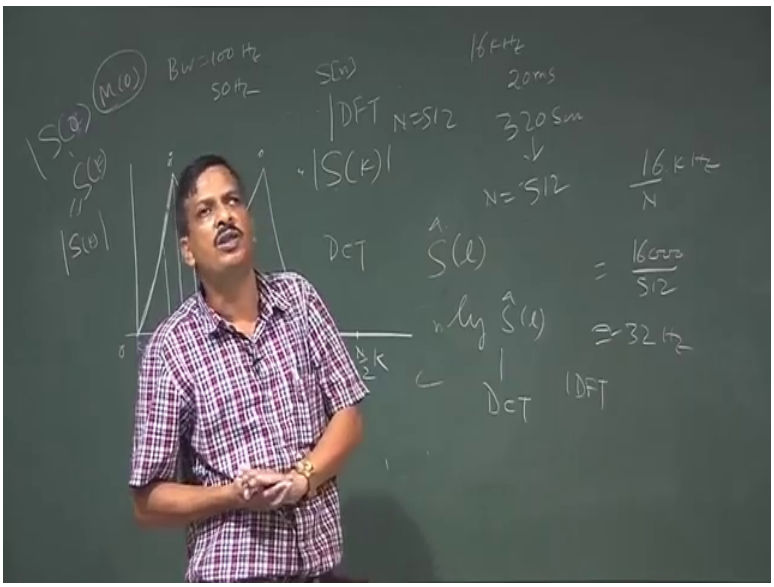
Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

Lecture – 34
MFCC Features Vector

So what we said is that if the cepstral coefficient is extracted from Mel cepstrum, then we can get the MFCC. So, how do you do that? Just little bit of detailed discussion how implement it.

(Refer Slide Time: 00:35)



So, you know that if I have a S n; let us I have signal speech signal which is samples as 16 kilo hertz, a speech signal is sampled at 16 kilo hertz. And then I take a window let us 20 millisecond. So, how many samples I should get? 20 millisecond 320 sample. So, what should be the DFT site 256? So, N is equal to length of the DFTs sorry, 512, 512 point DFT I have applied. So, what is the resolution? 16 divided by N 16 k kilo hertz divided by n. So, I say 16 thousand divided by N is 512 almost I can say 32 hertz. 32 to 33 hertz ok I am satisfied. The resolution is 32 hertz.

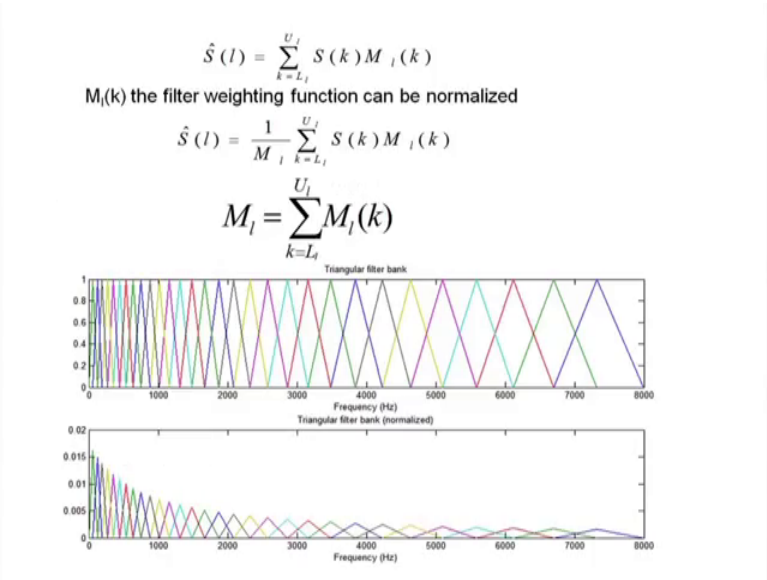
So, if I calculate if take the DFT, N is equal to 512. Then I can calculate I get S k, once I get S k I can take the mod of S k which is nothing but a real square plus imaginary square, I can take it. Once I take it and then I want to plot it. So, this axis is my k axis this axis is my mod of S k. Now if I take it, then what I will do? I will that I just S k I

have is here. So, what I get this is the k equal to 0 and then next k equal to 1. So, k equal to once means 32 hertz k equal to 2 means 64 hertz. So, that that is why it will going on going on upto N by 2 ok.

