Virtual Reality
Prof. Steve Lavalle
Department of Multidisciplinary
Indian Institute of Technology, Madras

Lecture – 17
Audio (physics and physiology)

Welcome back. Today, we are going to talk about the Audio Component of Virtual Reality. In the last lecture, we finished up with, what is mainly standard computer graphics material. But with a particular virtual reality perspective on it talking about, what the different kinds of problems are in our particular context.

And you know finished up with that going over, rendering techniques and eventually explaining some of the open challenges and problems that we face.

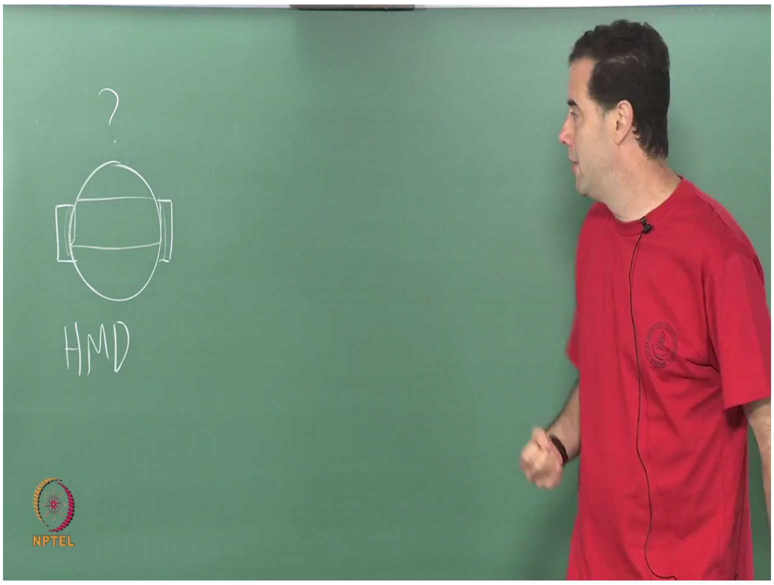
(Refer Slide Time: 00:47)

So, for today, let us look at audio, for VR. Now, remember we have said this before a vision is the most, fundamental and important, a sense that we rely on the human body and, there are, there are more neurons by far devoted to that sense than any other sense that we have. But nevertheless we can stimulate other senses as part of the virtual reality experience.

And audio is an important, component of that. So, you might imagine in the real world, our ears are hearing sounds that are propagated from various sources. We want to somehow in a virtual reality generate sounds synthetically.

So, if we are in some kind of cave like virtual reality system then, it is a matter of generating appropriate sounds that are placed on some kind of that are rendered, let us say on some kind of speakers that are placed around in the environment, in a fixed way, in the real world in a fixed way.

(Refer Slide Time: 01:51)

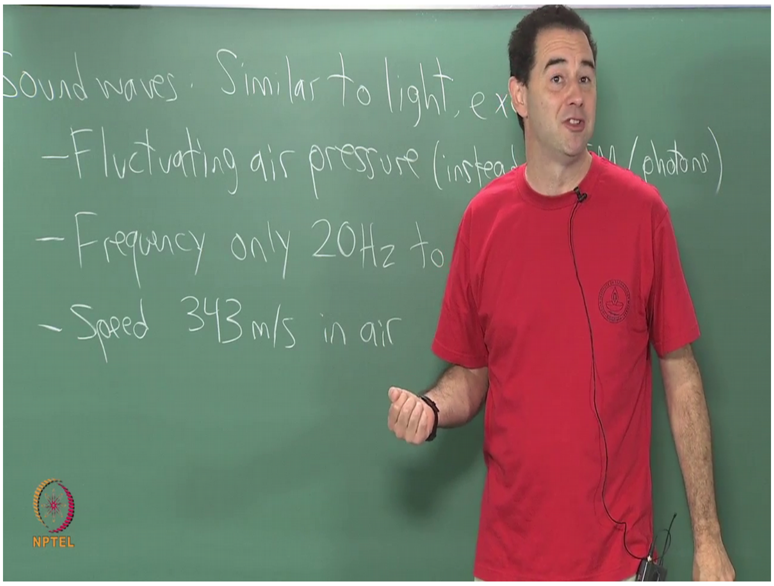


If we have a head mounted display, then the user has their ears covered through some kind of earphones and their eyes covered, through some kind of eye phones as we have said and so, this is what we get in the case of a head mounted display and we have to figure out what exactly needs to be rendered for the audio displays right.

