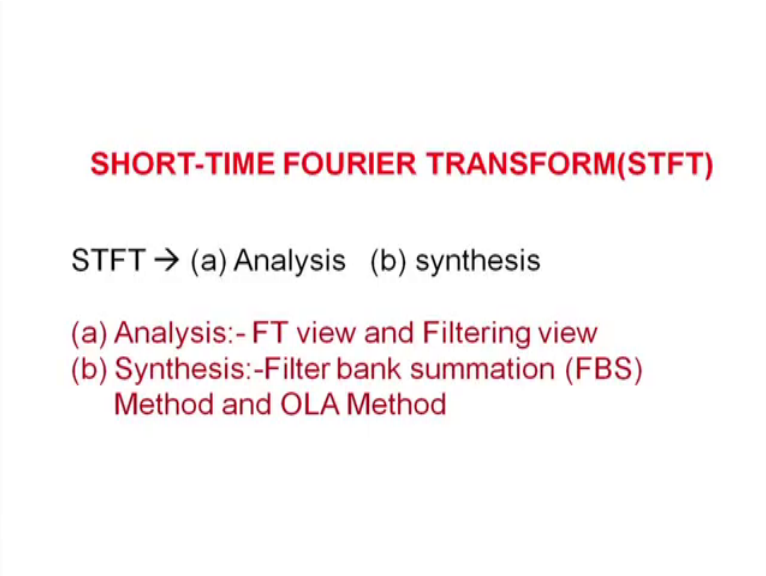
Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture - 26
Over View Of Short – Time Fourier Transform (STFT)

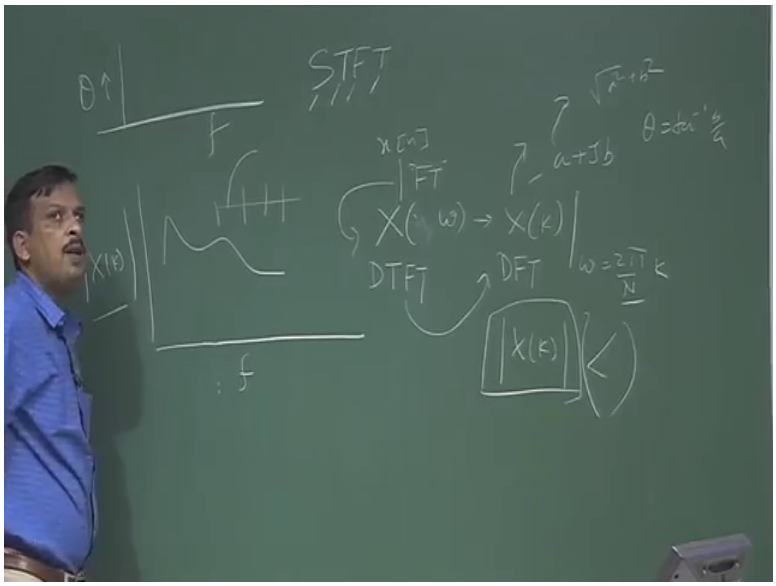
Let us start that new topic which is called Short Time Fourier Transform STFT; Short Time Fourier Transform.

(Refer Slide Time: 00:22)



So, in this whole week, we discussed about the STFT. So, STFT we have the two dimension one is analysis, another synthesis view.

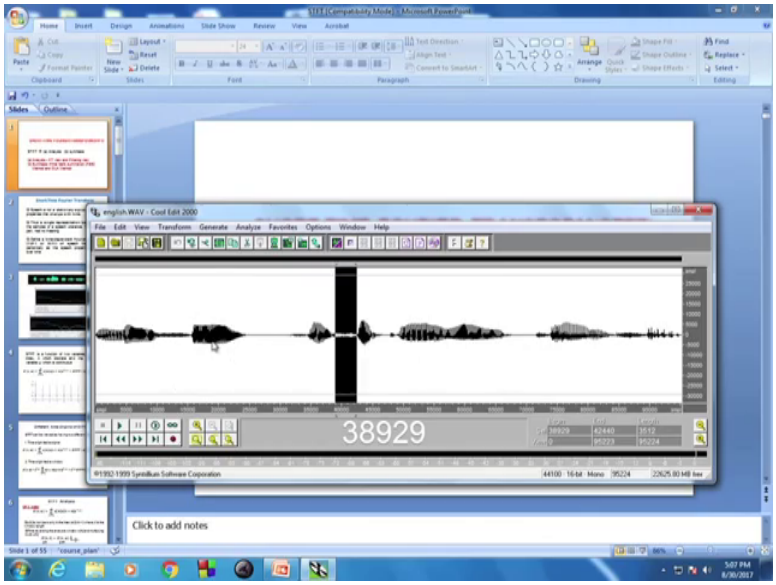
(Refer Slide Time: 00:26)



We will discuss and then do you discuss about that FT filtering view and that Fourier transform view. And then synthesis, why filter bank summation and OLA methods; overlap add methods. So, all those things will be discussed during this week.

So, what is STFT? Why STFT come into the pictures. Now, if I show you suppose this is my speech signal, this is my whole speech signal is this one. Now if I take the Fourier transform of whole speech signal, then I get a spectrum or frequency domain representation which consists of all average of all the variation that means.

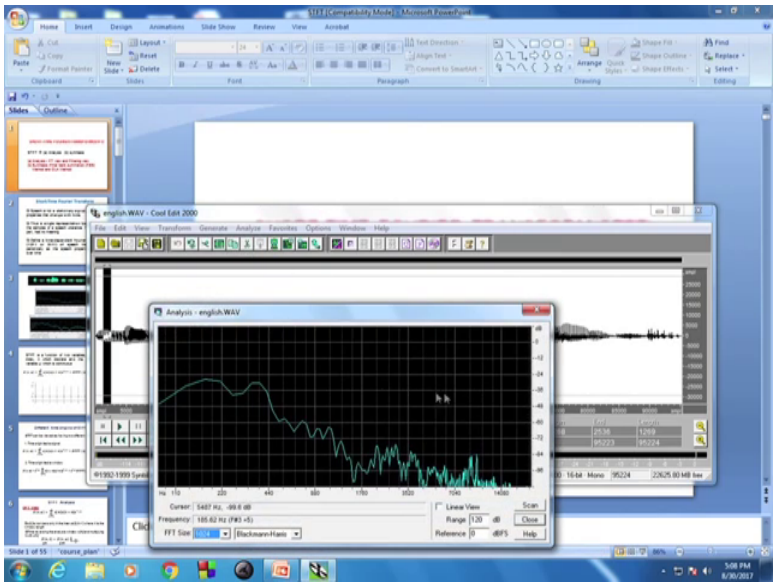
(Refer Slide Time: 01:28)