Now, I am calculating them I want to implement the Mel filter. So, based on the Mel scale let us design the initial filter bandwidth is 100 hertz, let us design 100 hertz and center frequency is 50 hertz. So, I can say this is 50 hertz, this is 50 hertz and this is 100 hertz. So, if it is 100 hertz how many k will be there? So, k equal to 64 k equal to c 96 hertz. So, k equal to 3 I can say almost k equal to 3. So, what is the power? So, k equal to 0 multiplied by the filter multiplied by the filter wave function. So, Mel k is the filter weight function. So, if it is if this is my this is the point fast filter I am designing. So, that required k equal to 0 to k equal to 3, so the value of S k multiplied by m k. So, m k means k equal to 0, S k is k equal to 0, mod of S k I can calculate the mod of S k. So, mod of S k I can get it. So, this capital S k here in this equation in the slide he represent the mod of S k original spectrum. So, if it is a original spectrum this is nothing but a capital S k I can write down.

So, original spectrum is multiplied by the m 0, what is the m 0 coefficient? If 0 then I can find out k equal to 1 then k equal to 2 and then k equal to 3. So, all those 3 will be multiplied by the m function. So, m is called filter weight function. So, that filter weight function multiplied by the individual power, then I take the sum I get the point this point. Next I design another filter with 50 percent overlap. Next I design another filter. So, that way I can design the filter.

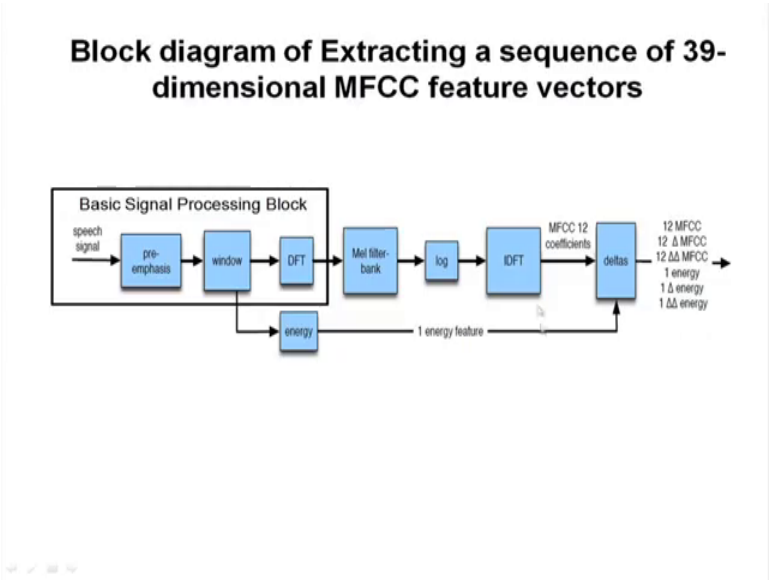
(Refer Slide Time: 04:52)