So, you might imagine, these are some kind of a speakers that are placed, close to the ears and as the user turns their head. How should the sound be adjusted accordingly, rights? All of the tracking methods that we talked about a few lectures ago, the information from that will also be useful here, when you have a head mounted display presenting audio

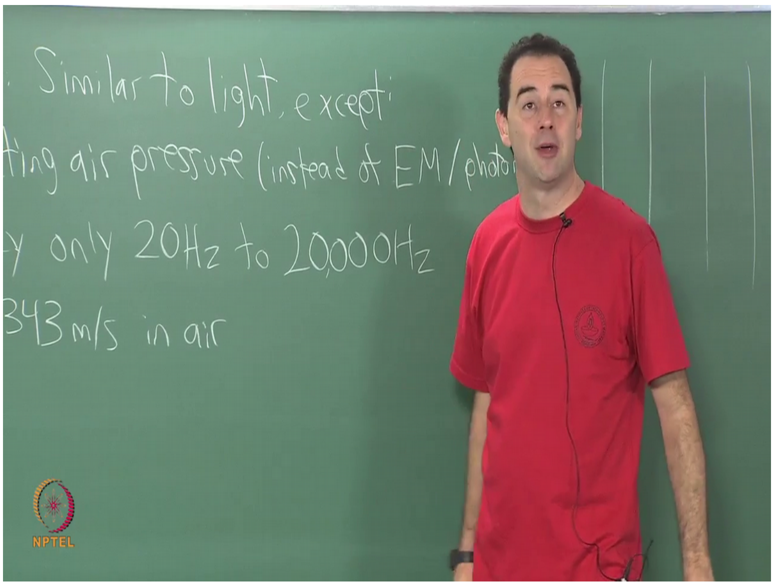
Because you have to change the stimulus that is presented to the ears, just like you have to change the stimulus, that is presented to the eyes and the same kinds of issues, such as, latency become important. There are resolution issues as well and other kinds of issues.

(Refer Slide Time: 02:55)



So, let us think about it. We have sound waves that propagate through the physical environment. We can think of them as being similar to light. In fact, I think it is valuable to draw parallels between the propagation of light and the propagation of sound and then compare the visual sense. I mean vision sense to our auditory sense.

(Refer Slide Time: 03:44)

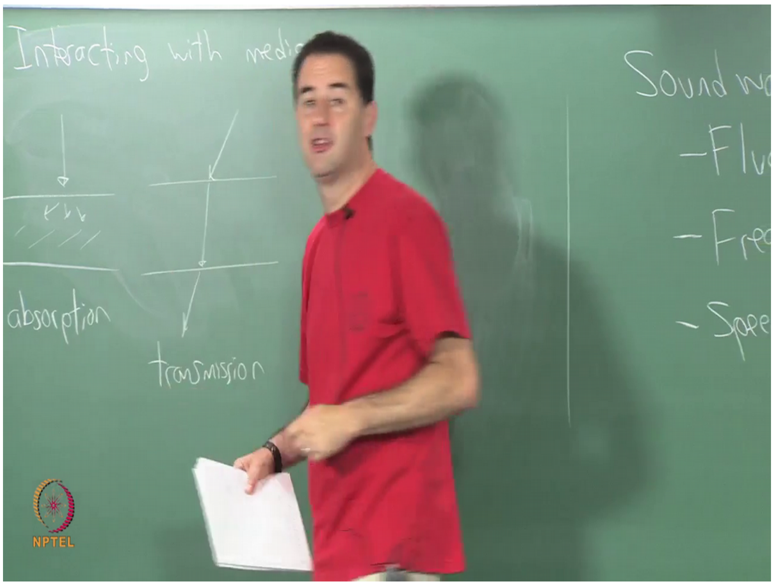


So, similar to light, except while, we have fluctuating air pressure, instead of, electromagnetic waves or photons and so, we are talking about fluctuations and air pressure waves, fluctuation in air pressure, which generates waves that have imaginary propagating waves. There is a compression part and there is a rarefaction part, where it is decompressed.

So, it is compression and decompression. These ways to move through space and the frequency and the frequency range is only from; let us say about as low as to 20 hertz to 20000 hertz, which in the case of light is up to 10 to the 14th hertz. So, much-much higher, frequency for a light; So, we are talking about low frequency and speed is much lower. So, say 343 meters per second in air, as opposed to three times 10 to the 8th for light.

So, the speed that the much slower a speed that sound propagates, compared to the light will make a significant difference, it will enable, certain types of perceptions that might not have been possible light. So, this ends up being, interesting and the human body uses, the fact that the propagation speed is slower in some cases, we will go back to this board.

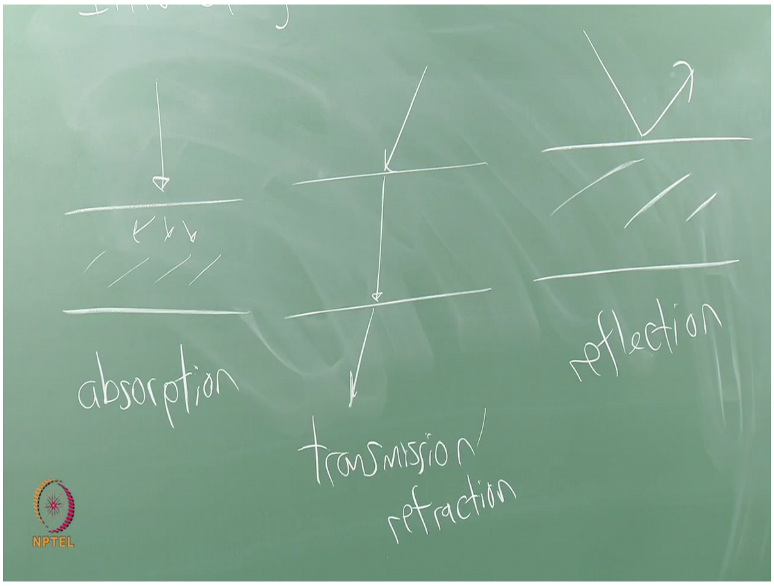
(Refer Slide Time: 05:35)



