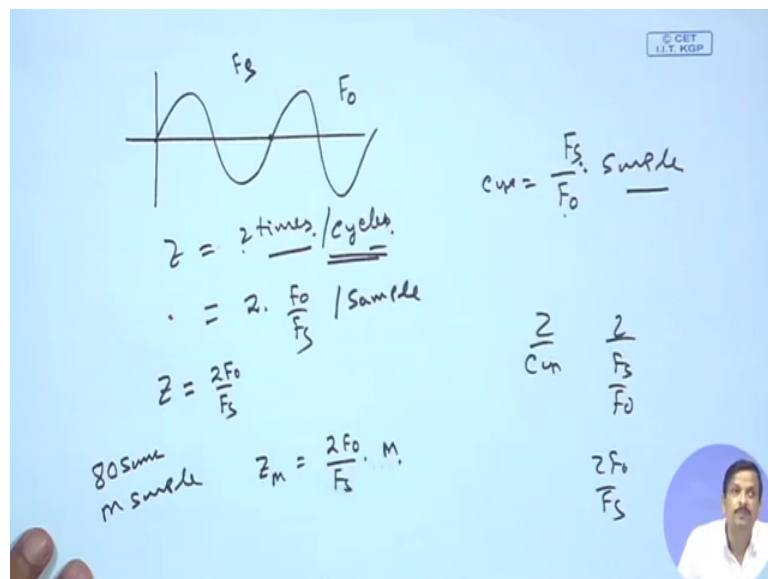


Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture – 20
Time Domain Method In Speech Processing (Contd.)

So, last class we are discussing about the 0 crossing.

(Refer Slide Time: 00:26)



So, we have saying that suppose I have a simple sinusoidal, simple sinusoidal. So, every cycle, every cycle a simple sinusoidal cross the 0 line 2 times. And if it is this is shifting let us correct the this is shift then find out the 0 crossing 2 times it cross the 0. So, 2 time pass cycle. So, for every cycles it cross the 0 line 2 times.

Now, if the sampling frequency is let us F_s and the cycle that the fundamental frequency of or this signal frequency is F_0 . So, for every cycle it cross the 0 line 2 time. So, sampling write F_s . So, if how many cycles are there in per sample. So, F_s samples are there by F_0 sample per cycle. So, for every cycle how many samples are there? So, if F_s is the number of sample and F_0 is the frequency. So, F_s by F_0 sample will be there per cycle which I have discuss already. So, 2 time per cycle. So, every cycle how many samples are there F_s by F_0 every cycle? So, cycles is equal to F_s by F_0 sample.

So, if I convert this 2 number of 0 crossing per sample. So, I can say 2 into it will be cycle convert to sample. So, it will be F_0 by F_s . Because cycle is 2 by cycle 2 per cycle. So, it is nothing but a 2 by F_s by F_0 . So, it is nothing but a $2 F_0$ by F_s number of 0 crossing per sample. So, I can say for a pure sinusoidal the number of 0 crossing per sample is $2 F_0$ by F_s . So, suppose I said find out the number of 0 crossing of a sign half sign wave for 80 sample, or for m sample. So, I can say the number of 0 crossing for m sample is nothing but a 2 into F_0 divided by F_s into m is clear.

(Refer Slide Time: 03:22)

Handwritten notes on a blue background showing calculations for zero crossings:

$$F_0 = 1 \text{ kHz} \quad F_s = 10 \text{ kHz}$$

$$Z_{400} = \frac{2 \cdot F_0 \cdot 400}{F_s}$$

$$= \frac{2 \times 10^3 \cdot 400}{10 \times 10^3} = 80 \text{ No}$$

Additional calculations and conversions:

- $\frac{400 \text{ sample}}{30 \text{ ms}}$
- $\frac{10000 \text{ ns}}{1} = 10 \text{ K}$
- $\frac{10 \text{ K}}{30} = 300 \text{ Sm}$
- $\frac{2 \times 10^3 \cdot 300}{10 \times 10^3} = 60 \text{ No}$

So now suppose I have a signal 1 kilohertz, signal F_0 is equal to pure tone of 1 kilohertz sampled at F_s 10 kilohertz. If I say find out the number of 0 crossing for 40 or 400 sample. How many time signal cross the 0 line for 400 sample. So, I can say it is nothing but a Z_{400} is equal to nothing but a 2 into F_0 by F_s into 400. So, I can say 2 into 1 kilohertz means 10 to the power 3 divided by 10 into 10 to the power 3 into 400 cancel. So, it is nothing but a 80 number. So, 0 crossing for 400 samples is 80 number. Even if instead of 400 sample I said how many time signal is cross the 0 line for 30 millisecond? 30 millisecond in the 30 millisecond signal how many time signal cross the 0 line?

So, I we say F_s is 1 kilohertz; that means, 1 this millisecond has 1 k sample or 10 k sample. So, one millisecond has 10 sample. So, 30 millisecond has 300 sample. So, instead of 400 I can say 2 into 1 k divided by 10 k into 300 60 number. So, I can find out the number of 0 crossing.