If you see along the time this axis is the time axis. So, along the time speech signal is not stationary. If you see the speech signal is changing, somewhere some different kind of voicing: then noise, then voicing, then silence, there is a different kind of voicing then know as again voicing. So, speech signal is changing along the time. So, if I had a long speech signal or I have a speech signal of a ward of speech signal of a sentence, then I say if I take whole sentence at a time or a whole word at a time it may consist of several variation of the signal; that means, signal is not stationary along the time.

So, if I take the whole signal and do the frequency analysis, I do not get that any conclusion any information of different speech event what is happening. So, what I want we want during the pronunciation different time speech signal property is different. I want to analyze that property, which how it is varies across the time. So, if I want to do it, then what I have to do? I have to take a short take a select a short segment of the signal and do the analysis. Next again I have to select the next segment and do the analysis. So, what I am doing? Instead of taking the whole signal I taking a small portion of the signal. Let us this is a 1269 sample I have taken then I take the frequency analysis ok.

(Refer Slide Time: 03:12)



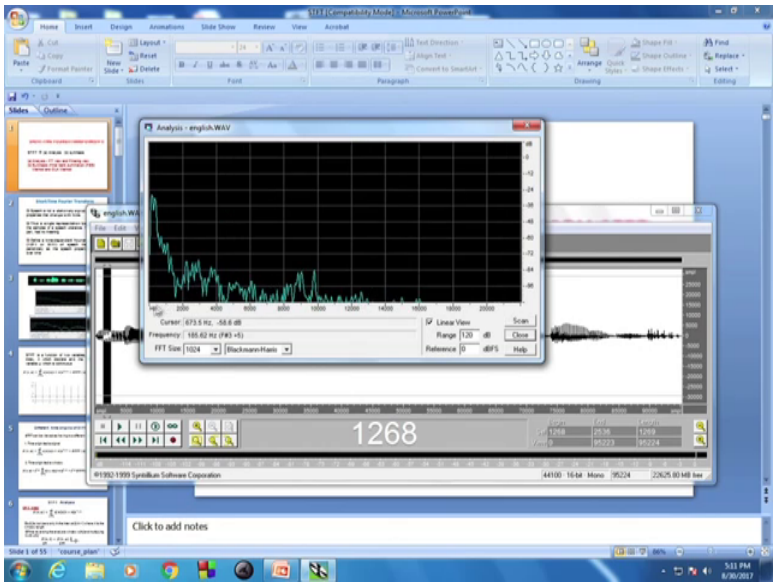
So, as you know in the recap; that what is frequency analysis we are doing? That if you know that if x n is my time domain signal, then if I take the discrete Fourier transform or

Fourier if I take the Fourier transform, then signal domain it is discrete, but frequency domain it will be continuous.

So, if I take the Fourier transform, then what is I get? I get X of n into omega or if it is single frame signal let us X omega. If it is discrete Fourier transform, then I get X of K that I have already discussed in the Fourier transform view. If x n is my signal and if I take the Fourier transform then in frequency domain it is continuous, but time domain it is discrete and that is called DTFT- discrete time Fourier transform. Once I both domain it is discrete, then I say it is DFT-discrete Fourier transform. So, FFT is nothing, but an algorithm of implementation of DFT. So, once I get that omega is discrete, then I say it is DFT; that means, here omega is nothing, but a 2 pi by N into k, where n is the length of the DFT analysis; that is recap of your review of the DFT digital signal processing. Now again you know; that if I get that X K has a complex number, X a is a complex number. So, it has a two property: one is mod of X K that is called amplitude and angle of X K tan inverse imaginary y.

So, this is phase and angle. So, if I take the spectra of magnitude only then it is called magnitude spectra, if I take the phase only then it is called phase spectra. So, if I take the time this axis is the magnet this axis is the frequency and y axis is the magnitude of discrete Fourier transform then I call this is a magnitude spectra. If I take the phase of phase so X K is nothing, but I in the form of a plus J b. So, what are the magnitude spectra? Root over of a square plus b square. If it is space spectra theta is equal to tan inverse b by a. So, if I plot the theta against the frequency, if this is the frequency axis and this axis is the theta then I call this is a phase spectra, so frequency analysis.

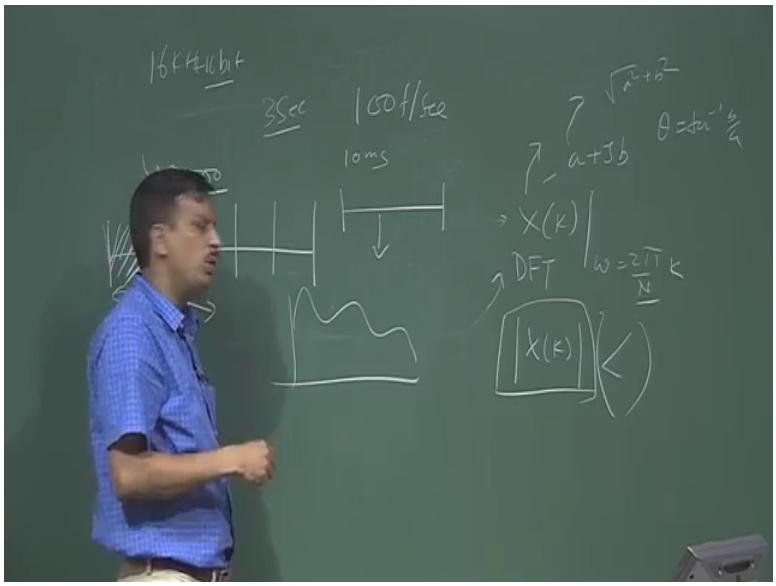
(Refer Slide Time: 06:25)



We find out two kinds of spectra. One is called phase spectra; another one is called magnitude spectra. If you see if I analyze it if I analyze it frequency analysis if you see this window this axis is the frequency, if it is in linear view; that means, frequency scale is linear; if it is not linear view, then I say the frequency scale is log scale and this axis is the amplitude of the particular component. So, this spectra is nothing, but a magnitude spectra ok.

