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

Lecture - 16
Speech Perception – Part I

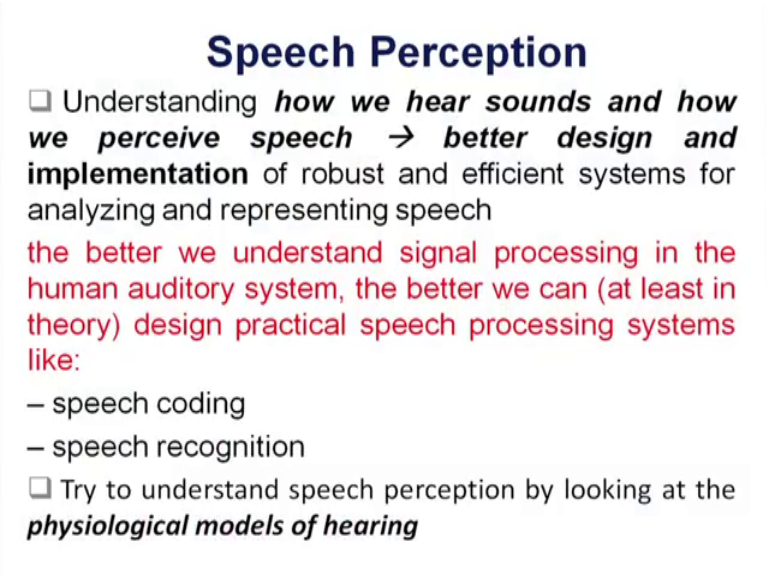
(Refer Slide Time: 00:21)

So, let us discuss that we have discuss about the speech perception. So, this class I will may be it is required 2 lectures. So, during these 2 lectures we discuss about the how human being perceive the speech. So, if you ask the why we have to know about the human speech perception, because if I talked about the digital speech processing then why you require that you should know that this digital speech perception kind of things. Now, if you see that human being ultimately we want to copy the human being through a machine or you can say we want to developed an algorithm or technology by which a machine can act as an human being or we can use the speech communication among the human being and that communication that you want to establish in machine.

Now, if I want to do that in speech communication if you see speech production and speech perception. So, 2 parts speech production, human being produce the speech and listeners perceive the speech. So, better we understand the how human being perceive the speech, we can better model that we can better process the speech signal that way. Or

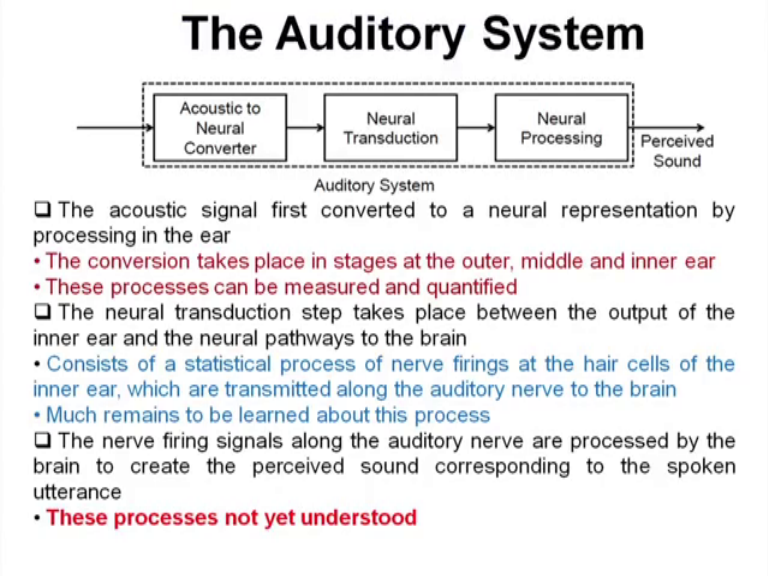
other hand if I say if I produce a speech and human listeners perceive the speech, how human listener process the speech to perceive it is very important to know, because unless we do not know what we perceive the development of technology will be not possible. So, we want to know what kind of speech processing happened in our brain and it try to follow that kind of processing in digital speech domain, so that we can develop the speech technology like the speech coding.

(Refer Slide Time: 02:24)



If you see that like the example of speech recognition and speech coding speech recognition lets the example of speech recognition. So, what listeners is produce what the speaker is produced that listen by the listeners and he perceive. Same technology we want to produce in machine that in front of a micro phone, we want to speak out that speech signal of the speaker output he want to give, and from the microphone and the signal processing technique, I want to develop how human being perceive the speech. So, if I want to develop that technology, we should know that signal processing involve in human brain, what kind of signal processing involve in the human brain, so that we want to know that is the speech perception, how we perceive the speech.

(Refer Slide Time: 03:25)



Forget about this speech chain, so if you see the auditory systems we perceive the speech to the auditory systems of the human being. So, if I see that a human being a human listeners first covert that acoustical signal to some nerval signal or nerve response we can say, so that required a transduction. So, trans or transducer which will convert the input acoustical signal to a nerves signal and that nervous signal that the signal which is in the nerve will goes to the brain and process it and perceive the speech.

So, we have there is we can sort of 2 part o or you can say the three part one is called acoustical to neural converter, neural transduction and neural processing. So, acoustical signal has to be converted to the neuron signal neural converter, and neural trans neuron has to be transmit the signal from that conversion point to the central brain for processing; so, here if you see the acoustical to this transverse, and the acoustical signal pass convert to neural representation by the ear. So, we have ear.

