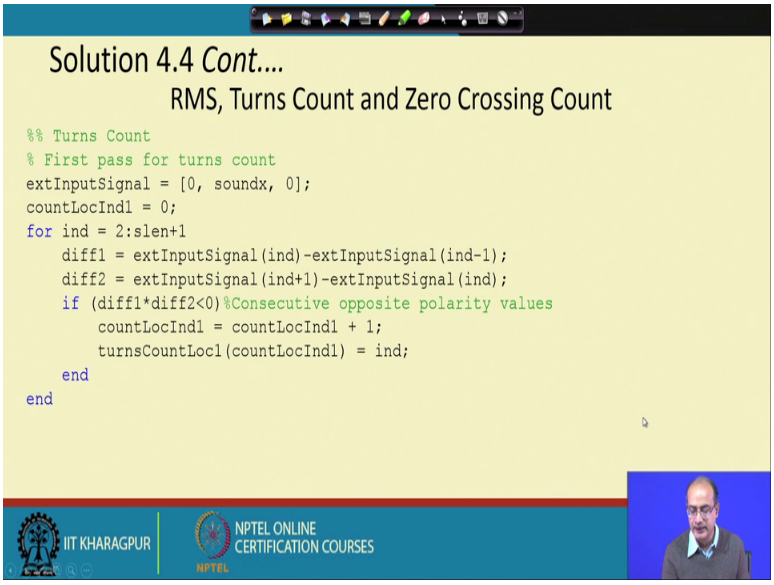
So, that is a the way we get that the different values of number of samples we can directly also get that, and the corresponding value out of that what we get that we get for 10 millisecond it is 80 samples, then 240 for 30, 480 for 60 and 800 for 100 millisecond, ok. That we could directly get also from the sampling frequency. So, here it is done in a little different way.

Now, we take some that threshold for turns count that whatever is the maximum value within this interval we are taking 0.7 of that, ok. That is the threshold taken for the turns count, we have already explained that why we take that threshold to give it the immunity from the that noise, so the threshold should be bigger than the noise pedestal. So, through experiment then that value is take.

Next, we compute the first the RMS value that we pad both the sides with 0s, and we keep the sound at the middle, and here we are not supposed to take actually the causal window. So, both the size the padding is required and the size of the that padding both the side is half of the size of the window that is if we are taking 80 samples, 40 samples, actually 40 zeroes at the beginning, 40 zeroes at the end are padded.

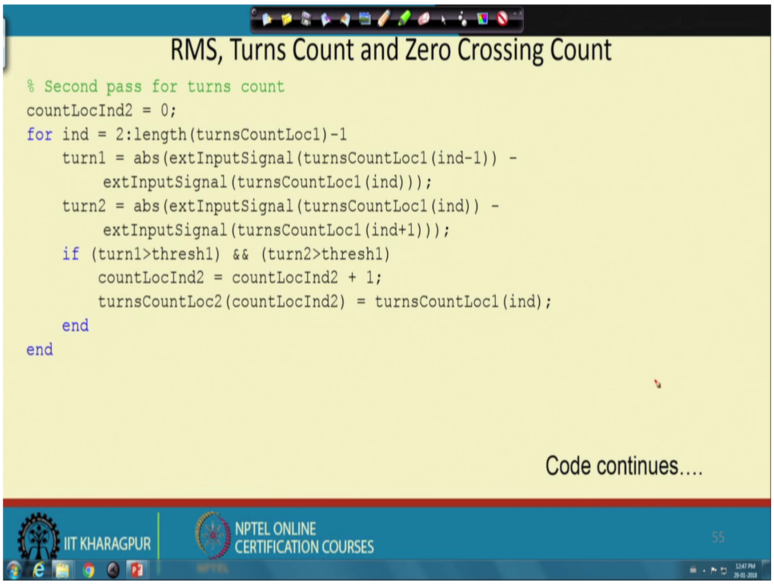
If you are taking 240 then 120 zeroes at the beginning, 120 at the end is padded. So, there is a way the padding is done. And we compute that sum of squared actually energy divided by the count; that means, the average energy instantaneous average instantaneous energy and taking the square root that that is giving that that RMS value at that instant. So, we do that for different sizes of the window.

(Refer Slide Time: 10:14)



Now, we move for the turns count. So, as in the previous case that we are going through the first pass to find out the turns; that means, we are going up and down or going down and up, ok. There is a change in the slope, so we are looking at that and we are finding out these are the terms. So, in the first actually phase we are getting that only.

