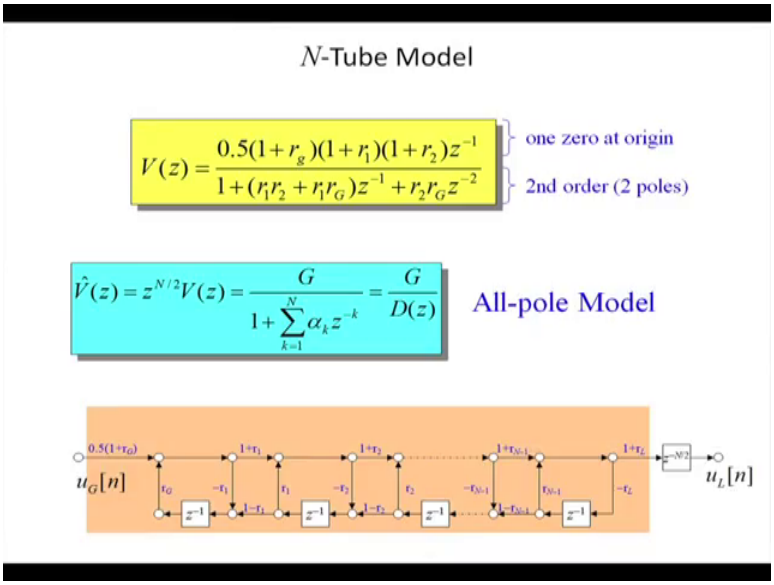
Digital Speech Processing
Prof. S. K. Das Mandal
Centre for Educational Technology
Indian Institute of Technology, Kharagpur

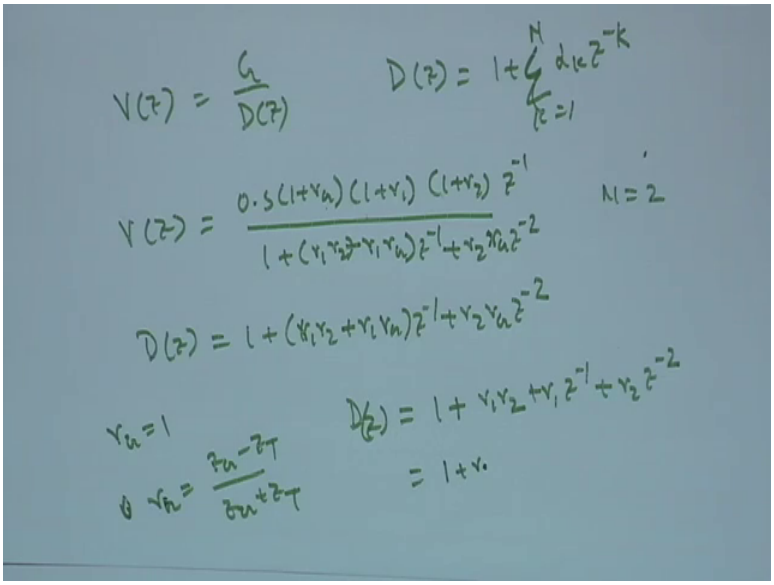
Lecture - 15
Uniform Tube Modeling Of Speech Processing Part – VII

So we have derived that the all pole model for vocal track. So, V z is nothing but a; I can assign a G by dz.

(Refer Slide Time: 00:25)



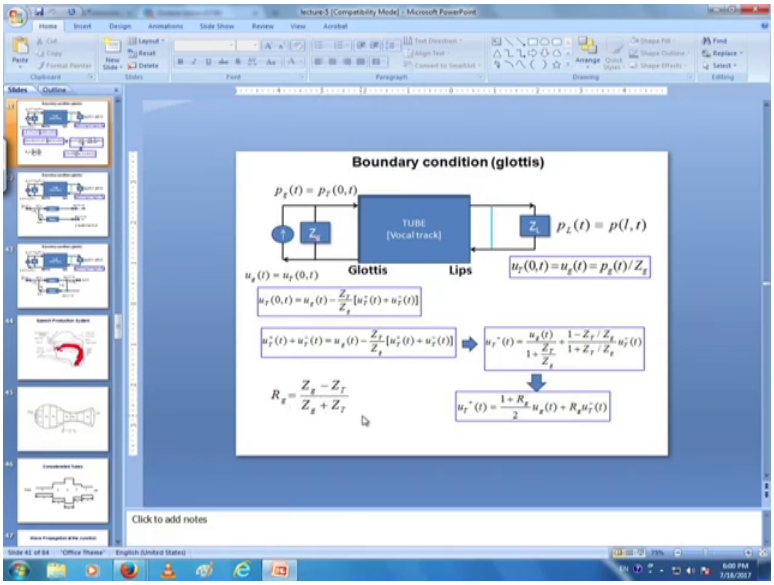
(Refer Slide Time: 00:26)



So, where dz is nothing but a 1 plus k sorry, summation of k equal to 1 to N alpha k z to the power minus k now if you see that d that second the for the 2 tube model, the equation vz is nothing but a 0.5 1 plus r G 1 plus r 1. So, for 2 tube model vz is N equal to 2 N equal to 2, is 2 tube model 0.5 1 plus r G 1 plus r 1 1 plus r 2 into z to the power minus 1 divided by, I can say 1 plus r 1 r 2 r 1 r 2 r 1 r G into z to the power minus 1 plus r 2 r G z to the power minus 2. So, dz for 2 tube model is nothing but a 1 plus r 1 r 2 plus r 1 r G z to the power minus 1 plus r 2 r G z to the power minus 2. Now if I say r G is equal to 1 if I say r G is equal to 1 what all impedance is infinite r G is equal to infinite then r G equal to 1. So, r G is nothing but a you know that r G is r rt minus r G divided by rt plus r G or whatever r G minus rt or rt minus r G; r G minus r t.

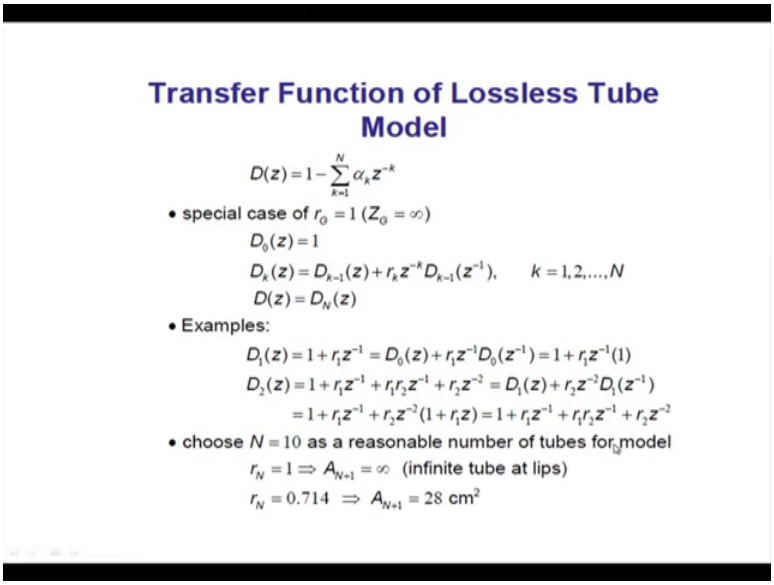
So, to that that is there in constant when when we model in that boundary condition. When you model in boundary condition we said r G boundary condition modeling where here. So, here I can say r G is nothing but a r G minus zd divided by r G plus.