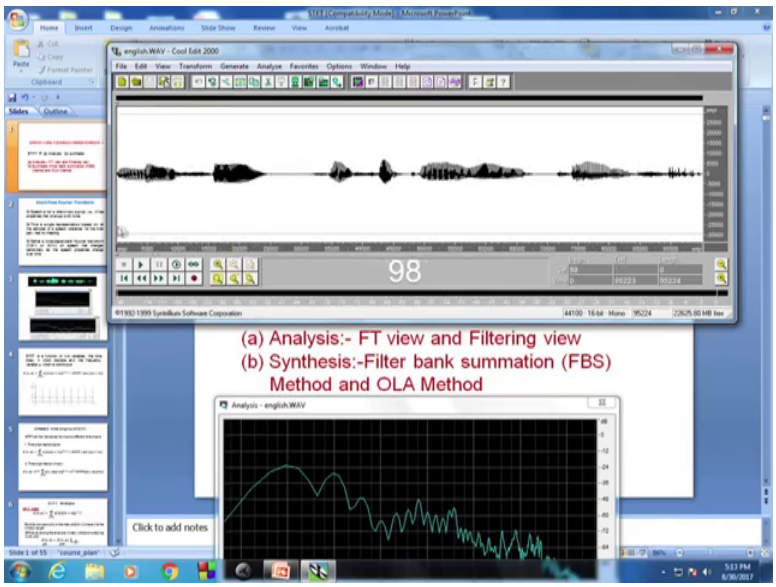
Now, if you see magnitude spectra of this portion of the signal if I select this portion of the signal it is different; different spectra. So, I can say the speech is not a stationary signal. So, I have to analyze the speech signal with a segment wise. So, instead of taking the whole signal, I have to take the signal part of a signal. So, let us I have recorded my voice of a sentence and it has one second it has sampled at 16 Kilo Hertz with 16 bit.

(Refer Slide Time: 07:27)



So, each sample is encoded with 16 bit and sampling frequency is 16 Kilo Hertz. Suppose I record a sentence which is 3 seconds long. So, if it is a sentence is 3 second long, how many sample will be there? 3 into 16 K sample so, 48 k sample will be there. So, 48 thousand sample will be there. If I take whole signal at a time do the frequency analysis. So, during the sentence speech is not same sometime I said some consonants sometimes some sometimes some consonants sometime consonant to (Refer Time: 08:12). So, all kinds of variation is exist.

(Refer Slide Time: 08:23)



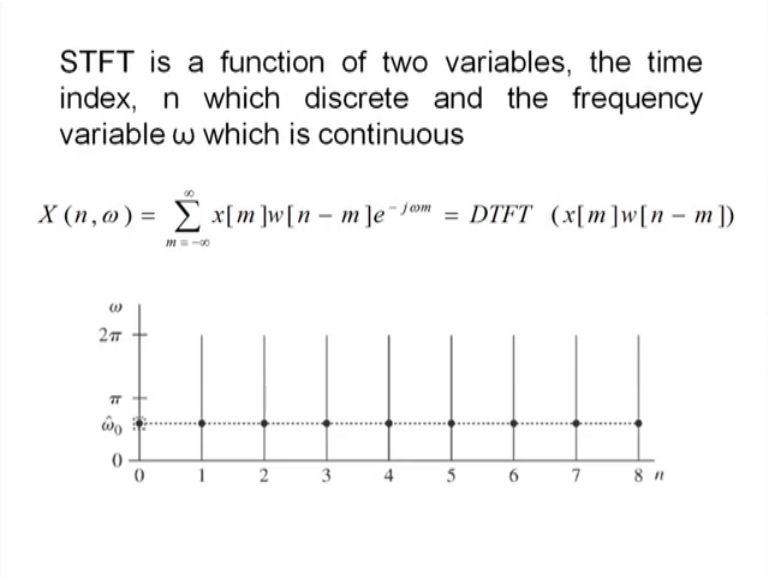
If I take the whole signal and draw the magnitude spectra, which is like this if I take the whole signal and draw the magnitude spectra; that means, whole signal and I draw the magnitude spectra that is nothing, but the average spectra of whole signal I cannot get the local variation. So, what I want instead of taking the 48000 sample at a time, I let us say I divide the signal with a some segment; those are the segment. So, let us I divide the signal with a segment which is called hundred frame per second; that means, in one second I will analyze hundred frame. So, let us hundred frame. So, I am I am a segment I want to analyze a segment, which is nothing, but a if it is 16 Kilohertz, it is 10 millisecond of each segment. So, 10 millisecond so, you know already know the framing. So, what I do? I cut a segment of the speech and do the frequency analysis.

So, once I cut a segment of the speech one segment, then it is called time is short time. And if I do the Fourier analysis, Fourier transform. So, this is called short time Fourier transform. So, I can say I frame by frame I take and analyze and I have to get back the same frame again if I take the inverse DFT. So, whatever modification I do in the spectral magnitude, then again I take the inverse transform to get the same frame back. So, that is called synthesis. So, analysis when I doing the transform synthesis when I get back the signal again is called synthesis.

So, after analysis after finding out that let us I get the magnitude spectra is like this, I modified the magnitude spectra; I can modify it; once I modified it and again I take the inverse transform I should get the signal back which is modified property ok.

