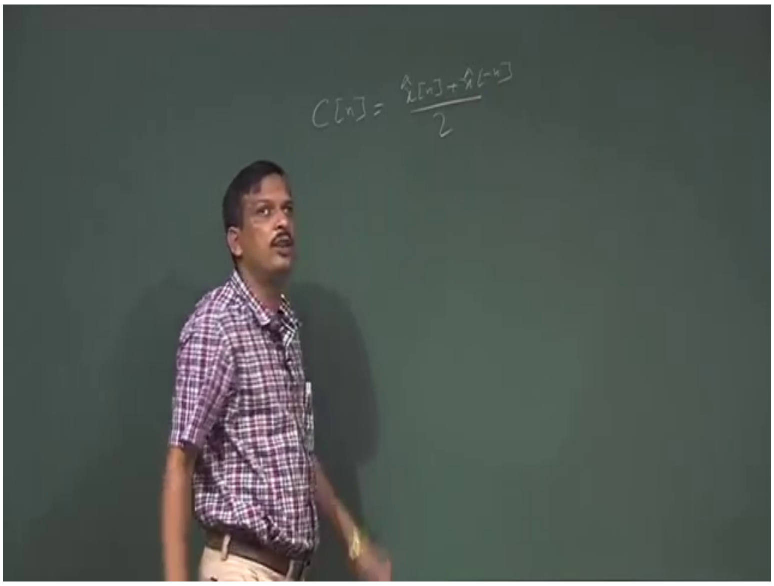
Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture – 33
Mel Frequency Cepstral Coefficients

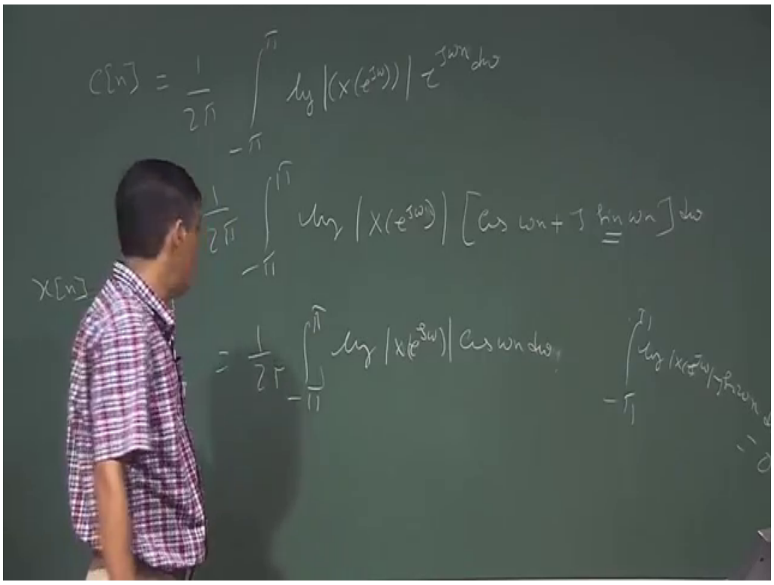
So, let us start that last class we have said that the real Cepstrum is C n, then it is nothing, but a complex Cepstrum is x n plus complex Cepstrum of minus n, then divided by 2. This we said if the x n is a complex Cepstrum and C n is the real Cepstrum.

(Refer Slide Time: 00:24)



Now we will go for the proof of it; can we prove it, yes, we can prove it. So, now, I going to prove it. So, what we will do first; think about that real Cepstrum; what is C n?

(Refer Slide Time: 00:55)



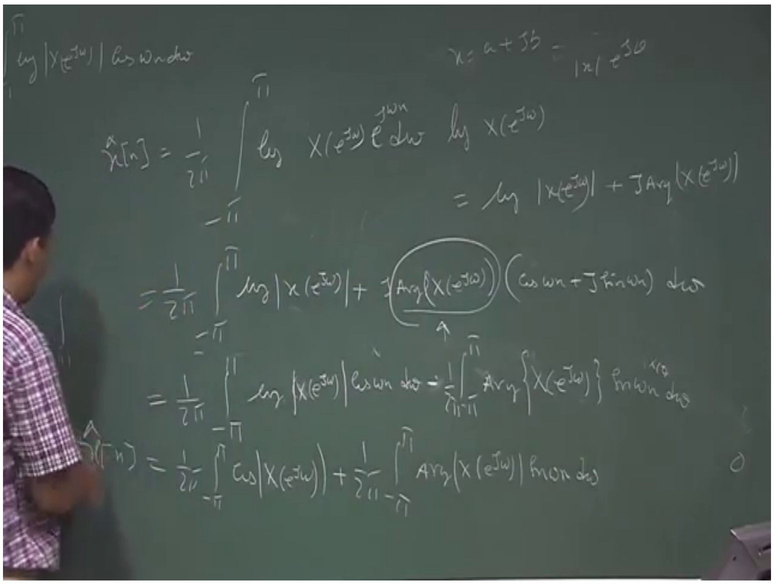
Real Cepstrum means inverse Fourier 2 pi minus pi to pi; what is the inverse Fourier of log magnitude log of x; I can say e to the power j omega mod of that log e to the power j omega m d omega or n sorry e to the power j omega n d omega this is a real Cepstrum.

