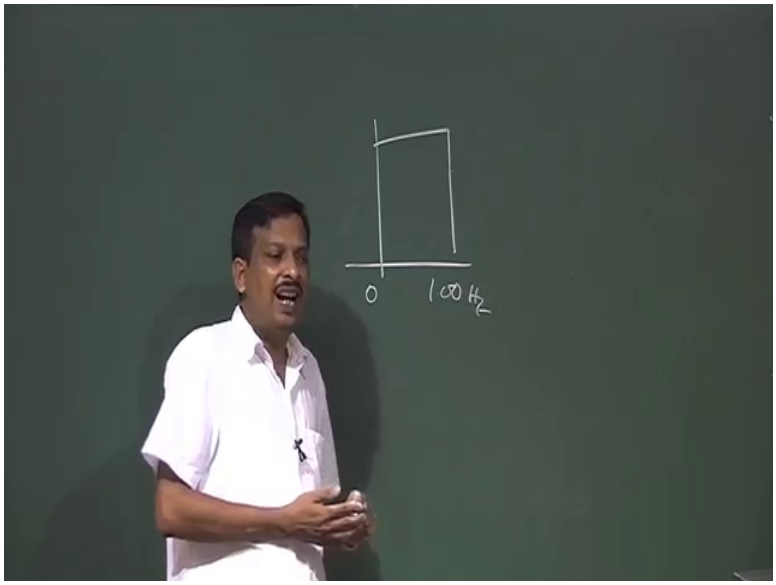
Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture – 32
Cepstral Transform Coefficients(CC) Parameters axtraction

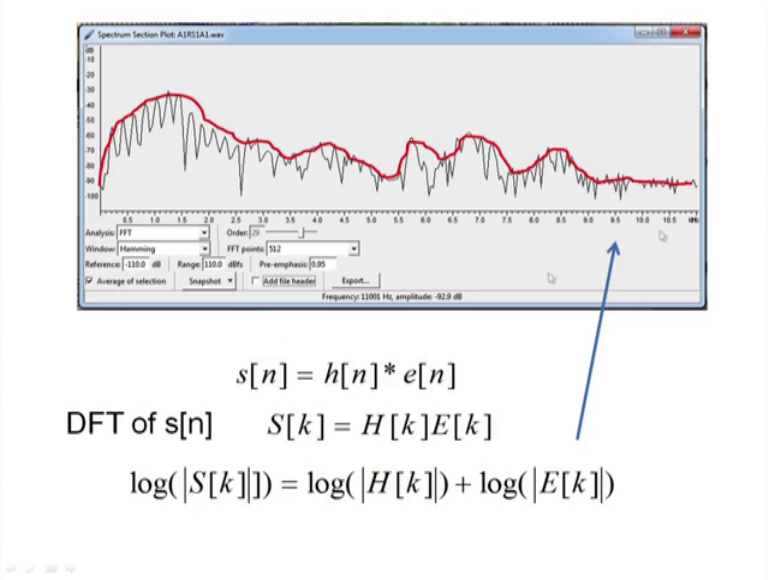
So, we discussed about the filter bank analysis. So, output of a filter is nothing but a frequency parameter that is why the filter bank analysis is call frequency domain parameter extraction methods.

(Refer Slide Time: 00:42)



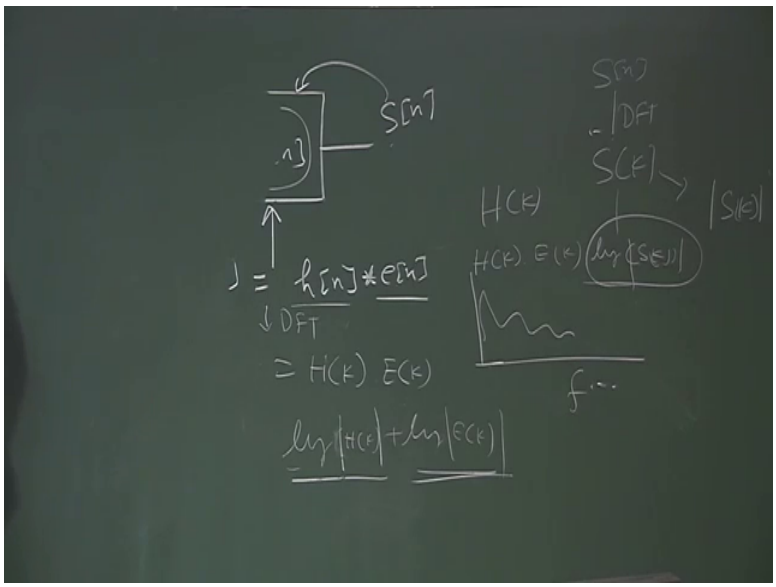
Now, design of those kind of filters has a little bit of problem means they are not that much of robust because design an ideal filter of pass band 0 to 100 hertz is very difficult you know that an DSP, how to design this filter. So, instead of doing that let us we try to represent other way.

(Refer Slide Time: 01:05)



So, what is that if you look at the slides I am interested about this envelop only the red line because what you know the speech is nothing but this is human vocal track.

(Refer Slide Time: 01:15)



Lets this is h n, and there is a excitation e n. So, en pass through the, so glottal excitation passed through the vocal track produce the different speech event. So, I can say this is nothing but a s n speech signal. So, different speech event who is responsible to produce different speech event h n; e n either e n maybe excitation is present or excitation is not

present. If it is excitation is present, then we call voice signal; if excitation is not present there is a random noise then we call unvoiced signal.

So, let us there is a voice signal. So, the excitation is present and that excitation is same for all voice signal only difference is that shape of the vocal track is different and different kind of speech signal is produced. So, our ultimate aim find out the actual representation of h n from the s n. So, I recorded the speech signal s n which is nothing but a convolution of h n convolved with e n. Now, I want that how do I extract h n eliminate en I want to eliminate en. So, if it is a convolution time domain, it is a convolution, now if I take the frequency transform of this signal. So, this is s k, this is h k, this is nothing but a multiplication of e k.