(Refer Slide Time: 02:26)



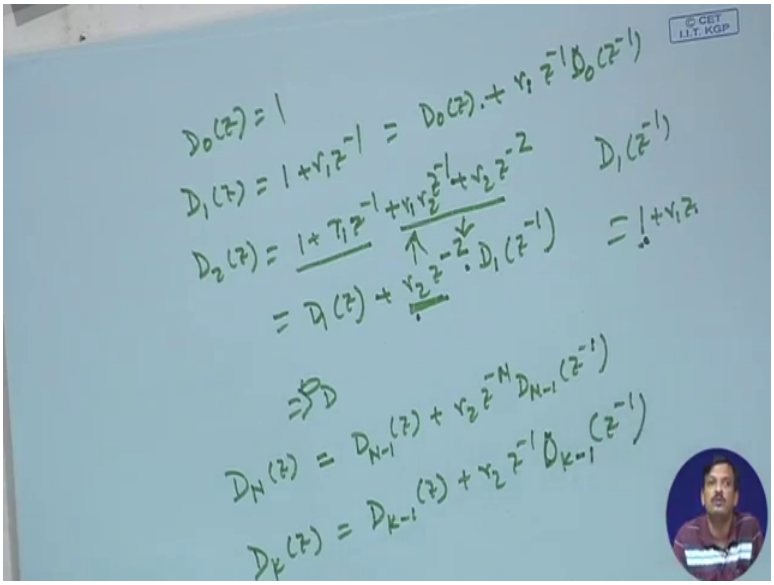
So, r G is nothing but a here I write capital r G let us consider small r G rg is r G minus zd divided by r G plus zd r G is infinite then r G is equal to 1. Now if I consider r G is equal to 1 then dz is nothing but a dz is nothing but a 1 plus r 1 r 2 plus r 1 plus z to the power minus 1 plus r 2 z to the power minus 2. So, if it is nth order I if I can write that if it is instead of 2 tube it is N tube or I can say it is nothing but a 1 plus r 1 1 plus r 1 r 2 r 1 z to the power minus 1 plus r 2 z to the power minus 2. So, if I say that I can want to write down the recursive equation. So, D 1 z If it is d 2 z.

(Refer Slide Time: 03:26)



So, D 1 z if I write D 1 z is nothing but A 1 plus r 1 z to the power minus 1 if r G is equal to 1.

(Refer Slide Time: 03:34)

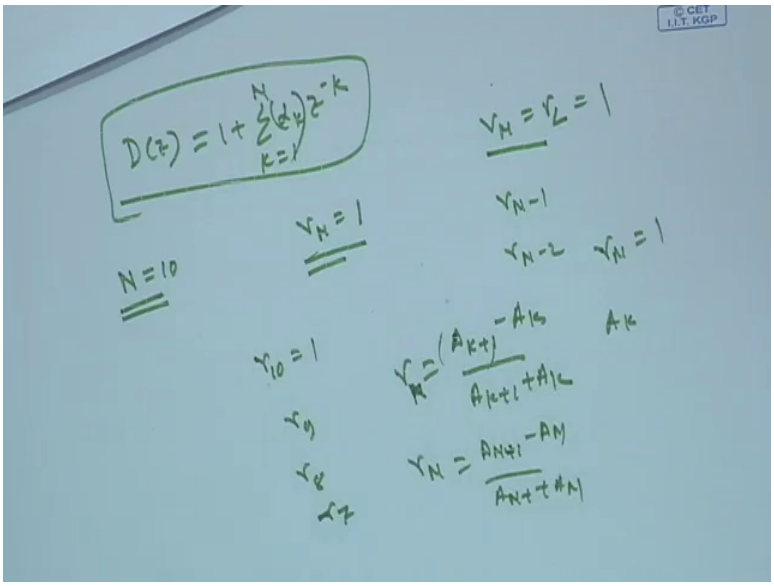


D 2 z is 1 plus r 1 z to the power minus 1 plus r 1 r 2 z to the power minus 1 plus r to z to the power minus 2, that I have derived here z to the power minus 2.

Now, if I say this is nothing but a D 1 z 1 plus r 1 z to the power minus 1 is nothing but a D 1 z. So, D 1 z plus r 2 z to the power minus 2 into D 1 z to the power minus 1. R 1 r 2 z to the power minus 1 plus r 2 z to the power minus 2 this term can be written like this way r 2 z to the power minus 2 into D 1 z to the power minus 1 because D 1 z to the power minus 1 will be 1 plus r 1 z to the minus 1 instead of z. So, z So, r 1 z into this one. So, r 1 r 2 z to the power minus 1 and r 1 z. So, this can be also written by d 0 z there is no junction no nothing is there is 1. So, it is nothing but a d 0 z into plus r 1 I can say z to the power minus 1 d 0 z to the power minus 1. Similarly I can write down this way because this is nothing but a D 1 we put D 1 value and this will come same value will here.

Similarly I can write dnz or dkz dnz I can write nth nth tube N number of tube this is nothing but a D N minus 1 z plus r 2 z to the power minus N D N minus 1 z to the power minus 1. Or I can say dk z is nothing but a d k minus 1 z plus r 2 z to the power minus 1 dk minus 1 z to the power minus 1. Which is same as if I write the equation dz is equal to 1 plus k equal to 1 to N alpha k z to the power minus k.

