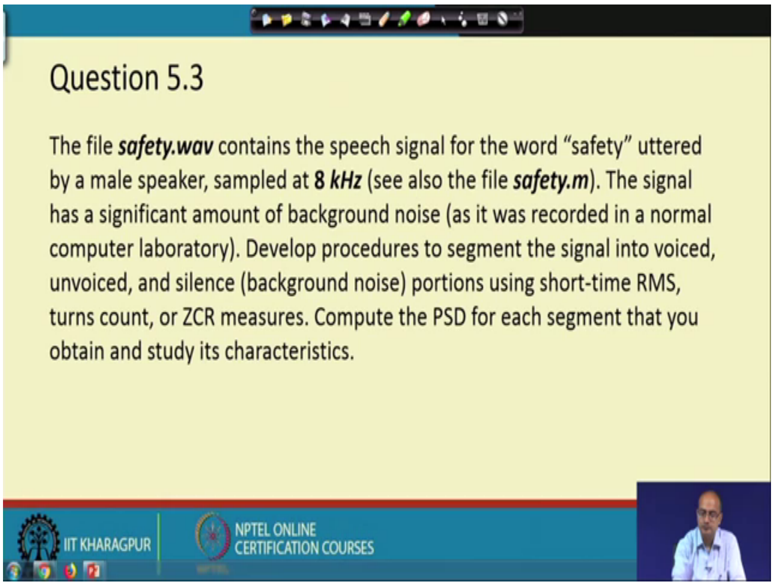
Biomedical Signal Processing
Prof. Sudipta Mukhopadhyay
Department of Electrical and Electronics Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 65
Tutorial - V (Contd.)

In the third experiment of the fief set we are given a signal safety dot wav.

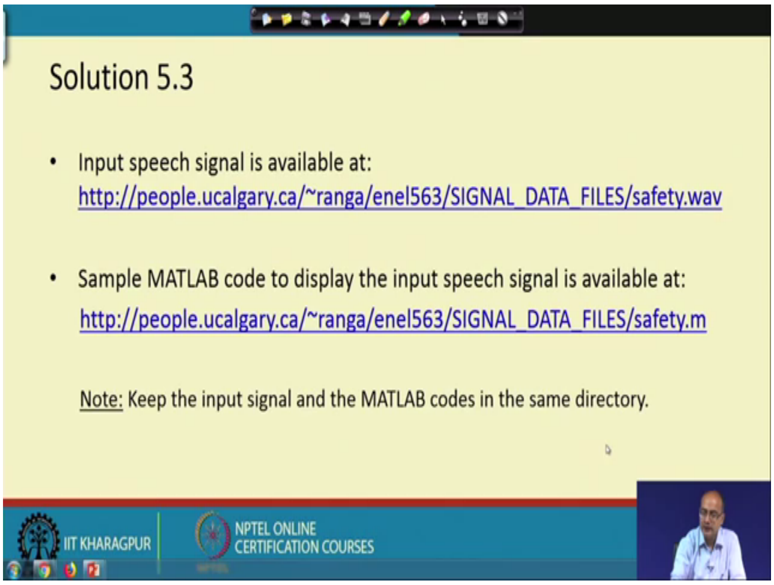
(Refer Slide Time: 00:26)



Is the occurrence of the word safety by a male speaker, and it is sampled at 8 kilohertz frequency. The signal also has some amount of background noise, now the first part is to segment the signal into voiced, unvoiced and the silence part, and for that we have to use the short time that RMS value turns count or zero crossing rate.