So, if I take the DFT here - discrete Fourier transform, or if I represent the speech signal in frequency domain, so this is nothing but a product of spectral representation of vocal track with the spectral representation of source excitation source. So, if I say excitation source is nothing but a impulse, then I want a speech or signal processing methodology by which I can separate h k from the product of h k into e k. So, what I can do if you see if it is a product, if I take the log, log of S k is nothing but a log of H k plus log of E k. So, I can take the absolute log or I can take the simple log also. So, in log domain the product is nothing but a additive. So, if it is additive then can I separate H k from E k from product of H k and E k from the addition of log H k and E k.

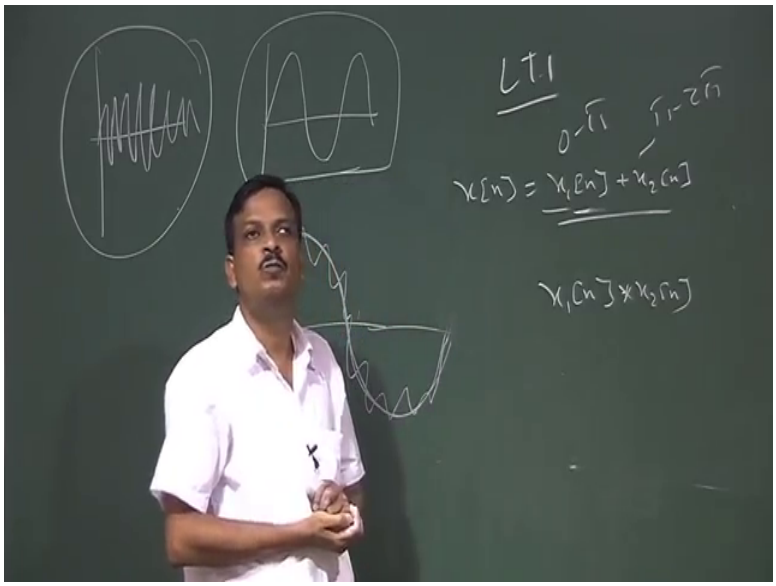
So, if you see if this is a log spectra that means, after DFT analysis I take the log this is log of H k this plot is log, I have taken a signal S n plane signal I take the DFT, and after DFT I what is get I get S k then I take the log magnitude or log of S k log of log magnitude of H k. Then if I plot it, it will be look like this, this is log frequency axis is in log, if you see I can show you.

(Refer Slide Time: 05:36)

If you see, this is a linear view. So, this is the frequency scale is linear. So, here instead of taking the log I find out that mod of S k, and I plot that mod of S k with respect to frequency that is spectrum that this axis is the frequency axis. And this axis is the mod of x, which is root over of x square plus b square because S k is in complex domain. So, I get this kind of response which is linear. Now, instead of plotting this if I plot the log of this with the frequency then instead of linear view, if I take the log view. So, if I take the log view, I get this one.

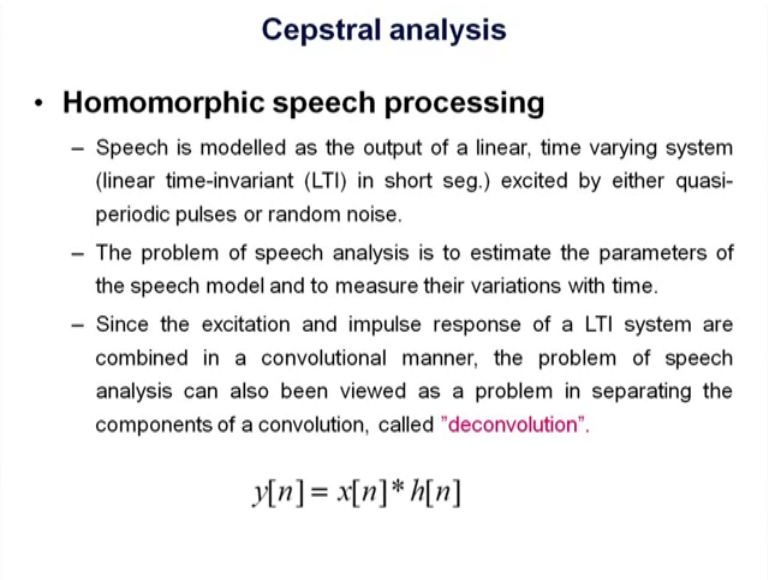
Now, if you see that I am interested about this envelop. So, envelop represent if this slides envelop represent the h n or H k and variation represent the E k. So, if I want to extract that if I consider this is a signal, let us consider this log plot is a signal. So, there is a signal if I told you that if you if you remember that suppose there is a sine wave there is a signal which is high frequency signal, and there is a signal in pure sine wave.

(Refer Slide Time: 07:12)



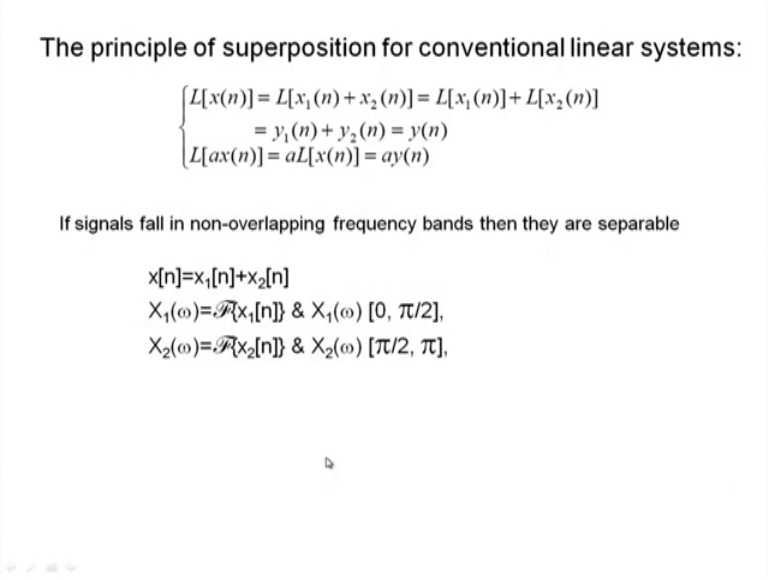
If I take the product of them, what I will get the sine wave impose will be the high frequency signal. So, now, I want to extract this smooth percent of the signal this sine wave. So, I can say the sine wave is a low frequency component, this is nothing but a high frequency component. If this is my complex signal, so I can say if I pass this complex signal through a low pass filter, so low pass output will be the smooth variation of the signal; and high pass output will be the high variation of the signal. So, I can say if I pass this spectrum with a low pass filter, this is consider a signal, and pass this through a low pass filter, then I can say the output of the low pass filter actually give me the envelop representation, so that is my extraction (Refer Time: 08:24). So, since I want the extract only H k, so after log if I pass this portion to the a low pass filter, then I can eliminate E k that is my target and to find out log of H k. This kind of signal processing has a special name, this is called homomorphic signal processing. So, this kind of signal processing has a special name which is called homomorphic signal processing.