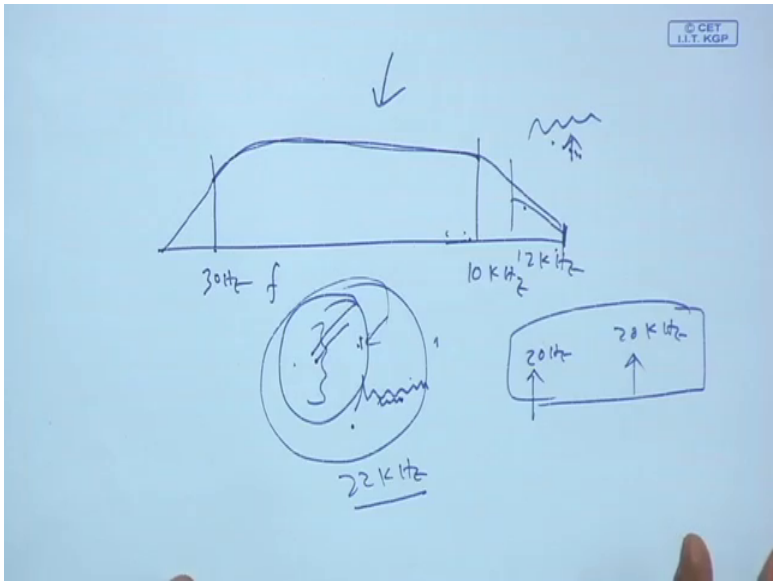
So, function of the ear to convert that input acoustic signal to the neural signal that is the process. Then neural transduction they take place between the output of the inner ear and neural pathway that means, whatever the output is coming from the neural transduction has to be taken away to the processing unit. And nerve firing signal, so how did; that is a human brain, so brain process that firing signal and perceive or understand the speech. So, perceive and understanding is happen in brain.

Now, we have to know what kind of transduction is happen and what kind of processing in happen in human brain, so that 2 things we have to know, to know the human perception. Or either I can say what kind of, so if you see the if you study the microphone the purpose of the microphone is that it has to be convert that acoustical signal to a electrical signal. Now, I have to know that this microphone has how efficiently this microphone can convert the acoustical signal to electrical signal, or I can say the properties of acoustical signal how accurately impose on electrical signal that is the electrical.

So, acoustical signal is equivalent to electrical signal, if the all properties of acoustical signal like that its frequency (Refer Time: 06:27) or you can say frequency composition or its amplitude all things how it is representing in electrical signal is the important point. So that means, how efficiently a microphone convert the acoustical signal to electrical signal is the property or you can say depends on the construction mechanism of the microphone.

Similarly, human being has an ear - 2 ear, if you see the 2 ears. So, how efficiently human ear can convert the input acoustical signal to the nerval signal or nerve signal. So, I want what kind of limitation this system is imposed on speech perception that means, human being or what kind of you can say that what kind of constrain it put on the human speech signal ear conversion, so that that kind of constrain we can exploit in speech processing. So, that human being cannot hear that error.

(Refer Slide Time: 07:51)



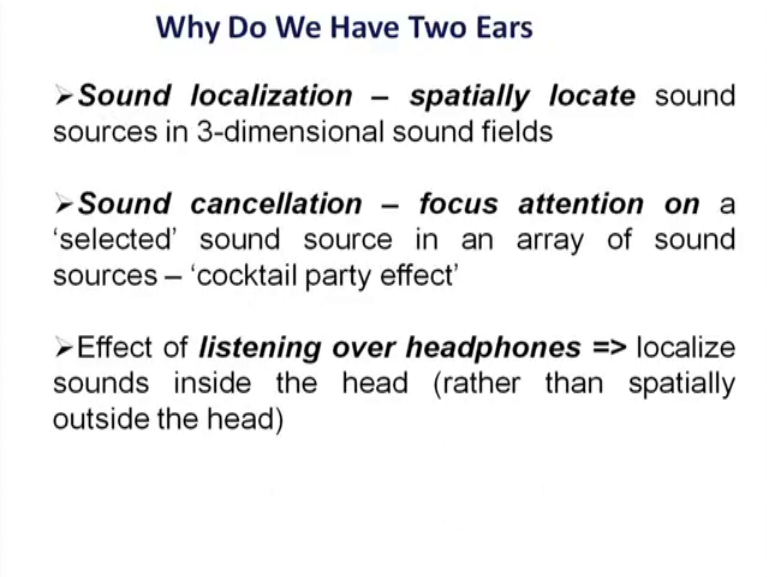
So, suppose a microphone has if you see the microphone has a frequency response. Let us discuss about the frequency response. Suppose, in microphone has an frequency response, if you know the frequency response because frequency response is nothing but a frequency versus amplitude plot that means for a particular intensity sound of different frequency how it represent in electrical signal. So, suppose microphone has a frequency response whose flat area is around let from here it is 30 hertz to 10 kilo hertz, so that is data set is given that this is the microphone. So, this microphone is efficient to convert the acoustical signal in between 30 hertz to 10 kilo hertz, and produce the electrical signal.

So, if an acoustical signal consist of 12 kilo hertz, the electrical signal should not contain that component because that response of the microphone at 12 kilo hertz is almost 0, may be almost 0 let us design down in here is 12 kilo hertz. So, it is 0; almost 0. So, I cannot get 12 kilo hertz response in here so that means, limitation which is imposed by the microphone in electrical signal is 30 hertz to 10 kilo hertz.