If it is a real Cepstrum, then if I compute this one, what is e to the power j omega n? E to the power j omega n can be replaced by 2 pi minus pi 2 pi log of mod x e to the power j omega n e to the power j omega cos omega n plus j sin omega n; e to the power j theta cos theta plus j sin theta into d omega. Now if you see it, x n is real, if the input signal is real, then we said the real Cepstrum C n is nothing, but the even function even function.

So, if it is even function, then I can say that if this only exist cost term sin term cannot be exist; that means, 1 by 2 pi minus pi to pi log of x e to the power j omega into cos omega n d omega n d omega because log of minus pi to pi log of x G e to the power j omega mod into j sin omega n sine omega n d mega is equal to 0 because this is the even part. So, even part is 0. So, only the odd part is exist. So, C n is equal to 1 by 2 pi minus pi to pi log of e to the power j omega cos omega n d omega this is proved.

Now, find out the; so, this C n; I can write down C n in here in one corner. So, later on we can use that.

(Refer Slide Time: 03:36)



So, I in one corner I write down C n is equal to minus pi to pi log of x of e to the power j omega then mod cos omega n d omega is the C n. Now what is complex Cepstrum. Now complex Cepstrum if say this is x cap n is my complex Cepstrum, it is nothing, but a 1 by 2 pi minus pi to pi x of log. So, there is 2 part; one is you can say log x e to the power j omega d omega or not e to the power j omega n do you make.

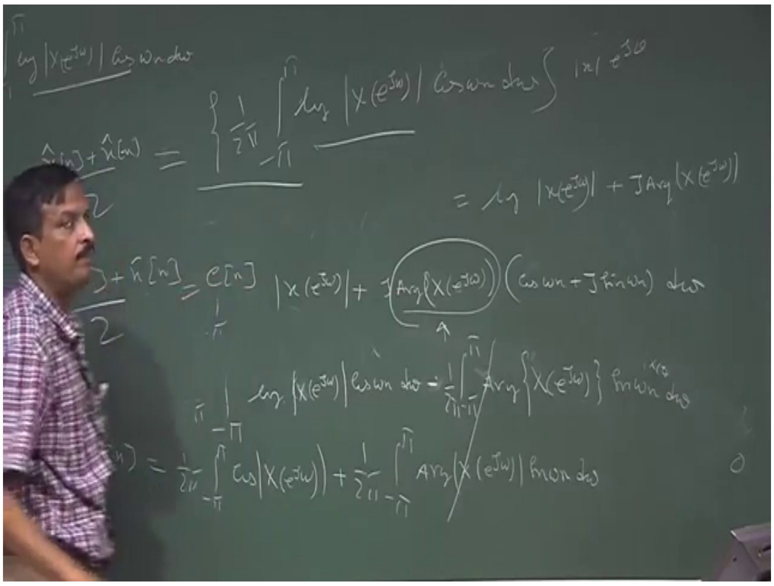
Now, what is this? This is a complex function. So, I can say log of log of x e to the power j omega is nothing, but a log of mod of log mod of x e to the power j omega plus. So, if I take the log of theta component. So, a plus j b can we express as root over of a square plus b square which is nothing suppose this is x. So, it is nothing, but a mod of x into e to power j theta. So, if we need to do a j theta then if I take the log. So, it is nothing, but a plus j arg; j arg x of e to the power j omega j omega. So, I can replace this x e to the power j omega by 1 by 2 pi minus pi to pi log of x e to the power j omega mod plus plus j arg x of e to the power j omega x arg e to the power j omega into e to the power j omega n. So, e to the power j omega again we can write cos of omega n plus j sin of omega n d omega.

Now, there are 2 term a plus b a plus let c plus d. So, both term will be there, if I multiply they are only a 4 term, again I can say the; for x n is real if x n is real, then all component minus pi to pi. So, I can that arg e to the power j omega u omega only the odd component exists only the odd function exists. So, this multiplied with cos omega n will be 0.