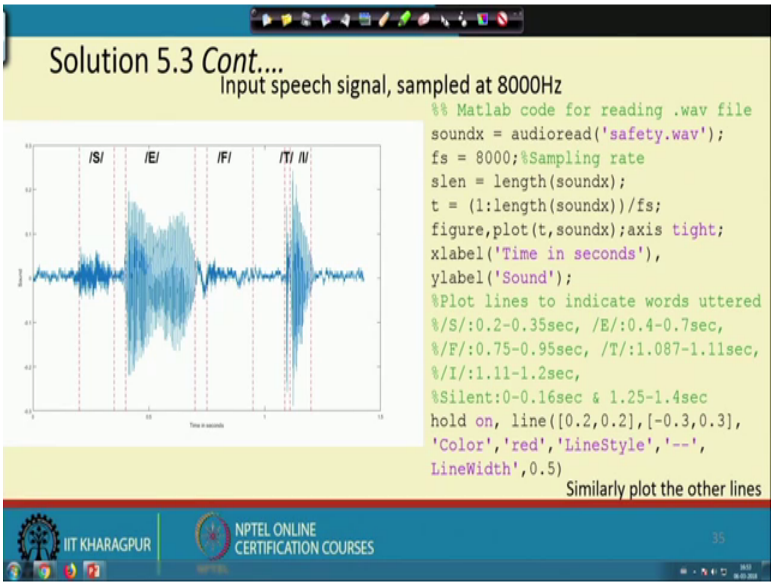
Next we have to compute the PSD of each of the segment and study its characteristics. The first part of it that is computing the RMS value, turns count and zero crossing rate we have already done. So, we would not repeat that part. We directly assume that we have those routines with us we will make use of them and apply it here for segmenting the signal into 3 parts, voiced, unvoiced and whatever is not taken into either of them that is our that silence.

(Refer Slide Time: 01:57)



So, first we start with collecting the signal safety dot wav, and the MATLAB code to read it that is safety dot m. And we kept both of them in the that the working directory of the MATLAB.

(Refer Slide Time: 02:21)



Now, let us take the signal it has read that signal for the purpose of processing. So, here we make use of a different command audio read to read the verb file ok, and sampling frequency is given 8 kilohertz. So, first we compute the that number of samples using the command length of the variable sound x where we have stored the verb file safety dot wav; And then using that that we compute the time axis to plot it and triangles.

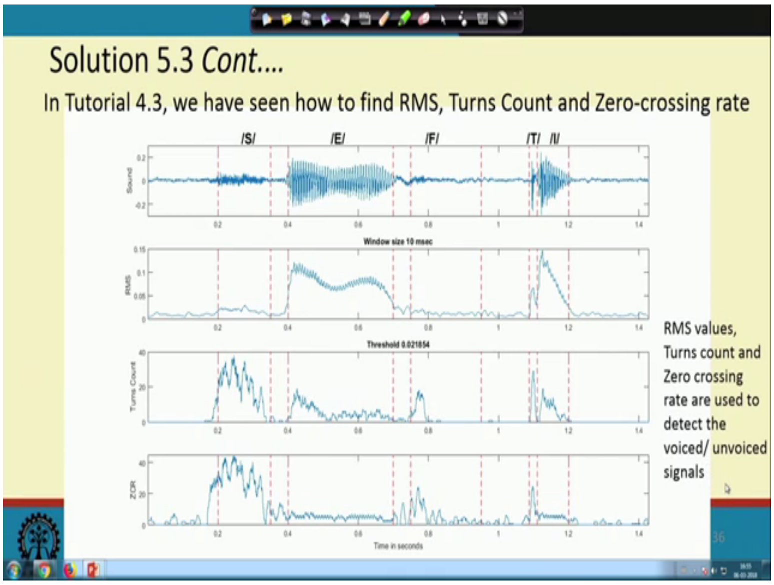
And here first part one colon this is you can say slen, the length of the variable sound x, that we multiply it with 1 by fs to get the time the first part it is giving the sample numbers only. Then we plot that with respect to the time and we label the two axis that time in second and y axis give us the sound.

And then we look at that from the book we get actually what should be the that the boundaries of the different phonemes that is first part is that S it is 0.2 to 0.35, then E 0.4 to 0.7, F 0.75 to 0.95 second, then T is 1.087 to 1.11 second and I is 1.11 to 1.2 seconds. And in between silencers are there at the beginning there is small silence, at the end also there is silence.

So, the same we can get, but we need to keep in mind that this is what we get from the book this is not what we have computed ok. So, here is the plot is given and the red lines as showing those boundaries. The initial part here is the silence, at the end we have silence. Here we have the phoneme S, then we have E, then we have F, then we have the small part that is T, then we have I.

Out of that that E and I they are voice sounds or vowels the other 3 that is S, F and T they are consonants or unvoice sound. So, our first task would be to segment this phonemes and then we would be able to compute the PSD and compare that the different kind of sounds and this part that is what we called silent, they are not exactly silent their background noise is present there. So, those 3 parts in terms of their PSD we need to compare, ok.