Similarly, suppose I have a human ear, so acoustical signal human ear convert into normal signal. So, perception of that signal that how we perceive what is the limitation of our per perception that is the limitation of conversion by the ear to another signal. So, limitation of perception how weak signal we can perceive how which frequency we cannot perceive all kind of limitation is impose in here. So, if I say first impose is the 20 hertz to 20 kilo

hertz frequency we can heard. So, ear can convert only the acoustical signal which is lie between 20 hertz to 20 kilo hertz. If I apply 22 kilo hertz to the human ears, it will not perceive with may not be cannot be converted to nerval system or nerval system cannot represent or cannot sensitize a response in the human brain.

(Refer Slide Time: 10:19)



So, I have to know how this signal or how this acoustical signal is converted to nerval signal, so that is the physiological mechanism we have to understand the physiological mechanism of the human ear. So, if you see I come this on later, human has 2 ears, if you see a human has 2 ear why a human has 2 ears, if I do not have one ear what will be the problem. So, sound the 2 ears help to localize the sounds, sound localization if you close your eyes, and if sound is come from some direction you can tell that sound is coming from this direction. If the sound is coming from this direction without seeing the source you can identify the direction of the source.

So that means localization of the sound is because of we have 2 ears, if you see that there is a home the stereo earlier there is a mono then comes stereo then comes surround sound. So, all are effect of sound localization because if it is stereo sound that means, we can identify the source of the direction of. So, source direction from the sound, Dolby - surround sounds. Suppose you are watching it a movie where a train is coming from left corner of the screen to right corner that means, it is a three-dimensional you are

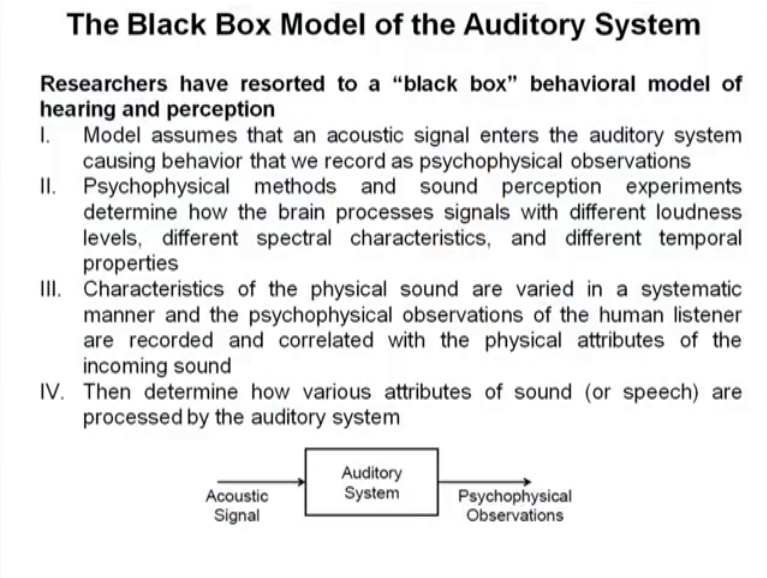
visualizing in a two-dimensional respects.

Now, if I say suppose I am standing in here, a train is coming from this direction, and it is going this direction. So, what will the effect of sound? If the train sound is coming from this direction, so I should able to localize the train sound is coming from my right hand this side, and the sound intensity will be increases when train come to my place and when train is going away then I can say the sound is coming from this direction. So, if I able to manipulate this effect then I can say it is a surround sound because sound is coming from this directions. So, I can understand sound is coming from this direction. So, sound localization is one of the major issues because human has a 2 ears that is why we can localize the sound.

Second one is the focus attention on a particular sound or you can say the noise cancellation. If you study the radar signal processing in electronics you know the how the cutter noise is rejected or kind of things. So, if we have 2 ears help human being to attention or focus on a particular sound. If in this class there is a 5 people are talking together even the 5 people are talking and they are the that is audio level is almost equal, I can make my attention to a particular student speaking so that means, I can focus on a particular sound that is one of the biggest property in human perception. I can easily cancel the noise or the signal.

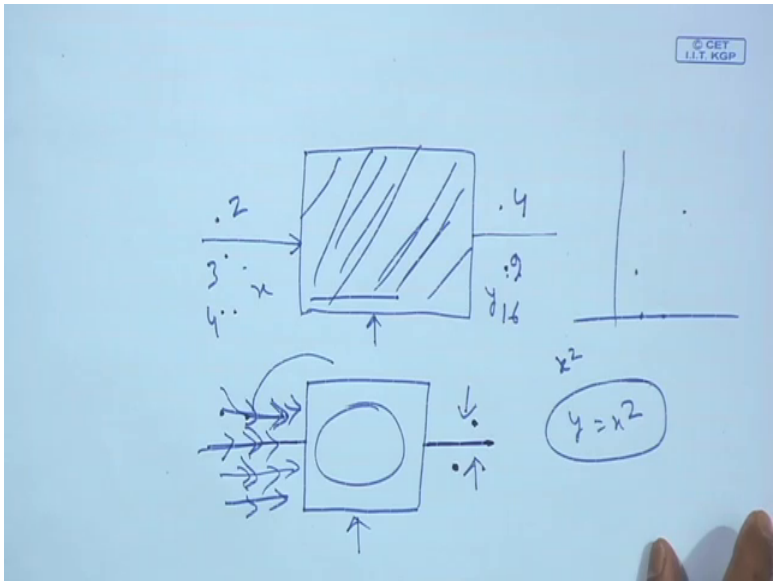
What is the difference of the noise, unwanted signal is nothing but a noise. So, suppose I want to listen the student A sound, so student B produce sound is noise to that. So, I can easily ignore the student B's voice and perceive the student A's voice. So, I can easily focus on a particular sound that is also because we have 2 ears. Now, once you put your headphone inside the ears then localization is within the headphone, it is not outside. Once you put the headphone sound is producing there only, so whatever the sound ear is perceived the localization happened in the headphone only. So, there is a sound localization issue.