(Refer Slide Time: 09:02)



So, I am not explaining again this slides. So, I want a methodology by which I can deconvolve the convolve signal. So, this type of signal processing represent as a homomorphic signal processing.

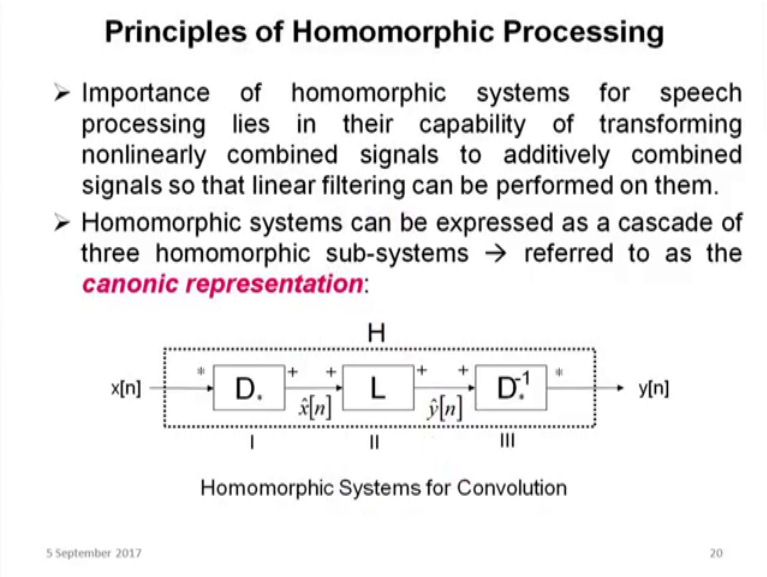
(Refer Slide Time: 09:19)



Now, what is LTI system LTI system LTI system (Refer Time: 09:30) linear systems or conventional linear system suppose support the superposition principle that means, if I apply L is a transform of x n is nothing but the L of x 1 n plus x 2 n. Or of I apply every input separately and output can be added up both will be same that is the LTI system

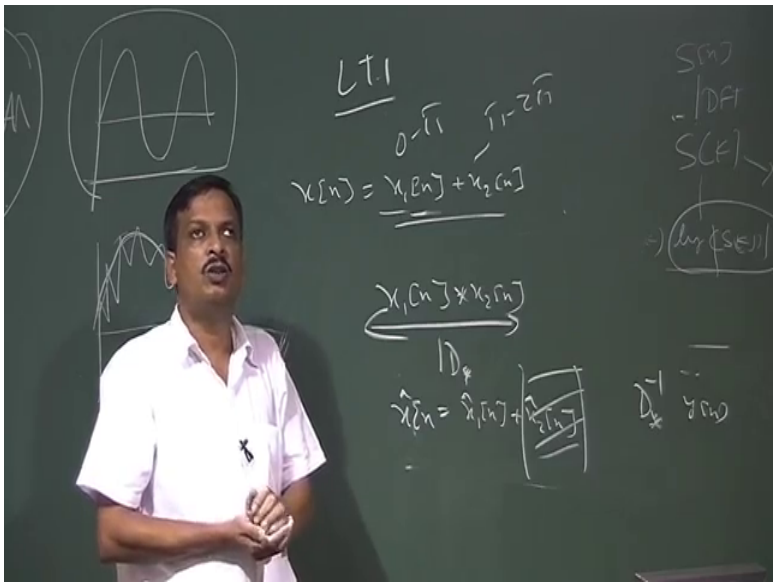
superposition principle. So, if the signal fall in non overlapping frequency band then they are separable. So, suppose I have the signal which is nothing but x 1, x n is nothing but a x 1 n plus x 2 n, addition of two signal. Now, if x one n consist of frequency zero to pi and x two n consist of frequency pi to 2 pi then I can easily separate by a linear filter I can easily separate. But if x 1 is convolved with x 2 n then it is very difficult to separate them or some of the x 2 is overlap with x n the it is difficult to separate.

(Refer Slide Time: 11:01)



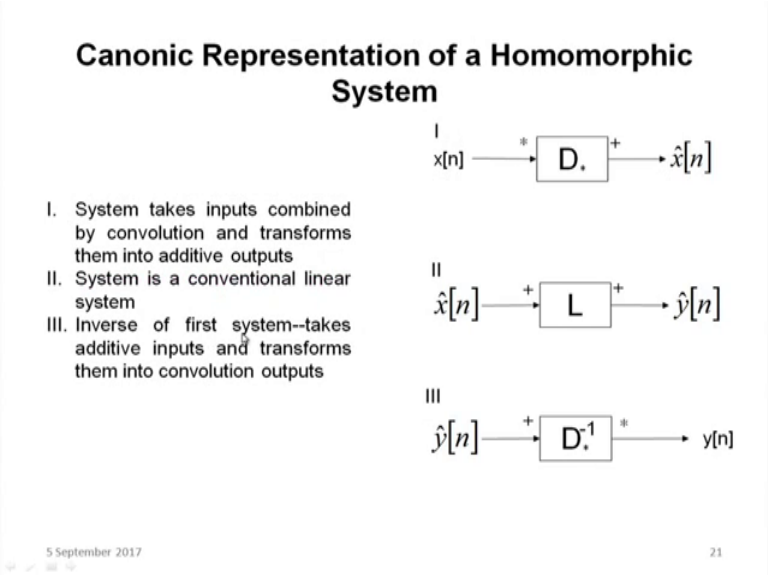
So, what I want we want in principle of homomorphic signal processing. So, importance of homo system is a speech processing lie in their capability of transforming non-linearly combined signal to additively combine signal. So, basic purpose is that transform the non-linearly combine, so I want that convolution signal can be represented by a additive signal. So, what kind of transformation I should do, so that convolution become simple addition. So, if you see the slide this figure represented the homomorphic system for convolution, so x n some kind of transformation D then it represents the instead of convolution I get x 1 n plus x 2 n. Instead of x 1 n convolved with x 2 n, I want a transformation D star which will represent x 1 so the which may be the x cap n which in the form of x 1 cap n plus x 2 cap n that I want.