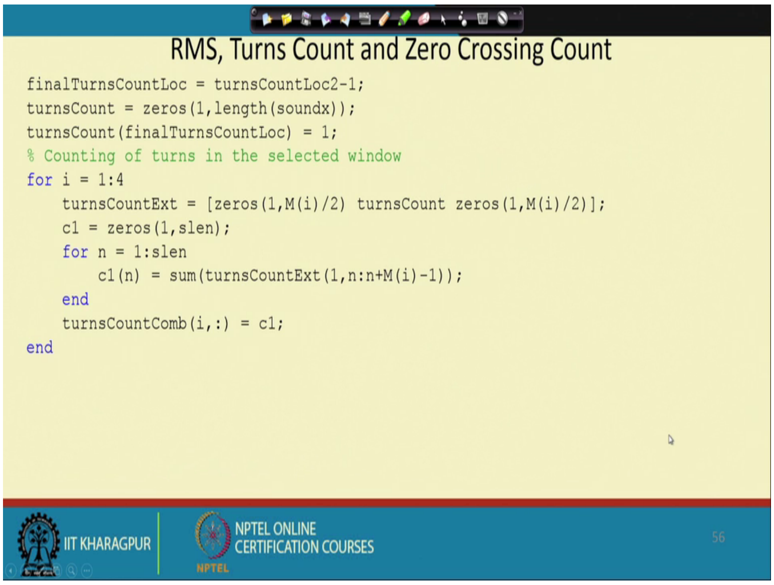
(Refer Slide Time: 10:45)



In the next phase we are looking at that applying that threshold it should be above the threshold. That means, if we have a code that if we have a turn like this which has sufficient height before the other turn both the side then we tell that both the side that there is a sufficient height is that will tell this is a real turn, is the real turn.

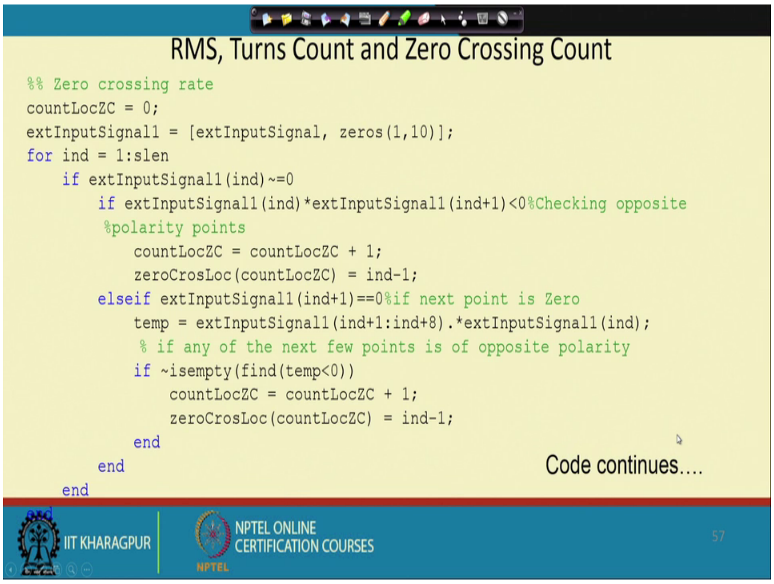
However, if we get it like this that then if we look at this is a turn, this is also a turn, but the difference between them is so small that height does not justify to call as that a significant turn. So, we should ignore them, ok. So, that is that purpose of the this piece of comb.

(Refer Slide Time: 11:48)



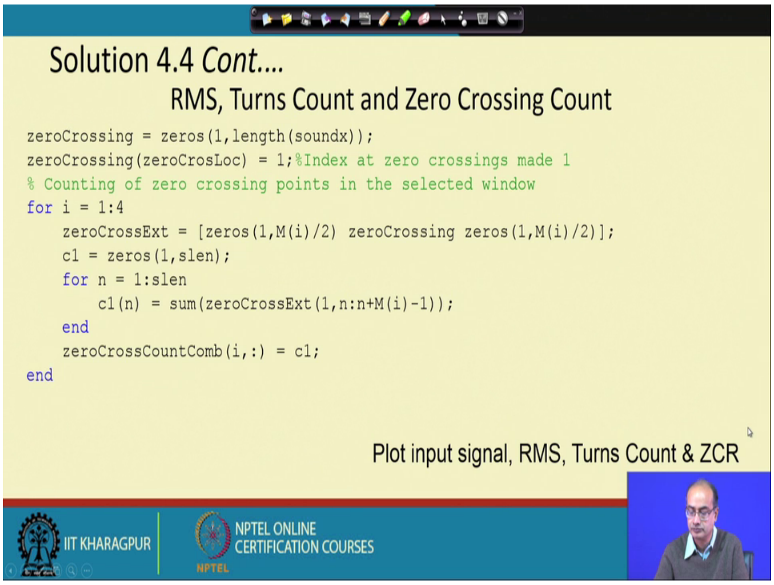
So, we are having that and we are selecting the turns count after cleaning those part after the thresholding.