(Refer Slide Time: 14:12)



Now, if I come that I want to make the auditory model auditory system model. So, I can say whole auditory system model, it include sound the acoustical signal to normal conversion and perception of that nerve signal by the brain. So, I can say human auditory system auditory sorry human perception of sound, I want to model. Now, if I say I want to model it, how can I model because a human being is perceive the sound, I cannot measure every and each and every point what is happening in the signal level, I cannot do that things.

(Refer Slide Time: 15:00)



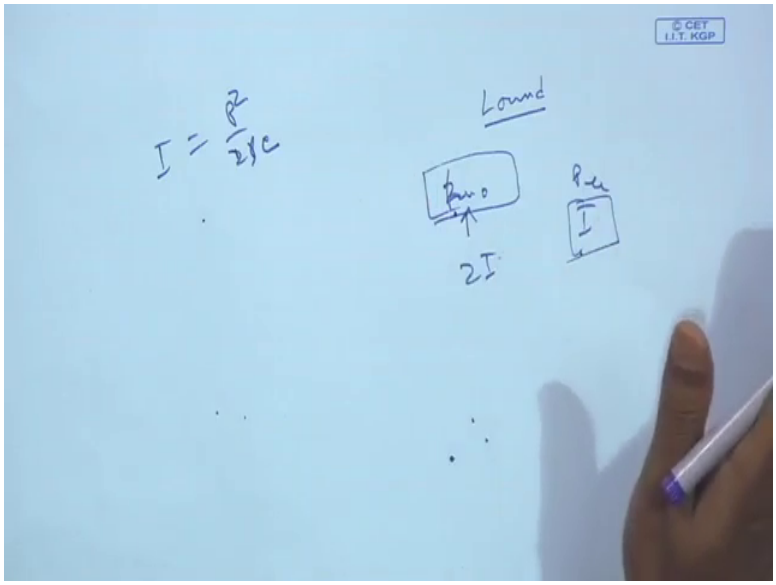
So, I can think let us the human sound perception is a black box. Now, how do you define how do you find out the black model property of a system? Suppose I have a system I do not know anything this is black box, I do not know system property system things nothing, now how do you determine we excited the system with a particular signal or known signal then try to find out the observation. So, let us I given a example I do not know the system if I put 2, it gives output 4; if I put 3, it gives output 9; if I give 4, it gives output 16. If I able to observe this behavior then if I able to plot 2 then 4 then 3 then 9 then I can know from the plot the behavior of the this black box, so that is nothing but a I can say this is nothing but a square. So, x square, if x is the input, y is the output y equal to x square. We can derive this system property from observing the known input what is the behavior.

Same black box model can be applied in human auditory system or human speech perception, suppose I apply a particular frequency sound with a particular intensity, then I want to observe the human perception. So, you can say that physiological observation and this is not a physiological observer, this is a physical input and physical output, but here I cannot get that physical output. So, I can say here I can give a physical input which is nothing but a sound acoustical sound and I observe the physiological observation and try to correlate what kind of stimulus, this is stimulus only input signal, what kind of stimulus what kind of physiological response is produced. Then try to find out what is this black

box how human being perceive the sound.

So, what I said I excited the this black box with a different kind of known stimulus and then I have to find out the physiological observation of human and then try to correlate what kind of stimulus producing what kind of observation, and try to draw a conclusion what kind of processing is done by the human being. So, that is why if you say there is a 2 dimension of sound parameter, one is called physical dimension, another is called perceptional dimension.

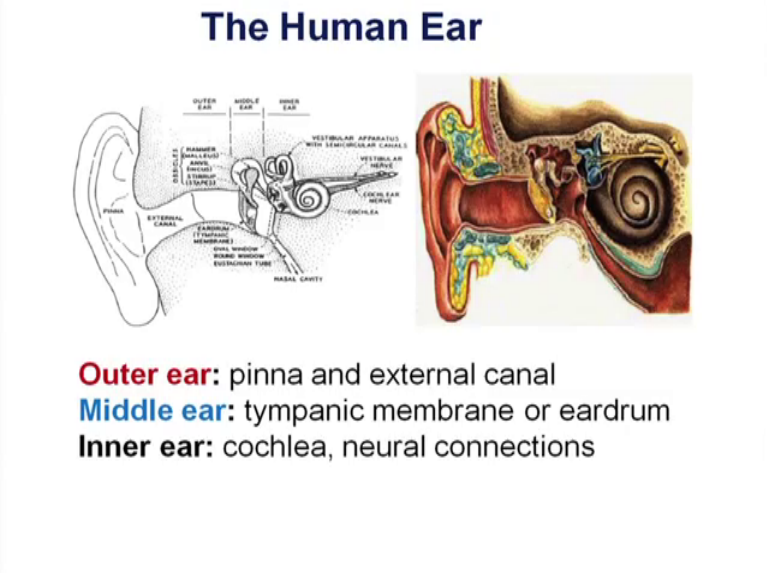
(Refer Slide Time: 18:09)



