

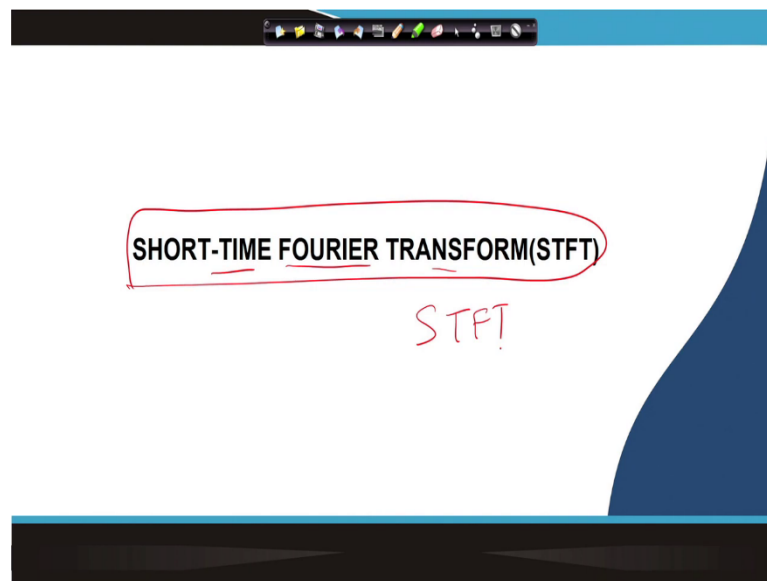
**Signal Processing Techniques and Its Applications**  
**Dr. Shyamal Kumar Das Mandal**  
**Advanced Technology Development Centre**  
**Indian Institute of Technology, Kharagpur**

**Lecture - 28**  
**Short-Time Fourier Transform (STFT)**

Ok. So, let us do it next. What are we discussing? So if I have a long signal infinite-length signal, I select a small portion of the signal, and then I take the DFT for analysis in the frequency domain.

So, basically, I am multiplying a window with the signal, and the frequency response is nothing but a window frequency response of the window convolved with the frequency response of the signal, ok.

(Refer Slide Time: 00:55)



So, now that the example, short term Fourier Short Time Fourier Transform, Short Time Fourier Transform which is known as STFT, is used in speech signal processing.

(Refer Slide Time: 01:08)



For example, let us see. Suppose I have a long signal, a long speech signal. Now, if you see the speech is a non-stationary signal, non-stationary means the signal changes its property over time.

Now, what is the meaning? This means that the colour of the signal in this portion is different from the colour of the signal in this portion. So, that is a non-stationary signal. Now, let us say you may say ok, say, can I take the whole signal at a time and do the DFT and take the IDFT and also do the IDFT and get the signal again.

But what will happen? When you take the entire signal, since the basis of the Fourier transform is that the signal is stationary, and it has a period with  $N$ , that is the basic assumption of the Fourier transform.

So, what will I get? I will get this Fourier transform, this spectrum. If I plot the spectrum, it will look like this. This is in an average spectrum of the entire signal.

So, some portion of the signal is noise, some portion of the signal is silence, and some portion of the signal is periodic. So, when all are added together, I get an average response, which may not be useful. I do not understand the composition of the signal in this portion, but I am interested in that.

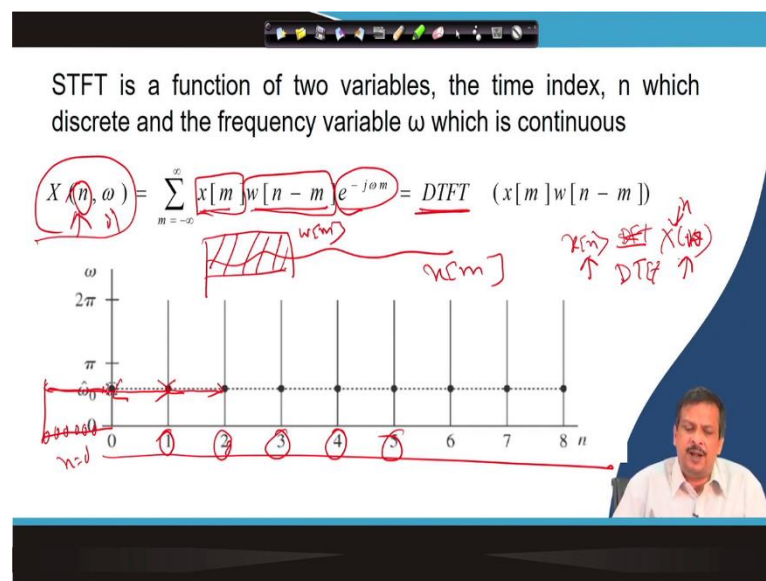
So, what will I do? I will select a small portion of the signal and do the frequency analysis. If I draw the spectrum, then I get this kind of response. So, what do I want to do? I want

