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

Lecture - 17-2
Audio (auditory localization)

So, next part I want to talk about is audio or auditory localization. How do we know where sounds are coming from? Let us compared with the vision case, how do we know where the light is coming from? I guess I can just turn my head and look right.

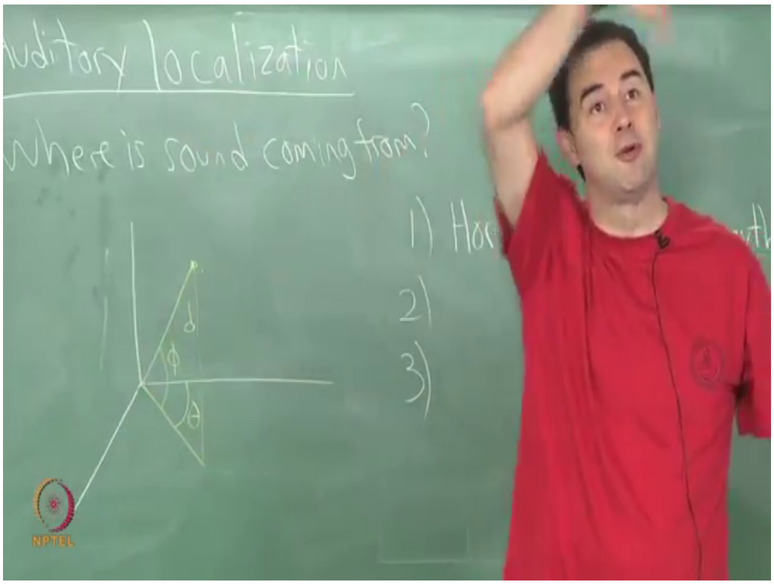
So, I someone talking to me, I look at them, I can see the, I can see exactly where their eyes are right. So, I have a lot of information I can use I have the information of how my head is oriented which our bodies are keeping track of, the eyes are oriented where the image appears on the retina usually it is appearing on the fovea I am looking at something.

So, I have all of this information. I know where the light is coming from the particular stimulus that I observe right if I am reading. The clock on the wall, I know exactly where that is at relative to me in terms of angles and might not know the distance, but we had depth perception right we talked about that.

So, if I know the scalar of something maybe based on the size of the image on the retina, I can estimate the distance. You can also do psychophysics studies to figure out how good you are at estimating the distance based on the size on your retina right. So, we can do these kinds of things.

How do we know where sound is coming from using only our ears? And if you close your eyes, can you determine the source of audio? All right. So, every morning when I wake up, I hear the loud Asian Koel the outside on the campus here, very loud bird. I feel like I can narrow down where that bird is to sit a degree or two. Do you feel like that? And you tell where you can usually find the bird in the tree right. How are we able to do that?

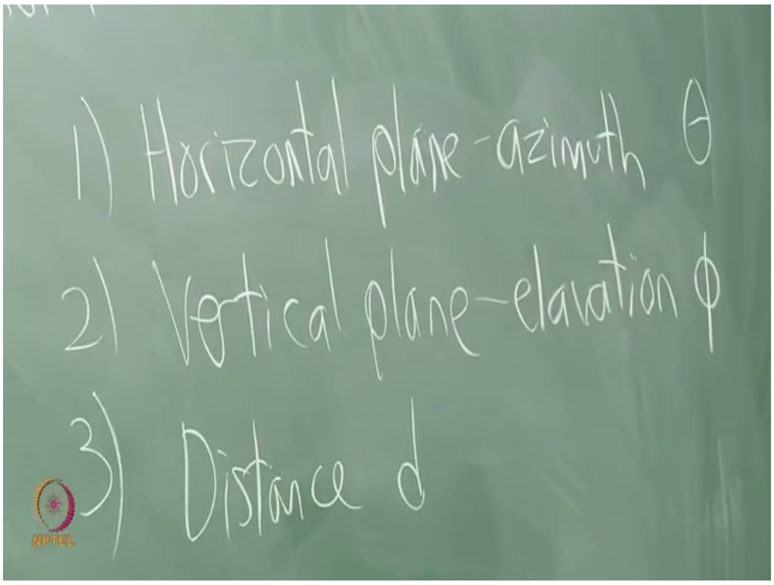
(Refer Slide Time: 01:52)



The auditory, so auditory localization is a very important part of perception. If we are going to make a virtual reality system that produces virtual sounds; and if in the real world, we have the ability to localize to figure out where sounds are coming from we better not mess that up right, we better not fail when we do it in virtual reality. So, how do we how we get these things? So, that is why understanding is going to be very important.