(Refer Slide Time: 06:57)

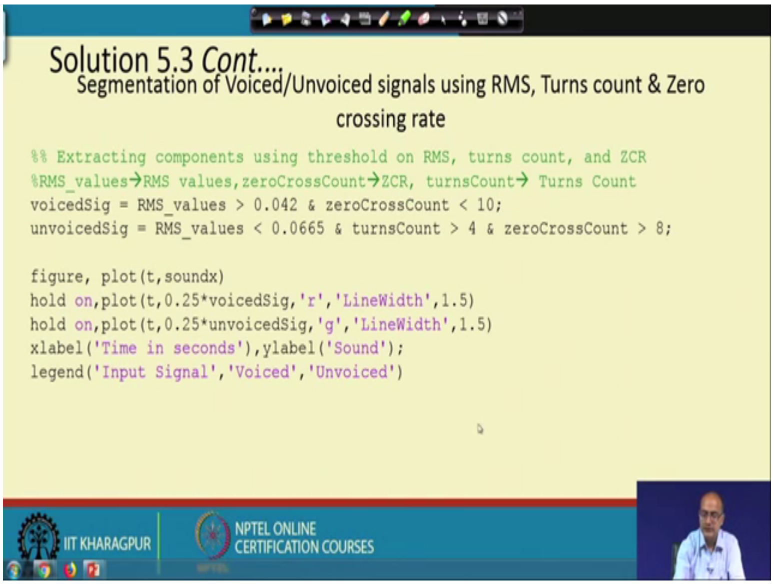


So, first we give the plot of the signal and RMS value turns count and zero crossing rate now already in tutorial 4.3, we have shown how to compute the RMS value turns count and zero crossing rate. So, we simply make use of them rather than explaining that once more.

So, first when you look at the RMS value what we know that for the vowel sounds E and I, the RMS value is high for both consonant as well as the silent period the RMS value is small. For consonant that turns count is high, that is small again for the that silent period. Its intermediate for the RMS that vowel sounds, between the voiced and the and unvoiced part zero crossing rate gives a better differentiation here the that zero crossing rate is low for the that vowel and it is high for the our that the unvoice sound.

So, that can be noted and we have to make use of these 3 to decide that which part is voice which part is unvoiced, ok.

(Refer Slide Time: 09:14)



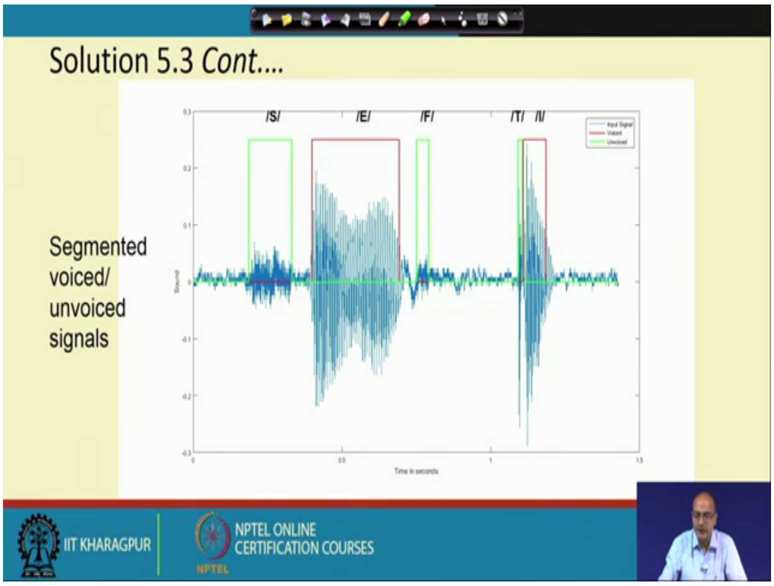
So, let us proceed with that. So, first what we do? You do the observation of the signal and we find out couple of thresholds. What we find that if we can take the RMS value RMS value is the vector which we have plotted in the previous page, if it is more than 0.042, and zero crossing rate is less than 10. Then it is voice sound.