to make a window, or I can say, I want to select some portion of the signal, which is nothing but a windowing, as we discussed in the last class.

So, I have a long signal, and then I window the signal, I have to make a window. So, I have to select a small portion of the signal. How do we do that? Again, as I explained, I have to define a window function  $W_n$ , which is 1 only within the length of the window; let us know if the length of the window is  $L$ .

So, 0 to  $L$  minus 1  $W_n$  is 1. If  $n$  varies from 0 to  $L$  minus 1 outside,  $W_n$  is 0. So, basically, I am multiplying a signal with the window function, ok or not. So, basically, if this is my  $x[m]$ ,  $m$  varies from 0 to or, let us say, minus infinity to infinity or 0 to infinity, whatever. Then, I am multiplying with the  $x[m]$  with a  $W_n$ , which varies from 0 to  $L$  minus 1, where  $L$  is the length of the window.

(Refer Slide Time: 04:51)



So, basically, what was I doing? I have a long signal then, so if I say I have an  $x[n]$  when I take the DFT, I get  $X(k)$ . So, here, it is only time that is varying; here, only the frequency is varying, but in this case, the time is also varying. So, here I may do another dimension, which is called time.

So, another time dimension is added here. That time dimension is discrete. Why is it discrete? Because I have selected the signal up to this, and that is window number 1. I have selected the signal up to this. So, first, I have to select the signal here to here. So,  $n$  is equal

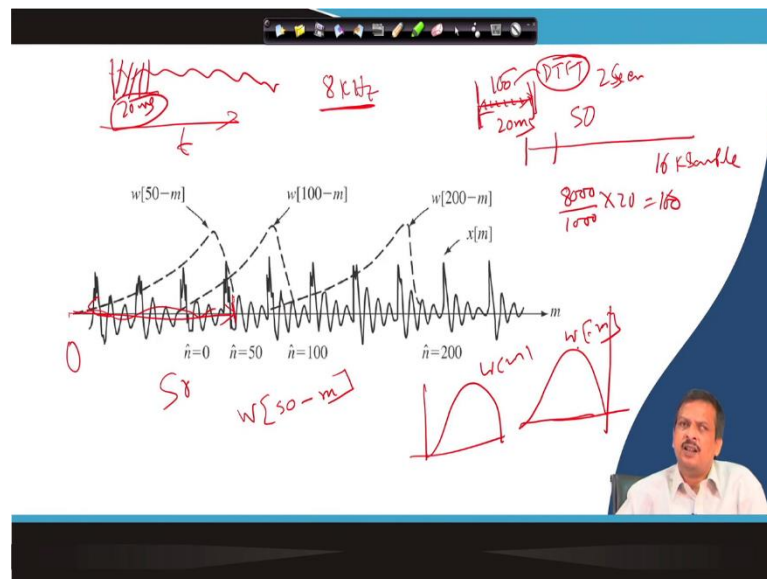
to 0, this portion is also there, but this portion signal is 0, outside the negative end signal is 0.

So, it is the 0th window. Then it is 1st window, 2nd window, 3rd window, 4th window, 5th window. So, my  $x$ , so if I take the  $x[n]$  instead of DFT, if I take the DTFT, then instead of  $k$ , it will be  $\omega$ . So,  $\omega$  is continuous. So, let us take the DTFT first, and then I will come to the DFT. DFT is nothing but a sampled version of DTFT, which is a frequency sample.