Similarly if this multiplied with psi omega n will be 0. So, then we once we get instead of 4 term; 2 term will be 0 if 2 term 0, then this will be 1 by 2 pi minus pi to pi log of x e to the power j omega mod cos omega n will only exist d omega plus 1 by 2 pi minus pi to pi j. So, j and j; j and j multiplied will give me minus 1. So, minus I can say r of x e to caps x e to the power j omega sin omega n d omega. So, this is xth.

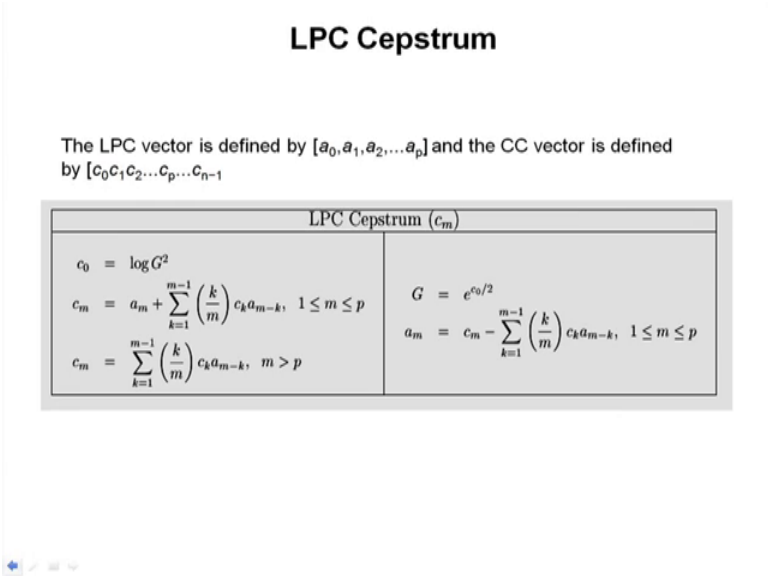
Now, what is my x of minus n, this will be only plus. So, it is nothing, but a 1 by 2 pi minus pi to pi log of x of e to the power j omega. So, this is capital X mod plus 1 by 2 pi minus pi to pi r x of e to the power j omega sin omega omega n d omega. So, if I add this x of k and x plus n. So, if I add these 2 term; what will I will get. So, if I add this; this term will be canceled, only these 2 term will be there.

(Refer Slide Time: 08:49)



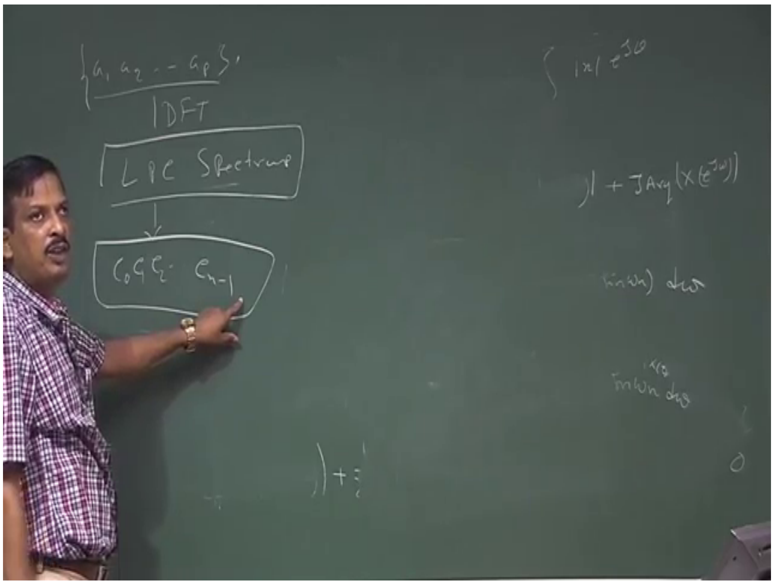
So, I can say it is nothing, but 2 into 1 by 2 pi minus pi to pi. So, I can say X cap n plus X cap minus n is nothing, but a 2 into this pi to pi log of x of e to the power j omega mod cos omega n d omega. So, if put make this by 2 is equal to this. So, if it is this, then I can this is equal to this. So, I can say X cap n plus X cap of minus n divided by 2 is equal to C n.