(Refer Slide Time: 12:11)



Then it can pass through the linear filter and I do the deconvolution inverse transform of this D, then I get the y. So, linear filter is that suppose I want to extract this one. So, I can extract this one. and I can inverse filter I can apply inverse transform I can apply, I can get the x 1. So, this is the purpose for homomorphic convolution.

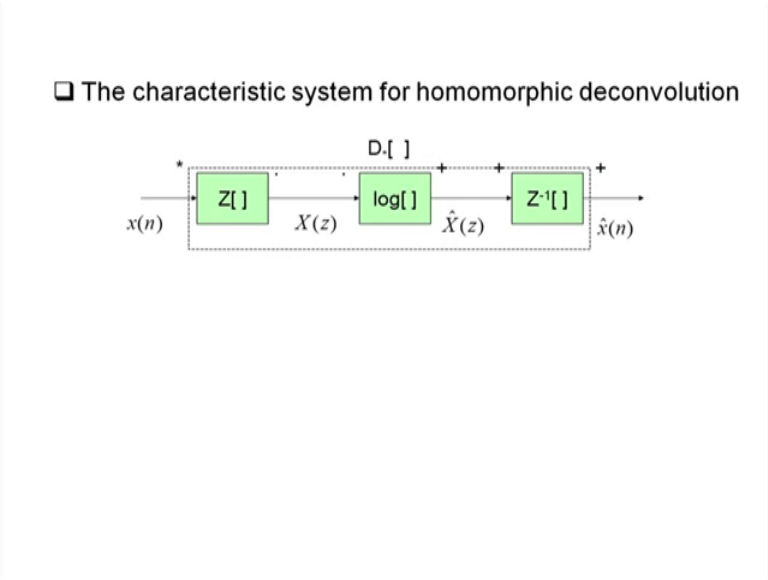
(Refer Slide Time: 13:00)



So, if you see this picture, there is a three 1, 2, 3 - three part. First part system take input combine by convolution and transform them into additive output. Second part system is conventional linear system I can suppose linear filter inverse if first system take the

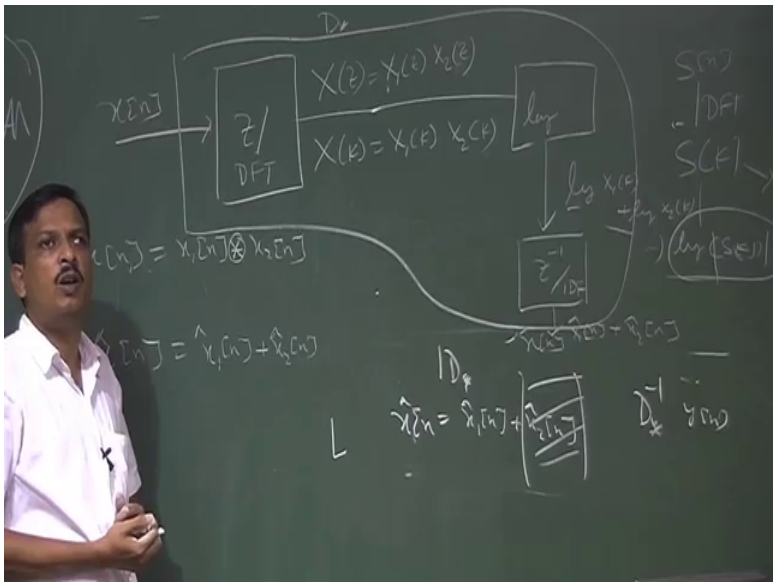
additive input and transform to them convolution output. So, whatever modification I will do in L part linear time in variance of those will be in additive signal. So, this is called canonic representation of homomorphic system. If I see this diagram this diagram for convolution, I can write a diagram for multiplication also.

(Refer Slide Time: 13:49)



So, if I want that I want that convolution should be represented in a some form, then the system transform function should be like this.

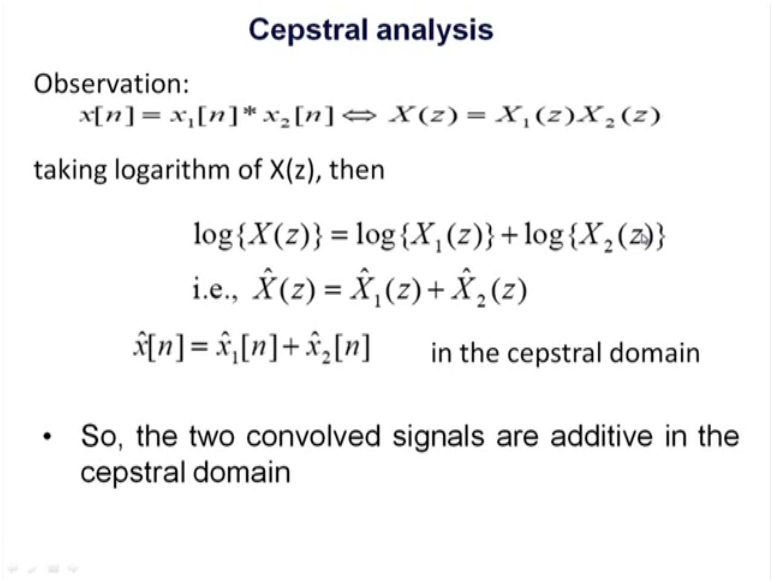
(Refer Slide Time: 14:12)



Lets I take x n, I take x n, I want that x n is nothing but a x 1 n convolved with x 2 n.