(Refer Slide Time: 05:15)

ZC Rate Definitions

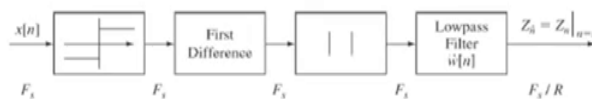
$$Z_{\hat{a}} = \frac{1}{2L_{\text{eff}}} \sum_{m=\hat{n}-L_{\text{eff}}}^{\hat{n}} |\text{sgn}(x[m]) - \text{sgn}(x[m-1])| \tilde{w}[\hat{n}-m]$$

$$\text{sgn}(x[n]) = \begin{cases} 1 & x[n] \geq 0 \\ -1 & x[n] < 0 \end{cases}$$

• simple rectangular window:

$$\tilde{w}[n] = \begin{cases} 1 & 0 \leq n \leq L-1 \\ 0 & \text{otherwise} \end{cases}$$

$$L_{\text{eff}} = L$$

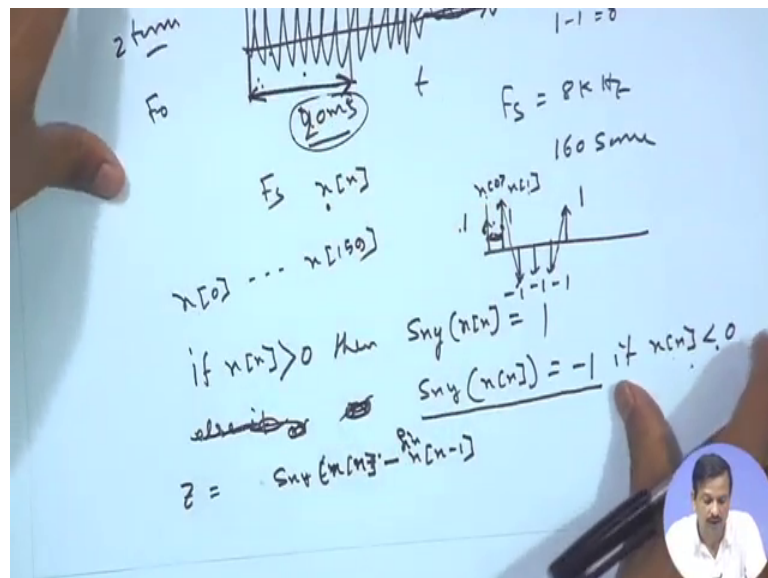


Same form for $Z_{\hat{a}}$ as for $E_{\hat{a}}$ or $M_{\hat{a}}$

Now instead of sign wave

Let us I have a speech signal. So, it is not I cannot say 2 times per cycle. So, I have a ARB signal. I do not know number of cycles, which is the period I do not know anything. There is the signal only. I want to know how many times signal cross the 0 line for a 40 millisecond segment. Or of a 20 millisecond segment for one window.

(Refer Slide Time: 05:23)



So, my problem is that I have to find out how many times signal cross the 0 line for 20 millisecond I take a window. So, that is why it is called short term 0 crossing rate. If I want to calculate how many times signal cross the 0 entire signal, that can also I calculate what that will be not use, because a signal which is time varying. So, this may

be voice this may be silence this may be noise. So, all kinds of steps or signals are there. If I want to know the which part is voice which part is noise. So, I have to instead of taking the whole signal at a time I have to take a small window. Same as and calculation of energy that is why called short term 0 crossing rate.

So, once I make a short term let us 20 millisecond window, I want to find out how many times signal will cross the 0 line. Here my formula will not work because I do not know where the cycle number of cycle F_0 I do not know. F_s I know I know only F_s . Sampling frequency I know I have sample the signal something else. Now let us this is an x_n is a signal, first I explain it then I put the generalized formula. So, this is the first window this is the first window and signal is x_n . So, x_n is 20 millisecond signal if F_s is equal to 8 kilohertz, then how many sample will be there in 20 millisecond? 10 millisecond 80 sample. So, I can say there will be a 160 sample value.

So, x_n has an 160 value if it is. So, if it is start from 0 then it will be x_{159} . Now I want to find out how many times signal cross the 0 line. So, if I say if I say generally drawn suppose those are the sample. So, when the 0 crossing will be happened this is sample once the signal sample will come this side. So, this is the positive sample this is the negative sample. So, there will be a 0 crossing. Similarly there is a negative sample negative sample then once it is a positive sample then there will be a 0 crossing. So, I can say the 0 crossing only happen if the signal value signal or sample, sample sign in change positive to negative sign.

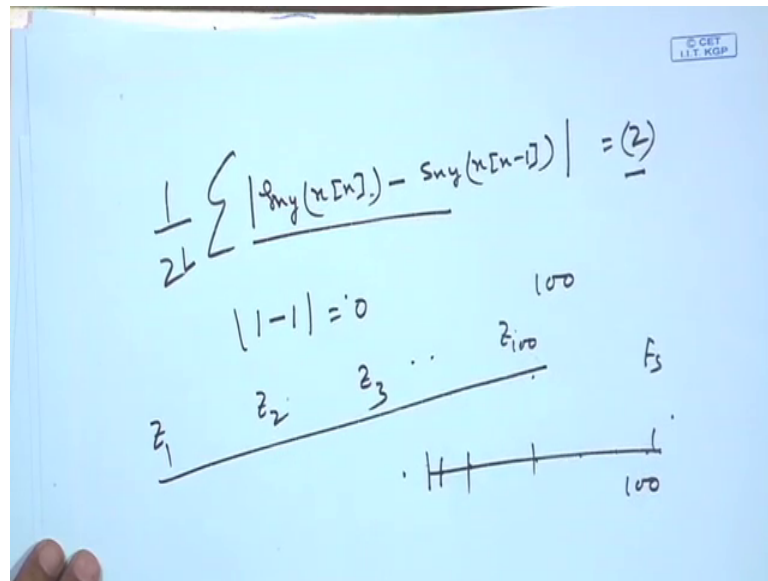
So, every time when a previous sample if it is positive next sample is negative 0 crossing occur, previous sample is negative next sample is positive 0 crossing occur. So, I can say let us I compare where the So, there is a 0 crossing will be occur in between 2 sample. So, within this 2 sample I do find out whether there is a 0 crossing occur or not. So, how do I find out? I do check within this 2 sample whether the sample value change the sign or not. So, I can check let us this is x_0 and this is x_1 . So, what I define? Let instead of $x_0 x_1$ let us write it 1, 1, minus 1, minus 1.