(Refer Slide Time: 10:03)



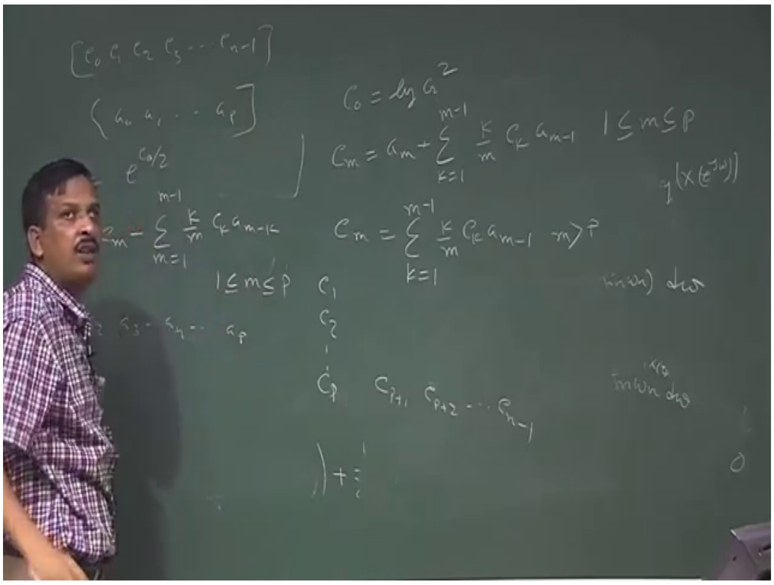
So, I can say the real spectra the real Cepstrum is nothing, but a sum of complex Cepstrum x of n and x of minus n divided by 2 is proved. So, proof is there in the slide also even refer to the slide also now come to another one which is called LPC Cepstrum which is called LPC Cepstrum. So, what is LPC Cepstrum. So, what we said that if I have a x signal and if a 1 a 2 dot, dot, a p are my LPC coefficient those are the LPC coefficient of pth order, then if I take the frequency transform of this, if I take that DFT of this what I will get I get LPC Cepstrum, I get an LPC Cepstrum which is nothing, but s I can get easily LPC Cepstrum, if I instead of depth if instead of signal if I use signal x n and take the DFT, I will get x k which is called a real spectrum signal spectrum and this is give me the LPC Cepstrum.

(Refer Slide Time: 11:07)



I can use this LPC spectrum to compute the cepstral coefficient c 1 c 2 c 0 c 1 c 2 c n minus 1, if I use the signal spectrum, then I get the cepstral coefficient C n; if it is real log magnitude, then it is real. So, real Cepstrum C n; now instead of signal spectrum, if I use LPC spectrum then it is also possible to find out the cepstral coefficient from the LPC spectrum. So, the parameter name is called LPC c LPC del cepstral coefficient LPC cepstral coefficient. So, if it is LPC cepstral coefficient how do we calculate it; it is said using this 3 equation which is given in the slides and I am writing. So, if a 1 and a p are the LPC coefficient and I want to find out the value of c 0 to c n minus 1 at the Cepstrum coefficient Cepstrum coefficient real Cepstrum coefficient.

(Refer Slide Time: 12:23)

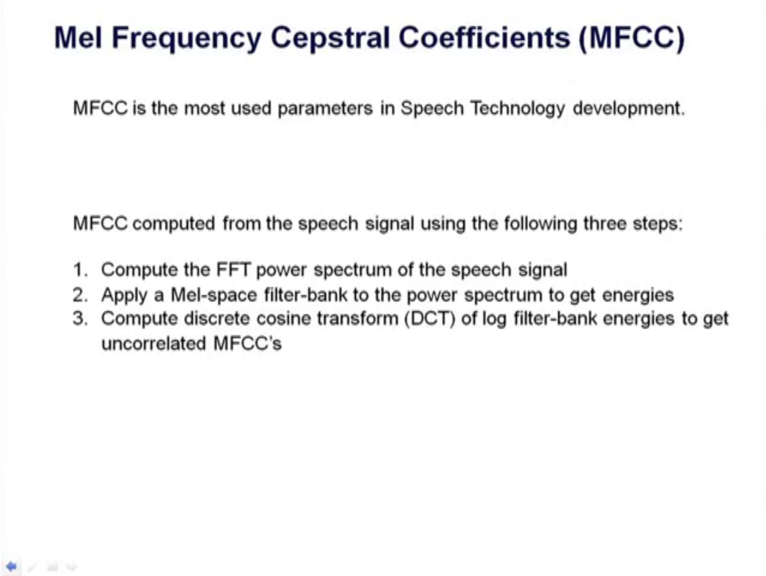


So, how do I do that we said that c 0 is nothing, but a log of G square what is G? G is called gain model gain LPC gain during the LPC analysis I have already discussed how to calculate the G for a given signal. So, if I know that G; c 0 is nothing, but a log of G square, then I can say c m is equal to a m plus k equal to 1 to m minus 1 k by m into c k a m minus 1 where 1 is equal to m. So, it varies from m varies from 1 to p, then I can get c 1 c 2 dot, dot, dot, dot, c p using this equation, then it is said c m is equal to k equal to 1 to m minus 1 k by m into c k a m minus 1 is if m is greater than equal to p; if m is greater than equal to p, then I get c p plus 1 c p plus 2 dot, dot, dot, dot, c n minus 1 using this formula.