We have the usual, as we had for the case of light, interacting with, various media. You may remember, we had the case of light waves coming in and we talked about absorption, where they just seem to disappear into the material.

So, same thing is true for audio waves. So, there will be some amount of absorption, where the vibrations just get absorbed into the material. We also get some, transmission. It is very natural, we can hear sound through walls.

(Refer Slide Time: 06:33)



Some refraction of waves occurs in that case as well and of course, we get reflection of sound waves. So, we can hear an echo back, as sound waves are bouncing off, of a buildings. For example and for the fourth case which we had before; we can get a diffraction of sound waves.

And so if you are, if a sound source starts here and you are standing here with your ears, you should be able to hear something right. So, there is diffraction as well, it will propagate. The waves will propagate around corners, bending, there is some significant difference at this point, when I went through light, I started talking about optical systems right. We talked about using the power of refraction, in order to make lenses that have some focal lengths. Why do not we talk about lenses for audio, you may thought of that.

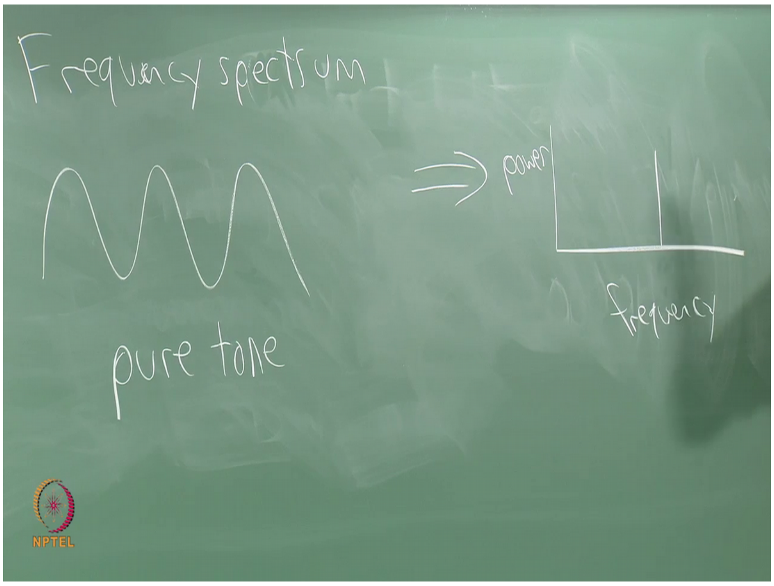
So, you could try to focus the, the sound waves in some way to generate some kind of image. So, that you can tell exactly, where the various, a sound sources are, right. Where the sound is coming from by location along some kind of audio receptors that you have right; So, we do not have that, we have not seen that kind of thing, before you may have seen cases of, generating, Let us say, constructing some parabolic surface and then focusing the sound.

So, when you put your ear right, at some special nodal point, you can hear, all of the audio, all of the sound waves, that are being propagated, within a large region. So, that you can hear someone whispering really far away; For example, you may have seen something like that, would be the closest thing to a lens, in, in this case right.

In the case of audio, but I just want to generally point out of that the human vision system and in the same way, that we handle audio in the real world. We do not make things, that are equivalent to images right. We do not have a audio images that go pixel by pixel and each audio pixel would be based in this hypothetical case, on the precise location right, that we are, perceiving the sound coming from right.

So, we do not have something equivalent to that, presumably, it is, because of the large size of the waves in comparison to the tiny size of the light waves. So, that makes it very difficult to have some sort of high resolution audio image.