So, I can say if x_n is greater than 0, then let us I define a function Sng of x_n will be 1. Or else of else or can write or Sng x of n equal to minus 1 if x of n is less than 0. So, if x of n is less than 0 then the value of Sng of x_n is minus 1 and this is 1. So, if I see this sample this is 1, this is 1, this is minus 1, minus 1, minus 1, 1. Now if I say So, I have to

check 2 sample at a time I have to check 2 sample at a time. So, I can say if this minus this if both are 1. So, 1 minus 1 is equal to 0.

If the sign is change then only I want one count. So, I can write down a function, that z n let us the 0 crossing count of 0 crossing is nothing but a change of sign Sng of x x. So, sample one so; that means, x of n plus 1 or if I write x n then minus x of n minus 1 previous sample sign of. So, I will let us use the separate slide.

(Refer Slide Time: 11:37)



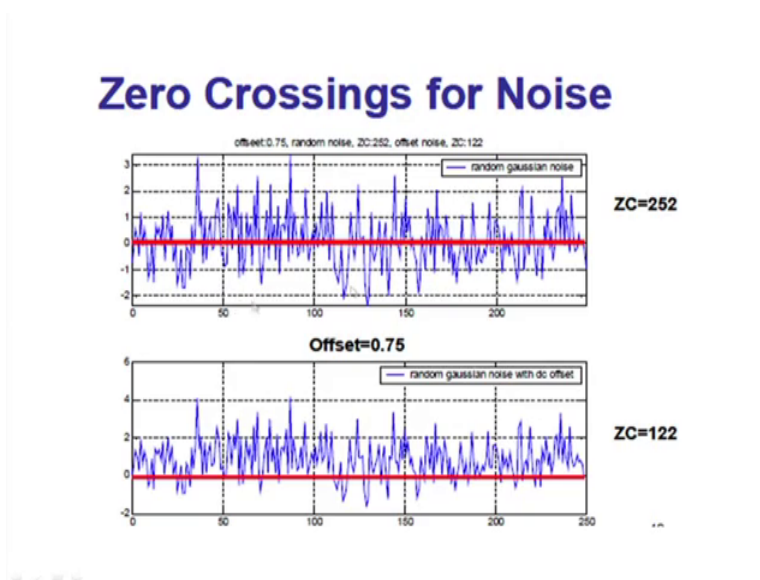
So, why s n Sng x n n minus Sng of x of n minus 1 previous sample then take the mod. Give me the every time. So, if I take the mod of this function what will give. This is plus 1. So, this in this case, this is minus 1. So, let us this case this case I take. So, coming here.

So, this case. So, x 3 minus and I take x 2. So, I taking the samp the x 3 and x 2. So, this is x 3. So, if it is x 3 x 3 is negative sample show minus 1 minus x 2 is positive sample minus 1 then take the mod then the value will be 2. If it is 1 and 2 value up 2 second sample is positive. So, one value of first sample is positive equal to 1/ take mod 0. So, if I take this function and check for all sample, take the sum. So, every time signal cross the 0 line I get the value of 2. So but what I want I want number of 0 crossing through the that window let us this is the 1 20 millisecond. So, I can say this will be 1 by 2 has to be normalized.

Give us 0 crossing is number 1 number 2. So, be instead of number 1 every time in this function I get value 2. So, 2 has to be normalized. Now if I want to normalized with respect to window length then I get normalized with 1, this is for first window. So, same thing we will happen for any window throughout the signal. So, generalized formula for calculating the 0 crossing rate is this one. And this is the walk diagram first signal pass through the minus 1 and one I make it then take the difference take the mod, then I pass through the window and get the So, what I get for every window I get a z value. So, for the window number 1, I get a z 1. For window number 2, I get a z 2 for window number 3 I get a z 3.

So, how many if the there is 100 frame per second. So, I can get 100 z value. So, instead of F s I can get 100 points of z value. So, F s is 1 second the 1 second is divided into 100 point. So, sampling frequency them come down to the instead of F s it is 100. Sampling rate the frame rate divided that sampling frequency. So, I get a one value here one value here that a 100 value I get. So, 100 z value like it. Then if I plot those 100 z value. So, if the signal is c b land or if the fricative.

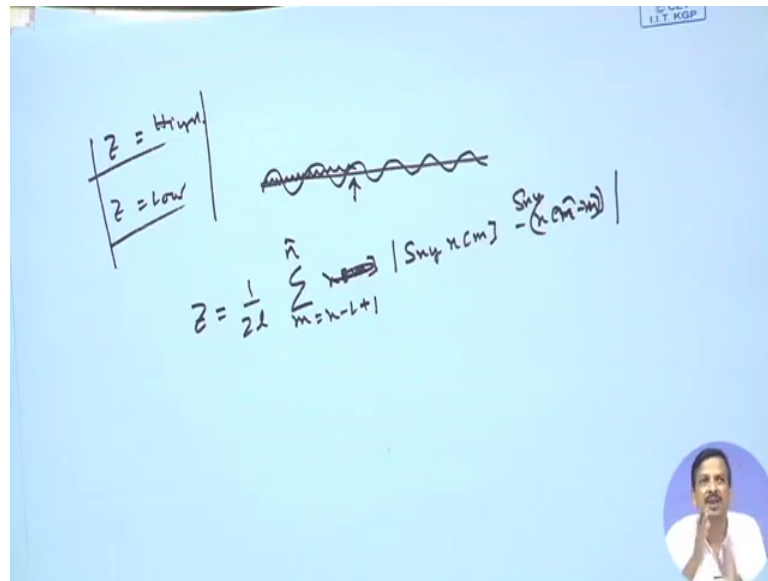
(Refer Slide Time: 15:09)



So, if I see that see then

example if the c b land signal. So, number of 0 crossing will be much high. So, z corresponding window z value will be very high. If it is signal is worst signal is worst kind of signal number of 0 crossing will be less. So, z value will be low.