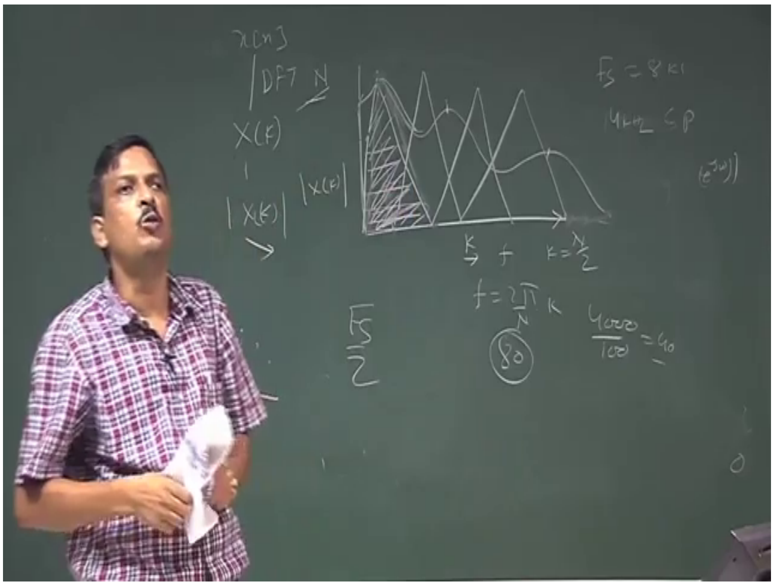
So, I get c 0 c 1 c 2 c 3 up to C n minus 1 using this similarly reverse also possible that if I know the c cepstral coefficient c 1 c 0 c 1 c 2 c 3 dot, dot, dot, dot, c n minus 1, then I can calculate the pth p order LPC coefficient. So, a 0 a 1 dot, dot, a p; how do I do that in that case gain is nothing, but a e to the power c 0 by 2, this is inverse; c 0 is equal to log G square. So, c 0 is equal to e to the power c 0 by 2, then, I can say am is equal to c m minus m equal to k 2 m minus 1 k by m c k a m minus k or 1 is equal to m. So, I can get a 1 a 2 a 3 a 4 dot, dot, dot. So, l if I calculate the cepstral coefficient of LPC spectrum using these 3 formula; I get the cepstral coefficient from LPC spectrum.

So, LPC cepstral coefficient; if I know the cepstral coefficient I can calculate the LPC coefficient using this formula. So, all are the speech parameter this cepstral is a speech parameter is a LPC coefficient is a speech parameters and l p. So, to cepstral cepstral to LPC conversion is possible.

(Refer Slide Time: 16:15)



(Refer Slide Time: 16:24)

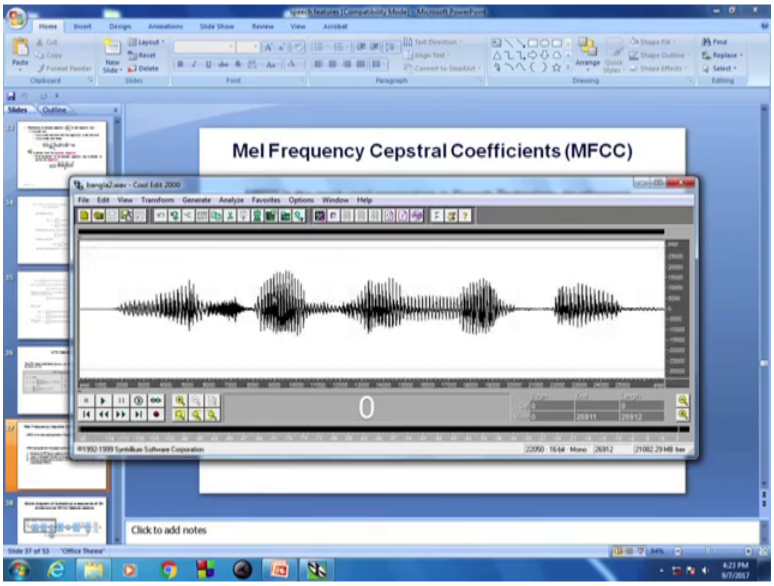


Now, we go for the next topic which is called MFCC parameter Mel frequency cepstral coefficient. So, there is a many information in our slide I am not going details of the slides I just give you the gist Mel frequency cepstral coefficient and MFCC; this is popularly known as MFCC which is the most used parameter in speech application MFCC are used in most of the cases most of the cases FCC is used. So, compute. So, even though the there is a 3 simple step in FCC compute effective power spectrum of the speech signal applied Mel filter bank and then compute DCT; why I will come why these steps we required.