And for the unvoice sound the RMS value is low 0.0665, on the other hand that turns count is more than 4, and zero crossing rate is 8. Why we are taking the both? To make sure that the silent period does not get included in the unvoiced sound. If we just look at the RMS value then the chances are that we may take silent period also in the unvoiced sound. So, now, this two parts we have actually separated and the two variables voice signal and unvoiced signal they are capturing the value in terms of 1 or 1 or the 0.

So, what do we do to see them that what is they are span. First we create the pin with comment figure then plot the sound with respect to the time, then hold the plot that means, we want to actually over right on the same plot and we plot the voice signal, we have a small scaling to place it in appropriate level and we use green color to draw the span of the voice signal. Wherever the voice signal is present it will draw it like this ok, rest of the part it will be 0.

The similar way whenever wherever the unvoiced signal is there it will have a draw a rectangle with green color, ok. And the x axis it is in seconds, y axis we have the sound and we have to 3 legends for 3 signals, one is input signal voiced and unvoiced, ok. So, with this you go for having the plot.

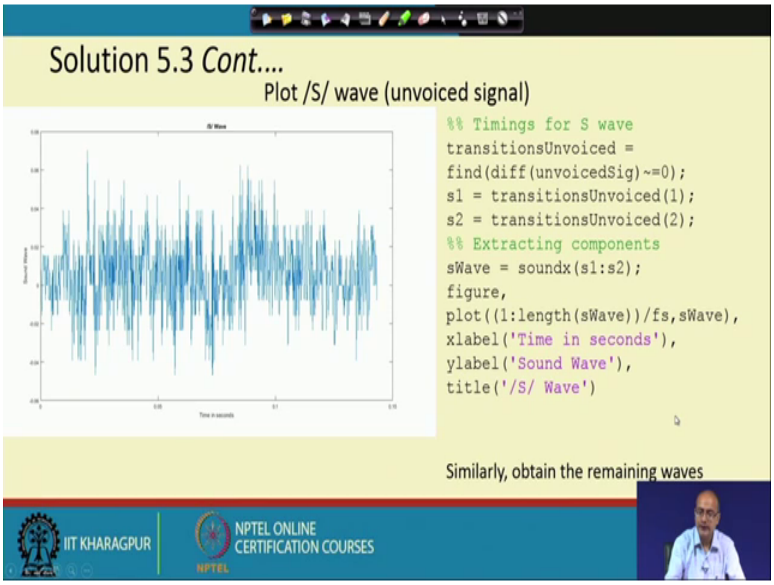
(Refer Slide Time: 12:48)



Here we show the that first the that actual signal and the ground truth then we draw the voiced and the unvoiced part, ok. So, we can just go back and forth we see for S, I think we have it more or less accurate, for E also we have close I think segmentation. On the other hand if you look at the phoneme F this is the actual boundary and we got actually much smaller, ok.

Let us look at the other two phonemes T and I. We get that the segments what we have created they are close to the what is given in the book. So, what we get that for out of the 5 phonemes we have good segment created except for 5 that F, and if we have to segment it in a better way we need to then fine tune those parameters and have a better segmentation of it. However, that given this formula this is a segmentation we get, so we will go ahead with this segments for further analysis.

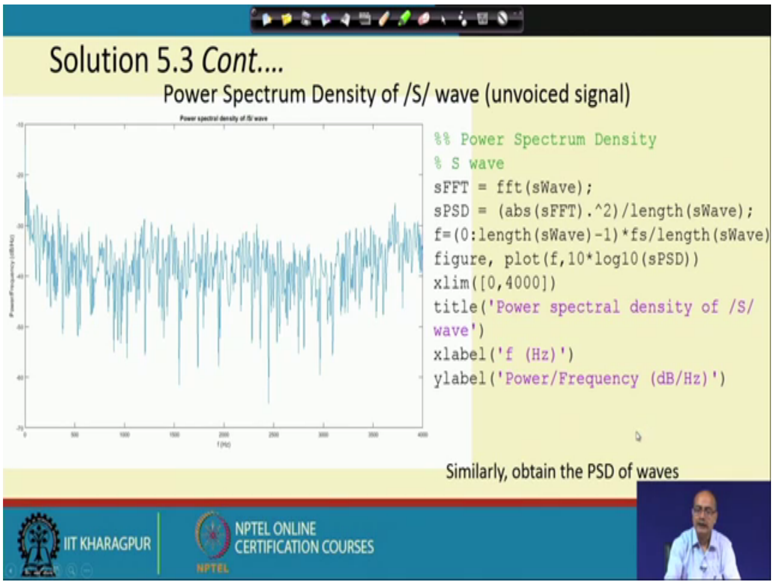
(Refer Slide Time: 14:49)