(Refer Slide Time: 06:18)

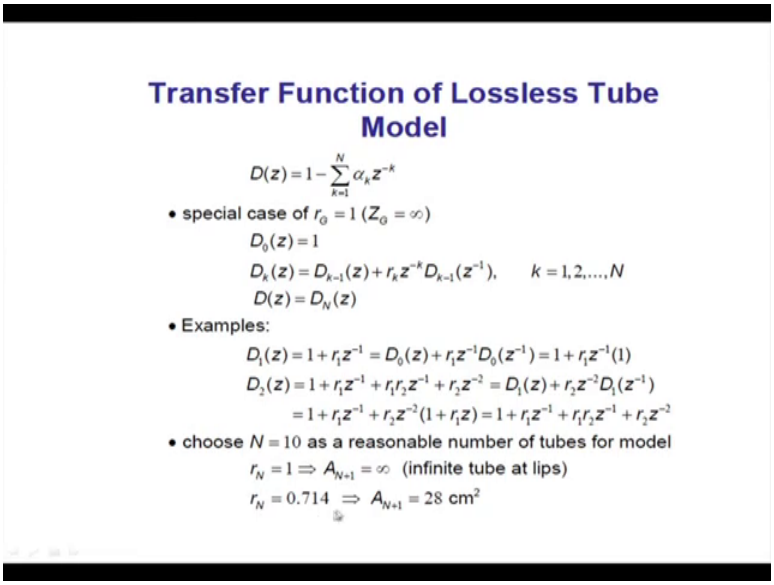


Same way I can write. So, this can be written as like this way. So, if I write this one if N is eq number of section is equal to 10 and let us r N is equal to 1 load, but r N is equal to r l. So, let us load is rl is equal to 1. So, no registive load for low frequency r N is equal to rl

which is equal to 1.

Now, I can say; I can find out the what is the value of if I know that r N is equal to 1, then from that side last side this a last last r N is equal to 1. So, I can find out r N minus 1 r N minus 2 r N minus 3 if it is 10 tu 10 tube model. So, r N means r 10 is equal to 1 I can find out r 9 r 8 r 7, how? Because using N is equal (Refer Time: 07:44) number of tube then model r N is equal to 1 r n is equal to let us this one N plus 1 is equal to 8 18 centimeter.

(Refer Slide Time: 07:45)

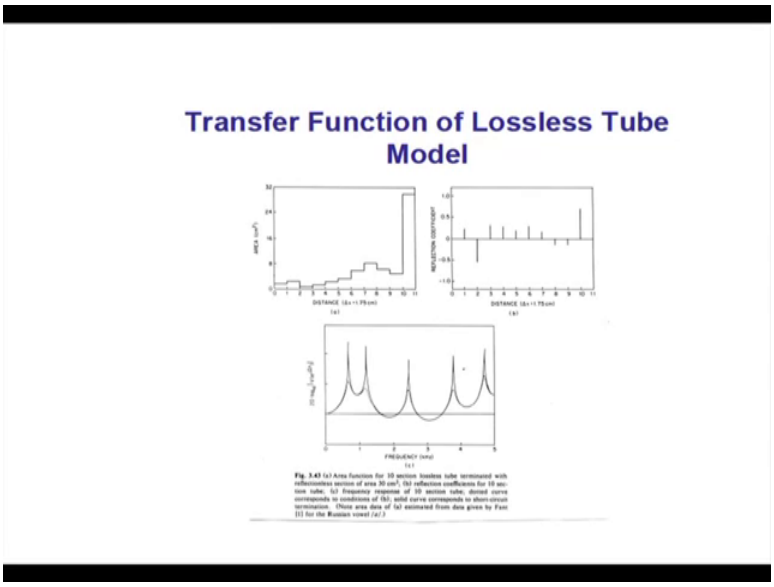


So, what is saying that that if I know the reflection coefficient I can implement the tube digital implement the tube. Or other hand if I know the signal which is coming out from the tube from there I can estimate the area function of different junction. Because if I know r r is r N is nothing but a what is r k? R K is nothing but a A K plus 1 minus A K divided by A K plus 1 plus A K. So, A K a are the area cross sectional area function of the tube different tube. So, if it is a 10 number of tube consist of the vocal track is model within a 10 number of tube, last r N rn is equal to 1 last one is equal to 1, then I can find out A K. So, r N I know r N.

So, this is r K. So, so I know r N, r N is equal to a N plus 1 minus an divided by a N plus

1 plus A N. So, if I know the cross sectional area then I can know the derive the reflection, coefficient and I can model the tube in digital domain. On the other hand if I know the pc signal and if I able to find out the value of reflection coefficient alpha k, in somehow I can able to find out the alpha k value. Then I can find out the cross sectional orea area of the different tube. I can give you one example from ravinand Say for that that is written.

(Refer Slide Time: 09:34)

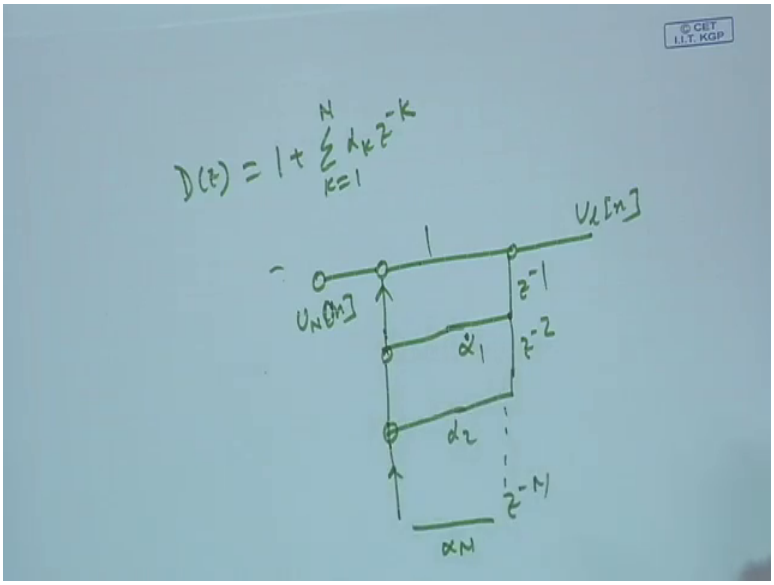


There is a different cross sectional area for a vowel I think the vowel which vowel I do not know that there is a vowel. So, this vowel area functions is written. Either I know the area function I can generate the vowel if I know the vowel I can derive the area function. So, how do we digitally implement this tube?

(Refer Slide Time: 10:00)

I want to digitally implement this tube. So, I know I have to implement dz is nothing but a 1 plus k equal to 1 to N alpha k z to the power minus k.