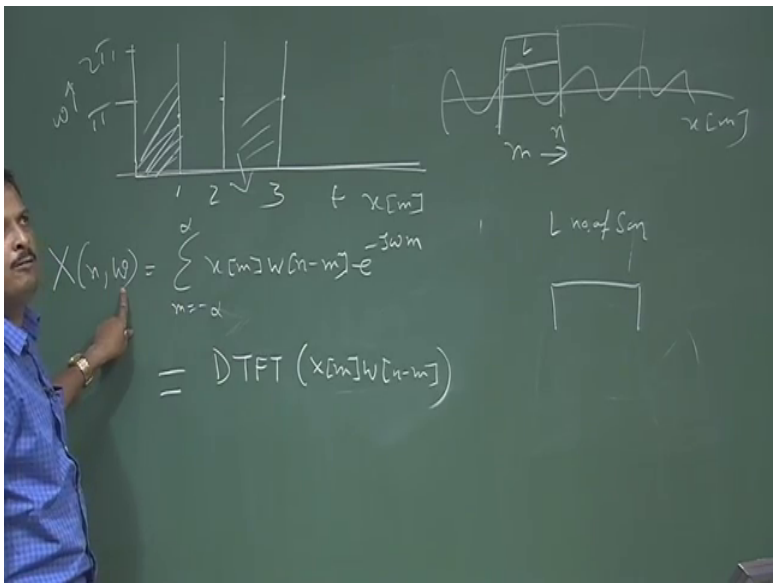
So, then once I get the signal back that is called synthesis the analysis part is called analysis. So, this is called that is why this is the short seg segment of time is analyzed that is why it is called STFT short time Fourier transform; STFT, what is STFT? Now we know; now next topic is that.

(Refer Slide Time: 11:15)



So, short time will I. So, what I am doing. So, this is a time scale, if you see this is my frequency scale.

(Refer Slide Time: 11:20)



So, what is the maximum frequency of if I need normalize discrete way? So, twice f this is omega scale and this is f pi, twice pi is the f pi is the maximum baseband signal. And I analyze let us for this time instant. So, this time instant I analyze. So, this portion of the signal I have analyzed I get the spectra. Then again I analyze for this time instant this

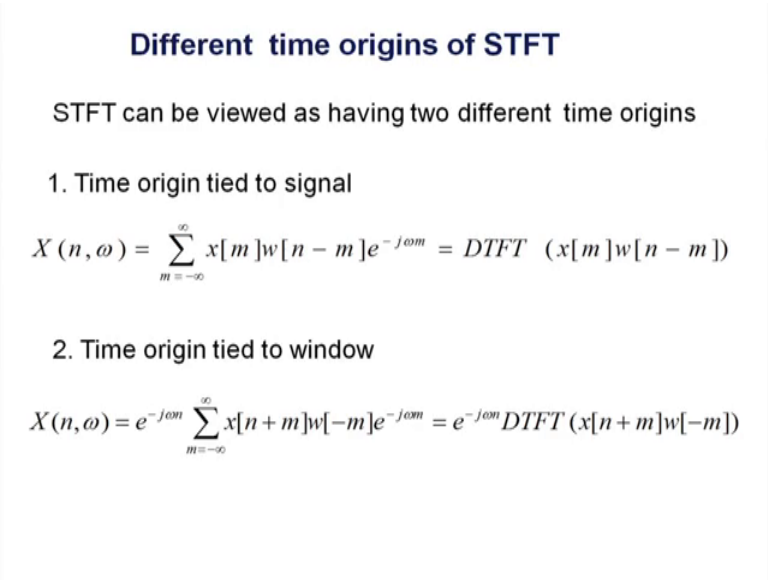
time instant I analyze again I analyze for this time instant, this frame I analyze and get the spectra. So, if you see this one is that.

So, now if I want to mathematically represent this thing. So, what is this? This is nothing, but a x n digital signal, let us it is infinite to infinite digital signal let us it infinite to infinite or 0 to 4800 in case of real signal. So, if it is infinite to infinite, then I can say x m x n omega, x n omega; I want to find out the frequency response of this portion. So, this portion is 1, 2, 3. So, this is n; I can say this is n equal to 1, n equal to 2, n equal to 3, this n will vary; for this omega is nothing, but a x of m, n equal to minus infinity to plus infinity I have cut the signal; that means, I am multiplied this signal with a window function which is n minus m. I just illustrate this one.

What is this? Let us this is my whole signal. This is whole signal x m. So, this is the nth instant n. I want to cut a portion of the signal, what I do? From the nth instant, let us I want to cut L number of L number of samples, L number of sample. So, what I do from nth instance I take the L number of sample this side. So, this is L length sample. So, once I cut it means I am multiplying a rectangular window of L length. So, suppose this is my rectangular window. So, what I am doing I inversing the time of the window this side and cut the signal.

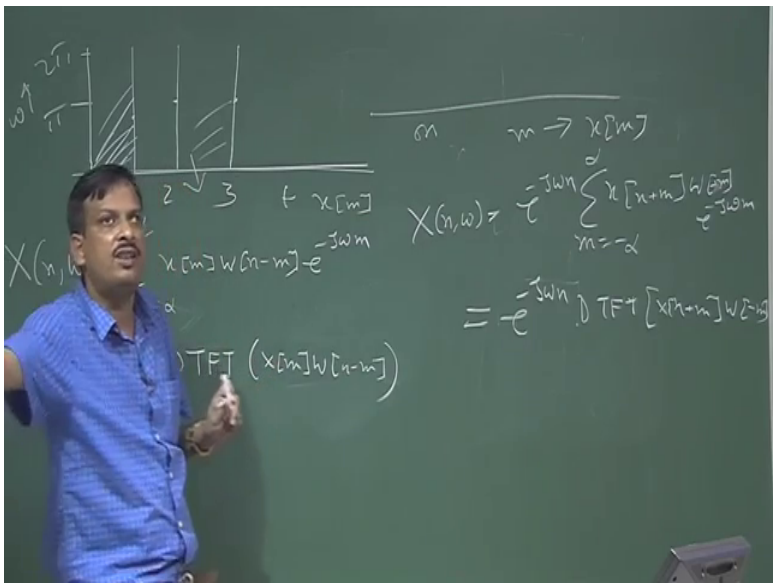
So, that is why it is n minus m; because m is varies in this axis. So, n minus m number of signal I cut and multiply it. So, what is this? This is nothing, but a frequency domain it is continuous. So, I can say this task form. So, it is frequency transform means this e to the power minus J omega m. So, this time domain it is discrete frequency domain it is continuous. So, I can say it is nothing, but a DTFT-discrete Fourier transform of x m multiply by w n minus m which is window function; this product. Discrete time Fourier transform of this one. Now, if I want to make this omega is discrete; then it is called discrete Fourier transform.