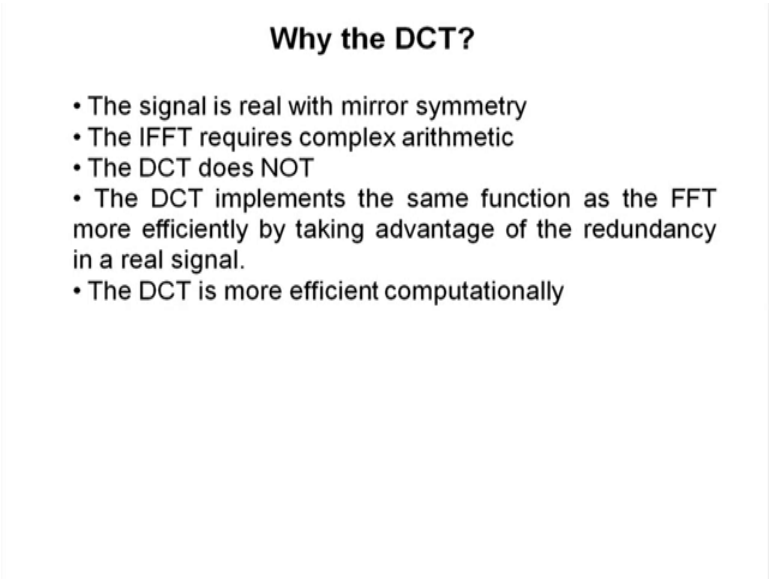
So, equation becomes. So, S k is the spectrum then ml k is the filter weight function multiply and k equal to l L l to ul L l is the k equal to 0, and upper frequency is k equal to 3. Similarly what the next filter of it is 50 hertz k equal to 2, almost k equal to 2 or k equal to 1. If it is k equal to 1, then k equal to 1 2 1 2 3 4 5 I can calculate that k. So, you know k u L k or L l is equal to 0 for the first filter and ul is equal to 3 for the first filter then second filter I can find out L l may be equal to 1, and I can find out ul equal to 4. That way I can design the filter and find out that ss cap l S cap l which is Mel scale. Mel scale spectrum.

Then I can take the log of S cap l because cepstral coefficient has to be calculated using homomorphic decomposition which require the log. So, log of sl and then I take the DCT instead of I DFT instead of IDFT, I take the DCT, so complete block diagram MFCC MFCC equations. So, I this block diagram contain IDFT IDFT can be easily replaced by DCT discrete cosine transform ok.

(Refer Slide Time: 06:22)



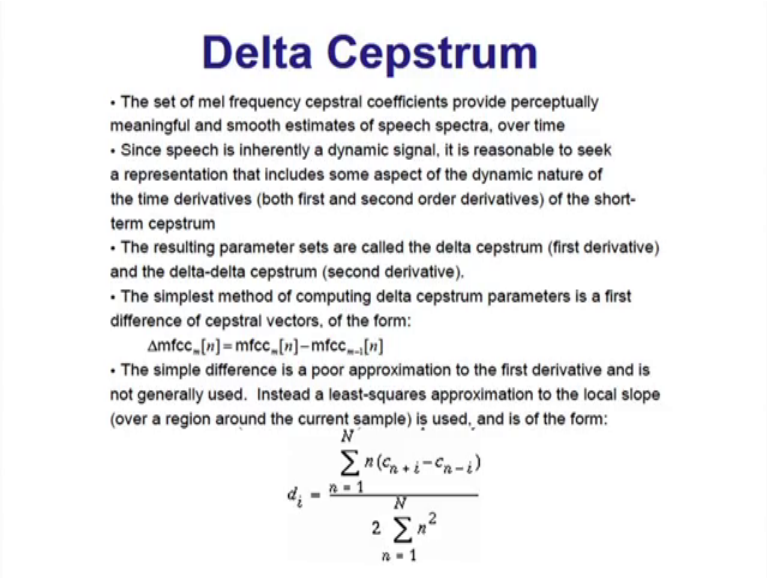
(Refer Slide Time: 06:35)



Why we use DCT? This is the example this is the reason the DCT is required to disseminate or differentiate between the uncorrelated the cepstral coefficient extracted cepstral coefficient.