(Refer Slide Time: 10:09)

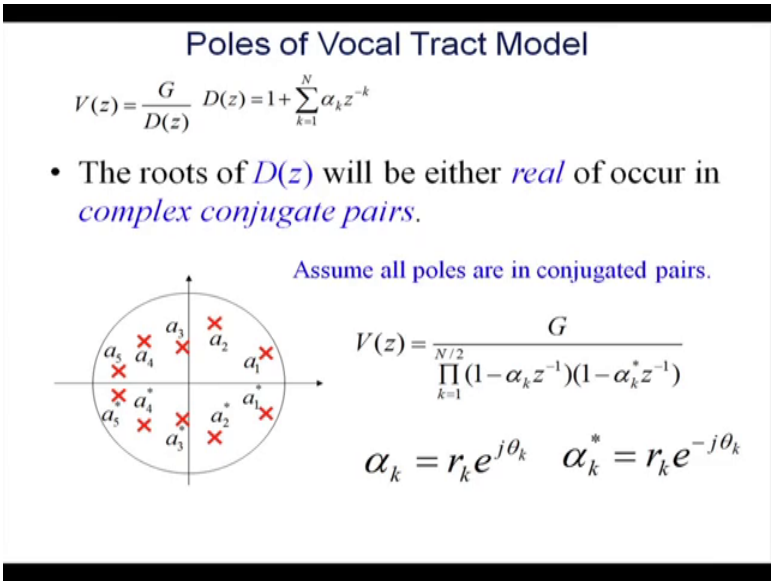


This equation I have to implement. So, if I able to implement how to implement very simple. So, there is a U N let us N unn impulse then I have a output which is u L n. So,

this will be simple delay by z to the power minus 1 alpha 1 and has to be added with here. Then z to the power minus 2 alpha 2 added in here dot, dot, dot z to the power minus N alpha N added here this is 1.

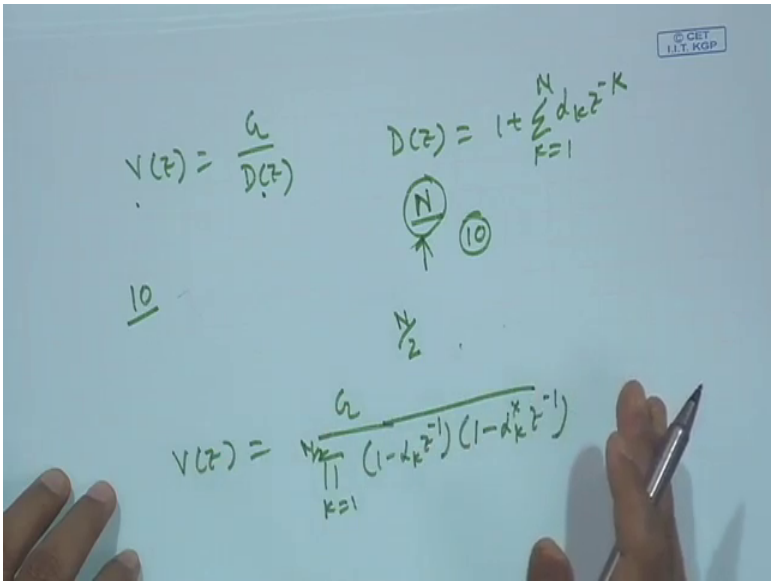
So, I can implement in digital systems I can implement it digital filter the line nothing but a digital filter the equation looks like a nothing but a digital filter. So, I can easily implement it using that things that vocal track I can implement using digital domain.

(Refer Slide Time: 11:22)



So now, there is some concept poles of vocal track. So, you that poles all pole models are vocal track tube.

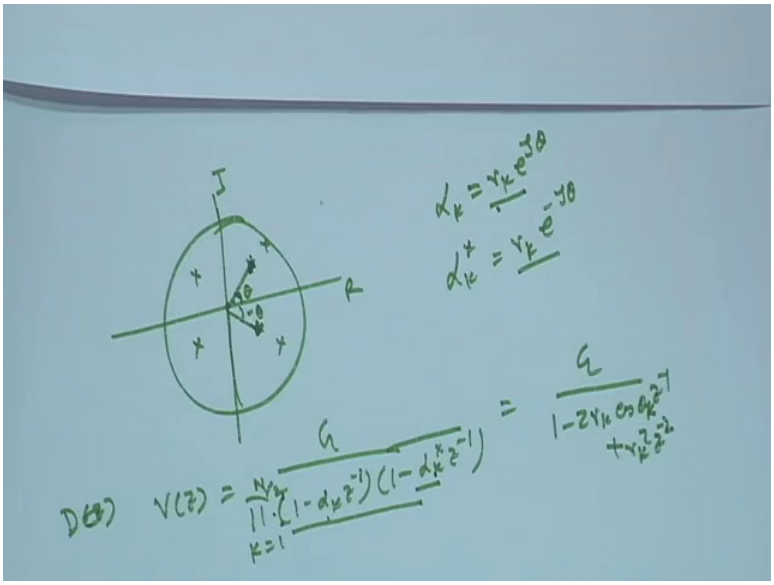
(Refer Slide Time: 11:31)



So, I can say the vz is model using G by dz where dz is equal to 1 plus k equal to 1 to N alpha k z to the power minus k. So, dz has an N if it is order is n. So, N number of pole. So, I can say vz is a all pole model which has N number of pole. So, if I have a 10 junction tube 10 tube 10 10 section tube then I can say 10 poles will be there 10 poles will be there. So now, if dz is real. So, the dz have either the all re real pole will be the real functions.

So, there is a real pole will be occur or there all pole have a complex conjugate in nature. So, if dz has a N number of pole then I can say that is a N by 2 complex conjugate pole will be there, N by 2 pair of complex conjugate pole. So, I can say vz is nothing but a G by k equal to 1 to N by 2 1 minus alpha k z to the power minus 1 into 1 minus alpha k star z to the power minus 1. So, it is nothing but a N by 2 complex conjugate pole product. So, what is the if you see that then if I say the this Is my unit circle.

(Refer Slide Time: 13:14)



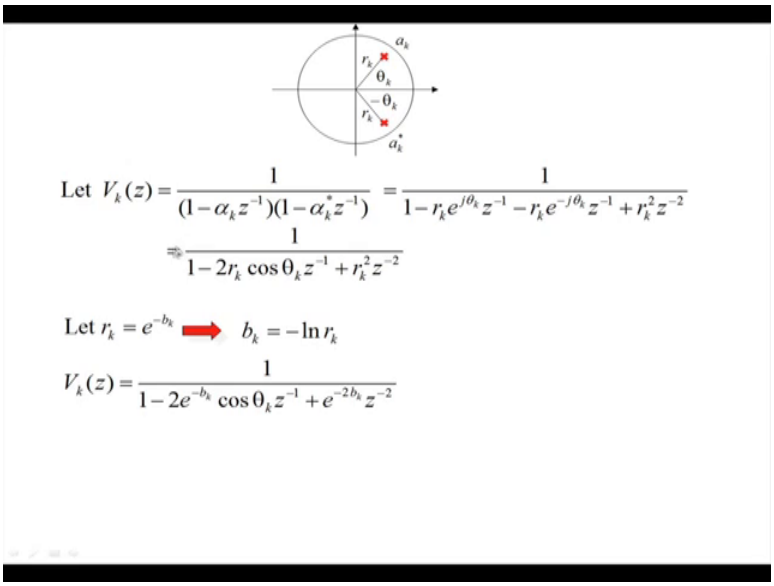
So, complex conjugate pole. So, this is the real axis this is the imaginary axis. If a pole occur in here with a angle of theta. There will be a another pole which is conjugate this pole minus theta and here. So, if there is a pole in here there is a another pole will be in here, if the pole in here there is another pole will be here.