(Refer Slide Time: 15:16)



So, depending on the z value I can say whether this signal is silent or this signal is silence. In the problem is that even through the signal is non silent let us silence, but there is a simple noise is there, then also I can get the 0 crossing rate very high or let us there is a 50 hertz frequency find the line frequency disturbance, then also I get the 0 crossing rate with the like the voice signal. So, that is why the 0 crossing rate is not a you cannot say this is a very rigid parameter for detecting the voiced or unvoiced detection, but yes I can do the voiced unvoiced detection using 0 crossing rate.

So, normalized 0 crossing rate means it is normalized with respect to window length. So, if it is. So, I can say normalized 0 crossing rate is z is equal to nothing but a 1 by $2L$ summation of m let us x m sorry, mod of Sng of x m minus x of m minus 1 Sng will be there mod. Now if I want to put the w in pictures normal value then m will be n minus 1 plus 1 . And here will be n and here will be n minus m . So, I have said the n th window, n th window. So, n equal to 0 n equal to 1 , n equal to 3 n equal to 4 should I know the which window it is part. And for each window I calculate the normalized 0 crossing rate with respect to $2L$ the here is the L is the length of the window.

So let us I have a signal. Let us this same see this example. For a 1 kilo 1 kilohertz sign rate as a input using 40 millisecond window length, with various value of sampling frequency we get the following. We get the same value of z m . So, window length is 320 . So, if you know that z 1 is nothing but a 2 into F 0 by F s . And if it is multiply for

particular 40 millisecond, then I know how many sample will be there in 40 millisecond and I can find out the z value, if it is see the z value is same z m is same.

(Refer Slide Time: 18:13)

ZC Normalization

- For a 1000 Hz sinewave as input, using a 40 msec window length (L), with various values of sampling rate (F_s), we get the following:

F_s	L	z_1	M	z_M
8000	320	1/4	80	20
10000	400	1/5	100	20
16000	640	1/8	160	20

- Thus we see that the normalized (per interval) zero crossing rate, z_M , is independent of the sampling rate and can be used as a measure of the dominant energy in a band.

But z 1 is different, per cycle per sample with the number of 0 crossing per sample is different, but number of 0 crossing for a particular number of sample is same. For particular number of time sorry, particular number of time is same.

So, this is I can say z m is independent of or I can say the z m is independent of sampling frequency, that is called normalized 0 crossing rate.

(Refer Slide Time: 18:51)

Autocorrelation

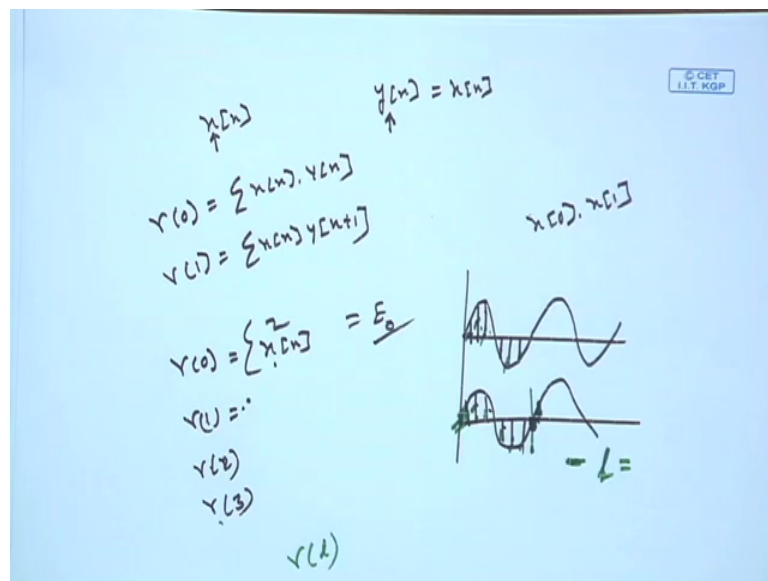
- Autocorrelation is a cross-correlation of a signal with itself.

$$\phi(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n + \tau)$$

- The maximum of similarity occurs for time shifting of zero.
- An other maximum should occur in theory when the time-shifting of the signal corresponds to the fundamental period.

Then there is another parameter we said time domain parameter is call autocorrelation coefficient. The details I will discuss during LPC analysis and F 0 analysis, but I just give you a hince what is autocorrelation that idea. So, what is correlation? You know suppose I want to find out the correlation between these 2 things. That is nothing but a similarity between the 2 object is called correlation. Similarity between these 2 object if 2 object are similar, then I can the similarity is similarity I have to measure the degree of similarity. So, there is a requirement of some parameter with respect to which I can find out the degree of similarity.