So, where is sound coming from? We can generally think of three coordinates for that. Let me now draw the a kind of coordinate system here. Suppose, this is the location of the ear for example, and we have some distance d that the source will be from this from the origin. And then project down into the plane here, we will have an angle theta; and an angle with respect to the plane coming upward which I call phi.

(Refer Slide Time: 03:30)

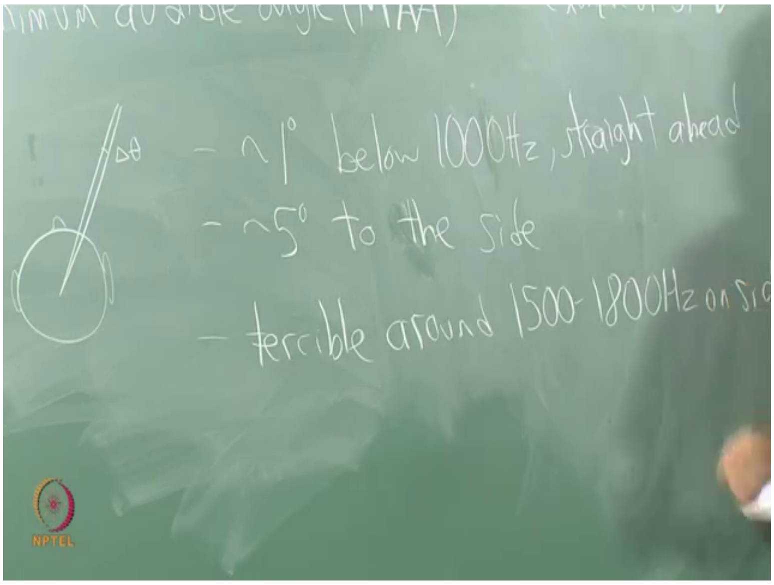


So, all these three components, so we have say 1, 2 and 3, the horizontal plane direction which is called azimuth which I have represented there as theta right. So, we have the horizontal plane, so just some direction is 0 to phi, where is the sound coming from. It seems related to yaw in the when the coordinate systems we have been talking about for head transformations.

We have vertical all right. So, how high or low is the sound, this will be called Elevation. I represented that with phi. And then we have distance which we have represented with d right. So, just using spherical coordinates does not matter right, all right.

So, this just ends up being convenient for where ears are varies in the type of information that we get and are able to infer that, that allow us to resolve where does sounds come from. I am going to give an example of a just noticeable difference with regard to audio localization.

(Refer Slide Time: 04:59)



It is called the minimum audible angle or MAA which is an example of just noticeable difference. This would be exactly as I said trying to localize where is the bird sound coming from.

And so in terms of the azimuth, so we have the head. And let us suppose it is I guess I will draw kind of a nose here. So, we are looking top down. And I want to understand what is the smallest angular change here delta theta that can be detected. And if you make a very small change you ask people to tell whether or not it is in fact moved.

So, one thing that is interesting is when it is closer to the front, we are much better added; when it gets to the sides, we are not as good added all right, so that is one thing to pay attention to. So, when we get up to looking straight ahead, so it is around 1 degree if the stimulus is below the 1000 hertz, and straight ahead. It is around 5 degrees to the side.

There is some exact path in the book that contain much more information I just want to give you the general idea to point out that just based on the geometry of the ears and the way the sense is developed there is no simple answer, it is not always 1 degree. So, it depends on the frequency, it depends on where the location is.

So, once you go over to the side it goes from 1 degree up to about 5 degrees, and it is still varies depending on how far over you get to the side turns out that we are terrible perhaps completely unable to localize the direction of sound around 1500 to 1800 hertz on the side right.