(Refer Slide Time: 15:08)



So, if I see that I am again this slide I have already explained ok. Let us explain this slide also. So, if you see there is a time origin which is m; either will be with the window or with the signal. So, if it is this one; then I call time origin is with the win with the signal because m is varies infinite m series m is where is an x m is where is infinite.

(Refer Slide Time: 15:35)

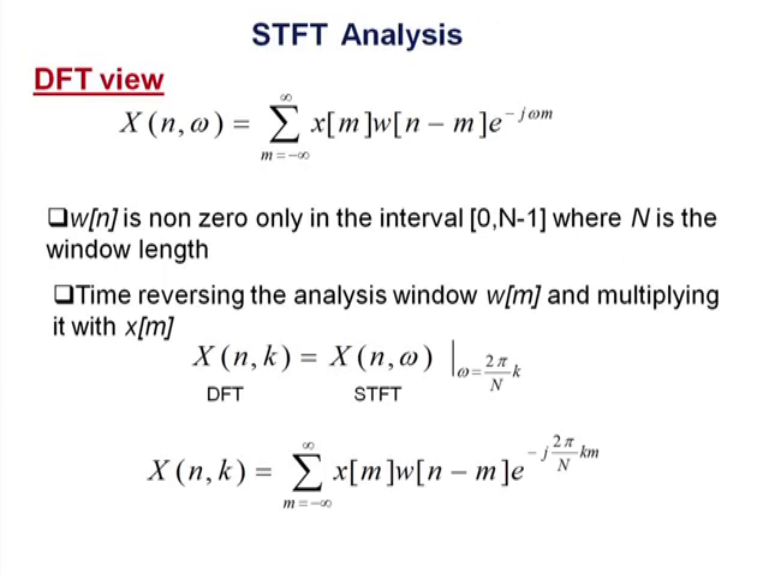


So, time origin with the signal itself that is why window is clap back and cut the signal. Now if the time origin tired with the window, then the same things I can write x n omega is equal to so, time will origin with the window. So, what I am doing? I am shifting the

signal of a shifting the window with the time. So, e to the power minus j omega n omega n m equal to minus infinity to infinity x of n plus m omega n sorry omega minus m minus m e to the power minus j omega m. In that case, I can say it is nothing, but a e to the power minus J omega n DTFT of x of n plus m into omega minus m time origin tied with the window.

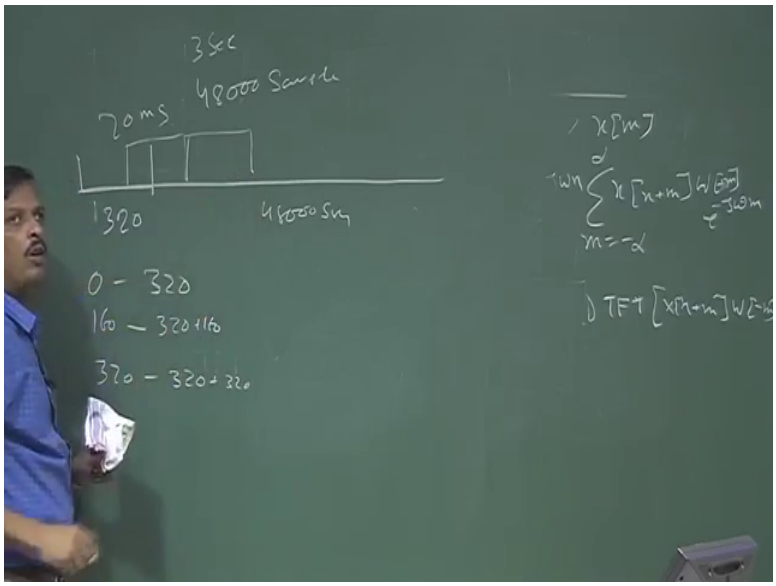
So, I said the time is shifting with the window function only. So, one is that I have a signal, I can cut the signal. Another is that I have a signal I have take the signal I have a long window I cut the window at a m number of time and multiply with the signal rest at 0 ok.

(Refer Slide Time: 17:21)



So, in practical what we are doing? In practical scenario what will happen time origin tied with signal. Let us forget about that part. So, suppose the example whatever I have given. So, I have 3 second recording.

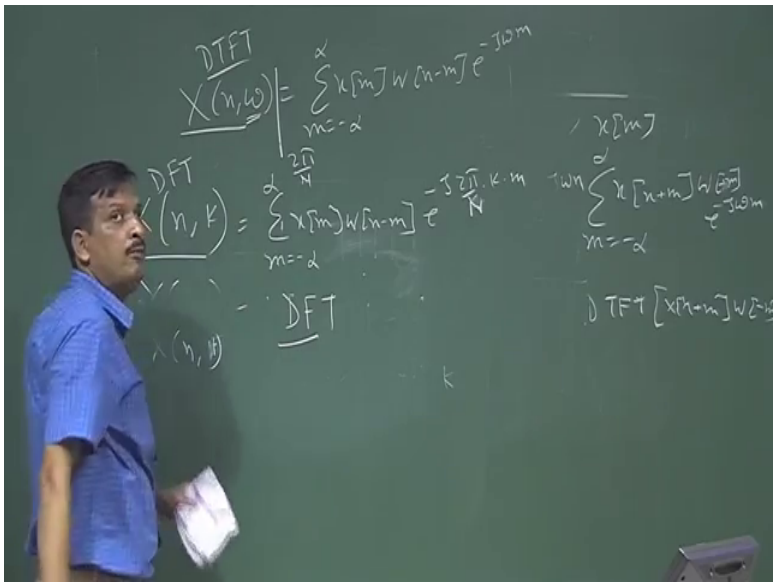
(Refer Slide Time: 17:38)



So, I have a 48000 sample. Let us I take 10 millisecond window, so; that means, 160 sample. So, in 160 sample I take and I do that DFT, then again take the another 160 sample then I do the DFT. Actually what we have done since there is a cutting effect, if I have a signal of long signal 4800 sample, I take a window size of let us 20 millisecond and then shifted the window 10 milliseconds that mean; 100 frames per second.