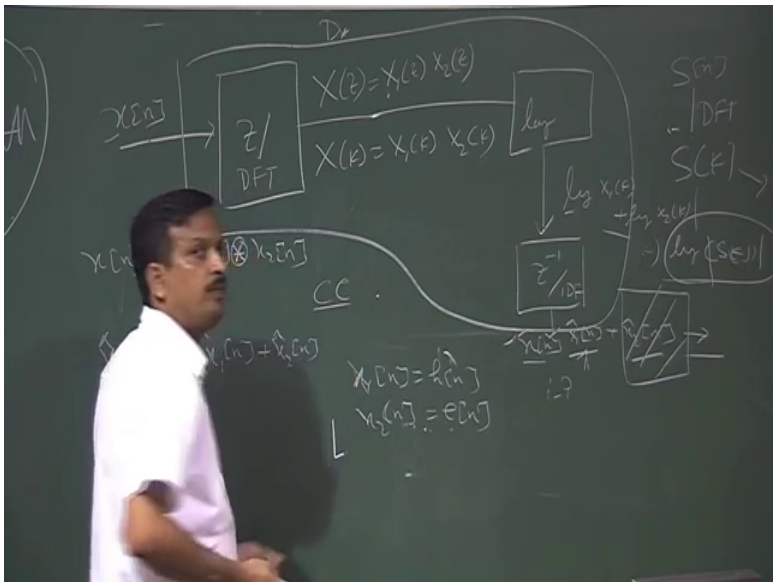
So, at the end, I want to find out the x 1 n instead of convolve, I want a x cap n which is nothing but x 1 cap n plus x 2 cap n that is my target and this is my input. So, I apply this input. Let us take the z transform, z domain. So, let z transform. What I will get at output I will get X z. So, X z is nothing but X 1 z multiply by X 2 z, z transform of frequency domain representation or I can say it is a DFT. So, I can get x k is nothing but X 1 k into X 2 k. Then I take log if I take the log, what we will get I will get log of x 1 k plus log of x 2 k, x 1 k and x 2 k both are complex, both are complex. Now, these things so this is in k domain. Now, I apply inverse z transform or IDFT. So, I get x 1 cap n plus x 2 cap n which is nothing but a x cap n. So, I can say this whole system can act as a D star which is nothing but a homomorphic system for deconvolution.

(Refer Slide Time: 16:55)



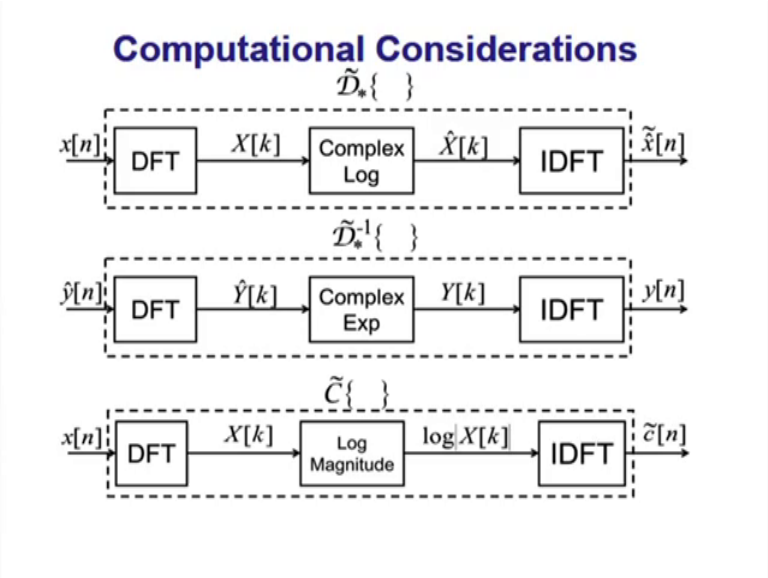
So, if you see the slides x n, z transform, so convolution become product dot, then dot becomes addition, then addition becomes addition, I take the inverse transform in time domain. So, same things whatever I do here, it is there in the slide x 1 n in z domain then I take the log I get cap take the inverse transform then I get the x 1 cap into x 2 cap, this is called Cepstral domain. If you see it is not exactly x n, I taking this log signal as a time domain signal and take the inverse DFT. So, this is not exactly x n. So, this domain is call Cepstral domain, Cepstral domain. Now, if you see it is a additive signal, so I can say it can be pass through a low pass filter to suppose I want to discard x 2 that can discard x 2 find out x 1, so that x 1 actually represent.

(Refer Slide Time: 17:57)



So, let us x 1 is equal to nothing but a x 1 n is equal to h n and x 2 n is equal to nothing but a e n then I can say I have remove e n, what about the x cap 1 n is present is represent the envelop of the cepstrum. So, x 2 n is the cepstrum of that envelop. So, actually x one n represent the time variant vocal tract, so that x 1 n those can be used as a parameter who will actually represent the envelop portion of the spectral signal. So, this kind of homomorphic signal processing is used to extract the Cepstral parameters or are called CC parameter I will come.

(Refer Slide Time: 18:55)



So, this is the computational consideration. So, if I say D star, DFT, log either complex log or I can take the log magnitude. So, if it is a complex log then I call complex cepstrum; if it is a log magnitude only then I call real cepstrum. So, if this instead of log I can take only log magnitude if you see the slides, if I see the only log magnitude hen I can this is the real cepstrum. If I take the complex log this is called complex cepstrum.

(Refer Slide Time: 19:44)