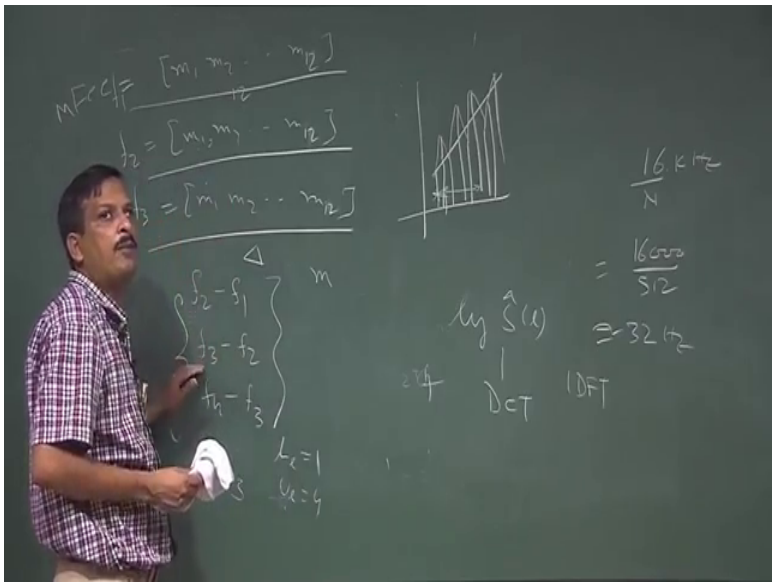
So, I use DCT to uncorrelate this filter out. That is why I use DCT and DCT acts like an FFT. Here I have use DCT, because this is a complex complex arithmetic is required. So, for DCT implementation is easy in this stage that is why instead of FFT we use DCT, discrete cosine task.

(Refer Slide Time: 07:23)



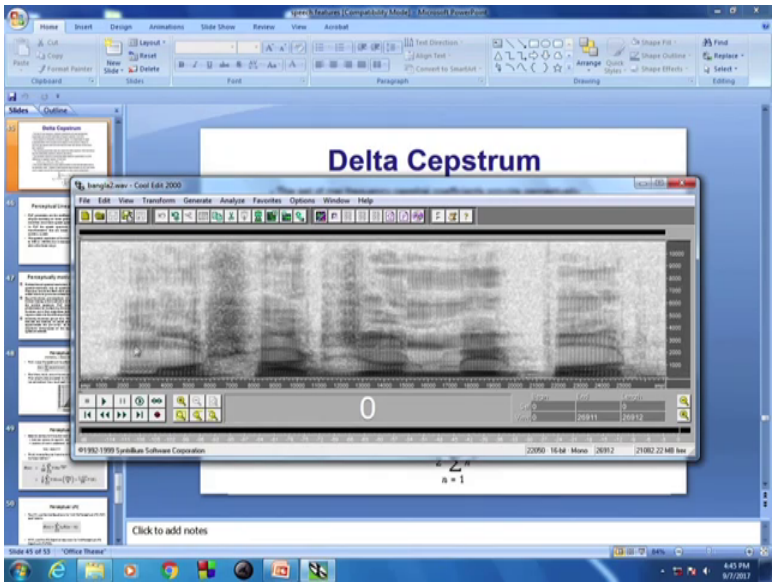
Now, I go for the delta cepstrum. So, suppose I calculate MFCC, MFCC coefficient m equal to 1 2 let us calculate 20 MFCC coefficient m 20. So, this is the MFCC, MFCC coefficient for single frame.

(Refer Slide Time: 07:32)



I can calculate for next this is for f 1 frame 1, I can calculate for frame 2, m 1 m 2 dot, dot m 20. I can calculate f 3 frame 3 m 1 m 2 dot dot m 20. So, it is said that out of m 20 if I take 12 number is sufficient 12 to 13 number is sufficient. So, instead of 20 I can write m 13, I can take. Let us I take m 13 or 12, let us take 12, 12, 12, 12.

(Refer Slide Time: 08:44)



MFCC a frame 1 frame 2 frame 3. Now if you remember if you or if you see this cepstrum of this signal, I can calculate if you see that I can just showing that spectrogram of his if you see the spectrogram of this signal. If you see look at this curser. This is the second formant. So, if you see the second formant are moving, even first formant is also moving. So, I can see there is a formant movement or I can say there is a dynamic of the parameter changing across the segment. So, that dynamics also can be a good features. So, here if I want to kept that those dynamics. So, this is represent the frame 1, this is represent the frame 2 this is represent the frame 3. So, what is the dynamics? Dynamics difference between the frame 1 and frame or frame 2 and frame 1 and difference between the frame 3 and frame 2 is the dynamics. So, if I say first order dynamics then I can say f 2 minus f 1. Frame 1 MFCC parameter minus frame 2 MFCC parameter ok.