So, let us a given example intensity, intensity on acoustic wave I can easily measure. You can know I is equal to nothing but a p square by 2 rho c, rho is the density, and c is the velocity of the sound, where this p is the pressure amplitude of the pressure wave then I can say intensity is nothing but a p square by 2 rho c. So, I can measure the sound intensity you can say the sound intensity measure may decimal the I will come dB meter I can easily measure the intensity of the sound. But once I say loudness once I say loudness this sound is louder than the previous sound it does not mean the intensity of the first sound is twice than the previous sound. If I say this sound is twice louder than the previous sound then I cannot say the current sound is not 2 I or previous, or if the previous sound is I it is 2 I. The perception of loudness is human perception, but intensity is a physical parameter.

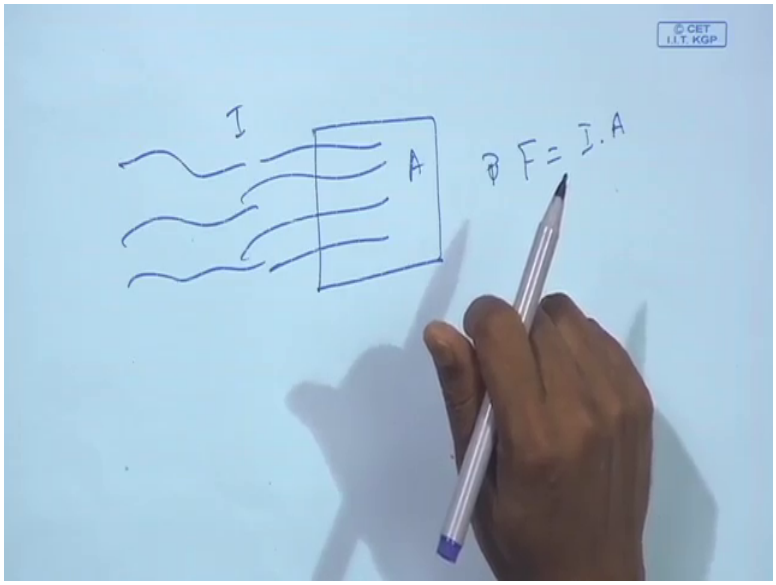
So, perception, so there is a 2 dimension one is called physical dimension where I can directly measure the quantity measure the parameters directly measure the parameter intensity frequency all are directly measurable, but if I say loudness is a human behavior physiological behavior of a human. So, I stimulus or stimulate the human being by intensity and I want to observe that his behavior then observation behavior is called loudness and input is called intensity. So, there is a physical dimension and there is a perceptual dimension. So, I will come details on that physical dimension and perceptual by dimension.

(Refer Slide Time: 20:02)



Now, let us start with the physiology or you can say the anatomy of human ears some sort of what is there in human ears. If you see this is a color nice pictures, which is that there is a human ears. So, there is a you consider outer ear which is this one only if you see this one is the outer ear. So, this is called pinna. So, what is the function of this pinna.

(Refer Slide Time: 20:40)



If you see if the sound wave is coming in this room in medium air in medium now if I put a like this, if the sound wave is coming in this medium if I put a area here then the sound wave will be strike here. Now, if the intensity of the sound wave is I then what is the total power of the acoustical wave which is strike in this plate if the area of the plate is A then I can say a power or force is equal to I into A. So, I can say the principle of the pinna working principle of the pinna is the or you can say the work of the pinna is that it collect that acoustical wave and channelize it to the middle ear. So, if I increase the pinna the channelize power will acoustical power will be increases that is why sometimes when you want to listen the very low sound, we put our hand in here.

Once we put our hand in here it increase the pinna area. So, more acoustical signal is channelize to their ear sometimes. If you see the rabbits to listen the sound rabbit can move his pinna, we cannot move my pinna, so that is why we have a head movement. So, once I want to listen the sound I can this I can rotate the head, but if you see the rabbit can rotate their pinna to the direction of the sound. So, the pinna work is that it should channelize the acoustical wave power to the middle ear that is the working of the pinna.

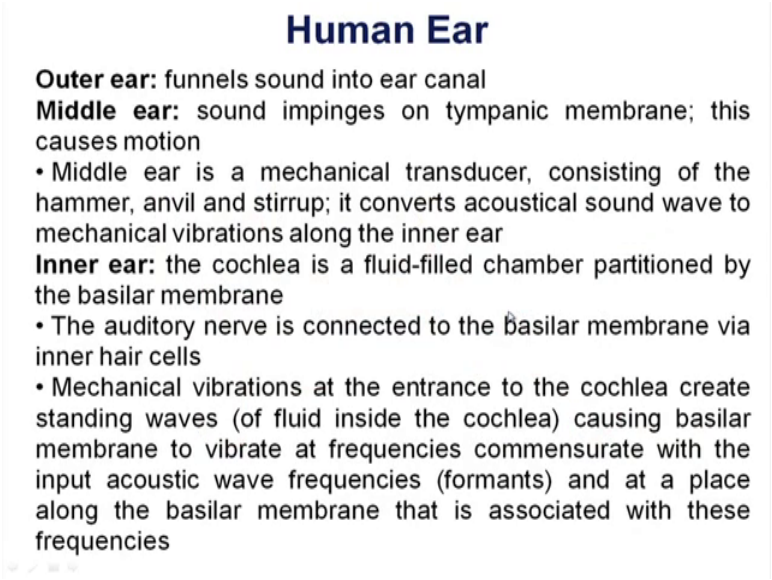
So, if it is large the more acoustical wave will entered, if it is small the area if the reduce then a power will be reduced, so that is the working principle of the outer pinna outer ear or pinna, pinna, we can say the pinna or penna. Then there is a middle ear middle ear is

what middle ear that acoustical signal passes in the middle ear and middle ear consist of a membrane and bone. I will come to what is the work and then inner ear has an cochlea, if you see there is an cochlea.