(Refer Slide Time: 19:41)



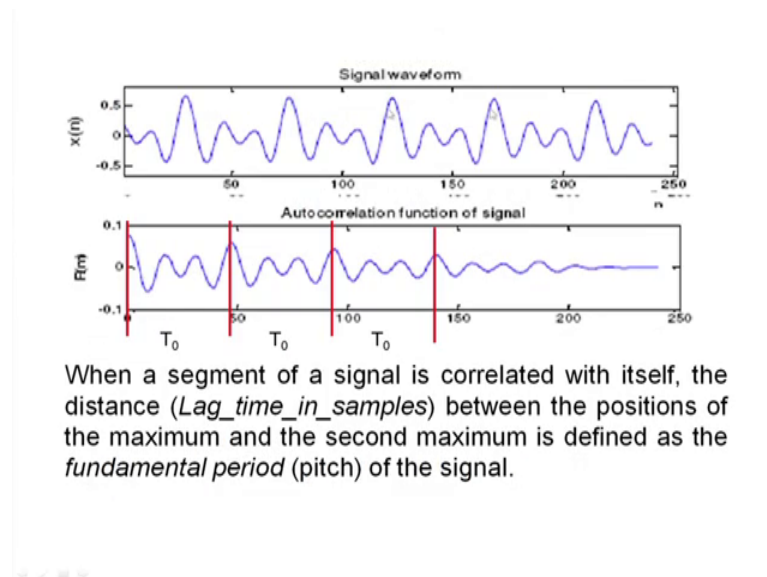
So, suppose it is a digital signal x_n another digital signal y_n , I want to find out the similarity between this signal and this signal. So, I want I will do for every sample similarity of the whole signal. So, similarity number r_0 , I have to find take the one sample of this signal take the first sample of this signal, take the product, then take the sum. So, it is nothing but a product x_n multiply by y_n . Now if I find out the similarity of degree one r_1 . So, I will do I will shifted the one signal by one sample. So, that is the correlation. So, if it is autocorrelation then y_n is equal to x_n it is similarity between the signal it selves. So, if that case the r_0 first coefficient is nothing but a x_n .

If sample has to be square and then some of. So, it is nothing but a $n \cdot E_0$. Similarly if it is r_1 . So, the signal is same signal is shifted by one then multiply. So, x_1 the x_0 will be multiply with next x_1 here x_0 is not x_1 x_0 x_1 is there as treated as a 0 and last sample

will be multiply with the 0, because if I take that the outside the vicinity of the window signal is 0. Then I get the value then I get the value r_2 then I get the value of r_3 just shifted. So, if I shifted it if you think in a sign wave manner. This is a sample, this is a sample, this is a sample, this is a sample, this is a sample, again sign wave 1, 2, 3, 1, 2, 3. Now once I multiplied these with this, these with these, these with these with these and add it similarity is maximum I get E_0 .

Once I shifted the sample with once is. So, the next time I extract this sample. Then I multiply this with this, this with this, this with this, and this with this, this with this, and take the sum. Now if you see seen there is a signals are not similar not in phase you can say the signal is not in phase. So, in that case the similarity becomes reduce again it will be multiply with the same signal when it will come in here. So, after complete one period again the signal once this same kind of similar kind of signal will multiply each other. So, summation will be increases. So, I can say r_0 r_1 r_2 r_3 if I calculate those r value with respect to different l . So, again I get the maximum value at r_1 where l represent the time of or if $F t l$ is equal to nothing but a length of the signal, where it is complete period it has a complete period.

(Refer Slide Time: 23:30)



So, l is nothing but a t complete period. So, if you see what the signal or not coming details I will cover it will look like this. So, if you see the rate color is the maximum. So, maxima occur at red color. So, what is the definition of a fundamental frequency. Signal

repeat itself. So, when it will be repeat if the sample value of similar. So, I can say the signal repeated in here, signal repeated in here, signal repeated in here. So, this is nothing but a F_0 this is nothing but a twice F_0 this is nothing but a twice F_0 . So, that can be used to calculate the fundamental frequency of the signal. So, that is the autocorrelation parameter for time domain processing.

(Refer Slide Time: 24:18)

Average Magnitude Difference Function(AMDF)

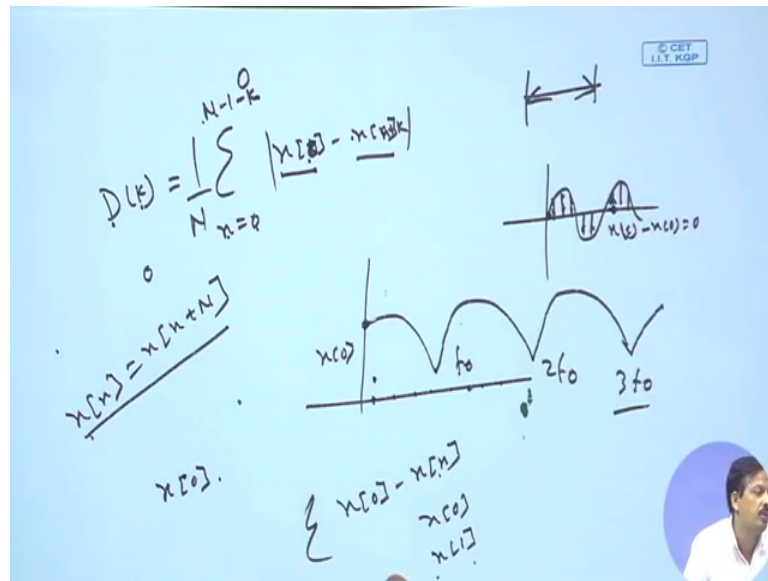
- It is an alternate to Autocorrelation function.
- It compute the difference between the signal and a time-shifted version of itself.

$$D_x[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} |x(n) - x(n+k)|, \quad 0 \leq k \leq K_0$$

- While autocorrelation have peaks at maximum similarity, there will be valleys in the average magnitude difference function.

Details again I will discuss during the F_0 extraction. Next is called average magnitude difference function AMDF. Instead of taking the multiplication of the signal, I just take the mod magnitude difference mod, mod of the magnitude difference.