So,  $\omega$  is continuous, and  $n$  is discrete. So, my frequency domain signal not only varies along the frequency, but it also varies along the time. So, how do we do that? I multiply a signal with a window.

So, I have a signal, and I multiply it by the signal. So, this is my  $x[m]$ ; I multiply the signal with a window function  $W[m]$ . And in the frequency domain, when I do the DTFT, it is nothing but a convolution of the frequency response. So, that is why I said  $x[m] x[n-m] e^{j\omega m}$  is the DTFT of the selected signal. Is the DTFT of the selected signal clear?

(Refer Slide Time: 07:35)



So, in a real-life example, let us say I have taken a real-life example. Forget about this picture. Suppose I have recorded a speech signal with a sampling rate of 8 kilohertz. Carefully listen to it. I have recorded a signal with a sampling rate of 8 kilohertz. Now, I am saying I am selecting a window whose length is 20 milliseconds.

So, I checked take a portion of the speech signal, which is 20 milliseconds. Why did I choose a portion? Because of who is along the timeline, the speech signal is changed with the property.

Now, I am considering if the length of the time is very small, then I can say within that period, the signal is stationary. So, change is negligible along the time axis, and within 20 milliseconds, change is negligible. So, I want to create a 20-millisecond window.

So, suppose I have recorded my name, which is 2-second length 2-second length. So, in 2 seconds, how much data will I get? I will get a 16 k sample if the sampling frequency is 8 kilohertz. So, I get 16k samples.

So, out of 16k samples, I selected the first 20 milliseconds. How many samples will be there? So, in 1 second, 8 k samples, it is nothing but a millisecond, which means 8 samples. In 20 seconds, I get 160 samples, so there will be 160 samples.

So, I have to select 160 samples, and then I take DTFT Discrete Time Fourier Transform initially, let us say Discrete Time Fourier Transform. But, in the frequency domain, it is convolution.

So, if I say this is my signal, and this is my  $n$ ;  $n$  equal to 50. So, 0 to 50 samples is my first window. Let us say this is 50 instead of 160 samples; there are only 50 samples. So, I can say  $W_n$  minus  $m$ , so 50 minus  $m$   $n$  is the index of you can say, the time because it is convolution in the frequency domain.

So, if I have a window function like this, it will be reverted; this is  $\omega_m$ , and this is  $\omega$  minus  $m$ . Is it clear? Then, if I slide over the signal, I get the window 50 samples, or if I slide over the window 160 samples, I get the 160 samples. Is it clear?

(Refer Slide Time: 10:58)

origins of STFT

STFT can be viewed as having two different time origins

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{-j\omega m} = \text{DTFT}(x[m]w[n-m])$$

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x[n-m]w[m]e^{-j\omega(n-m)} = e^{-j\omega n} \sum_{m=-\infty}^{\infty} x[n-m]w[m]e^{j\omega m}$$

$x(n, \omega) = e^{-j\omega n} X(\omega)$

$X(n, \omega)$

So, now, what will be the DFT? So, let us say STFT can be viewed as having two different time origins. One is time origin tied with signal, and the other is time origin tied with window.

So, in this case, the time origin is tied to the window, and here, the time origin is tied to the signal. So,

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{-j\omega m}$$

The same thing happens if I change the index. So, instead of  $x[m]$ , I said you know that the convolution index can be changed. So, I can change  $x[n-m]$  and  $w[m]$ .

So,  $m$  is nothing but a  $n$  I changing to  $n$  minus  $m$ , time origin tied with window. Here, the time origin time with the signal is  $n$  minus  $m$ , so if we do that,  $e^{j\omega n}$  since there is no variation in  $n$ .

So, it will be outside and  $e^{j\omega n}$ . So, if this is  $X \cap n \omega$ , then I can say  $x$  of capital  $X \cap n \omega$  is nothing but  $e$  to the power minus  $j \omega n$  into  $X \cap n \omega$ , which is nothing but a shifting.

Properties of discrete Fourier transform shifting, ok. So, this can also be an STFT of  $x[m]$ , and this is also an STFT of  $x[n]$ . So, when we use that view of STFT, we can use any one of the equations.