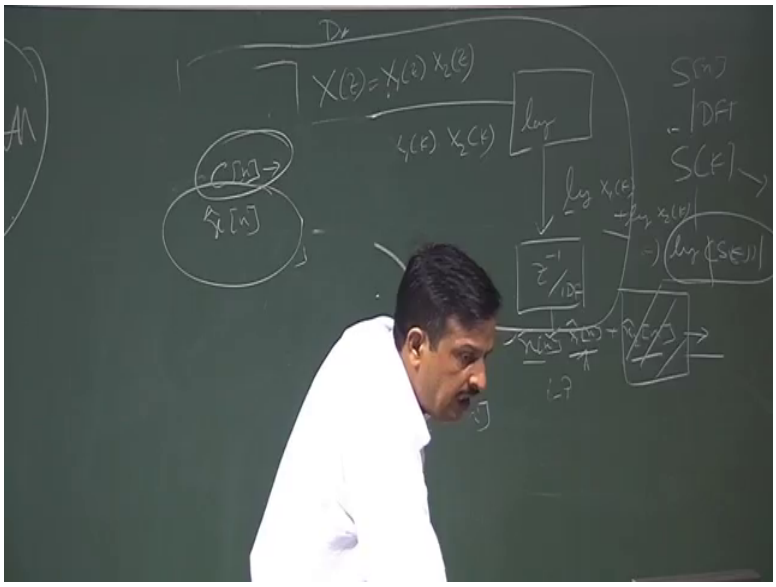
Cepstral analysis

Real cepstrum *c[n]* is the even part of $\hat{x}[n]$

$$\begin{cases} \hat{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{X}(e^{jw})e^{jwn}dw \\ \quad = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\{X(e^{jw})\}e^{jwn}dw \qquad \text{complex cepstrum} \\ c[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|X(e^{jw})|e^{jwn}dw \qquad \text{cepstrum} \end{cases}$$

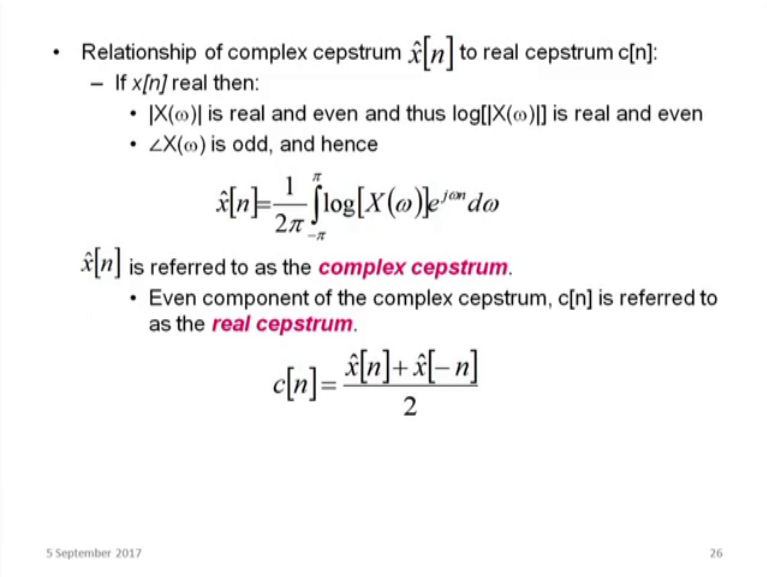
This is called D star inverse because I take the log I take the x complex exponential.

(Refer Slide Time: 19:53)



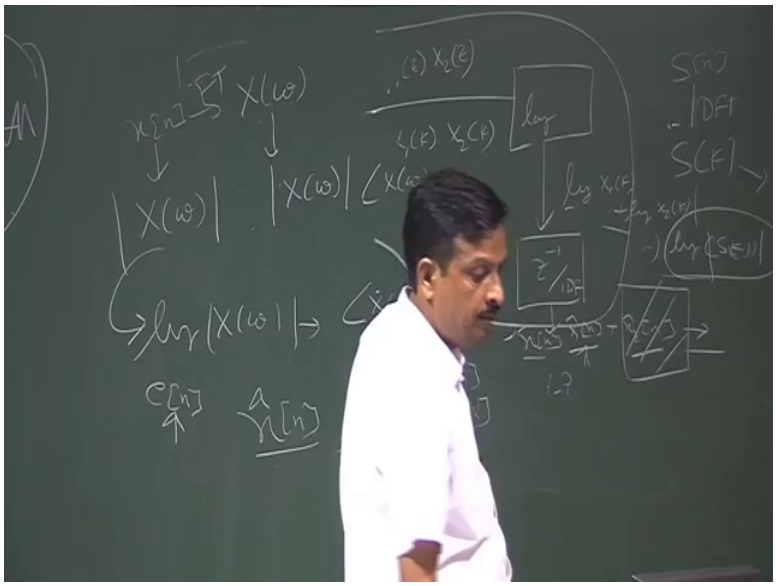
So, Cepstral analysis, there is a two kind of Cepstral one is called real cepstrum, which is represented by a c n is called real cepstrum or it can be a complex cepstrum which is x cap n. If I only take the log magnitude part then it is real cepstrum, if I take the whole signal this is a complex cepstrum. So, complex cepstrum 1 by 2 pi x cap e to the power j omega j omega n which is nothing but the log of x of e to the power j omega e to the power j omega have you understand all log of this complex.

(Refer Slide Time: 20:43)



So, this complex cepstrum consist of phase and magnitude both, but real cepstrum only the magnitude part phase part is not there. So, in the end next day or one class I will take or I can say relationship between the complex cepstrum to a real cepstrum that is if x n real if my signal is real then.

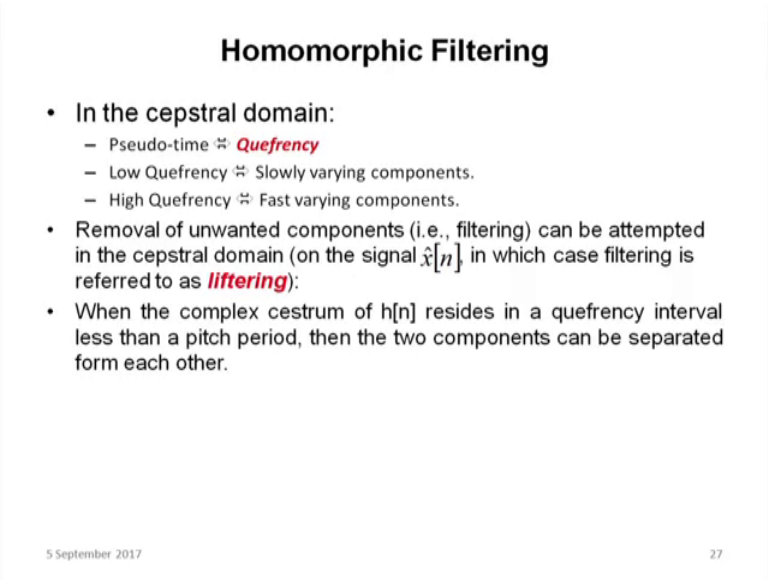
(Refer Slide Time: 21:01)



So, if lets the signal is real signal, x n is real signal, so if I say the spectrum log magnitude of the frequency equation of the signal, x of omega mod part is nothing is a real is a real part. And even thus the log of x of omega magnitude part will be also a real and even part of the signal. And if you know that, if I take the DFT of this or frequency transform F T of this I can get x omega x omega is a complex things which has a magnitude which is mod of x omega and an angle of x omega a plus j d.