So, next f 3 minus f 2, then f 4 minus f 3, so this is a first order dynamics first order dynamics. If I want to calculate the second order dynamics, so deference between the first order dynamics is nothing but a second order dynamics. So, first order dynamics is called delta coefficient. So, old delta is nothing but a difference of difference. So, I can say double delta coefficient f 3 minus f 2 minus f 2 plus f 1. So, it is nothing but a f 3 minus twice f 2 plus f 1. So, I can calculate the first order differentiation, then I can comes repeat the same process to get the double delta coefficient and input is this one. So, delta coefficient is the first order dynamics. So, if I say the formant is moving. So,

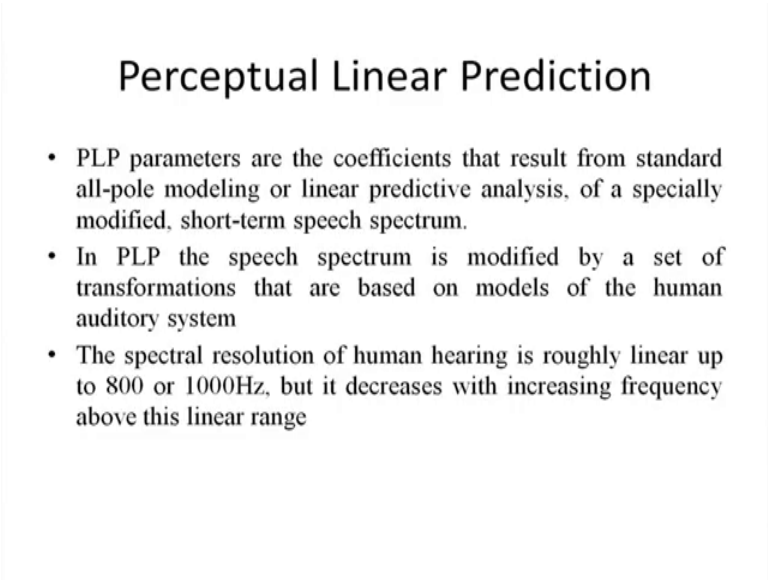
first order dynamics is nothing but a velocity of the formant. If I take the second order dynamics that is give me the acceleration of the formant.

So, how formant are moving velocity, how quickly formant are moving acceleration. So, delta double delta can be a speech parameter. Now if you calculate just differentiate that frame 2 minus frame 1 and frame 3 minus frame 2, suppose you have a signal which is randomly vary which variation is very high. Let this one, this one, this one, and this one then this one, then this one, then this one. So, if you see the variation is vary random if I want smooth variation what I should do? So, instead of taking the just difference between the 2 consecutive sample, I can take the average difference of few sample, that can be smoothly. So, instead of taking these difference I will take I will take calculate the delta coefficient, or double delta coefficient across the pure frame instead of single frame. So, suppose if on the first one instead of taking m 1 of frame 2 minus m 1 of frame 1 I can say that few frame m 1 and find out the delta coefficient.

So, this is the equation. To get the smooth movement of the format. So, delta coefficient is just difference between the frame. It is which takes several frame and calculate the average difference, and then move the filter like a smoothing function. So, that I can get smooth variation of the formant frequency so that is called delta coefficient and if I take the delta, delta the one delta coefficient.

So now if I if you remember in the first slides, if I make the MFCC vector features vectors. So, what I said? I take 12 MFCC parameter. Then I can take 12 delta parameter delta MFCC. Then I can take 12 double delta MFCC. Now of you see if you remember this must this last discussion, when I put the first filter I multiply k 0 k 0 th coefficient of this cepstrum, if the 0 weight function is 0. So, I have not taken the energy of the signal. So, what I do I will take the energy of the signal which can directly compute computed after windowing the energy, average energy of the signal I can directly compute and take energy first energy of the signal. Once I take the energy I can take the delta energy I can take the double delta N energy. So, that give me 12, 12, 12 - 1, 1, 1, so 36 plus 3. 39 dimensional feature vector. So, MFCC feature vector with a 39 dimension. 39 dimensional feature vector. So, this is called MFCC feature vector.