So, that first I take 1320 sample from 0th sample to 0 to 320 samples. Next analysis window I will shifted the window only 10 millisecond so, next I said 162 again 320, so 320 plus 160. Again I said 320 to 320 plus 320 to 320. So, that way I have taken that window shifted the window with 50 percent overlap. Again I will come that later on briefly. So, STFT analysis with DFT view discrete Fourier transfer view.

(Refer Slide Time: 19:14)



So, what I said x of n omega; omega is continuous is nothing, but a m equal to minus infinity to infinity x of m w n minus m e to the power minus j omega m. Now this is DTFT discrete time Fourier transform. So, omega means continuous; now I want to make omega is discrete. So, how the omega becomes discrete? Omega is continuous in the frequency plane. So, what I do? Let us I have a frequency scale, I have a DFT; DTFT the DFT discrete Fourier transform, then the length of the discrete Fourier transform is involved. So, DFT length if you remember the review of the DSP there is a n is called DFT length so.

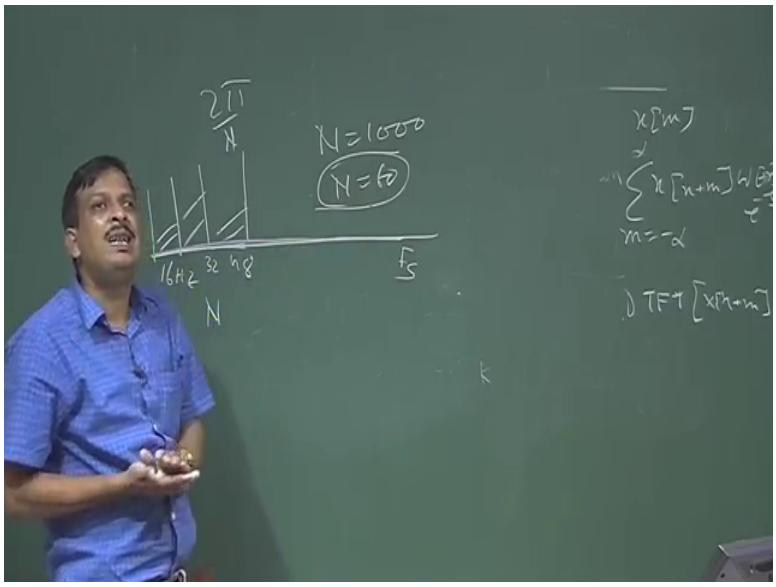
If I take that n number of DFT n is the DFT length; that means, 0 to 2 pi whichever the discrete frequency that 2 pi frequency scale I have divided in n number of sample. So, there is. So, each division is nothing, but a 2 pi by n 2 pi is the highest frequency which is f s. So, if it is a 16 Kilohertz signal and length of the DFT is length of the DFT is 1000, then I can say each of the segments is 16 Hertz. So, it is 0 to 16 Hertz, 16 to 32 Hertz. So, each of the segment is 16 Hertz which is twice pi by n nothing, but a fs by n. So, I can say I can divide this omega with respect to k in term of discrete frequency which is called k.

So, instead of omega I can write x of n k where I can write m equal to minus infinity to infinity x of m will be same w n minus m will be same. So, instead of continuous omega I have divided the omega in term of 2 pi by k e to the power minus J 2 pi by k 2 pi by n

into k into m; that means, this omega has divided into 2 pi by n. So, x of n 0 first component is nothing, but a dc; second component k equal to 1. So, it is 1. So instead of k I put one first frequency, second discrete frequency, third discrete frequency. So, k has a relationship with analog frequency f; f is nothing, but a 2 pi by n is the resolution into k.

So this is discrete frequency discrete once I do that, then this process is called DFT discrete Fourier transform. So, I can write this representation is called DFT this representation call DTFT discrete time Fourier transform and discrete Fourier transform now what is 2 pi by n? If you see 2 pi by n 2 pi by n is my frequency resolution.

(Refer Slide Time: 23:14)

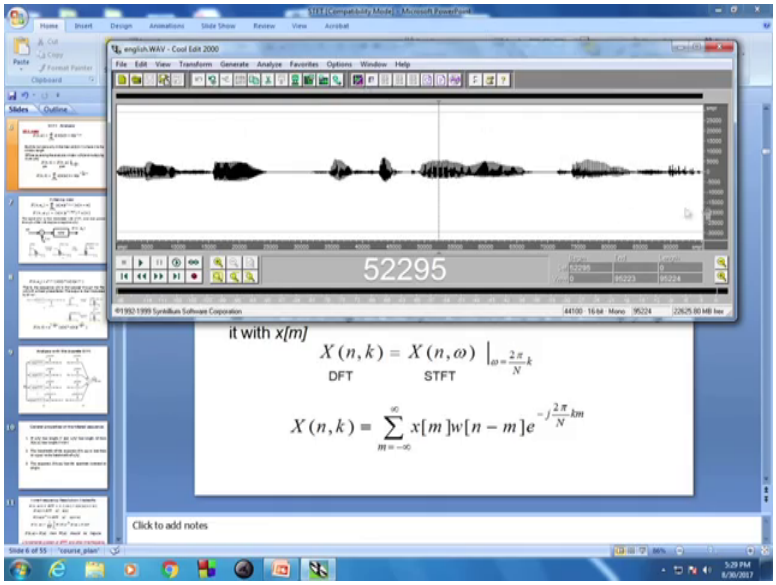


So, if I have a 16 Kilohertz signal, 16 Kilohertz is a sampling frequency and n equal to 1 Kilohertz 1 k then I said every band is nothing, but a 16 Hertz. So, this is 16 Hertz, this is 16 to 32 Hertz, this is 32 so 32 to 48 Hertz. So, I can say, but f s the frequency scale is divided in a band. So, number of band is divided by number of band is represented by n. So, if it is n equal to 1000; that means, number of band is 1000. If it is n equal to 64, then I can say 0 to fs signal is divided in 64 band 64 4band. So, sometime this is called band number also n is called band number also.