(Refer Slide Time: 22:54)

Now, so in the auditory canal this is called auditory canal middle ear auditory canal, acoustic wave is coming in here, now it is strike in this membrane. Once this acoustic waves strike in the membrane, it produce the vibration that vibration transmitted to the cochlea by a conduction process. So, there is a bone and that bone this acts as an mechanical resonator or you can say not resonating a amplifier we can mechanically amplify the sound and transmit to the cochlea. If you see the guitar not that electronics guitar, if you see there is a mechanical amplifier or you can resonator that you can say resonator that if it is not that is not there, if I strike a string the sound will be less. But if it is there then it acts on mechanical resonator, then the sound is EMA a larger. So, that is that is the effect in ear also that is this acts in the mechanical kind of things and that transfer the sound to the cochlea. So, that vibration goes to the cochlea.

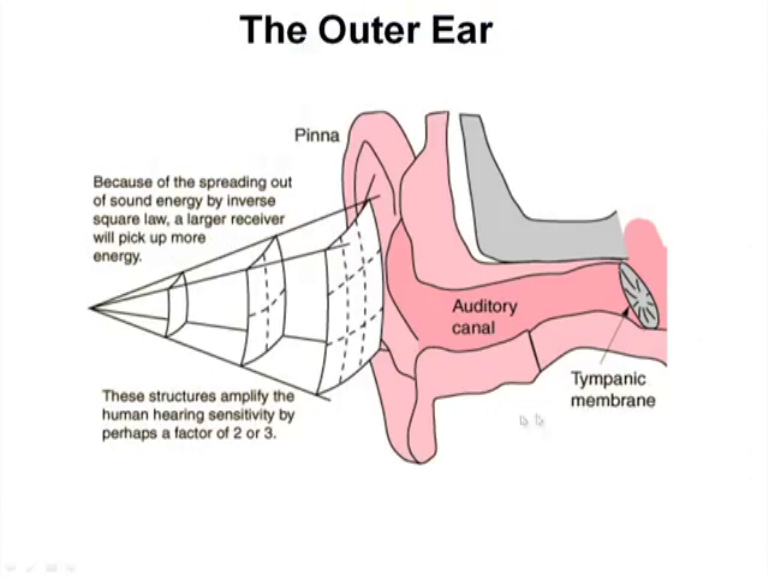
(Refer Slide Time: 24:08)



Now, cochlea, so I am not reading the slide, you can read the slide. So, inner ear is the cochlea. So, this is the semicircular canal and there is a cochlea. So, you know that the ear has 2 function, this whole system has 2 function one is balance the body and another one is that listening. So, in inside the cochlea there is a basilar membrane, which is responsible for listening the sound and this semicircular canal. So, the cochlea is fill of liquid. So, you can say the it is a nothing but a packet of liquid.

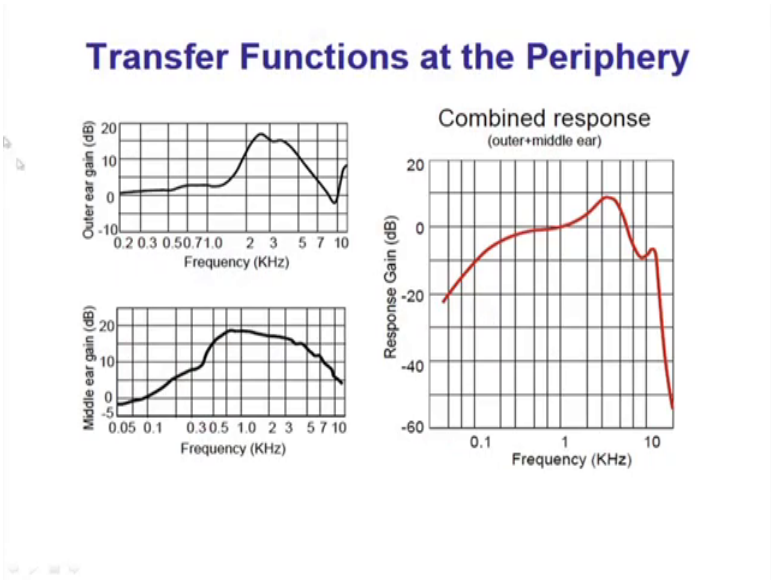
So, if I vibrate that outside wall of the liquid then what will happen that vibration will transfer to the liquid a liquid is spread the vibration to everywhere, so that is the mechanism for the inner ear to perceive the sound, and this semicircular canal is full of liquid that acts the body balance. So, this is nothing but a you can say if you see that sometimes when we are balancing or you can say leveling some floor, we use the water balance that to make a tube with full of water and then (Refer Time: 25:17) the water level balancing we said whether the plane is correct or not, where is the slope or kind of things. Similarly, this thing is used we have 2 semicircular canal in 2 side and which is responsible for find out the body balance like that water balance same things is in here also.