(Refer Slide Time: 24:35)



So, suppose I have a $x[1]$ take the $x[0]$ minus $x[0]$ take the mod. Then instead of $x[1]$ $x[0]$ take let us $x[k]$ and n plus k . So, I can say n equal to 0 to capital n minus 1 minus k which is nothing a $D[k]$. So, I can say difference number 1 . So, k equal to let us 0 . So, I take the difference between the $x[0]$, I take the difference, difference of first sample with all other sample and there sum. Then I have to there is a n number of sample. So, I can normalized it 1 by n . So, I take a n number of sample signal then take the difference. So, this is there is a n number of signal in my signal, and take the first sample. I take collect the difference of the first sample with other sample, and take the sum and divided n .

So, see the philosophy. If this is my $x[0]$, let us my $x[0]$ is in here. So, I take the $x[1]$ take the difference of $x[1]$ is in here, let us $x[1]$ value is here. So, if I if I say $x[1]$ $x[2]$ $x[3]$ I take the differences. So, differences will be minima, when the value of this one this $x[0]$ which matched with similar kind of period will come. If you see if you take the plot, I take this first sample, I take this first sample. Take the difference between this sample, this sample, this sample, this sample, this sample, this sample, this sample and sum it up. Then take this sample, then take this sample with difference of this sample this sample difference of this sample, this sample, difference of this sample. So, $x[0]$ minus x of n which n varies from 0 to n minus 1 k . So, k equal to 0 means $I \times x \times x[0]$ minus $x[0] \times x[0]$ minus $x[1] \times x[0]$ minus $x[2] \times x[0]$ minus $x[3]$ I have taking it and take the sum.

Then I take x_1 minus x_0 , x_1 minus x_2 , x_1 minus x_3 , x_2 minus x_3 and take the sum. So, if I do that way when it will be minimum? When I take the difference from here, same signal with the same signal. So, x_0 minus x_0 . So, this is nothing but a look like a x_0 all though it is a not x_1 . So, it is a x_k when x_k minus x_0 is equal to 0. Then the difference will be minimum almost 0. So, I can say the definition of a phase is that if the 2 samples are identical after n delay then n is call period. So, the if you know that x of n is equal to x of n plus n . So, x of n after n sample if the same sample is appear then I can call the period of the signal is x_n . So, difference will be minima when the signal repeat itself. So, if I say this kind of plot I will get.

So, first minima we will occur at F_0 . Second minima we will occur at twice F_0 . Third minima occur thrice F_0 . So, details I will discuss that what extraction of F_0 what is the drawback of this algorithm or call things I will. So, this is call average magnitude difference. So, all those time domain parameters can be used to detect the speech and suppose I want to detect the speech and non speech.

(Refer Slide Time: 29:05)

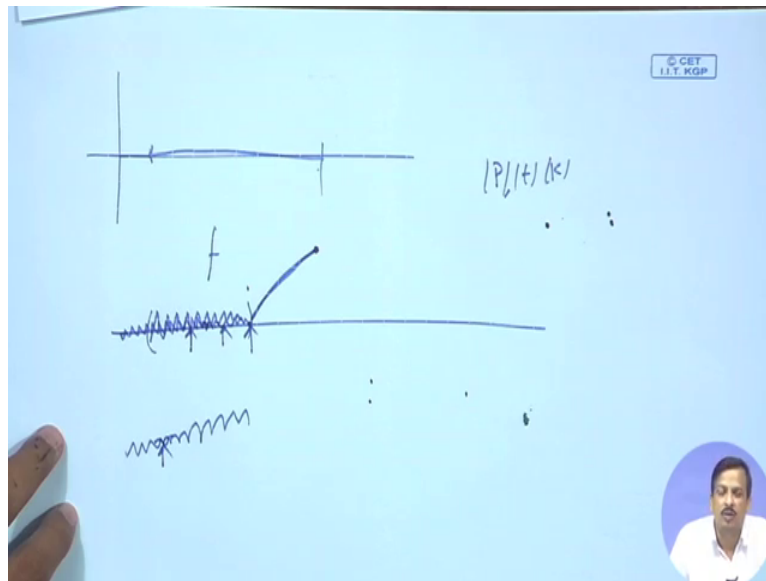
Lecture-10

Speech/Non-speech Detection

So, what is the important? If

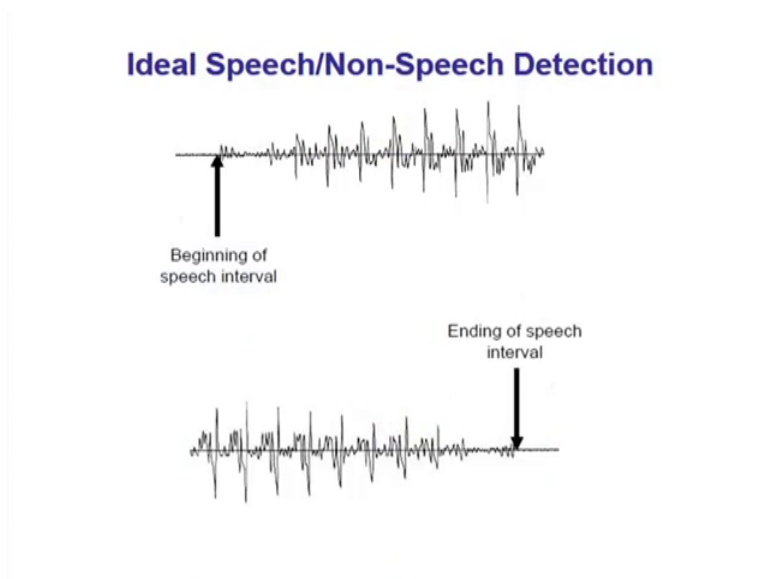
I not talk suppose I want to developed and which is a speaker identification and systems, and I want to find out why are the speech event is started and where the speech event is ended.