So, every pole has an complex conjugate pole. If it is complex number then alpha k I can say it is nothing but a r K e to the power j theta. So, a complex number has a 2 a pole let this is the pole. So, these as a theta and this as a amplitude r. So, alpha k can be represented using r K and e to the power j theta, this r K is not reflection coefficient. So, there is a amplitude of the pole and e to the power j theta. So, if I say alpha k complex conjugate part it will be nothing but a r K amplitude will be same e to the power minus j theta, value of alpha complex conjugate pole. Now interpreted the value of value of alpha.

So, if I do that if I put the value of alpha in here then the dz will becomes or you can say vz will become G by 1 minus alpha k z to the power minus 1 into 1 minus alpha k star z to the power minus 1. Let us this right down this product I am not I am product of k equal to 1 to N by 2. So, if I do this 1, instead of alpha k I put the value of r K e to the power of j theta. Instead of alpha star k I can put the ar r K e to the power minus j theta if I put that value it will come G by 1 minus 2 r K cos theta k z to the power minus 1 plus r K square z to the power minus 2, for this complex conjugate product.

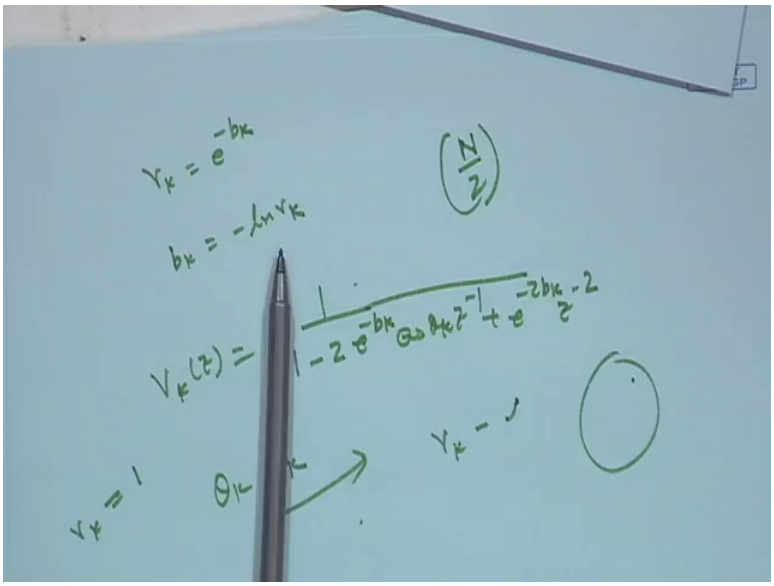
So, I can now if you see the vz vkz in term of this one.

(Refer Slide Time: 15:43).



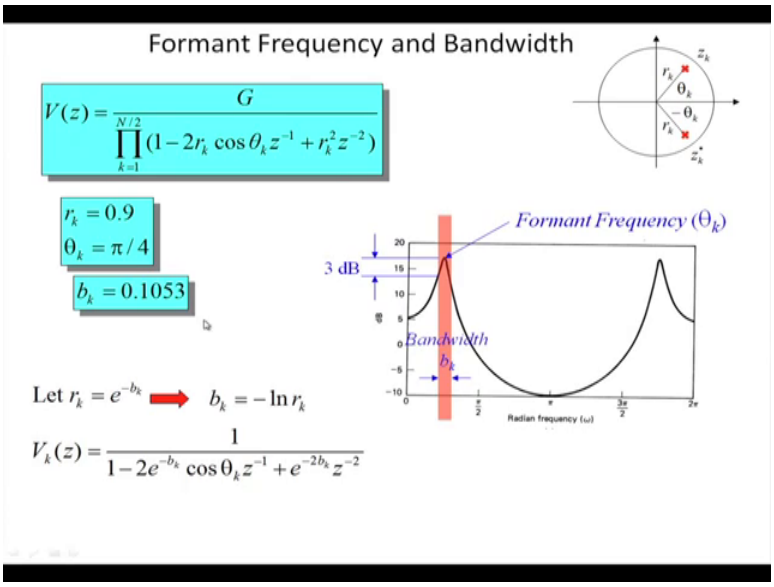
This is V k z, V z if I want vz right vz then there will be a product of pi equal to k, k equal to 1 to n. So, V k I can say this 2 only this is v if I write vk this one then I do not write this product term. Now if it is that then what is contribution of r K and bk let us r K.

(Refer Slide Time: 16:07)



Is equal to e to the power minus bk and bk is equal to. So, bk is equal to minus ln r K if I put that value then vkz will become 1 by 1 minus 2 r K. So, 2 e to the power minus bk cos theta k z to the power minus 1 plus e to the power minus 2 bk ask r square k. So, z to the power minus 2. Now importance of this bk is nothing but a produce a band, width bk produce a band width.

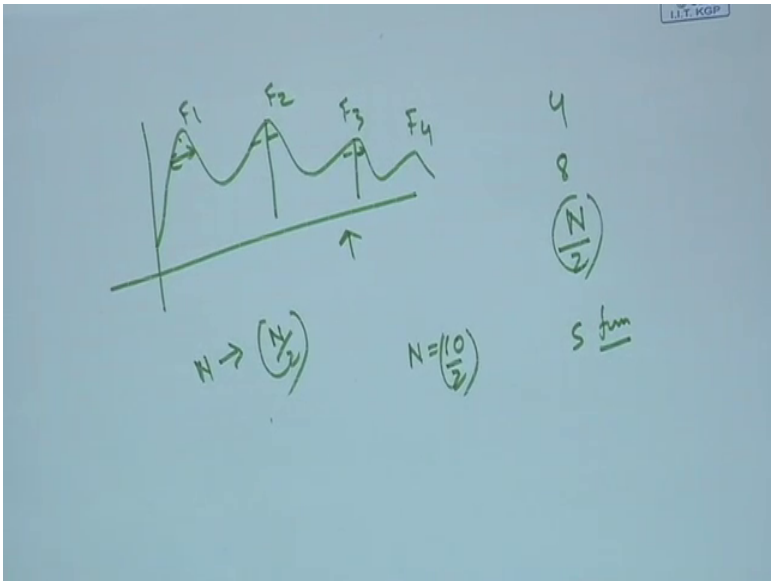
(Refer Slide Time: 16:56)



And theta k give the formant position. When the value of r K approach to unit I can get the formant frequency resonance frequency. So, if it is r K approach to unit cir close to unit circle then I can get the resonant frequency. And if the bk right bk value is non 0 then I get a band width pole, there will be a pole there will be a formant which has a band width. So, bk provide me the band width theta k provide me the formant position.