(Refer Slide Time: 25:43)



Now, in inner ear, so this is basic description of the how human being what the function of that nature in part of the human ear. Now, I go to the details. So, this is ok, this is physiological we understand. Now, in engineering model, so I should know what kind of action or what kind of things is happened in pinna that is outer ear, and what kind of response is happened in the middle ear, and the final is the inner ear.

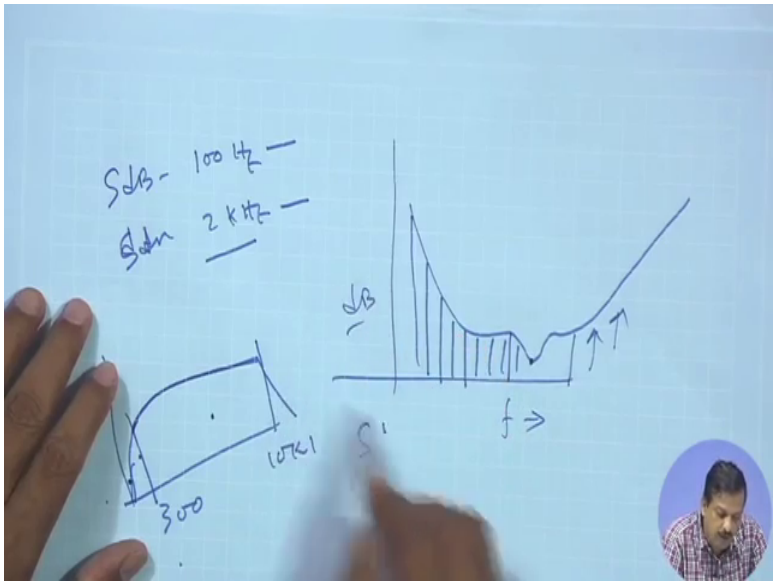
(Refer Slide Time: 26:18)



So, if I see the frequency response of the outer ear that is the pinna, so pinna if you see that it is not channelize all frequencies same amplitude I can say. So, what I say the outer ear if I see the frequency this, this axis is the frequency and this is the response in dB. So, all frequency 0 dB is not same intensity or you can say the 0 dB of 0 forget about 0 as let us see 0.3 kilo hertz; that means, lets 300 hertz. So, 300 hertz sound of let us 0.5 dB not less than 0.2 dB and 1 kilo hertz sound of 0.5 dB will be the same response because it is amplify or you can see the amplification of the power or I say them all frequency by the outer ear is not same.

Similarly, amplification or you can say the frequency response of the middle ear is nor flat throughout the frequency it also different frequency is a different response. If I combine this 2 middle and inner, not middle, and outer ear the combine response is come as a red color. So, I can say human ear is not equal sensitive to all frequency.

(Refer Slide Time: 27:58)

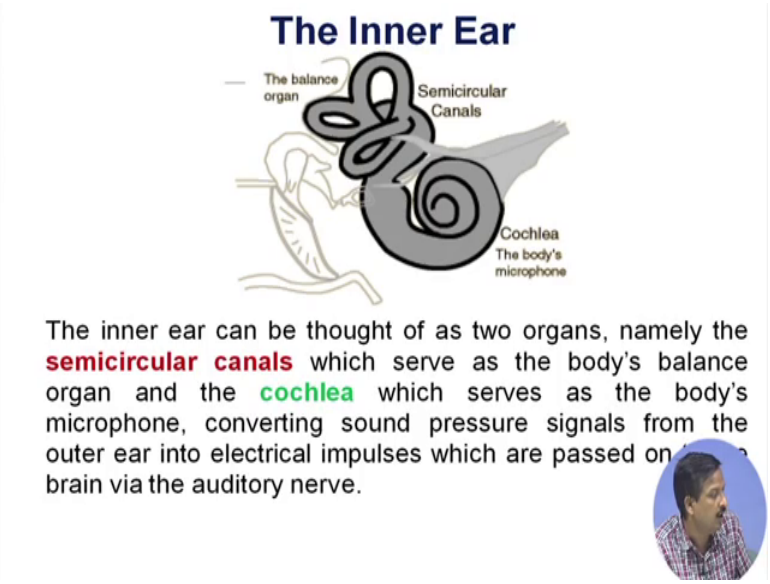


So, I can say we are not equally sensitive or I can say let 5 dB sound of 100 hertz and 5 dB sound of 2 kilo hertz is not produce same intensity sensation or you loudness sensation in our human brain. So, this may be louder compared to this because of this response. So, if I inverted this curve, if I inverted this curve, the curve will come like this. So, all if it is this is the frequency, this is the dB amplitude. So, all frequency has not equal response in case of human ear that is the frequency response of human ear that is the

frequency response of the microphone. If it is flat for 300 hertz to 10 kilo hertz that means, the microphone can produce the output or you can as equal sensitive the frequency between the 300 hertz to 10 kilo hertz. But other frequency the conversion from my acoustical signal to electrical signal is very low wither very low or almost 0.