(Refer Slide Time: 11:51)



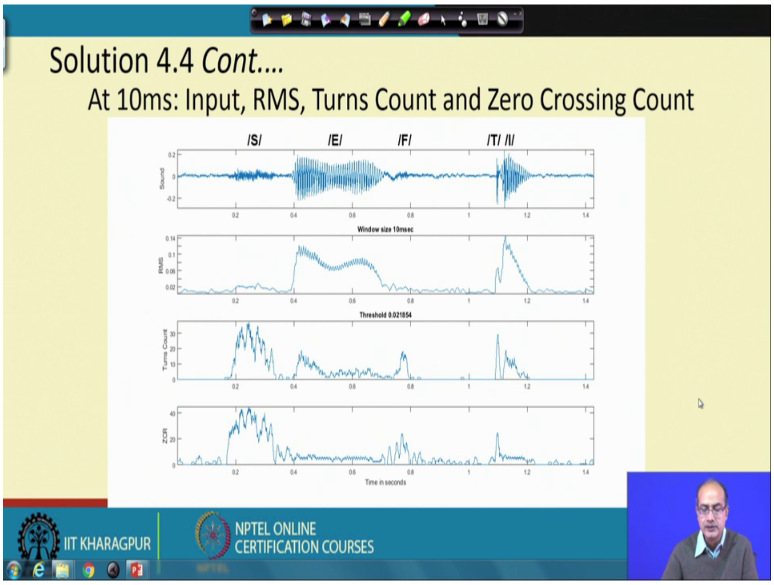
Then we go for zero crossing rate. We are finding out that how many actually change in signal is happening and we are marking that. And then we are actually counting that within that window size how many such actually change in sign is happening, that gives us the zero crossing count ok, or zero crossing rate.

(Refer Slide Time: 12:35)



So, the that last part of it is to take that count that how many zero crossing has happened.

(Refer Slide Time: 12:51)



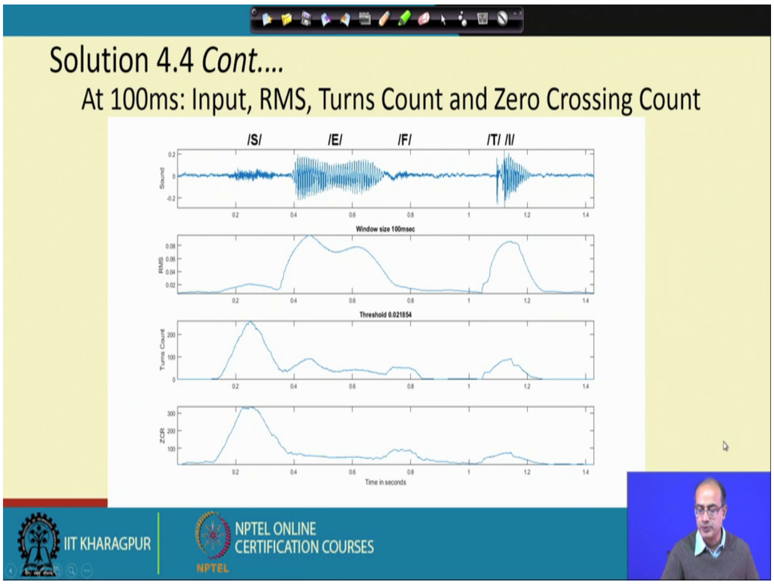
Now, this means we plot these signals for that input signal the sound we are giving here. And we have taken 10 millisecond window that is 80 samples for that the RMS count is given, turns count is given ZCR has been given. What we notice that that RMS is giving the envelope of the signal.