(Refer Slide Time: 13:01)

Analysis

DFT view STFT


$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{-j\omega m} \quad \text{DFT}$$

$w[n]$  is non zero only in the interval  $[0, N-1]$  where  $N$  is the window length

Time reversing the analysis window  $w[m]$  and multiplying it with  $x[m]$

$$X(n, k) = X(n, \omega) \Big|_{\omega = \frac{2\pi}{N}k} \quad N$$

DFT STFT

$$X(n, k) = \sum_{m=-\infty}^{\infty} x[m]w[n-m]e^{-j\frac{2\pi}{N}km} \quad \text{DFT}$$


So, let us say STFT as a DFT view. So, I said STFT is a discrete Fourier transform view discrete Fourier transform. So this is my DTFT, discrete-time Fourier transform. Now, when the discrete-time Fourier transform  $\omega$  is continuous.

Now, I make the  $\omega$  is in discrete  $2\pi$  by  $N$  into  $k$ .  $N$  is the length of the DFT. So, instead of  $j\omega N$ , it is  $2\pi$  by  $N$  into  $k m$ , which is the DFT equation. Now,  $\omega$  becomes  $k$ . So now, it is discrete in time and discrete in frequency. So, that is the DFT view.

(Refer Slide Time: 13:59)

Filtering view

$$X(n, \omega_0) = \sum_{m=-\infty}^{\infty} (x[m]e^{-j\omega_0 m})w[n-m]$$

$X(n, \omega_0) = (x[n]e^{-j\omega_0 n}) * w[n]$

The signal  $x[n]$  is first modulated with  $e^{-j\omega_0 n}$ , and then passed through a filter with impulsive response  $w[n]$ .


$X(n, \omega_0)$   $\omega_0$

$X(\omega)$   $\omega_0$   $\omega$   $\pi$   $\omega$

Modulate  $\rightarrow$  Filter

$X(\omega + \omega_0)$   $W(\omega)X(\omega + \omega_0)$   $W(\omega)$   $\omega$   $\pi$   $\omega$

$\omega + \omega_0$   $\omega - \omega_0$



Now, I come to the filtering view. Forget about these slides. Let us come here only. So, let us say I am saying that what is the equation of the DFT?

(Refer Slide Time: 14:15)

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x[m] w[n-m] e^{-j\omega m}$$

DTFT

$$= \sum_{m=-\infty}^{\infty} x[m] e^{-j\omega m} * w[n-m]$$

$$= X[n] e^{-j\omega n} * W[n]$$

Block diagram:  $x[n] \rightarrow [w[n]] \rightarrow y[n]$ , where  $y[n] = x[n] * w[n]$ .

Let us say I have a signal whose frequency is  $\omega_0$ . I have a long signal whose frequency is  $\omega_0$ , as I have taken in the last class  $\cos(\omega_0 n)$ . So, let us say I have a signal with only one frequency component, which is  $\omega_0$ , a long signal.

I take a small portion of the signal, and then I compute the STF short-time Fourier transform, so I compute the DTFT. So, I can say

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x[m] w[n-m] e^{-j\omega m}$$

is the frequency transform.

So, it is if it only consists of one frequency, I get only  $\omega_0$  here, which is this one. So, it is only the  $\omega_0$  component will be there.

So, how do you get that Fourier transform? You are creating a different frequency component and convolved with the signal. If that matches the signal, then the power will be high; that is the procedure. So, let us I have only I have a signal which only has a frequency  $\omega_0$ . So, that is the frequency transform, or DTFT discrete-time Fourier transform.



Now, if I do it in DFT also, I can say that this is nothing but a  $m$  equal to minus infinity to infinity  $x[n]$  multiplied by  $e^{j\omega_0 n}$ , then convolved with. So, instead of a summation sign, I can say that this is convolved with  $W_n$  minus  $m$ .