(Refer Slide Time: 09:17)

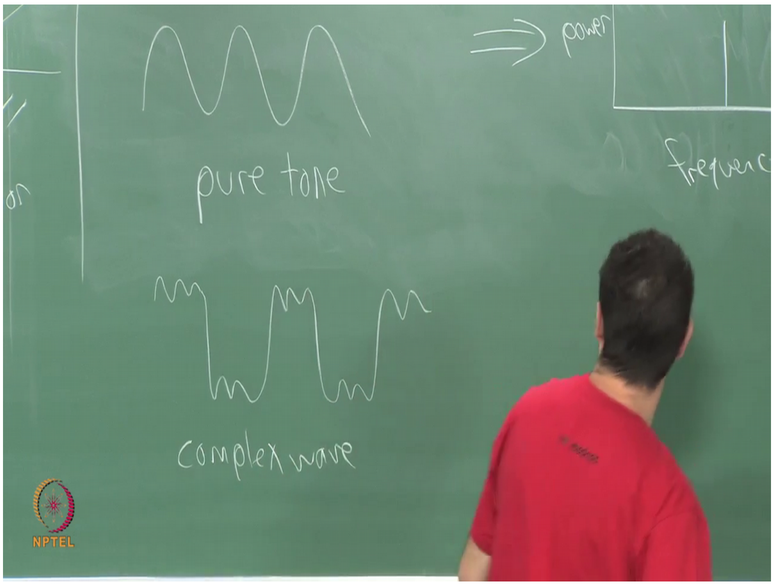


We do end up though with frequency spectrum, a spectrum becoming quite important, which is just like a light. So, we may have a, a pure, a pure tone of sound, which corresponds to a single sinusoid and in the, frequency spectrum. We may get then some particular spike that corresponds exactly to the frequency of the sinusoidal, wave front.

So, this would be, let us say, maybe the amount of power on this axis. So, the larger the amplitude, the higher this gets, but the location here, along the x axis, the horizontal axis is based entirely on the frequency right.

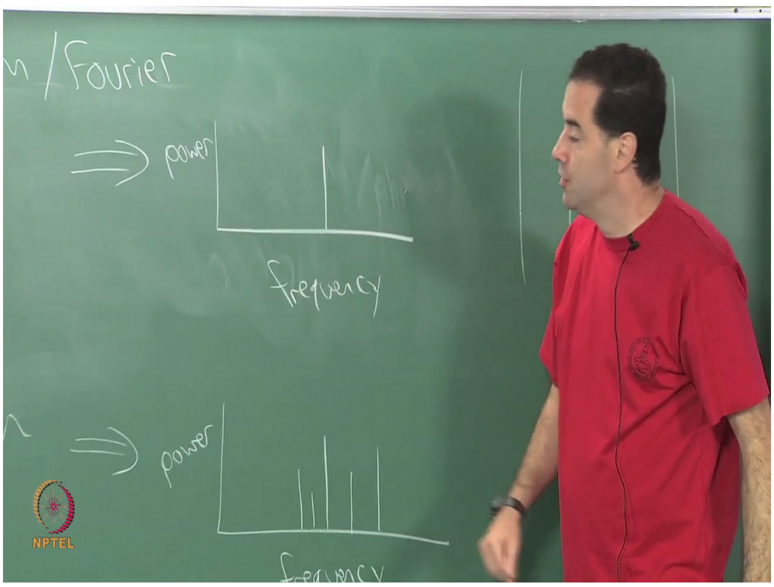
So, this would be analogous to the colors of the visible spectrum; that we talked about right and if you have a more complex wave.

(Refer Slide Time: 10:27)



So, maybe, some more complex wave, then the frequency spectrum will be more complex.

(Refer Slide Time: 10:40)

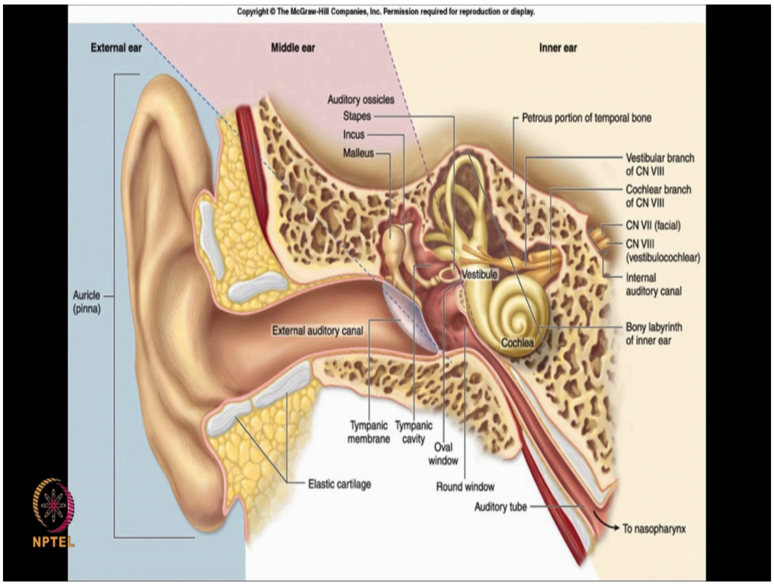


It may have a several components. So, a more complex wave can be considered as using, a frequency spectrum or Fourier analysis, as some linear combination, some adding of sinusoids that have various amplitudes and there can also be phase shifts as well.

So, I am not really showing, how they are shifted across time. So, you can imagine, taking some number of sinusoids different frequencies, different amplitudes and different phases, adding them together to get some picture like this right. You can eventually go and even represent square waves, if you put together an infinite number of these right. So, maybe, you have had some background in signal processing maybe not, but if not I just wanted, at least leave it at this stage, where just be aware that just like for light.