I can say this is a filter band of band frequency band this band this band this band. So, sometime n is called frequency band also ok. So, now I give you a real problem that

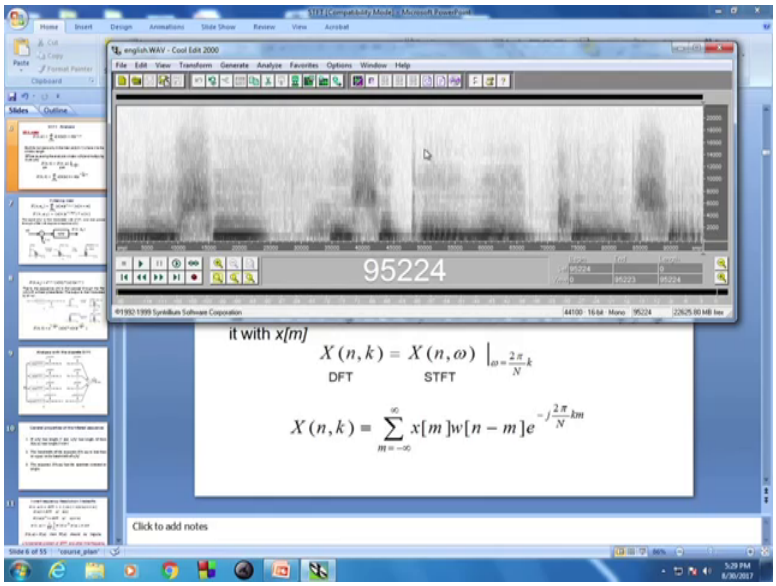
problem is that if you see in here I will filter bank analysis come later on let us come in here.

(Refer Slide Time: 24:51)



If you see there is a plot; we have already discussed which is called spectrogram.

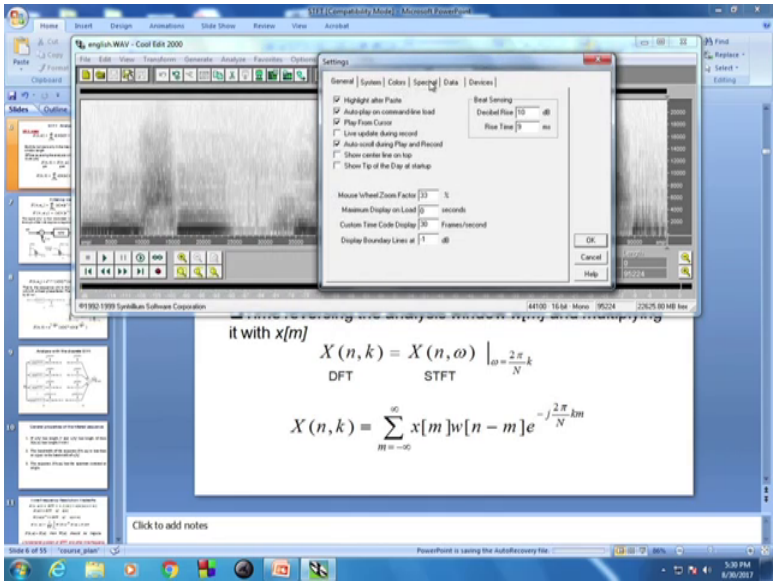
(Refer Slide Time: 24:57)



So, this plot is called spectrogram; this plot is called spectrogram. So, what is spectrogram consist? If you see this axis is the time, these axis is the frequency and the intensity of the particular frequency is represented by a intensity represent the amplitude of the particular frequency. Amplitude of the particular frequency is represented by

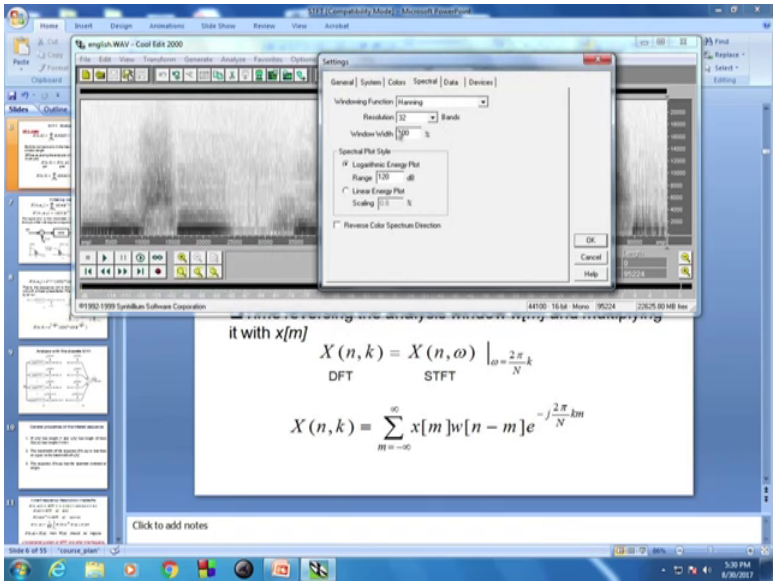
intensity. Suppose, I want to write a program for this pictogram analysis, How do I write? it if you see. So, I have to know.

(Refer Slide Time: 25:51)



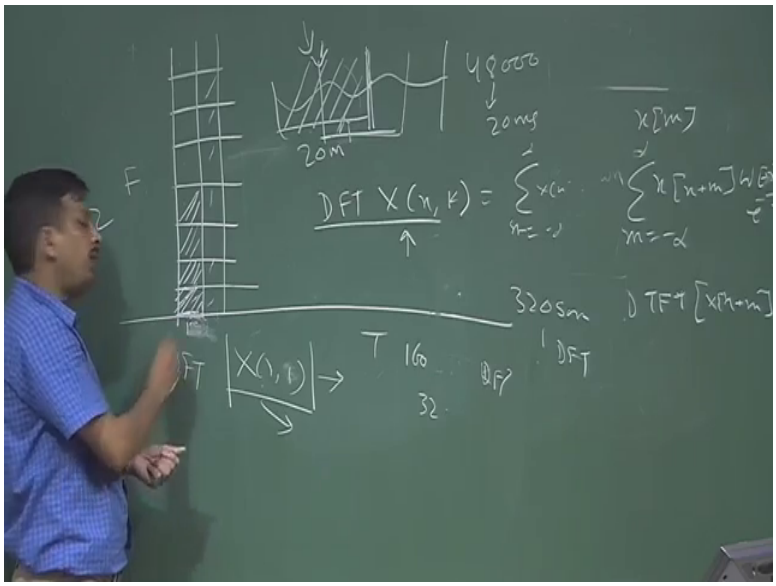
So, what are the parameters I have to set? Window I will come later on window option what kind of window I should say.