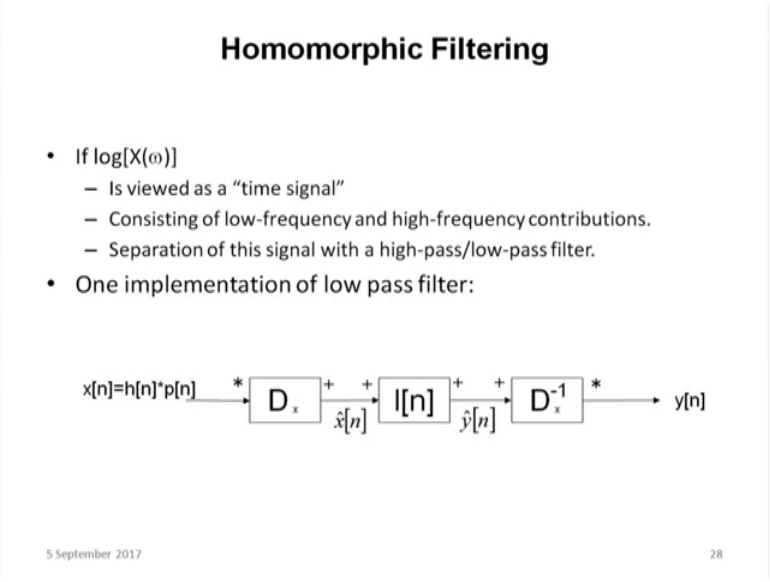
So, mod, so if I take the mod part, so mod part if it is x n is real this mod part is nothing but the real. So, if it is real then it is has a even function; and if it is even function if I take the cepstrum of take the log of this part, then what about c n, I will get cepstrum I will get this is nothing but the even part of the signal. And all the angle of that part this is nothing but a complex part. So, this is nothing but the odd part of the signal. So, if x cap n is my real cepstrum then I can say the. So, if complex cepstrum then the real cepstrum is nothing but a x cap n plus x cap minus n divided by 2, this proof I will do in the next class. So, this can be proved real cepstrum is nothing but a complex cepstrum plus x cap minus n divided by 2.

(Refer Slide Time: 23:00)



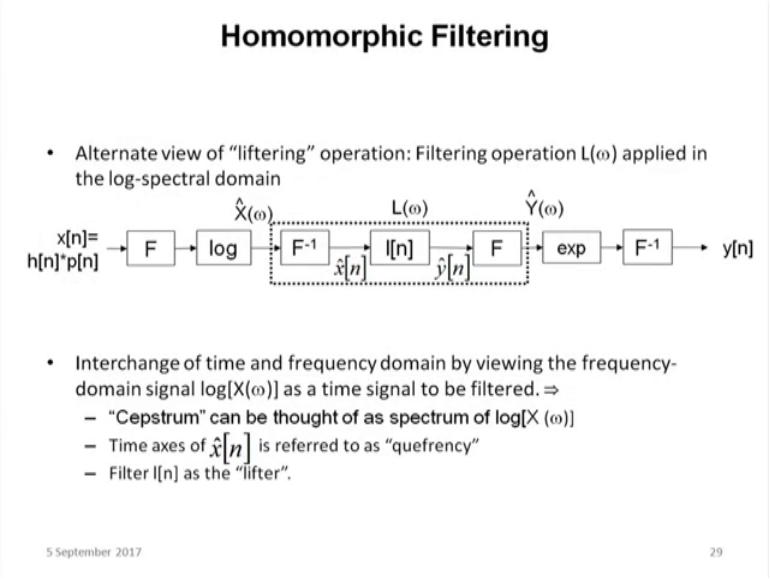
Now, homomorphic filtering what I said that these has to be passed through a filter to eliminate x 2 n. So, if I say this plane in cepstral domain, so I said this signal treated as a time signal and this signal is nothing but a frequency domain of that time signal, it is not actually frequency domain or also it not a actually time domain, so I can say it is called q frequency quefrency. So, low quefrency slowly varying component, high quefrency fats varying component. So, removal of the unwanted component filtering can be attempt in Cepstral domain it selves and that filtering calls liftering instead of filtering we say it is liftering because this is this filtering of this here this one the time domain signal then this filtering come liftering.

(Refer Slide Time: 24:07)



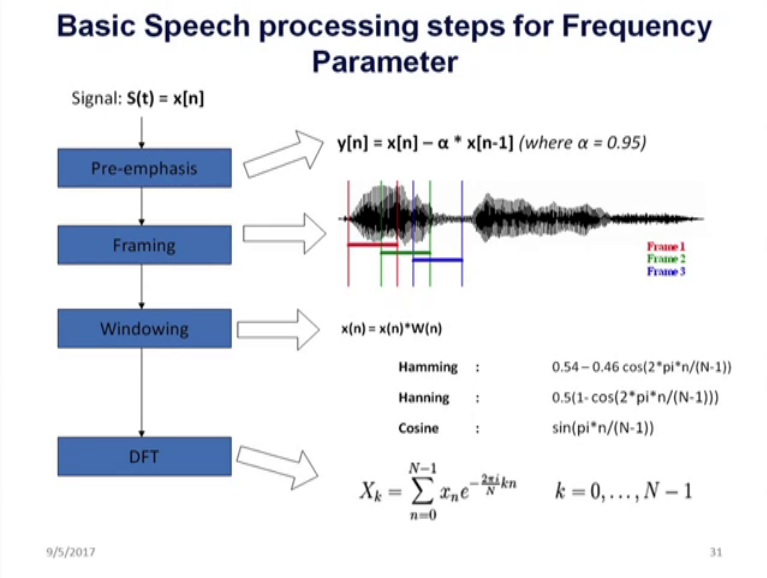
Then you can discuss the homomorphic filtering. So, if log x omega then I can pass through a. So, in view of the time signal, I can pass the low pass filter and I can find out the slowly varying component. So, I can say I am not going details of that in the filtering technique of here.