So, this information will be used when we develop the model using linear prediction model for speech production system. So, if I know the theta k if I know the theta k and value of r K then I can model the system, theta k give me the formant frequency position and if r K tends to 0 tends to unit circle amp amplitude of that pole is tends to your close to unit circle that give you the formant frequency. So, if there is a N number of pole. So, N by 2 complex conjugate pole will be there. So, each pair of complex conjugate pole give me a formant frequency. So, if I have a 5 formant frequency if I give you a spectrogram or let the I told you the spectrum of the speech signal is like this 1, 2, 3, 4.

(Refer Slide Time: 18:46).

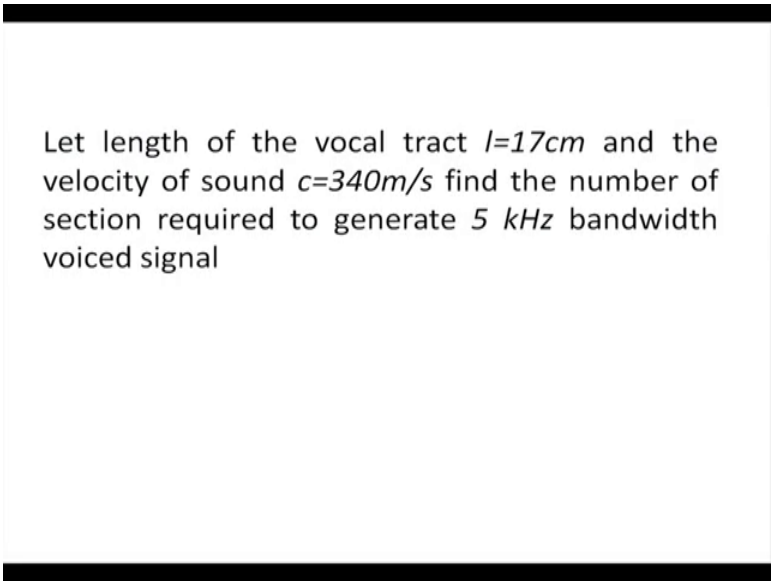


Let us f 1 f 2 f 3 f 4 then if I aks how many complex conjugate pole will be there 4 pair. How many pole will be there complex pole will be 8. So, if N tube model is there N number of pole is there in a transfer, function then I can say N by 2 complex conjugate pole will be there. So, literally I can get N by 2 formant frequency in a spectrogram. So, if I give you the formant frequency and formant band width if I able to find out for a speech event I can able to derive the transfer function for that speech event the vocal track transfer function. Now if I say that that event is given.

So, speech event let us say speech event the steady state vowel r if I take the steady state vowel and analyze the spectro frequency analysis, and find pout the formant frequency and formant band width then I can able to find out the transfer function of that vocal track. So, complex conjugate num, So, do not confuse to the number of complex conjugate pole and pole number of normal pole. So, if there is a N tube model N tube model is there.

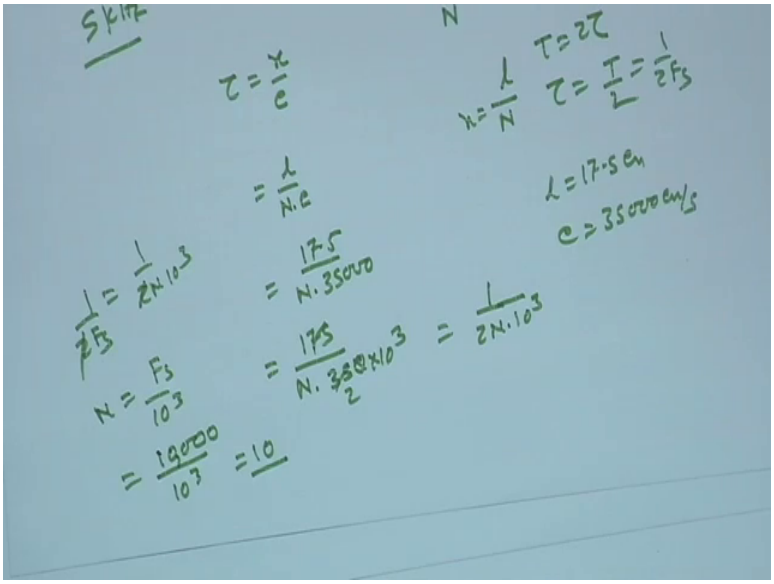
Then I can get N by 2 complex conjugate pole. So, there will be a N by 2 format. So, if I able to let us there is a N equal to 10 tube I take 10 10 tube to model this the 10 this whole vocal track system is divided by 10 parts and model it then I can get 10 by 2 complex conjugate pole. So, I can get 5 formant frequency. On the other hand let us I derive that mathematics that is a example, I have given in the I in the slides.

(Refer Slide Time: 21:02).



Let us the length of the vocal track l is equal to 17 centimeter and the velocity of sound is 340 meter per second find the number of section required to generate 5 kilo hertz band width voiced signal. So, I have to generate 5 kilo hertz band width voiced signal.

(Refer Slide Time: 21:19)



What should be the sampling frequency? F s should be 10 kilo hertz or not F s should be

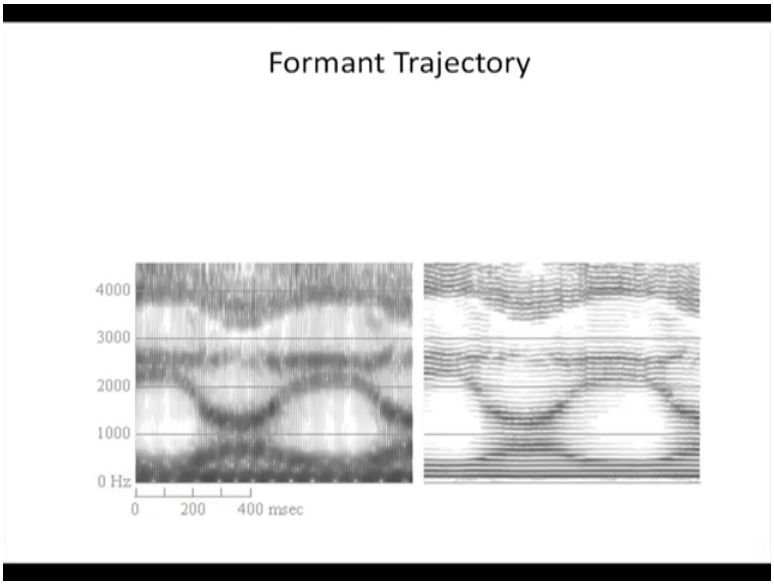
10 kilo hertz, then what is my tau? Tau is nothing but a x by c. So, what is x? X is a length of each section, I have to find out number of section required.