So, this is nothing but a convolution. So, I can say  $x[n] w[m]$ . So,  $x$  of let us say I said  $n$  here instead of there will be  $m$ , this is  $m$ . So,  $x[m]$  multiply by  $e^{j\omega_0 m}$  or  $j \omega_0 n$  whatever you can take; instead of here  $m$ , I have taken because the  $n$  index is there.

$$(x[m] \cdot e^{j\omega_0 m}) * w[m]$$

Here,

I am writing the same equation in terms of convolution form instead of mathematical convolution form. So, let us know if  $m$  is replaced by  $n$  time because my picture is in  $n$ . So, I change the index to  $n$  minus  $j\omega_0 n$  convolution with  $W_n$ .

So, what does it mean? This means that if I say I have a signal  $x[n]$  passes through a filter, the system  $w$  frequency impulse response is  $W_n$ , and then the output is  $y[n]$ . So,  $y_n$  is nothing but a convolution of  $x[n]$  convolved with  $w[m]$ . Here also, this can be a signal, and this is my filter, or I can say this is nothing but a system. So, the system is nothing but a filter.

So, I can say the signal  $x[n]$  multiplying by  $e^{j\omega_0 n}$ , what does it mean? It is nothing but a modulating; when you multiply a  $\cos \omega_m t$  multiplied with  $\cos \omega_c t$ , it is nothing but a modulation. So, when you are multiplying  $e^{j\omega_0 n}$ , that means  $x[n]$  is modulating with the frequency  $j\omega_0 n$ .

Then, it passes through a filter whose impulse response is  $W_n$ , and I get the frequency response of  $x[n]$  for a particular frequency  $\omega_0$  because my  $x[n]$  consists of a single frequency  $\omega_0$ .

Now, consider my single  $x[n]$  consists of a frequency response like this. So, this is  $\pi$ . So, the highest frequency component should be  $\pi$  by 2 and  $2\pi$  by 2. So, it is nothing but a  $\pi$ . So, this is below  $\pi$ . So, now, I have an I have a frequency, so  $X[n]$  has a frequency response which is  $X(\omega)$ .

So, this is my frequency response of  $X(\omega)$ . Now, when I say that I only pass  $\omega_0$ , that means this is my  $\omega_0$ . Let us see if this is my  $\omega_0$  here. So, when I say multiplying  $e^{j\omega_0 m}$  modulating,

that means  $\omega_0$  is transferred to the centre. So, that is nothing but an  $X(\omega)$  plus  $\omega_0$ . So, there is a two-component  $X(\omega)$  plus  $\omega_0$  one is  $\omega$  minus  $\omega_0$ , ok.

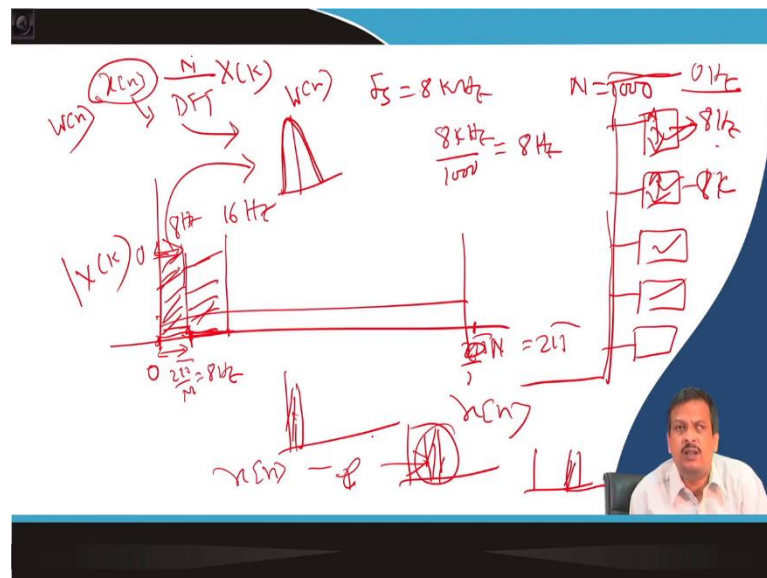
Let us say the negative frequency I am not considering. I am only considering one sideband. So, it is nothing but the  $X(\omega)$  plus  $\omega_0$ . Modulating then, let us know this is my filter frequency response.