(Refer Slide Time: 24:23)



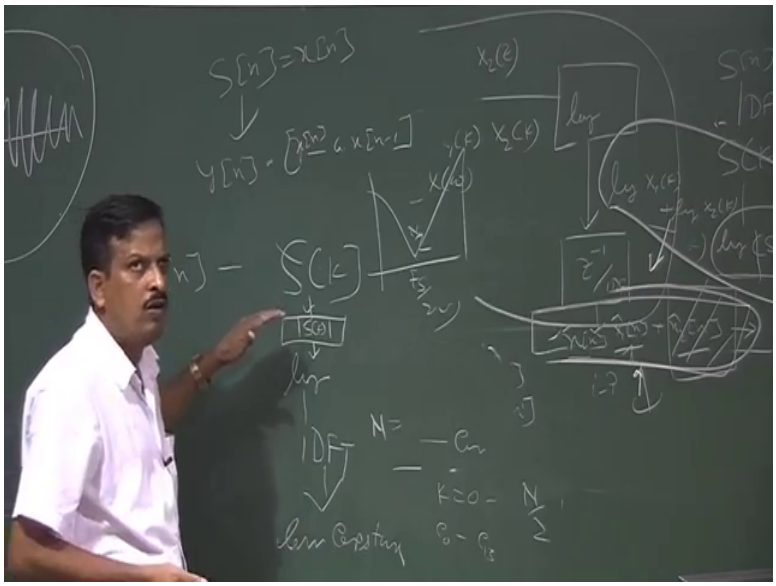
So, what are the step to find out the Cepstral coefficient. So, basic processing step for frequency parameters, I get a x n is the signal.

(Refer Slide Time: 24:39)



Then what I will do in this speech signal I do the pre emphasis to emphasize the high frequency component. Why it is required, if you see in the speech signal low frequency are emphasized, but high frequencies are not that emphasized.

(Refer Slide Time: 25:02)

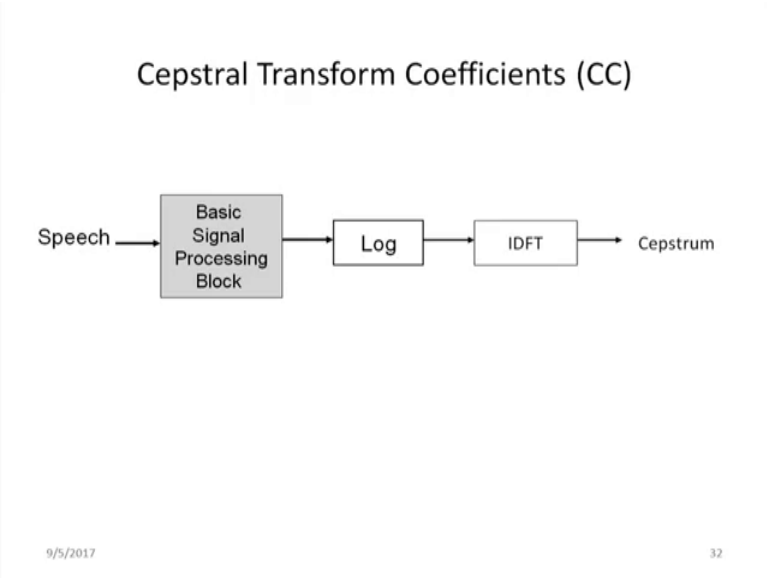


So, what I will do, first step I will get the speech signal lets s n, I do the pre emphasis. So, it is nothing but a or I can say it is s n. So, y n is nothing but a here I have s n is equal to x n I have made then I can say it is nothing but a 1 minus a into x n, x n minus a into x of n minus 1, so that is given. and a value may be or alpha value maybe 0.5, 0.96, 0.95 to

0.96 to 0.97 those are the variation. Then what I will do I do the framing what kind of framing I have a speech signal I extract I put the window length one window red color to red color one window then I shifted the window if it is 50 percent shifted if it is 20 millisecond and 10 millisecond shifting then 50 percent overlap. So, for each window for each frame I can get frame one frame two frame three, and this each frame after windowing pass through a window.

So, what I am doing I just priming the signal I am not doing the windowing there then after I cut the red to red I multiply W n. After I cut green to green, I multiply with W n which is window signal then what I will do windowing hamming, hamming cosine or rectangular if I not multiplying anything that means, I am multiplying with one which is nothing but a rectangular window then I do the DFT. So, the up to this process is convert S n to S k that conversion has to be done.

(Refer Slide Time: 27:01)



Then what we will do for power spectrum cepstrum analysis, this s k will be go to the log then I do the IDFT. If I do that, then output is call complex cepstrum. If I want the real cepstrum then what I will do instead of S k, I do another block here which is nothing but a mod of S k, then take the log. So, if I mod of S k is omitted complex cepstrum, if I take the mod then it is called real cepstrum. Then beginning, so then I can be pass through a low pass filter; that means, beginning portion is a slowly varying component. So, those

are called cepstrum coefficient or cepstral transform coefficient - CC or cepstral coefficient, you can say the cepstral coefficient.

So, if I extract the, so whatever I get here, so after IDFT, lets I take the n length IDFT. So, k will varies from 0 to n minus 1, so beginning component all the beginning component will represent the slowly varying component. So, you can say let c 0 to I can say c 13, let n is equal to something else c 13 or c 20 those actually it represent the envelop those are called cepstrum or cc- cepstral coefficient, CC - cepstral coefficient. So, basically this is done n by 2, because k equal to n minus 1 is not required because DFT analysis is symmetric property. This is F s by 2, you know that this is nothing but a n by 2. So, I can say k equal to 0 to n by 2 is sufficient. So, I can say take 0 to something some length I can take which represent the envelop cepstral coefficient. So, this is the extraction of the cepstral coefficient. So, what we have done, we are consider the spectrum has to be passed through a homomorphic signal processing or what one kind of homomorphic signal processing we have done by which we can extract the envelop of the spectrum which is called cepstrum.

Thank you. So, next class we will discuss about the MFCC.

Thank you.