So, let us I require a N number of section. So, if the length of the tube is l then the x is equal to l by n. So, then tau is equal to l by N into c. Now if it is l is equal to 17.5 centimeter, and c is equal to 3 5 centimeter per second. Then I can find out it is nothing but the 17.5 divided by N into 3500, which is nothing but a 175 divided by N into 350, let us right 350 into 10 to the power 3. So, it is 2; so, it is nothing but a 1 by 2 N into 10 to the power 3; now what is tau? So, what is t t is equal to 1 by F s. So, that t is nothing but a t is equal to 2 tau. So, tau is equal to t by 2 tau is equal to t by 2. What is t? T is nothing but a 1 by F s. So, it is nothing but a 2 by F s. So, I can write 1 by 2 F s is equal to 1 by 2 N 10 to the power 3 2 2 cancel.

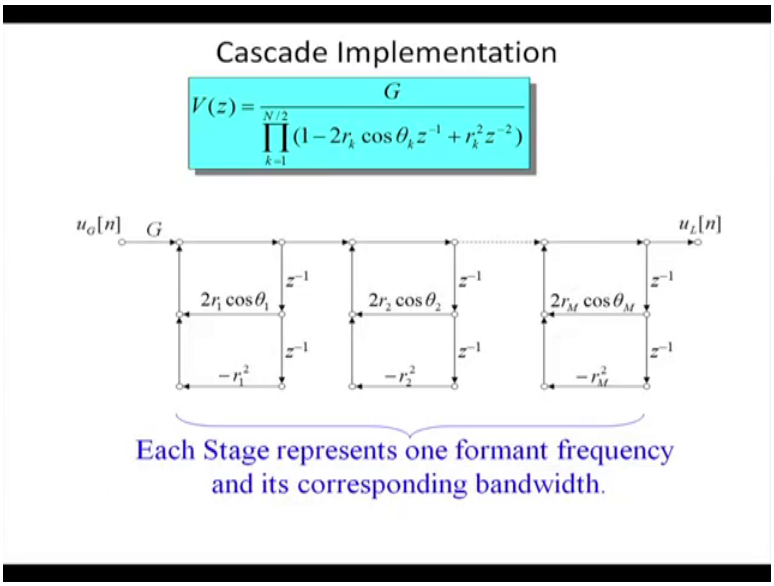
So, it is N is equal to F s by 10 to the power 3. So, F s is equal to 10 kilo hertz which is 10 kilo hertz nothing but a 10 sections. Or I can say N is nothing but a F s by N by 2, I can say N by 2 is nothing but a band width if it is a F s is equal to 2 b 5 kilo hertz. So, 10 10 section is required to model that tube 5 kilo hertz band width I require 10 sections. So, similar kind of mathematics you can expect that if I want to generate 4 kilo hertz band width signal and F s is equal to let us 8 kilo hertz sub (Refer Time: 24:32) frequency is 8 kilo hertz then find out the number of section is required, minimum number of section is required. So, I can find out the number of section is required. So, this way I can find out and this way you can model that signal.

So, since the vocal track can be model as a all pole model I am not going details of that Things this is the formant trajectory.

(Refer Slide Time: 24:56)



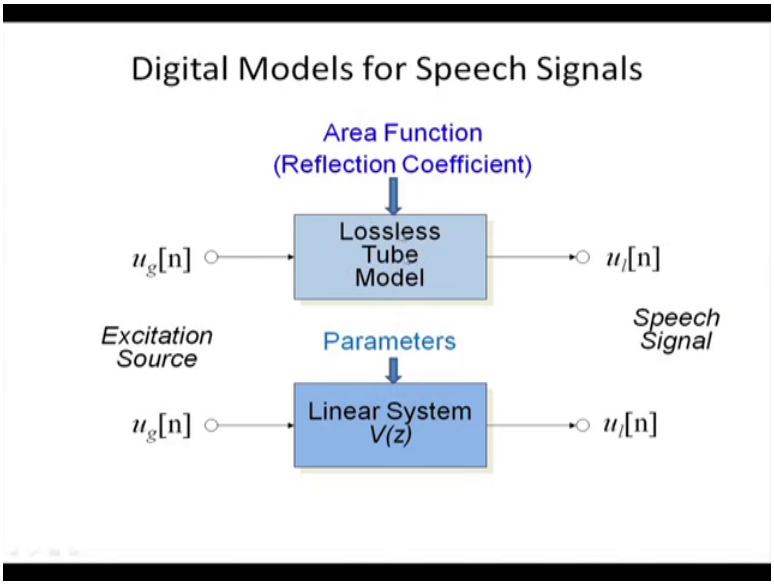
(Refer Slide Time: 24:57)



So, I can see that whole vz can be. So, this is the vk is the product of this if it is like vz then product will be here. So, I can say this digital implement is an again same each of the chunk represent the formant frequency. This is the first formant 1 formant 2 formant 3. So, that will come. So, each stage represent one formant frequency, each stage represent one formant frequency. So, that we N by 2 number of stage will be there. So, if it is N by 2 number of stage then N by 2 formant frequency will be there. So, I can say that loss less

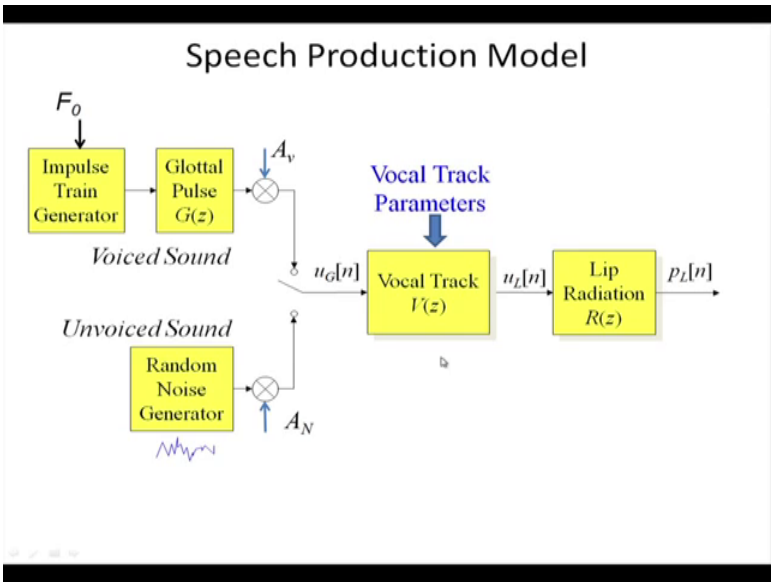
tube model, lossless vocal tube model can be done by a linear system which is nothing but a vz is a linear systems