So, this is nothing but a  $W \omega$ . So, this is my filter frequency response. Then, when this thing will pass when this will pass through this filter frequency response, it will be changed to the original signal, and this is the window.

So, since the signal does not have this portion, only this portion will be there if this portion is not there. So, this band is coming due to the filter response  $W_n$  frequency response.

Then, I can say instead of  $\omega_0$ , there is a  $\omega_1$ . So, I can say that DFT or discrete Fourier transform is nothing but a passing. You can pass signal through a several bandpass filter, ok. Let me give you an example, then you will understand clearly.

(Refer Slide Time: 21:33)



Let us say I give you a practical example. I have a signal  $x[n]$ , and I take the  $N$  point DFT, and I get  $X(k)$ . There is no short-term Fourier transform; it is a depth discrete Fourier transform of whole  $x[n]$ .

So, you know if the signal is sample  $F_s$  is equal to 8 kilohertz and  $N$  equal to 1000, then what do you know? The resolution is nothing but 8 kilohertz divided by 1000, which is nothing but 8 hertz.

So, what does it mean? This DFT means that I am creating an 8-hertz component and then convolving it with the signal. If that 8-hertz component exists in the signal, I get the output then I create a 16-hertz signal, and then I create a 24-hertz signal, which is the meaning of the DFT.

So, if I plot the spectrum of this  $X(k)$  mod of  $X_k$ , what I will get is 0, which will be  $k$  equal to 1. So, this is  $2\pi$  by  $n$ . So, this is  $N$ .  $N$  is equal to  $2\pi$ , and I divided the  $N$ . Each division length is  $2\pi$  by  $N$ , which is, in this case, 8 hertz. So, this portion is 8 hertz, which is another 8th hertz.

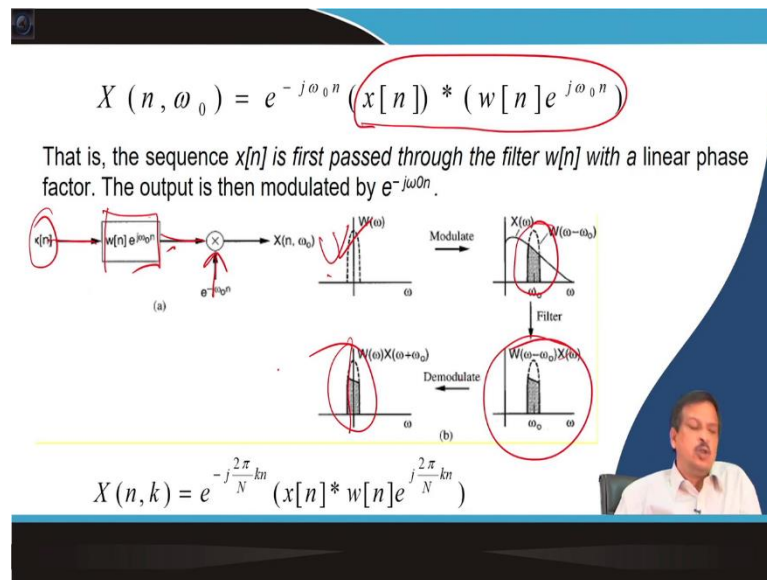
So, this is 0 to 8 hertz. This is 8 to 16 hertz. So, I can say my signal passes through a set of bandpass filters, a set of bandpass filters. So, this is my signal  $x[n]$ . Passes through a set of bandpass filters. Each filter bandwidth is 8 hertz. So, this is so: 1st filter is 0 hertz this is, 2nd filter is 8 hertz bandwidth, 3rd filter bandwidth is 8 hertz but shifted to 8 to 16 hertz.

So, these all are a band pass filter. So, this filter is the 0th filter, so the 1st filter will look like this, the 2nd filter will look like this, the 3rd filter will look like this, and if I sum all the filters, I get back the spectrum. This is the normal DFT. Now, when I say the short-term Fourier transform, there is another block; that block is that  $x[n]$  is multiplied by a window function.