So, first we plot the that unvoiced sound. So, for that what we do that we find out the transitions the first one is this is unvoiced signal. So, we look at the transition of the that unvoiced signal wherever it is not 0 when we take the difference signal, ok. So, we are getting the transitions there and the first two transitions we note that gives us the duration of the signal S.

Next that we pick up that the corresponding part from the variable sound x we take that part and we plot actually that part ok, sWave and here we are showing that time axis. So, we get the time domain plot of the that the signal S, ok. We get that it is a looks like a random signal, jagged signal amplitude is not very high and now we will go for the other waves in a similar way.

(Refer Slide Time: 16:54)

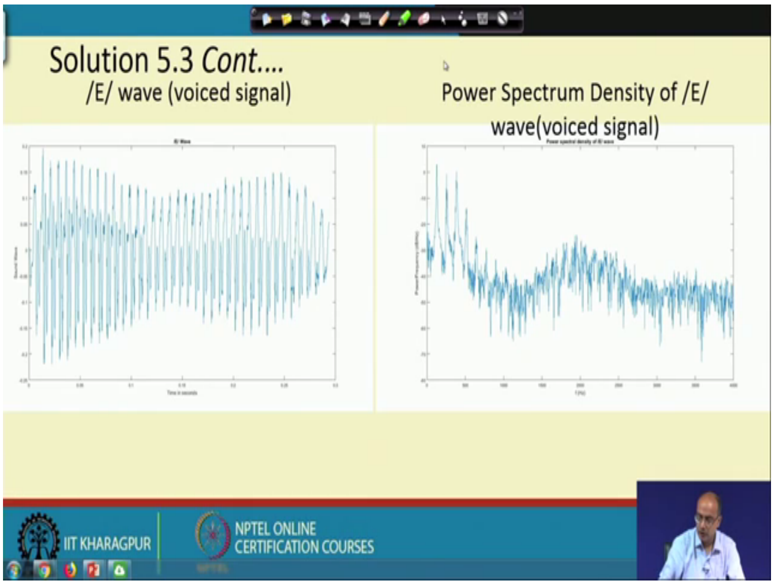


So, before that you look at the power spectrum. We have already seen that how to get the power spectrum we need to take the FFT, then we need to take the square of the absolute of each of the coefficient, normalize it with the length of the signal and then we need to plot in the db scale and for that we need to have the frequency axis that is what is done here.

And here we get the spectrum of S we get its more or less flat kind of spectra no prominent peaks are there though undulations are there compared to the DC value you see that the peaks are usually that what peaks what you are getting here that is about 20 to 30 dB below. So, they are not actually peak.

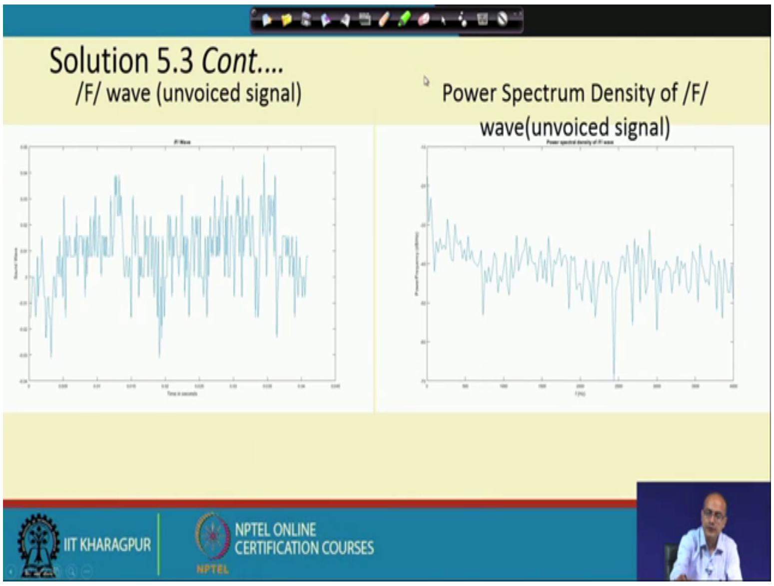
So, same way we can get the spectrum of other phonemes.