(Refer Slide Time: 29:26).



So, let us I quite for sometimes and start recording. I want to find out the time when I start speaking and when I end speaking, speech and non speech.

(Refer Slide Time: 29:48)

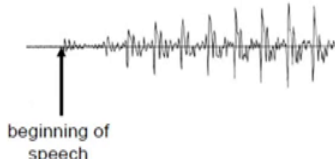


So, beginning of speech interval and ending of a speech interval that have to detect. So, how to detect it?

(Refer Slide Time: 30:00)

Speech Detection Issues

- key problem in speech processing is locating accurately the beginning and end of a speech utterance in noise/background signal



- need endpoint detection to enable:
 - computation reduction (don't have to process background signal)
 - better recognition performance (can't mistake background for speech)
- non-trivial problem except for high SNR recordings

(Refer Slide Time: 30:08)

Problems for Reliable Speech Detection

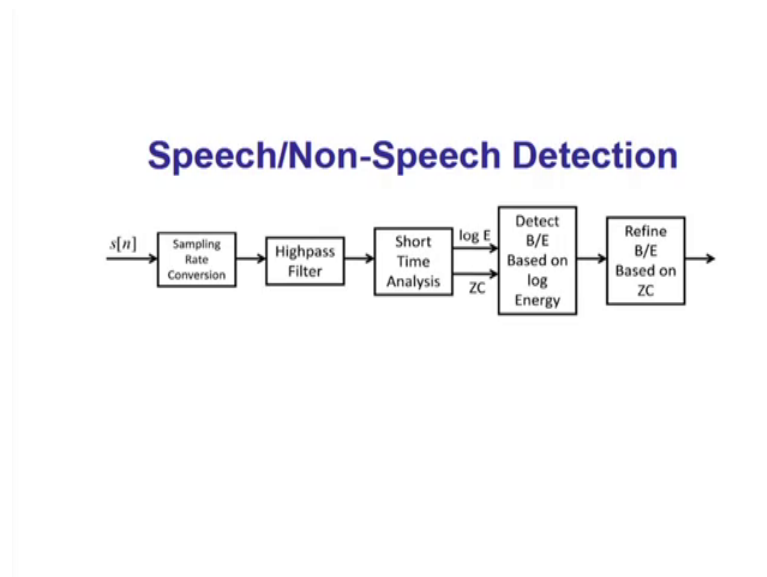
- weak fricatives (/f/, /th/, /h/) at beginning or end of utterance
- weak plosive bursts for /p/, /t/, or /k/
- nasals at end of utterance (often devoiced and reduced levels)
- voiced fricatives which become devoiced at end of utterance
- trailing off of vowel sounds at end of utterance

the good news is that highly reliable endpoint detection is not required for most practical applications; also we will see how some applications can process background signal/silence in the same way that speech is processed, so endpoint detection becomes a moot issue

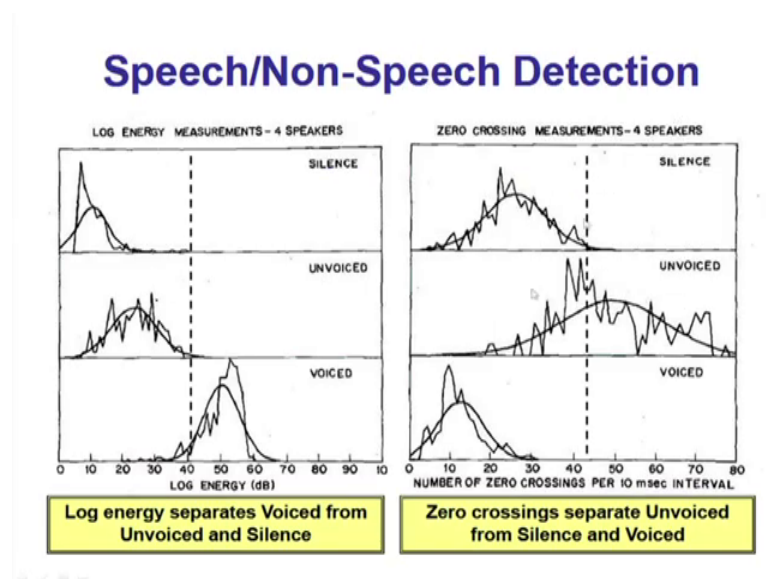
I have to accurately locate the beginning and end of the speech, with a noisy background something will be there. So, those kind of things will be there. So, what is the problem? Now suppose there is a some noise, and I start the speech with a plosive consonant pa, ta, ka this kind of consonant. If you see the (Refer Time: 30:26) period of pa is nothing but a silence. So, I do not know where the pa is started. Because when it start voicing I know here the voicing is started. So, that is nothing but a power to next voicing transition, but I do not know where the pa is begin whether it is begin in here whether it is begin in here I do not know.

Similarly, if there is a sum fricative with fricative like fa then also detection of that fricative is very tough we weak. Plosive fricative weak fricative even nozzle, suppose you started a nozzle consonant the voicing is started like this, and suddenly like this. So, why exactly it is started? It may be concede with the noise. So, those are the problem in voicing and non voicing detection. But using this energy or you can say the average magnitude, and 0 crossing rate I can detect the voice you know beginning and end point of the voicing.