So, it is not that the whole 8 hertz will pass if this portion passes through a filter whose bandwidth looks like this. Is the bandwidth of the  $W_n$  clear? So, what I am doing is I am generating this kind of thing. So, how do I generate this one? It's nothing but a frequency shift of 8 hertz.

So, I have an  $x[n]$  multiplied by a frequency shift, and I get this one. So, here also the same thing I am getting. Modulating this one, I get the filter, and then I get the signal. I can also do it in reverse.

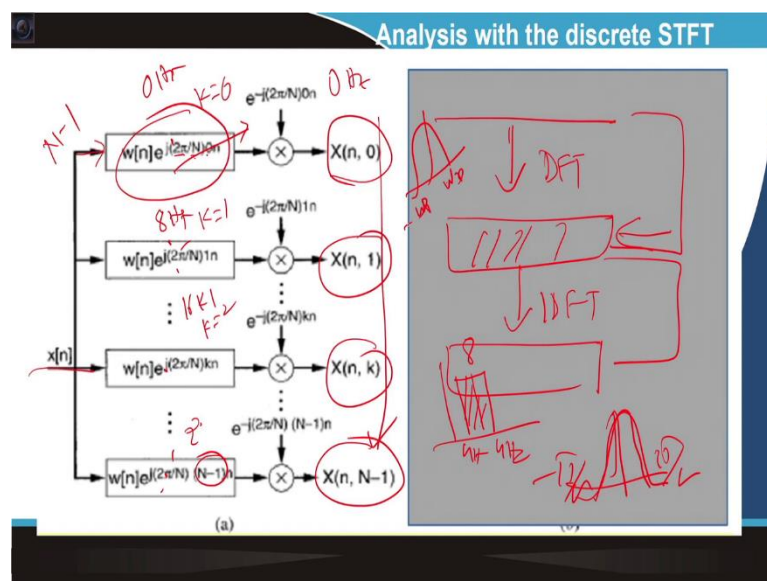
(Refer Slide Time: 25:32)



Instead of this, I can say first, the convolution will be done, and then shifting will happen, with no problem. So, first, I pass the signal with the system and then demodulate it. Same thing I will get. So, this is the window frequency response.

So, when this signal passes through this, I get this one. So, after the output of the filter is this one. Now, this one has to be centred at  $\omega_0$ . So, I shifted to the here I get the frequency response.

(Refer Slide Time: 26:17)



So, if I say what the analysis of discrete-time short-time Fourier transform is, I can say I have a signal pass the signal with window, whose frequency response is  $j 2\pi N$ . Here,  $k$  is equal to 0,  $k$  is equal to 1, and  $k$  is equal to 2. In my practical example, this is 0 hertz, this is 8 components, this is 16 hertz component, and this is 24 hertz component if  $k$  is equal to 3.

So I get all the  $k$ ,  $k$  varies from  $N$  minus 1. So, I get  $N$  minus 1 component. So, the  $N$  minus 1 filter will be there, and they are demodulated and get the frequency response here. So, that is called STFT synthesis.

Why STFT synthesis? Because I have a speech signal, I have to synthesize, and I have to task; when I say the discrete Fourier transform, I get the frequency domain, but again, I have to go back to the time domain.

So, I have to take inverse DFT, and I get back to the time domain. This portion is called analysis; this portion is called synthesis, so STFT analysis and STFT synthesis are two different things. So, whatever I explain, it is nothing but an STFT analysis. Is it clear? So, what is the bandwidth of this filter bandwidth of the filter is 0 hertz this filter.

So, I get a 0-hertz component, then an 8-hertz component centred at 8 hertz. So, 8 hertz, I will get like this 4 hertz this side, 4 hertz this side. Because of the window function, the window function main lobe; main lobe is  $\pi$  by  $L$ , which is minus  $\pi$  by  $L$ . So, this will be the defect of the window function that is the window function frequency response, this is  $W_p$ , and this is minus  $W_p$ .

So, this is STFT. So, this is used very much in speech signals. Even in the spatial domain also, you can use STFT. Instead of taking an entire image at a time, I can take a small portion of the image by multiplying a window, which is the same as multiplying a window and getting back the frequency response. So, in the next class, I will talk about STFT synthesis.

Thank you.