(Refer Slide Time: 18:22)



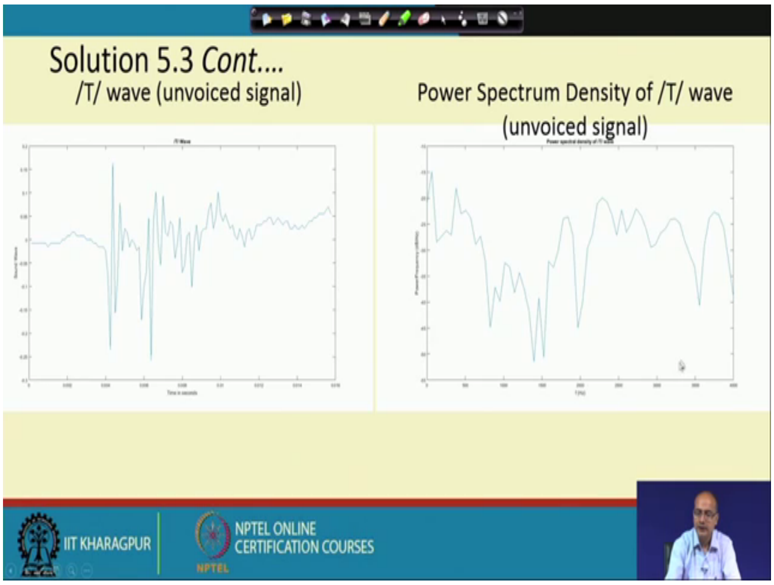
And here we show it for different phonemes. First we look for the voice signal E. You see the amplitude is more in this case and for that we see that when we look at the spectrum we get couple of peaks are there, very prominent peaks are there. So, that is the specialty of that phoneme E what we notice.

(Refer Slide Time: 19:07)



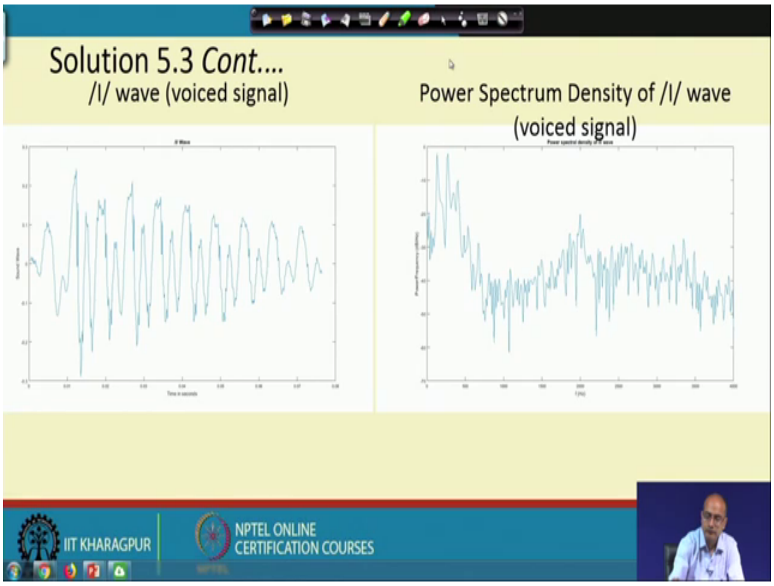
Now, let us move forward go for that next phoneme that is F, F is in the signal amplitude is low. The corresponding that PSD what we get, PSD also it is having not very prominent peak ok, though it is jagged we do not get any high peak, ok.