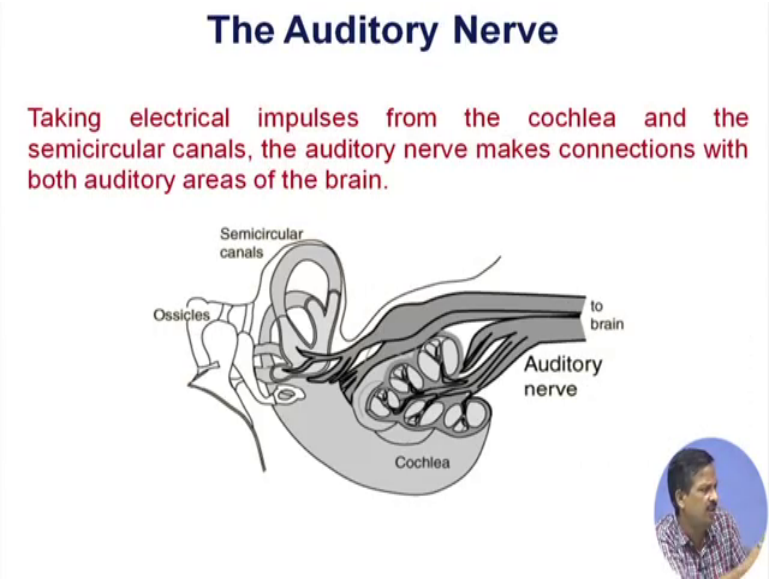
Similarly human ear is not sensitive or intensity sensitive to all frequency equally. So, may be here is 1 kilo hertz, here is after 5 kilo hertz it also increasing. So, 5 dB of 100 hertz, 4 dB of 2 kilo hertz the intensity perception by the human being will be different that is why we say because of the this combined curve. I will come this is called the threshold of hearing. I will come later on that one.

(Refer Slide Time: 29:48)

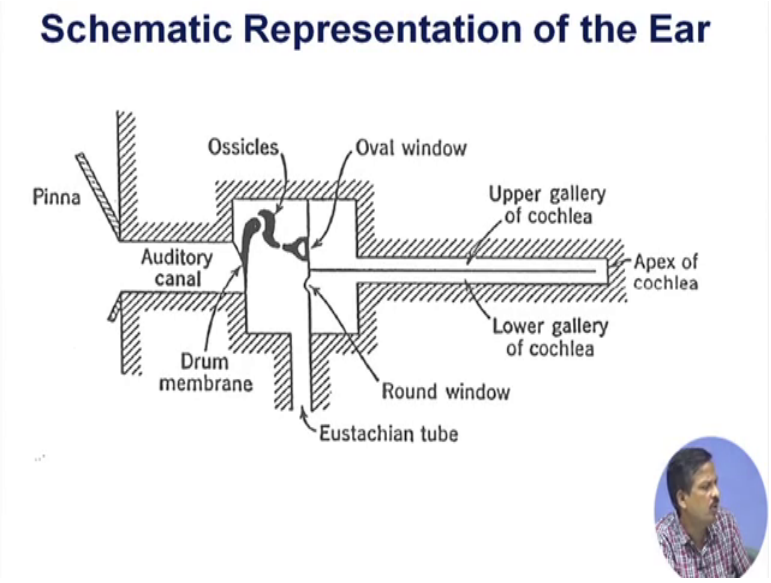


So, now, I am not a describing that function of cochlea and all those things you can read this, this is that read this slide that is nothing is there.

(Refer Slide Time: 29:55)



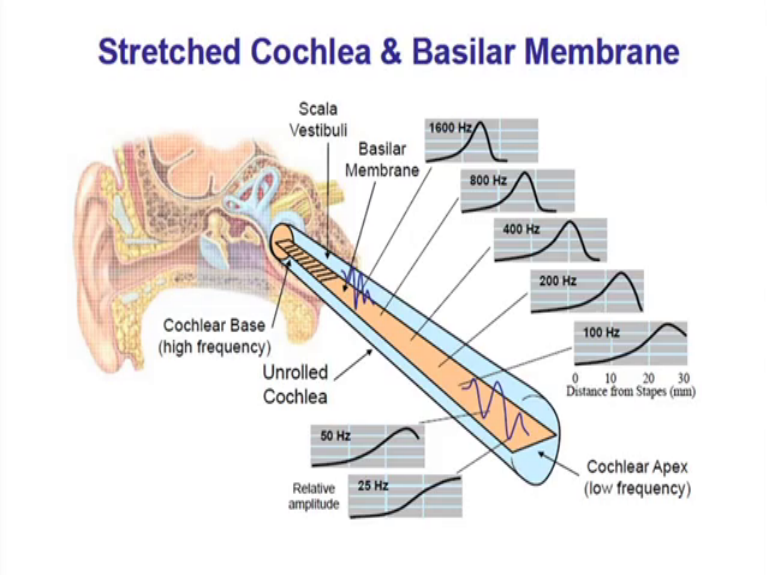
(Refer Slide Time: 29:57)



So, I can say this is the schematic representation. So, pinna collect that acoustical signal channelize to auditory canal, and middle ear convert that acoustical signal to a mechanical vibration and that mechanical vibration transfer to the cochlea, inside the cochlea there is a basilar membrane. So, this is apex and this is the beginning. So, this is the beginning of the cochlea, this is the apex of the cochlea. So, inside the cochlea, there is a basilar membrane which is responsible for speech conversion of that vibrate mechanical motion

to nerval signal.

(Refer Slide Time: 30:43)



So, how it is done it? So, I am not discussing here lets in the next class I will discuss how the basilar membrane is converted or you can responsible for or convert that vibrant or motion mechanical vibration to the nerval signal.

Thank you.