We have, a frequency spectrum corresponding to the waves, that are propagating and, we have the simple case of pure tones and we have the more complex case, which of course, is more, common in general for very a complex ways. There is more, there are more components in the frequency spectrum. All right, I want to describe the human auditory system. Now, I am going to go over some, pictures for that and then we will get into the perception of sound.

(Refer Slide Time: 12:10)



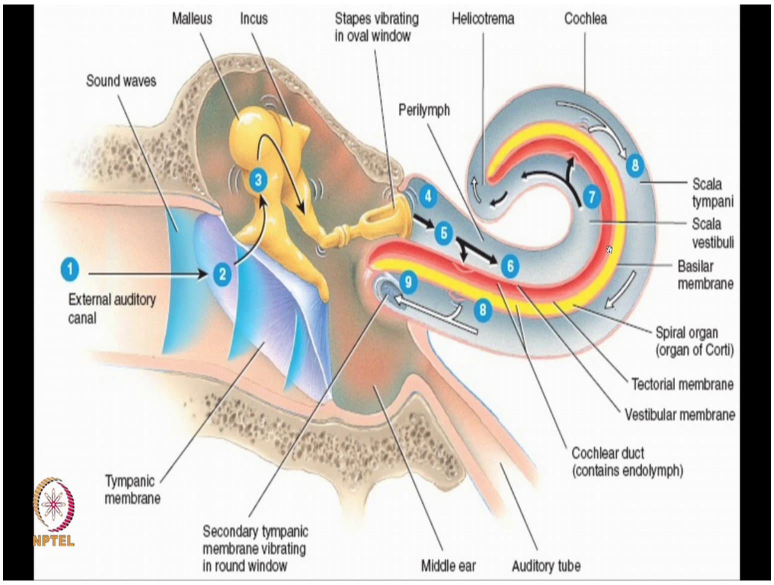
So, the outermost part of the human ear is called the, the pinna, this is like a kind of funnel that directs the sound into the, external auditory canal. You have a tympanic membrane here, which we also know as the ear drum informally. So, the air pressure waves come in, they cause the eardrum to, vibrate as a result of the sound pressure waves, that have be, come funneled in and then there is a sequence of a three bones here. So, this is some kind of kinematic structure, that in some ways translates or transfers. These, vibrations of the eardrum or tympanic membrane over to this part here, which is the entryway to the cochlea and inside of this part, this the cochlea part, here is a fluid.

And so, what are waves vibrating, corresponding to air pressure waves out here. They become fluid pressure waves on the inside here and there is a sequence of bones here and some kinematic structure to connect these. So, that there is impedance matching between the waves on the outside the propagate, through the air and transmitting those waves into the fluid that is inside of here. So, if you do not do some kind of, impedance matching, then the air waves would very likely just bounce away, rather than being transmitted. So, this complicated bone structure that we have inside here is to transmit the ways from air to fluid.

So, once they enter inside of this region, which I should point out by the way. There are two different parts here and I will say a little bit about this whereas, the cochlea part, which is for, auditory sensing and then there is this upper part up here and you can see, what I called the semicircular canals and this is part of the vestibular organ.

So, you are hearing and vestibular senses are very closely located and followed a very similar evolutionary path. There are a lot of similarities between them in many ways, you may be able to consider the vestibular part as being the very-very low frequency component of your hearing.

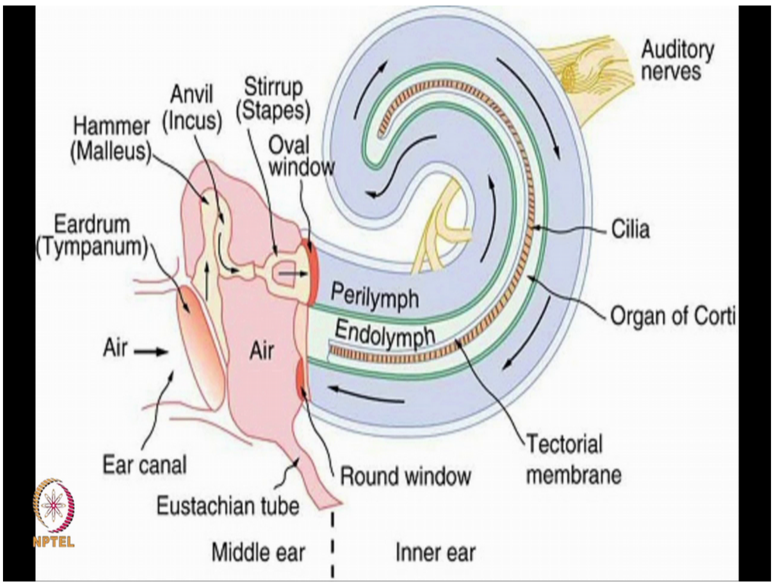
(Refer Slide Time: 14:19)



So, here is again another depiction. You have this external auditory canal; you have the tympanic membrane vibrating.