(Refer Slide Time: 31:38)



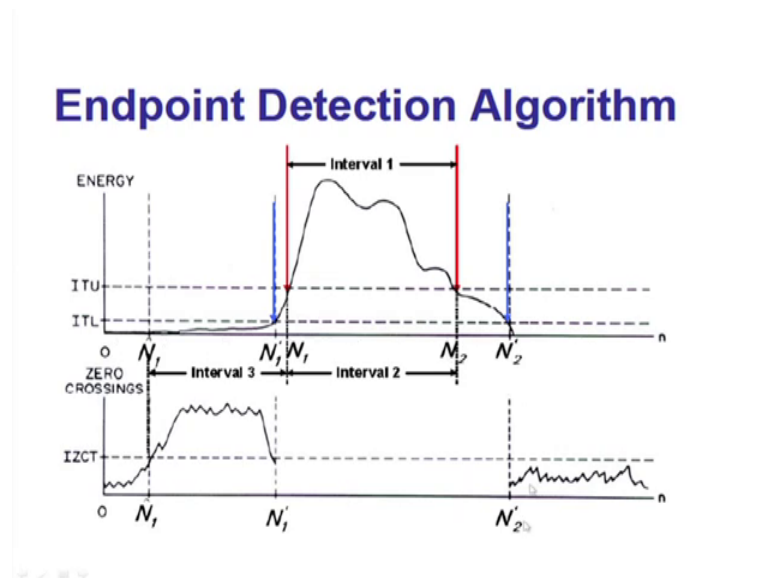
(Refer Slide Time: 31:40)



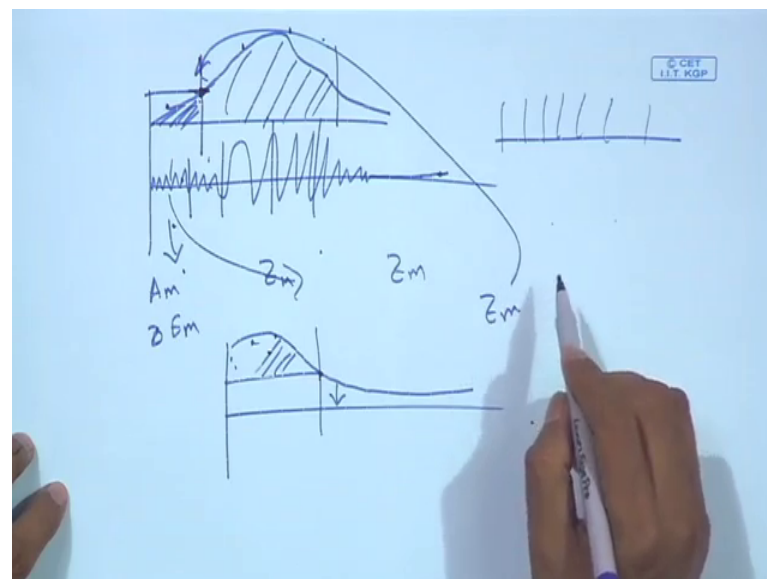
So, what I will see, if you see is a loss energy loss energy separated voice from unvoiced and silence. So, this is the silence, this is the unvoiced, this is the voiced. Log energy in D b. So, this is the 0 crossing silence unvoiced and voiced.

0 crossing pa for 10 millisecond interval. So, I take the window side is 10 millisecond. So, suppose I can developed an algorithm like this, this is my energy. So, I can get that the interval this is the this block line is plotted for every 10 millisecond. Every 10 millisecond there is a point. So, I can say I can explain it better here.

(Refer Slide Time: 32:13)



(Refer Slide Time: 32:36)



So, suppose I have signal been silence. So now, suppose this term. So, what I will do I take for every 10 millisecond I find out the average magnitude. Or I can say that average magnitude A_m or I can find out the E_m energy. And 0 crossing rate, for every 10 millisecond I get that one. So, if it is 1 second signal if it is hun 10 millisecond is the frame rate. So, I can get the 100 100 frame value.

So, with concede with the signal. So, I can say this frame corresponding to this energy, this frame corresponding to this energy, this frame corresponding to this energy, this frame corresponding to. So, I can get a energy plot like this. Now if I have a define a threshold, that if it is within this threshold value, then I call it is a silence if it is above the threshold value then I call it is a voice. So, I normalize the amplitude because your according then you say that if I recording the high volume then threshold value will be change. So, what will do? You can normalize the speech signal, with respect to some sample you can say that this is the maximum level of So, I get a signal. I normalize the amplitude of that whole signal, and take find out the threshold value was some from previous study. And I just the threshold value to find out the voice and un un voice detection.

Similarly, for every 10 millisecond I get a z_m value. So, if it is noisy. So, z_m value will be very high. So, I can get the z_m plot will be look like this, again down down down. So, again I can say the z_m value if it is high above this threshold value I can say it is a sivilent below this threshold value I can say it is a voiced or silence. If there is a silence can also have a high threshold value, but if you see the z_m and energy if I combine, then I can say whether it is a silence or sivilent. Sivilent amplitude will be high at least voicing sivilent there will be the some power, but in silence only background noise will be there. So, power will be reduce. So, that way I can find out this parameter can be used for voicing and un voicing detection.

For PDA and VDA voice detection or PDA can PDA and VDA, I discuss later on the speech detection algorithm and voice detection algorithm. So, this can be act as a voice detection algorithm, but every PDA is nothing but a VDA. If I able to find out the speech, speech is exist only for voicing signal. So, I can say VDA PDA is exist only for voice detection. So, VDA PDA PDA, P is speech detection algorithm PDA voice detection algorithm. So, I will discuss those later on details of when a discussing about

the F 0 extraction. So, time domain methods is complete. So, next class I will discuss about the LPC modeling.

Thank you.