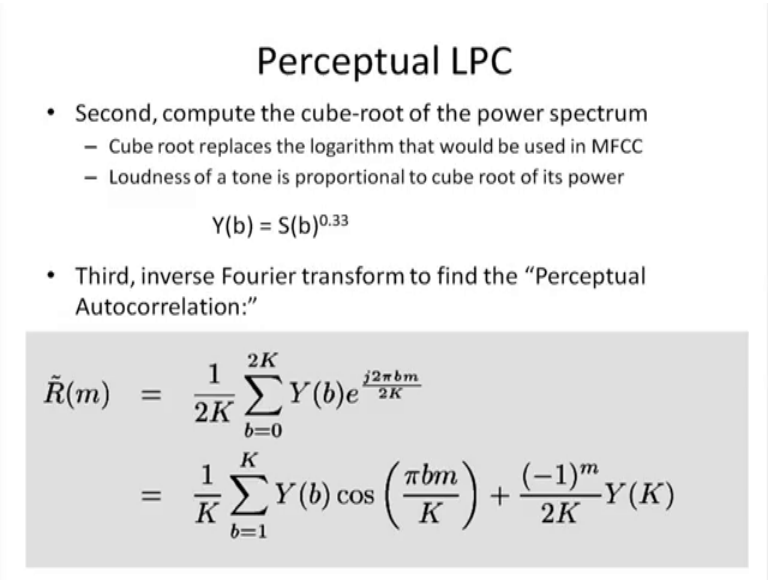
(Refer Slide Time: 14:48)



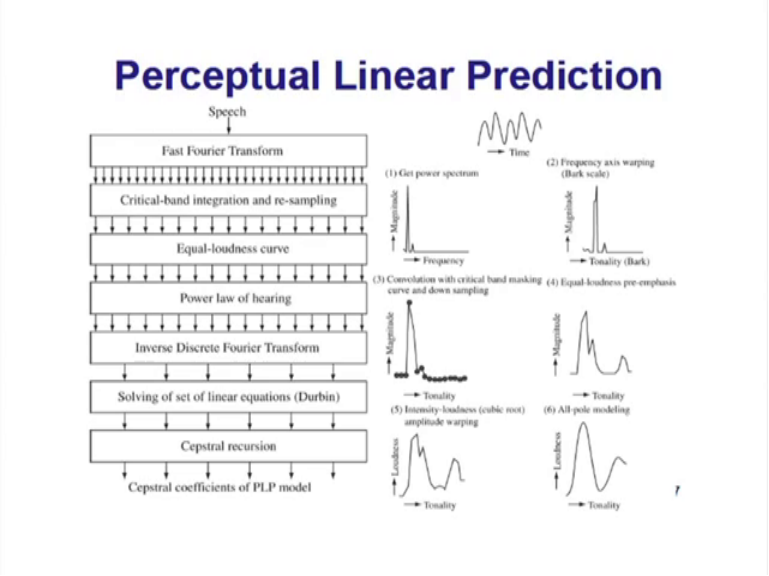
Next is PLP very much perceptual PLP perceptual linear prediction, PLP parameter PLP perceptual linear prediction. So, what is that part? Linear prediction we know lp linear prediction you know what is linear prediction you know. So, what is perceptual linear prediction, so already said that if I modify my signal as per the human perception. So, if you see the cepstrum of the signal is not at the scale of human perception. So, I modify the spectrum of the signal as per the human perception, behavior human has a 2 perception of frequency and perception of loudness. We have not percept equally loud all frequency because the LPC coefficient analysis is amplitude dependent it time domain analysis.

So, across the frequency if I want to normalized it. So, I normalize the frequency parameters with respect to human perception. If you see the block diagram I am not detail discussing the detailed slides are there.