So, what we said that if I take the Cepstrum; if I take the Cepstrum of a speech signal. So, how to draw the spectrum of the speech signal you know if x n is my time domain signal if I apply DFT, I will get x k I am not explaining that framing and windowing. So, that also included the speech processing trimming and we know instead of taking the whole signal at a time I have a whole signal I cut the signal I pre emphasize the whole signal then cut the signal in a frame in a window then shifted window by a frame. So, framing in doing all things I have done and after that if I take the DFT I will get x k.

Now, if I calculate mod of x k and if I plot with the frequency then I get spectrum spectrum of signal x n what is that. So, what is spectrum? So, this axis is the k axis discrete frequency k and this axis the amplitude of the amplitude mod of x k. So, this is mod of x k axis. So, k equal to 0 I get here. So, I get this kind of speech spectrum note that smooth I am drawing it smooth if we draw it real then if you say you get this kind of speech spectrum I have already explained in last class also.

(Refer Slide Time: 18:43)



So, if you see if I take a signal this; this is my signal then if I select this portion, if I analyze it then I if I can this is the spectrum. So, spectrogram plot of spectrum of the signal selected signal. So, this axis is the frequency axis this axis is the amplitude axis of the mod of x k.

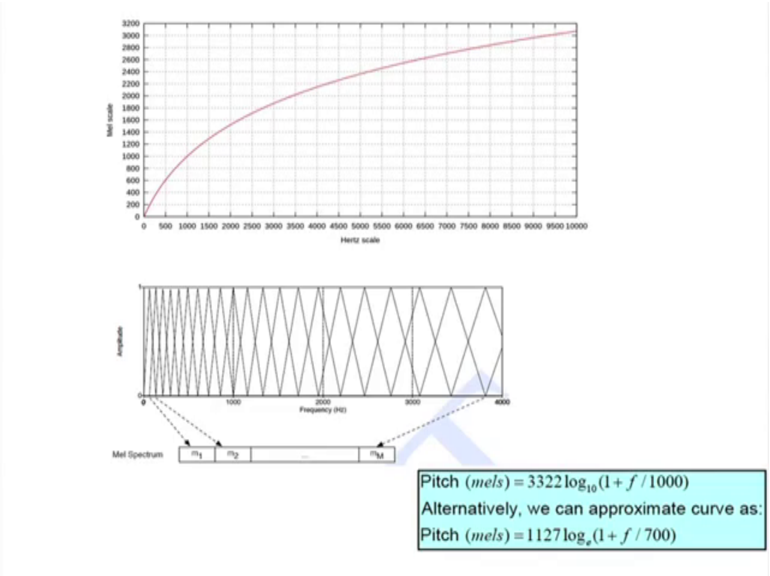
So, instead of discrete frequency, I can write f also what are the relations f is equal to 2 pi by n into k 2 pi by n is the resolution where n is the length of that DFT which I have taken now what is my interest if I want to extract the speech signal speech parameter I interest to find out this resonant frequency my interesting find out this resonant frequency and those resonance frequency actually define the speech event which is produced by the vocal tract. So, if I want to find out those resonance frequencies what I can do let us instead of taking the whole spectrum I can take a filter kind design some filter kind of things. So, this is not perfectly done. So, I can just shifted this one here. So, I can design some filter let us triangular filter with 50 percent overlap let us 50 percent overlap.

So, this kind of filter I can design let us uniform bandwidth filter I can design if I design that then I am assuring assuming that if this is the speak of my envelop, then power of this average power of this filter will be very high. So, actually I am kept try to capturing the formant frequency using spacing some filter along the spectrum. So, that can be possible. So, suppose I have a. So, what is the best band thing in the same sampling frequency is F s how much n, I should take I should take up to n by 2 because it is symmetric we have t is a symmetry. So, I can say if it is a F s, then maximum bass band frequency is F s by 2. So, if I have a depth in length is n then I can take up to n by 2 k equal to n by 2 is sufficient for design the filter.

So, suppose I have a pore eight kilo hertz F s is equal to eight kilo hertz then I know that maximum speech frequency is 4 kilo hertz. So, suppose I take 100 hertz filter that us linear all filter on 100 hertz bandwidth all filter are 100 hertz bandwidth. So, if it is not overlap how many filter is required 4 kilo divided by 100. So, 40 filter is required; now if it is 50 percent overlap, if it is 50 percent overlap, then I can easily find out how many filter is required. So, how many filter is required I can easily find out. So, now, once I know that is how many filter is required then I can say let us instead of all k design this filter. So, divided let us 40 filter, let us I have said if it is 50 percent overlap how many will be there 50 percent of overlap means instead of shifting 100 hertz I am shifting 50 hertz. So, I can say 80 filter is required let us.