So, for the part E that is a voice sound we are getting high value, again for safety and that is again a voice sound we are getting a high value. On the other hand when the unvoiced parts are coming or the consonants are coming like S or F we get the zero crossing rate is higher, ok, that those part and that zero crossing rate are giving higher.

And same nature we get sorry turns count is higher and same nature is followed by the zero crossing rate with a difference, the difference is when we are supposed to get silence still zero crossing rate is giving some value here, ok.

So, that is the difference we get, ok. But if you look at the turns count and the zero crossing rate they are giving the same phenomena that a, an index of the inherent frequency of the signal. So, they are very similar to each other, and that is a thing we could get for a single plot.

(Refer Slide Time: 14:57)



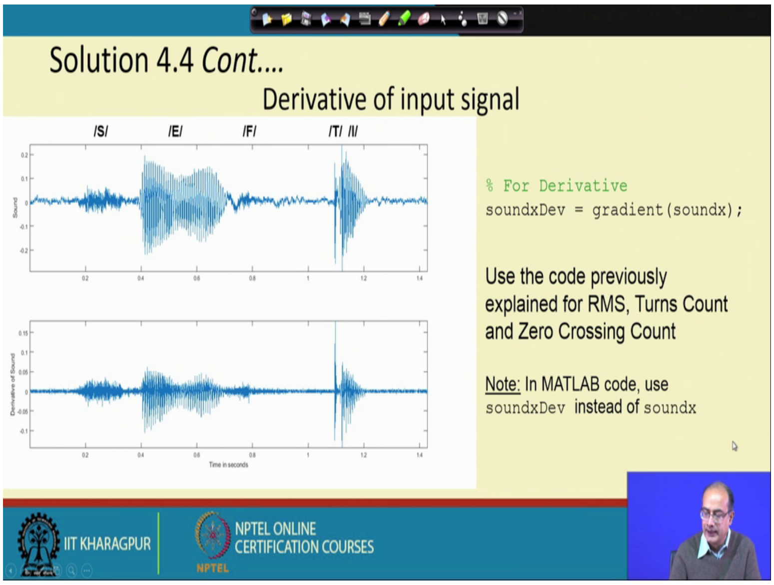
Now, let us go for other window size, going for 30 milliseconds we get that envelope has become much more smooth it looks better, same way with the turns count it has become smoother, looks better. Same with the zero crossing rate. All of them they have improved they have become much more smooth.

And let us proceed for 60 millisecond, it has become more smooth, but now, we see that the change are actually happening in a way that is sometimes not giving the true picture, ok. If we proceed for that for 100 millisecond we will get it more. I would like to draw your attention for the sudden changes where the plosive is there that is T, please look at that point ok.

Let us go back, at these place you look at, ok. So, earlier that when we are going for 10 milliseconds we could get the abrupt change and zero crossing rate also could give that peak and turns count at that, but when we are increasing that window we see that it is the height of this envelope has reduced that the turns count and the zero crossing rate it does not give that peak itself 40, ok.

So, that is becoming more and more actually smoothed out, ok. So, that smoothness comes as a cost that is what we can say, ok. It is in terms of envelope and all it is good, but when the sharp changes are happening we are not getting further actually that when the wind time window is bigger.

(Refer Slide Time: 17:25)



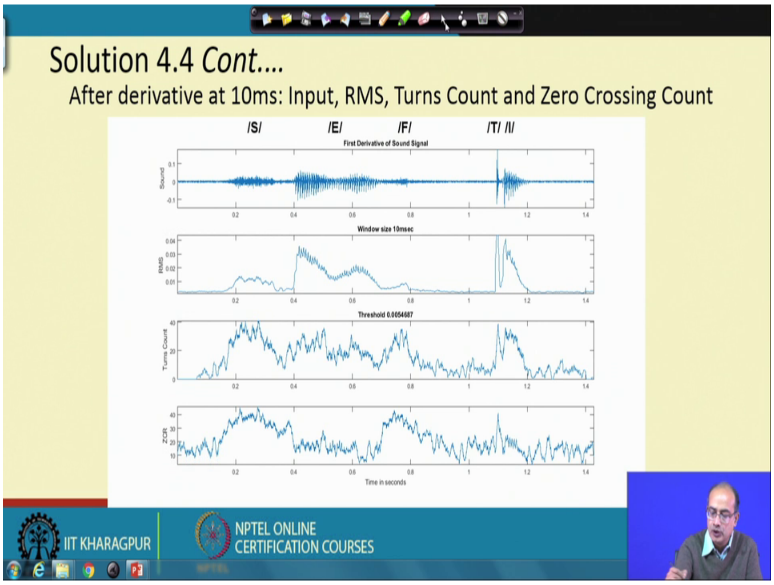
Now, let us go for the derivative input signal. So, we compute the derivative and that we for that purpose we are taking this the gradient of that signal. So, that is the difference of it, it is nothing special and we have plotted that thing, and we get the signal here.