(Refer Slide Time: 16:14)



(Refer Slide Time: 16:17)

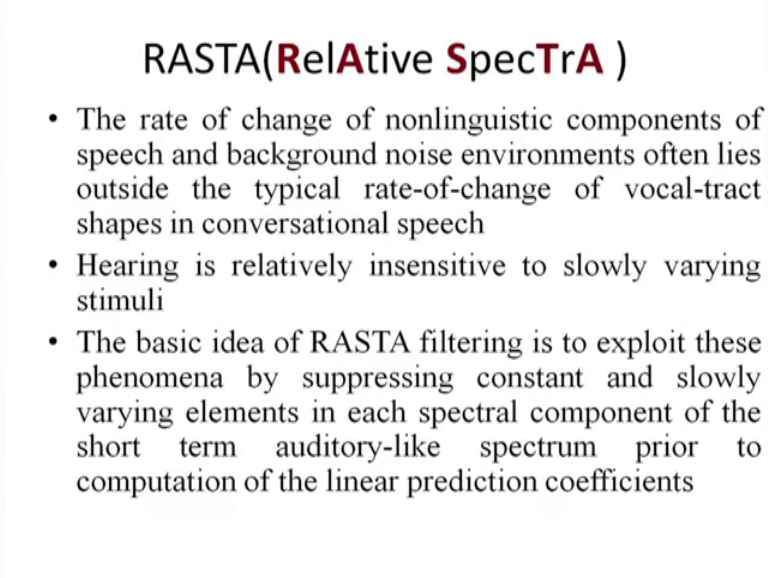


So, what is said take the speech signal find out the first Fourier transform then calculate the cepstrum. And modify the cepstrum as per the critical brand or Mel frequency cepstral cepstrum or bark scale, then modify the equal the modify the loudness. Loudness of the amplitude of the cepstrum as per the equal loudness curve. Then apply the power law of hearing what is power law. If you remember that relations between the loudness and intensity, I think it is 4 4 5 into I to the power 0.333.

So, what it is that? It is a cube root of intensity. So, intensity and loudness relationship is the cube root of intensity, this is constant, this is constant. So, modify the loudness using cube root of intensity. Then take the inverse Fourier transform or you can say the take the inverse Fourier transform, or then what I will get I will get the cepstral coefficient of the cepstral coefficient I can calculate the a 1 a 2 a 3 a 4, which I have already discussed, that is called TLP, coefficient perceptual linear prediction.

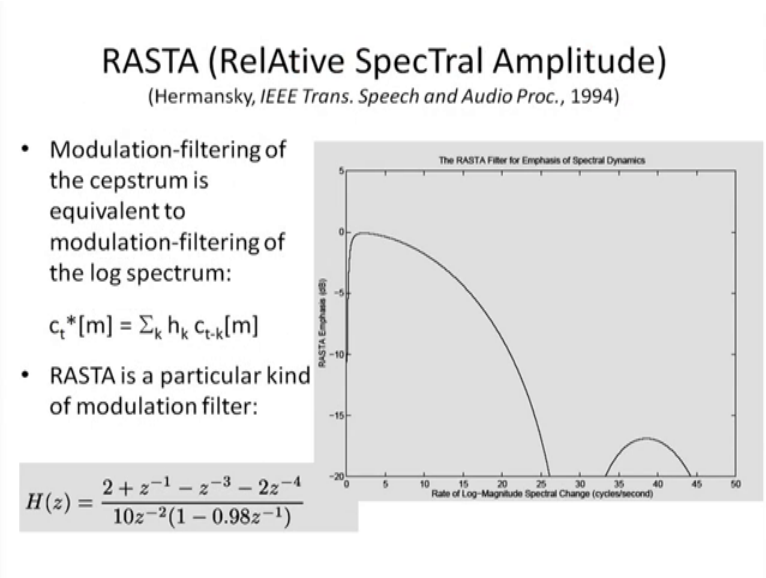
Then there is another speech parameter which is called RASTA relative spectra.

(Refer Slide Time: 18:00)



Relative spectra, So the rate of change of nonlinguistic component of a speech and background, noise environment often lie outside the typical rate of change of vocal track shapes in. So, it is one kind of modification of spectrum and then find out the parameter.

(Refer Slide Time: 18:25)



So, RASTA also there is a 1 another side. So, it is a modulate, relative spectral amplitude modulation filter of spectrum is equivalent to modulation filter of log spectrum.

So, those are the modification of the spectrum frequency parameter. So, this is the all about that parameter extraction the speech signal.

Thank you.