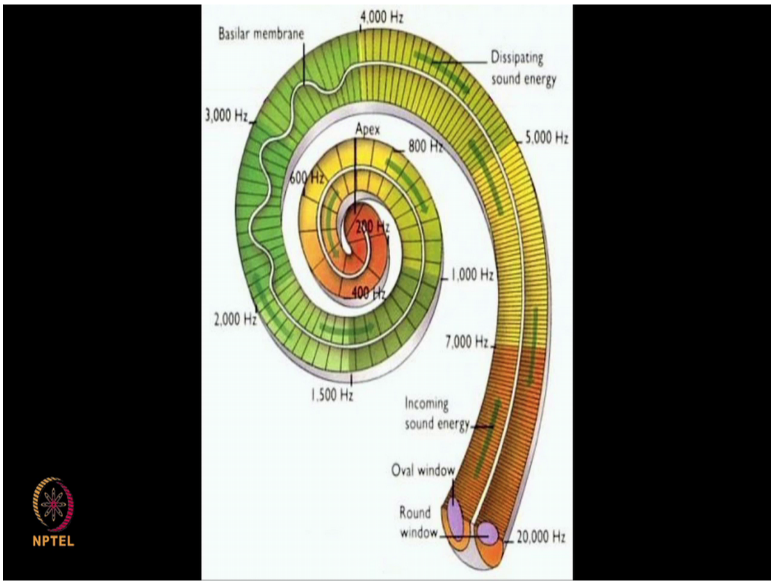
You have this sequence of three bones and then that causes a vibration here and then there is this long coiled canal of fluid that has an inner part and then the outer part and this is coiled up, much more than the picture indicates and then along the interior inside of here, there is a, a basilar membrane that contains, sensors.

(Refer Slide Time: 14:51)



Here is another picture of this as well and. So, I want to talk about, what goes on as their fluid vibrates here. There are, there are pressure waves, being sent through the fluid and these can be at varying a frequencies, just as I said there is a frequency spectrum as you see on the board still there.

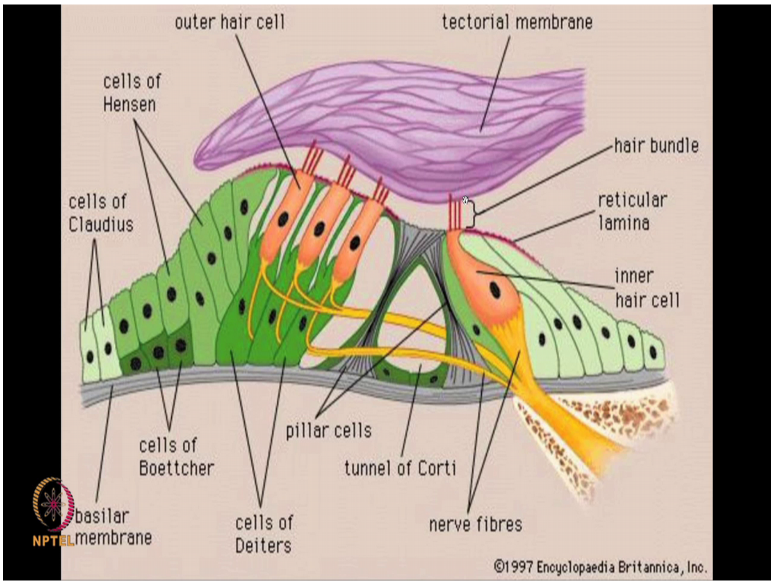
(Refer Slide Time: 15:11)



And so, there are cilia or hairs inside that are responsive at different frequencies. So, this particular center part inside of the cochlea, this membrane and it goes along, is narrower and stiffer. I do not think this picture exactly shows it, but it is narrower and stiffer, um, near the beginning and as it goes further in, it becomes wider and more flexible that makes it more responsive to lower frequencies; When you are all the way near the end and responsive to higher frequencies, when you come out to the, to the outside.

So, the cochlea is very much like a spectral analyzer and it has sensors inside. Let us say that respond. These are mechanoreceptors instead of photoreceptors, because it is responding to vibrations. So, it has mechanoreceptors that respond to sound by producing this perfect or beautiful. Let us say, frequency decomposition of the ways based on what location, the waves are inside of this, inside of the cochlea um.