What we see after taking the derivative that we get the amplitude of the voiced part it has reduced much, this part S so, it is not reduced, ok. E it is reduced, F part it is not much reduced, again T it is not reduced in fact, in a relative scale it is increased, then E that again that amplitude of the envelope has reduced, ok.

So, that is what we notice. Initially the signal was more or less that it is actually going through 0, so not much change in the DC value.

So, we after getting that derivative signal will apply all these techniques on the derivative signal and I appreciate that what are the changes we are getting.

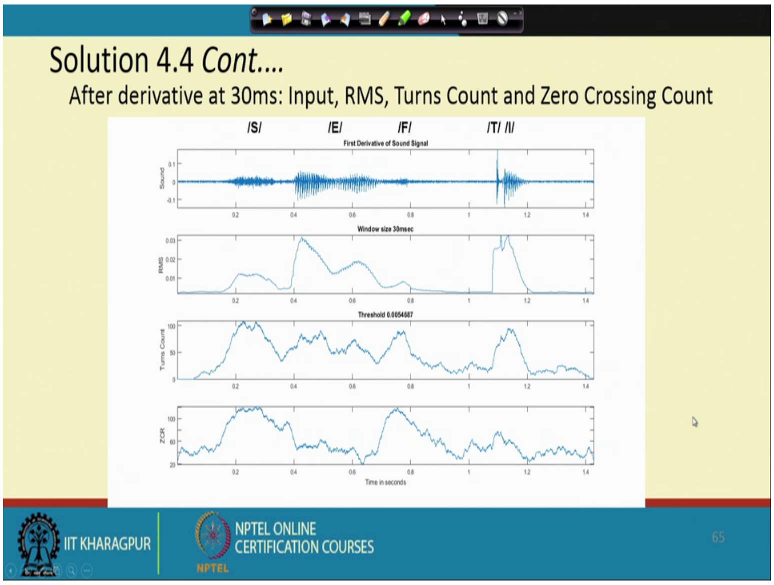
(Refer Slide Time: 19:05)



First we go for that 10 millisecond window, we get the envelope here, followed by the turns count and the zero crossing net. What we get because of derivative the noise is actually aggravated and. now, the zero crossing rate and turns count it is giving actually higher counts, ok.

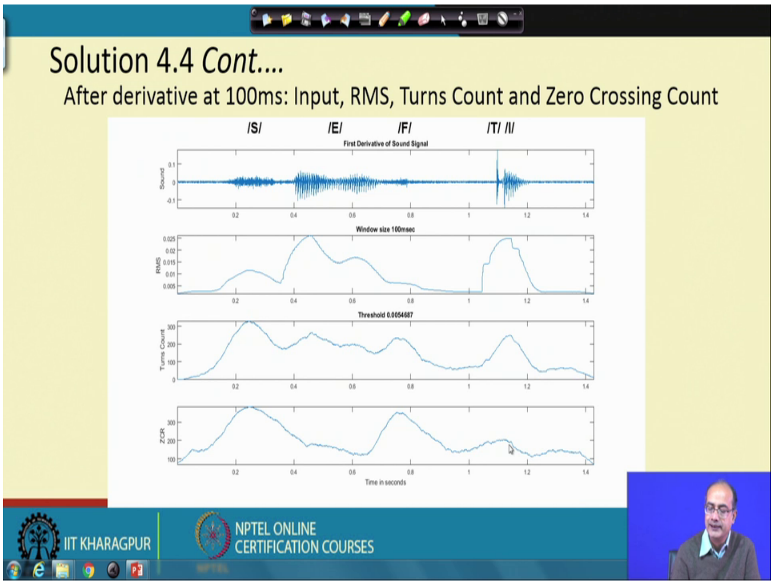
That if we see that overall there is an increase in the turns count as well as zero crossing rate. Earlier the difference what we had between the that peak for a unvoice sound and the voice sound, now the difference in that peak has reduced, that is the first thing we note here.