(Refer Slide Time: 25:52)



If you see the settings and then spectral, resolution is 32 band; 32 band means that whole frequency if it is f s is divided into 32 band. So, f s is divided into 32 band so; that means, that if I want to make it let us this is my y axis.

(Refer Slide Time: 26:20)

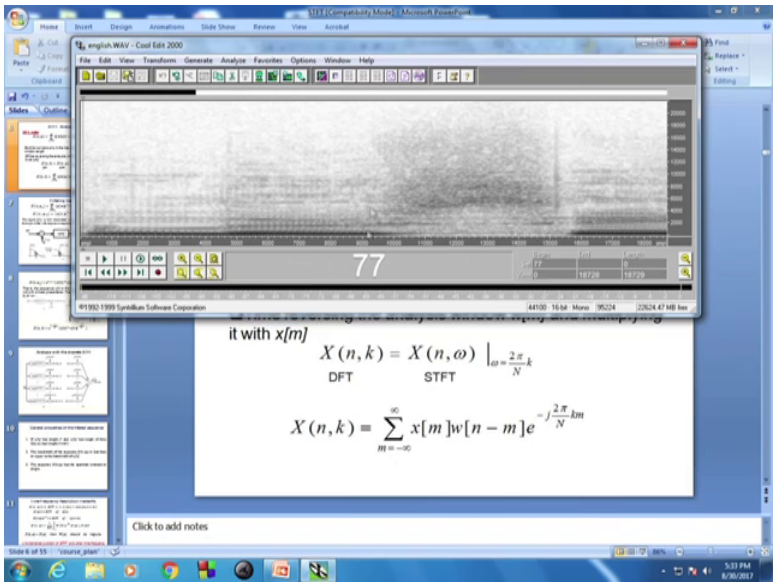


Which is frequency; this is my x axis which is time. I take a small segment of the speech let us I take this time. So, this time this is the time I have taken.

Now, if I say this frequency f s is divided in 32 bands; that means, these axis I can divided whole frequency scaling 32 band. So, n is equal to 32 because it is implemented in fft. So, 2 2 to the power something it has to be. So, either it will be 32, next is 64, next is 128. So, that way it will be increase. So, 32 band. So, now, for this time this band. So, I taken this signal and analyze DFT, what I will get? I get x let us this time instant is one. So, it is 1 k I will get for this time instant one k I get once I get one k if I take the mod then I get the amplitude spectra. So, this amplitude is decoded into a scale which scale color scale. What is the color? If it is black if I say if it is black then intensity is maximum. So, let us maxima is one if it is intensity is maximum max max intensity is black and if it is white then it is intensity is mean intensity is white.

So, the value of x 1 k will come a value. So, within that value I can assign appropriate color of this band for k equal to one this is the first band. Now for k equal to 2, second band k equal to 3rd third band k equal to four4th band I can find out this value and color this block again I shifted the time next block I color this block. So, if I plot that way I get get this spectrogram if you see 64 if I take it to 64 if I zoom time domain if you see there is a band, band will be visible sorry; band will be visible.

(Refer Slide Time: 28:57)



If you see the line; vertical line and horizontal line, so this is called spectrum spectogram plot, how do you plot the spectrogram? So, depending on the intensity or the particular frequency component the color will come and the band is called band resolution. So, that is will come resolution trade off. So, 2 pi by N or F s by N is nothing, but a resolution of that particular band

So, this is the STFT analysis on DFT view. So, I in DFT view I can say if I want to do that. So, I will say take a window take a window any kind of window effect I will discuss later on. So, if I have a signal take a small portion of the signal and analyze the frequency using DFT technique the DFT view of the STFT analysis we are doing the activity once I do the I g f t then I can say I can get back this frame again. Now what will happen, if I just separate this portion, once I get back I will have a problem in this junction, because next band is here.

So, the signal is continuous so, but this beginning this end of the window I get a problem, because if I say this DFT of X n k is nothing, but a frequency response of the signal multiplied by the frequency response of the window. If you see that it is nothing, but a m equal to minus infinity to infinity X n multiply by the omega n. So, I am multiplying a window function with this thing. So, so it is frequency domain it will be convolution. So, I can say that frequency response whatever I get it is a convolution of

frequency response of the signal and the frequency response of the window. So, the what about the frequency response I get it is a convolution window effect will be there.

So, at the boundary there is a window effect. So, if I do segment by segment and there will be a window effect, so instead of doing that, what we will do? We take a window of let us 20 millisecond and then shifted the window 10 milliseconds; that means, 50 percent overlap. So, I analyzed for 50 percent overlap. I will discuss why the 50 percent how much amount of overlap we should allow? How much amount of overlap we not allowed, so that we can get the signal back again so, that we will discuss in synthesis part.

So, if I do the 50 percent overlap, then I get the no problem. So, that I will discuss, why it is 50 percent? Why not it is 60 percent, 70 percent? How much is allowable? How much is allowable r not that will discuss during the ft, ft synthesis that is the o l a method and f b s method. So, STFT analysis what I will do if I recorded for 4800 sample. So, what I will do I take a window of 20 millisecond; that means, 320 sample, then I analyze the DFT and then I shifted the signal shifted the window for the next frame by 10 millisecond means; 160 sample. Then I get the next frame do the DFT, then I shifted the again 10 millisecond I analyze the window and do the DFT then I get the signal. Similarly for doing the spectrogram also, what I can do? I can take a signal for 20 millisecond; and analyze for 20 millisecond and shifted the time shifting of the 10 millisecond or 5 millisecond. So, this will be 5 millisecond.

If it is 10 millisecond, this will be 10 millisecond. If it is shifted by only single sample then single sample delay. So, if I single sample then see the computational complexity 48000 sample for 48000 sample of the analyze 48000 times. So, I will later on I will say what should be the trade off? What is the redundancy is there? That we will discuss then the for synthesis purpose also. So, now, depending on the shifting you get the resolution I will discuss. So, next class we will think about what is the filtering view of STFT then you go for the time frequency trade off and then we go for the synthesis part. We draw the analysis diagram and then go for the synthesis ok.

Thank you.