(Refer Slide Time: 25:49)



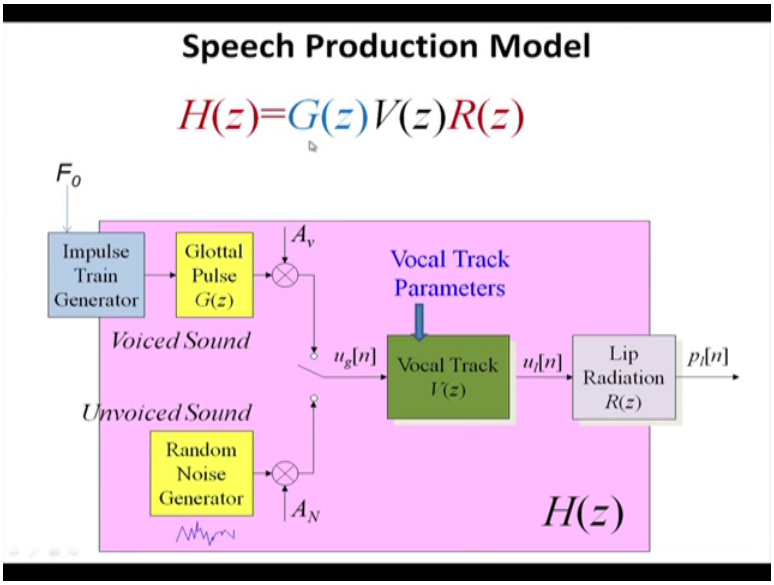
V z is equal to 1 plus k is equal to 1 to N n alpha k z to the power minus k lt. So, will discuss the about that lta system.

(Refer Slide Time: 26:07)



So, this I can say in summary this is the vocal track vz radiation is rz and glottal pulse is G z. So, if I say who what is the total transfer function is nothing but a H z.

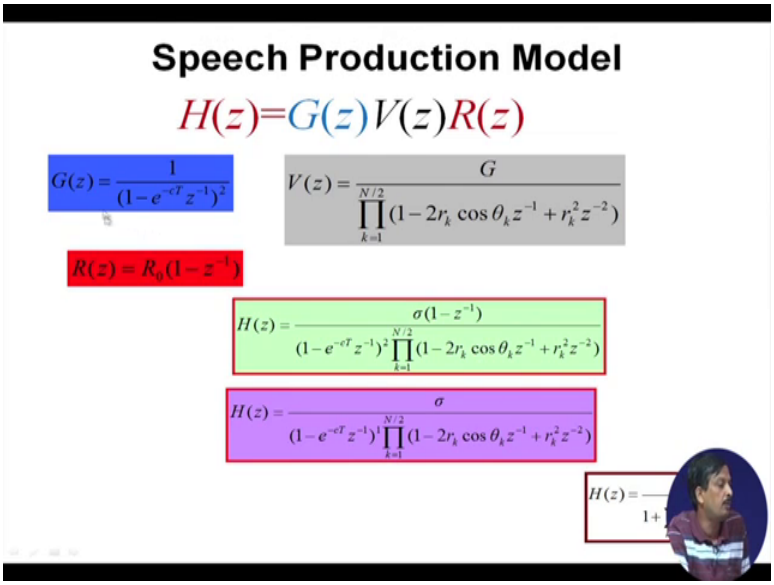
(Refer Slide Time: 26:26)



Whole speech production transfer function gz vz into rz whole hz the thus if I say thus from the speech I want to find out hz which is the product of glottal pulse transfer function, glottal generate glottal transfer function plus vocal track transfer function into vocal track transfer function into vocal track transformation into lip radiation transfer function. So, if there is a voice in impulse train will be there that will be modified by (Refer Time: 27:00) glottal transfer function and that modified signal fed to the vocal track and after lip radiation I get the speech signal. If it is unvoiced speech then I can say it is connected to the random voice and it will only modify by the vocal track and lip radiation and I get the speech signal.

So, random noise passes through vocal track and included lip radiation produce the speech for unvoice speech. If it is voice speech impulse 10 will you modify by the glottal pulse generator and that will pass through the vocal track and lip radiation produce the speech. So, if you see 3 transfer function gz can be approximate by the second order this 2 pole 2 pole function for radiation there will be a one single pole is required.

(Refer Slide Time: 27:39)



So, 2 plus 1 3 pole if I do that, then I can get hz. So, will discuss about that how many pole (Refer Time: 28:00) what are the linear prediction analysis that things during the linear prediction analysis function, what are the how do you get what are of the linear prediction that time we use this information. So, this is the whole vocal track tube modeling. So, in summary I can say the vocal track this human vocal track can be model using a digital signal processing or all can be implemented using a digital linear filter based on that requirement.

So, we have shown that if I consider this vocal track is a fun or you can (Refer Time: 28:38) is a simulated using a number of junction or number of section lossless tube section let us N number of lossless tube section, that then it can be implemented using a linear time linear system which is V z. And which can be implemented in digital domain. So, that is why this is called uniform tube modeling. The throughout the section if I say N number of section throughout the each section the cross sectional area of the vocal chord I said uniform. So, I can say throughout the whole vocal track can be single tube whole vocal track can be 2 tube or whole vocal track can be N number of tube, but if I say N number of tube each tube cross sectional area is constant. That is why uniform and loss less also because if it is loss lossy then all kinds of complex, it will come up. So, if it is loss less uniform tube model then it can be implemented using digital system that is vz. Once I know that then I know yes if it is implementable by a vz then can I think that

output speech which has collected using a microphone cannot be can it be analyzed using linear signal processing; yes.

So, that from there the concept of linear prediction analysis come. So, this system can be linearly model. So, from the signal if I want to find out the transfer function that sys the that is why I want to predict the vocal track constriction. So, if I know the signal yes it is possible to imple find out the area cross sectional area of different section if I say it is N tube model N different section cross sectional area is possible to find out from the behavior of output speech. Or if I know the area function I can implement it digitally it is possible to implement it, if I excited by impulse response I can able to produce the speech. So, this is the tube modeling. So, this is called uniform tube model or loss less tube modeling of speech production system.

Thank you.