(Refer Slide Time: 20:11)



Next when we increase the that window going for that that the 30 millisecond window, then we see that envelop is becoming smoother for all the 3 signals or 3 outputs, and I think that between voiced and unvoiced the difference are becoming little more clear, ok.

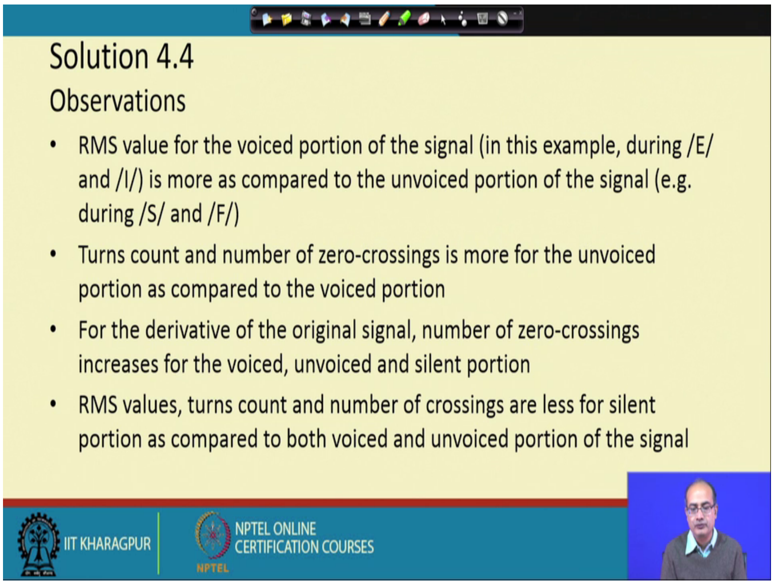
(Refer Slide Time: 20:42)



And going for 60 millisecond again all the 3 envelops they are actually the values they are becoming more and more smooth. But if we look at that what was the change we were expecting for the that T that is the this time, we see the envelope is showing a sharp change beyond that the turns count and the zero crossing rate because of higher window I think they have lost that part, ok. They could not capture that there is an increase in the zero crossing rate or so, that is not shown there. And same is true with the 100 millisecond. And with 100 millisecond another thing we know that the position of this change is coming from here. So, there is a actually change in the position. If T is here we see that our change is happening much before because the window is bigger, ok.

Same way that if a starts from here our signal that the envelope is starting before the zero crossing rate is also before it is nothing but that spread of the that energy over the window that is why it is happening, ok. So, we are getting the smoothness at an expense that much we can say.

(Refer Slide Time: 22:17)



So, here we conclude about it. The RMS value of the voiced portion is more as compared to the unvoiced portion of the consonant pass part, ok. In safety that word, that the first phoneme and the last phoneme that sorry, second phoneme and the last phoneme these two are the voiced that these two they are giving more actually RMS count or higher envelop compared to the consonant part that is S, F those portions.

In terms of turns count and the zero crossing both are higher for the unvoiced portion as compared to the voiced portion, ok. That is another observation we can get. And these two observations they are pretty vital that these can help us to separate which portion is the voice which portion is unvoiced.

Next is that for the derivative of the original signal the number of zero crossings increase for voiced unvoiced and silent portion, ok. That that when we take the derivative signal and the zero crossing is increasing it could be that because the DC bias which may not be visible in the plot, but small DC biases are there which are removed and that helps to get actually that more zero crossings, and everywhere we get an increase and that makes the thing more uniform, but which is not a good actually thing. We wanted to have them separated so that we can use that zero crossing rate to separate the 3 things out.

The RMS value turns count and the number of zero crossings our least for the silent portion as compared to the both the voiced and the unvoiced portion of the signal. So, what we get primarily that the turns count and the zero crossing rate they are similar, they are giving the frequency part of it. So, all the 3 RMS value turns count and zero crossing they are less compared to that voice and unvoiced portion of this signal.

So, using that we can separate out that this part that speech is there, this part is silence is there that means, only the background noise is there, we can separate that out. And once we could take out the portion which is the speech portion using the RMS value and turns count we can separate out the voice portion and the unvoiced portion. The voice portion will have high RMS and low turns count and the zero crossing whereas, the unvoiced portion will have high turns count and zero crossing and low RMS value.

Thank you.