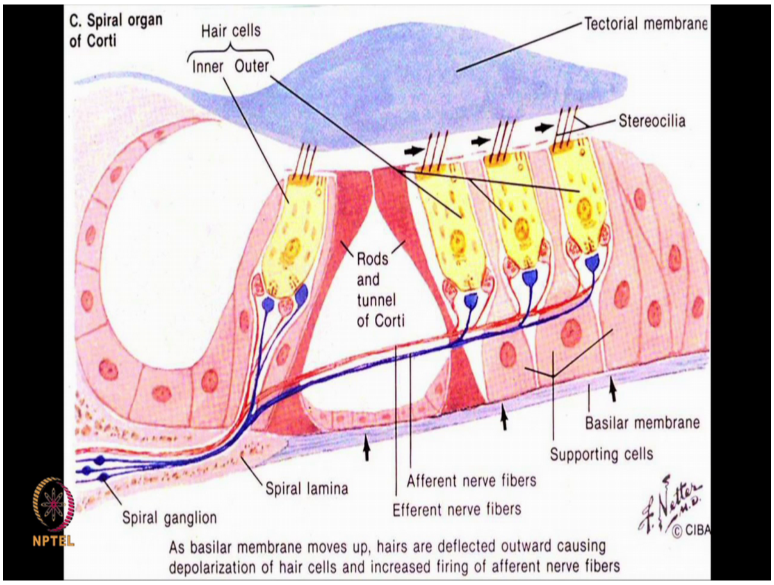
(Refer Slide Time: 16:09)



If we zoom in significantly to the center parts, if we are looking at a cross section of this inside, then there is a displacement that occurs, due to the vibration, inside of this a fluid channel and these hairs bend a bit and this causes, based on the frequency of vibration, causes neural signals to be sent to the brain and just as there is an optic nerve, there is an auditory nerve and just as there is a visual cortex. There is an auditory cortex that transmits the information.

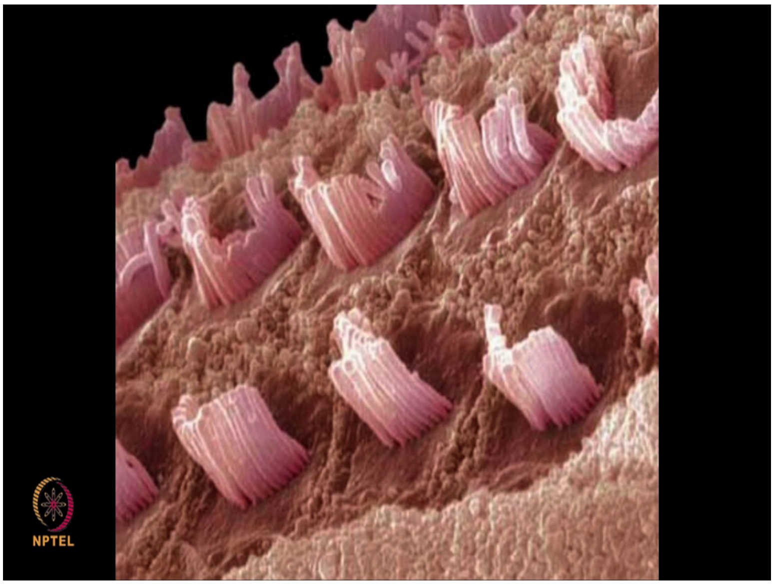
And so, based on motion between this plate and this upper tectorial membrane; These vibrations are transmitted through, this, what I, which, what I said is a, mechanoreceptor as opposed to a photoreceptor in the case of the eye.

(Refer Slide Time: 16:57)



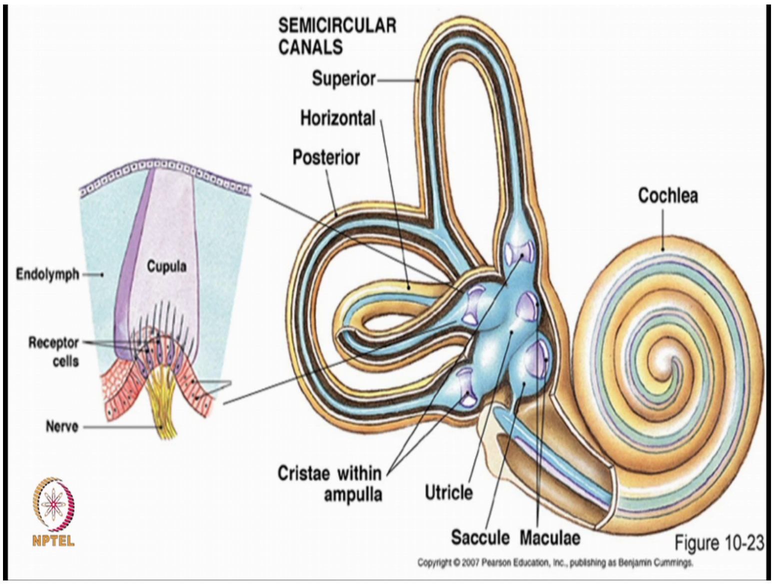
This is another depiction of it as well.

(Refer Slide Time: 17:01)



And, this is what it looks like, under electron microscope.

(Refer Slide Time: 17:06)



And I just want to mention this, vestibular part as well, is everything ok. There I am hearing sounds outside of the audio track. So, everything is fine all right. So, I also, I also want to mention the, on the vestibular part as I showed in my early picture back here., the vestibular organ is here as well I just want to use this opportunity to mention this additional sense, which becomes extremely important for virtual reality.