(Refer Slide Time: 19:41)



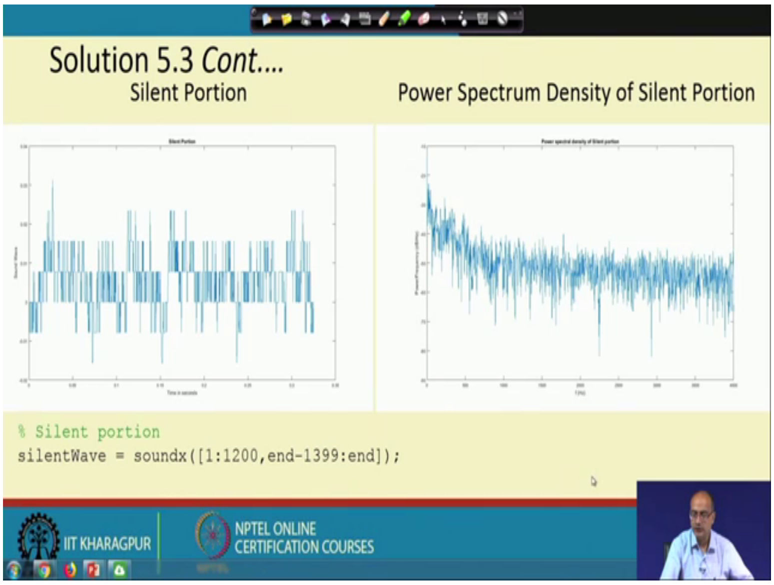
Next we go for the next phoneme that is T again it is a unvoiced signal. T comes with a actually huge change that that we see that suddenly the signal appears ok, and then it goes down. When you look at the PSD we do not get any sharp peak here.

(Refer Slide Time: 20:21)



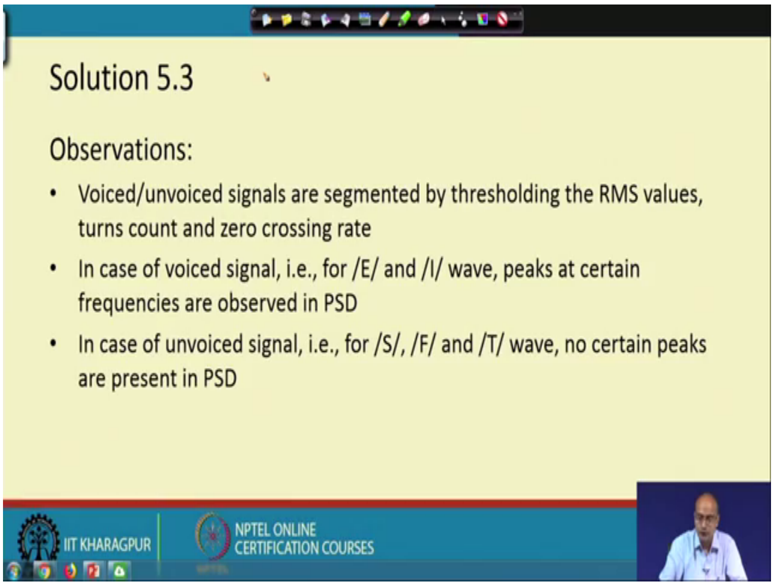
Now, the next phoneme would be I, here we get that higher amplitude and more or less regular shape in the time domain for the signal I. And if we look at the spectral domain again we get couple of peaks are there, ok. So, that is the specialty of the voiced phoneme I.

(Refer Slide Time: 20:57)



Now, next look at the silent portion you have taken a silent portion we see silent portion is really random and if you look at the PSD it is very much jagged that will not get any pattern at all in it. So, with that that we complete actually that our observations that the different kind of spectrum we get now we conclude upon what we have seen.

(Refer Slide Time: 21:34)



The first thing what we get that voiced and unvoiced signals can be segmented by thresholding the RMS value turns count and zero crossing rate. So, using those 3 we can have separation between the voiced and the unvoiced signal. However, that in case of the voice sound E and I when you look at the spectrum we get certain frequencies we have the peaks, in fact, for both of them in time domain we get some repeated waveform and that gives rise to the concentration of energy at certain frequency in the PSD.

On the other hand for the random signal kind of time domain waveform we get for the unvoiced signal here we have 3, S, F and T, all 3 cases we get that we have that in the time domain there is nothing specific we can get and same way in the frequency domain we do not get any peak in the PSD, ok. So, that is the signature of the, that unvoiced sound or the consonants.

Thank you.