So, instead of taking whole spectrum I am taking eighty point of the spectrum. So, I am taking eighty point instead of whole spectrum I am taking eighty point of the spectrum and find out the cepstral coefficient I can treat that is a signal pass through the inverse DFT and cepstral coefficient will be generated, cepstral coefficient can be generated. So, information is reduced. Now the problem is that if it is a linear filter this does not matched with human perception human frequency perception if you remember in perception during my lecture on human perception of frequency and amplify speech perception.

(Refer Slide Time: 23:23)



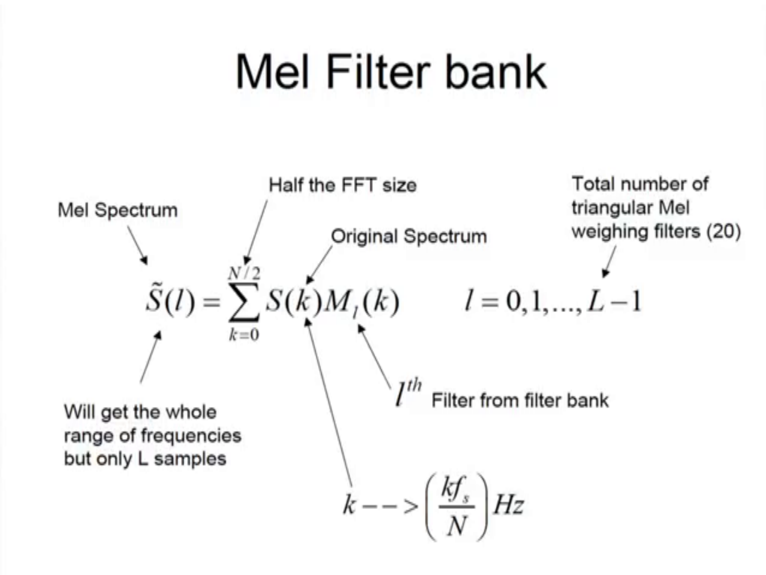
We said the signal the frequency perceived by human being is not in linear scale this is in Mel scale. So, instead of taking the linear filter now I convert the filter bandwidth as per the Mel scale what is the Mel scale you know this using this equation we have already derived the Mel scale. So, the bandwidth of the filter is depends on the Mel scale. So, instead of uniform bandwidth filter, I take Mel scale filter.

So, what is human perception in frequency lower frequency our resolution is very high; that means, lower frequency we can perceive linearly along the physical frequency, but at the higher frequency we have a some low resolution; that means, the band of frequency the frequency band we perceive as a same frequency is larger. So, I can say at the high frequency region the bandwidth of the filter will be large so; that means, who do not require that much of course, resolution. So, half estimation is sufficient for high frequency.

So, that is why we take the bandwidth of the filter will be larger. So, bandwidth defined by the Mel scale. So, if I take the Mel scale filter then let us I take the Mel scale filter if you find out the dividends 4 kilo hertz I think 20 filter will be sufficient to cover the 4 kilohertz frequency or entire frequency range and. So, I can get m 1 m 2 up to m 20 because every filter give me a single point bandwidth single bandwidth. So, every filter has a single bandwidth which is m 1, m 2, m 3, m 4, I can find out 20 Mel point.

So, since filters are Mel scale filter are designed in Mel scale then we can say we have locked the l p or the frequency spectra in Mel scale instead of hertz I want this spectra in Mel scale. So, what I get instead of hertz I get Mel scale in here and here is let X cap k instead of x k, I take X cap I average I take the average energy. So, I instead of may spectrum I get Mel scale spectrum once I get the Mel scale spectrum if I analyze cepstral coefficient, then this is called Mel scale cepstral coefficient. So, since my spectrum is frequency warped in Mel scale that is why it is called Mel frequency cepstral coefficient. So, what I am actually doing any mathematics I am designing the filters I design I will explain in the next class first I designed and described the equation.