I would say more. So, as a nuisance right, because, when you have visual stimuli that are in conflict with your vestibular organ as I said that is the, the path towards simulator sickness and discomfort in virtual reality. So, what is quite interesting is that there are mechanoreceptors as well in this portion.

So, the cochlea part is for audio, but this other part here is for vestibular sense. There are three semicircular canals. They have the names here, superior horizontal and posterior what is interesting about these is that, they are mutually orthogonal circular canals. They are only off by I would say within a couple of degrees of being orthogonal.

So, I find that really impressive. So, that is a three axis sensor, that essentially is an alternative to the gyroscopes that we have in our men sensors. It is measuring angular acceleration in three independent axes close to orthogonal.

Basically, what happens is as you turn your head around, there is a fluid that compresses and decompresses along here and let us say it is not a wave vibration like in the case of sound, but it is a very let us say low frequency wave measure in the, what the flu is sloshing back and forth along the canal, that causes pressure on these membranes here and there are again sensors that are mechanoreceptors that, have little hairs that move back and forth as a consequence of this fluid moving. So, that transmits angular acceleration, because of the fluid displacement.

So, it is not fascinating, there is some viscosity to the fluid and. So, as it moves back and forth, it causes some very low frequency pressure waves in addition to that, there are these two, parts here, which are comprised what is called the otolith organ. There is the utricle and saccule and these have this again the same kind of mechanoreceptors, but these correspond to measuring of, some amount of pressure that is due to gravity or linear accelerations, which remember those two cannot be separated.

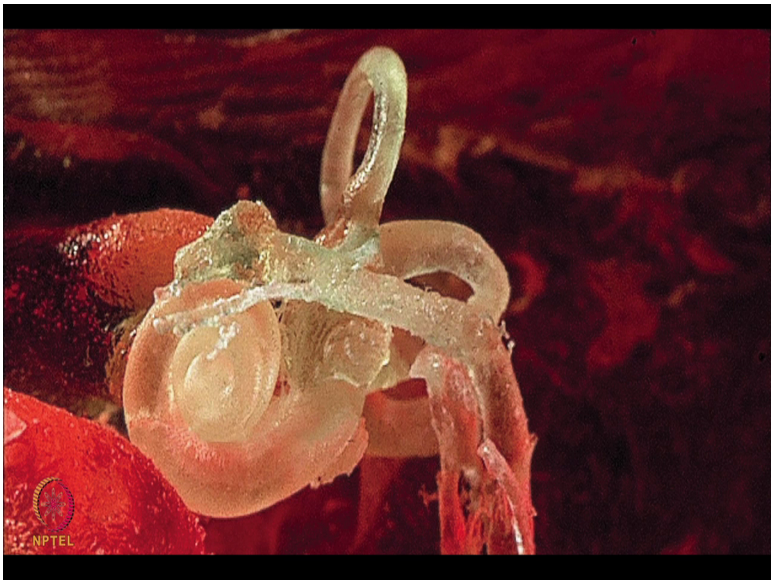
So, it turns out the utricle and saccule are the accelerometer components, um. They are oriented 90 degrees with respect to each other. So, that, there is a planar surface and the cilia are hairs displace.

So, you can measure two axes with one of them and you can measure two axes with the other. You get four axes total, but of course, you only need three independent axes. So, there is a little bit of redundancy. So, you get the ability to measure both linear acceleration and angular acceleration, remember that the gyros, we use measure angular velocity. So, it is a little bit different, but it is, it is essentially the same information via integration of angular acceleration. We get angular velocity information and by integrating that, we get, orientation information.

So, I find this amazing, there are two initial measurement units a one behind a each one of your ears right. So, we have this while, at the same time when we put an HMD on our heads. We put an inertial measurement unit inside of that, to measure orientation and then we seem to get upset, when the brain using it is own inertial measurement unit has recognized that.

There is a mismatch with the visual signal that we are sending right. So, I think it is very interesting that we have an engineer IMU and a biological IMU and they are measuring pretty much the same thing, which I find really fascinating.

(Refer Slide Time: 20:47)



All right, that is what it really looks like, if you were to cut out the vestibular organ and the, the cochlea as well. So, locally it is not right before lunch; all right. So, that finishes with the pictures. Now, I want to talk about the perception of sound remember, when we covered a vision, we had photoreceptors that are explaining the physical part.

You have the light waves coming in, you have photoreceptors and then you had layers of neurons that are performing some kind of low level filtering and the filtering or inference gets higher and higher level as it goes further and further back into the brain, into eventually the visual cortex in the vision case.

So, a similar thing happens with audio and the information goes back to the auditory cortex and as usual there is seems to be some hierarchical processing and higher and higher level information, gets inferred or perceived as, as these signals get further from the source from the actual sense.

So, I want to now, talk about perception of sound.