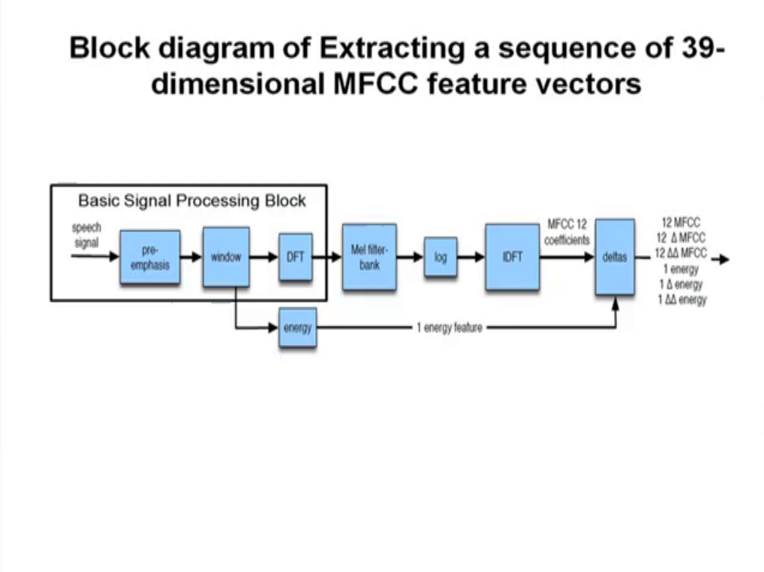
(Refer Slide Time: 26:57)



So, I calculating X cap L; L is the number of filter S cap L using Mel filter bank over the Mel filter, I am calculating Mel spectrum and once cepstral coefficient is analyzed based on the Mel shaped spectrum then is called Mel frequency cepstral coefficient a MFCC have you understand a MFCC that is why it is called MFCC Mel frequency cepstral coefficient.

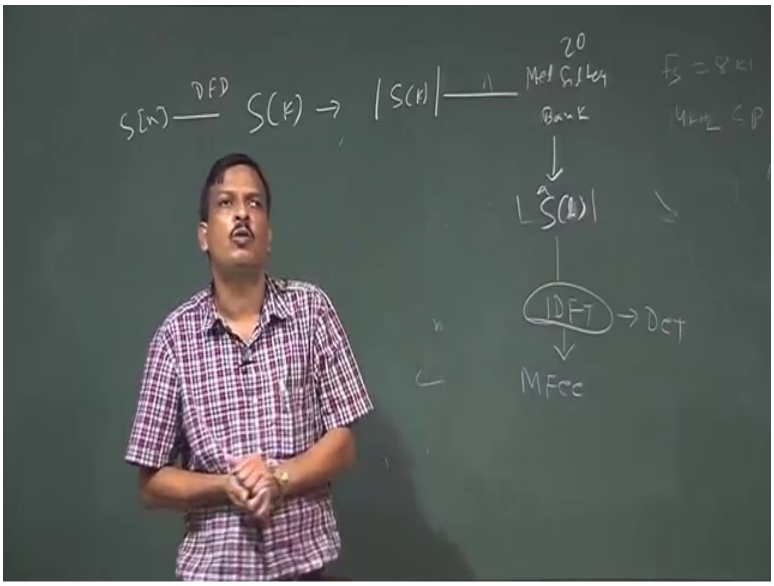
Now, how do you design those things? So, what should have or what should be the block diagram. So, I am I am describing the complete back walk there on Mel frequency.

(Refer Slide Time: 27:50)



So, if you see the basic signal processing block it remains same pre emphasis window DFT, after I get DFT what I can do after.

(Refer Slide Time: 28:04)



So, if it is S n is my signal, I pre emphasized it and after DFT let us I am not doing the; I am not writing the pre emphasizing on windowing part. So, after DFT I get S k. So, S k is the normal frequency cepstral; so, then from S k, if I calculate mod of S k and plot with frequency. So, mod of s k is nothing, but a frequency versus power plot which is the spectrum in normal frequency scale; now I want the frequency scale in Mel scale Mel frequency or you can say the Mel filter bank I pass that with the Mel filter bank when I pass them with the Mel filter bank what I will get; I get some spectrum, but in Mel scale instead of frequency I get S cap k which is in Mel scale or instead of k; I can here, right, l is the number of analysis filter. So, if it is a 40 filter or if it is a 20 filter. So, l varies from 0 to 19. So, I can get number of filters.

So, once I pass through the Mel filter I get Mel spectrum once I get the Mel spectrum I can pass through that IDFT to get the Mel Cepstrum OMF; I can write MFCC. So, those delta double del come later on. So, I get MFCC; MRL frequency cepstral coefficient. So, the mathematics is this is the mathematics and instead of DFT what I will do? I will pass instead of DFT this DFT; IDFT can be replaced by a DCT discrete cosine transform IDFT can be replaced by a DCT. So, this DCT equation is this. So, S cap m or l whatever cos pi by l m minus 0.5; I equal to 0 to c minus 1 number of cepstral coefficient is or not.

So, next class I will discuss how we implement it or I can say that let more details discuss on implementation issue on MFCC, then we discuss about the delta double delta, then PLP and then Rasta, then the features extraction will be completed we will go for the F G extraction.

Thank you.