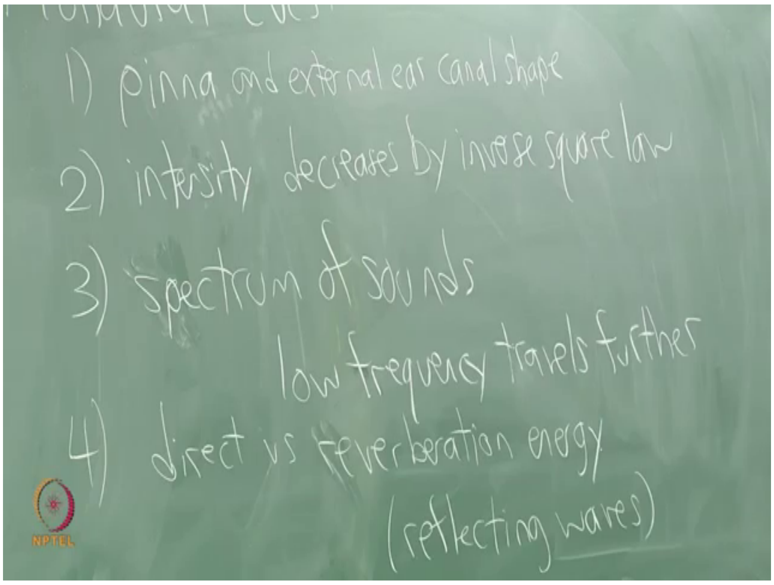
So, if we have sources on the side, we have a very difficult time localizing them in that frequency range. So, so there is no simple answer right, you have to take into account the frequency and where the direction is at (Refer Time: 07:52) your head, and then you can answer questions about the minimum audible angle. So, how much can you change the direction of the sounds or sensitive be able to tell that. So, I find that interesting.

So, if you were going to try to reproduce that in virtual reality and wanted to do some experiments, you could have people put on a virtual reality head and close their eyes listen to where the sound is coming from. And see if you can match that from the real world; maybe I record the Asian Koel making it sounds and then I try to somehow reproduce that in a speaker system. And see if I can get humans to respond and give minimum audible angles in the same way right that is how you would know if you have done it right. Knowing that you have done it right in audio seems significantly more challenging than with video right.

In the visual case, you look at the images and you see all I see pixels or all the colours do not look right it is very easy for us to just give simple feedback. And make some kind of heuristics or hacks that seemed to work well enough. Here it may be quite complicated things might sound right it may sound all right, but you might not be completely sure unless you have done systematic experiments to make sure you have reproduce the sound in a way that is as close to possible as in the physical world.

Now, remember in the case of visual we had we had depth cues for example, right. So, for depth cues we had monocular depth cues and binocular depth cues correct. And I said that I am with an emphasize binocular depth cues, and I really had to emphasize to you the monocular depth cues. So, the same thing is going to happen here, we have monaural cues for localization, and we have binaural cues.

(Refer Slide Time: 09:43)



So, let me go over some monaural cues. One, we get a significant amount of information from the pinna that is the shape or geometry of this outer part of our ears, and the external ear canal shape. So, basically the funnelling part provides a significant amount of information about where is sound is coming from. So, a kind of signal processing filter or transform is performed by outer ear.

I will get into more details of that shortly. But I just want to point out that is a significant amount of information that let us determine where sound is coming from. It is distorted in different ways across the frequency spectrum depending on where the sound is coming from just on how it the sound waves propagate through your pinna and external ear canal.

Two, the intensity decreases by the inverse square law we talked about that in tracking systems for light. So, same thing for audio; this may be equivalent to a monocular depth cue that has to do with the retinal image size right.

So, if you know how loud something should be again maybe it is the Asian Koel and you know how loud that bird typically is. If I can barely hear and it is probably far away right I do not need two ears to determine that I just need one ear to determine that. If it is a very unusual sound, you have never heard before, maybe this cue will not be so good because you are not sure how louder supposed to be anyway.

Three, the spectrum the spectrum of sounds so; it turns out that when you look at the frequency spectrum of sounds lower frequency components tend to travel further through air.

So, if you hear thunder in a lightning storm, I think we had one last night the thunder when it is very far away just sounds like a low frequency rumble. When it is very close, you hear the high frequency components, so that is a distortion in the frequency spectrum that gives you a cue as to how far away it is right. So, low frequency travels further or there is less dissipation let us say for low frequency as there would be for high frequency so that is a kind of distortion. So, that is as there is a kind of filter applied to it.

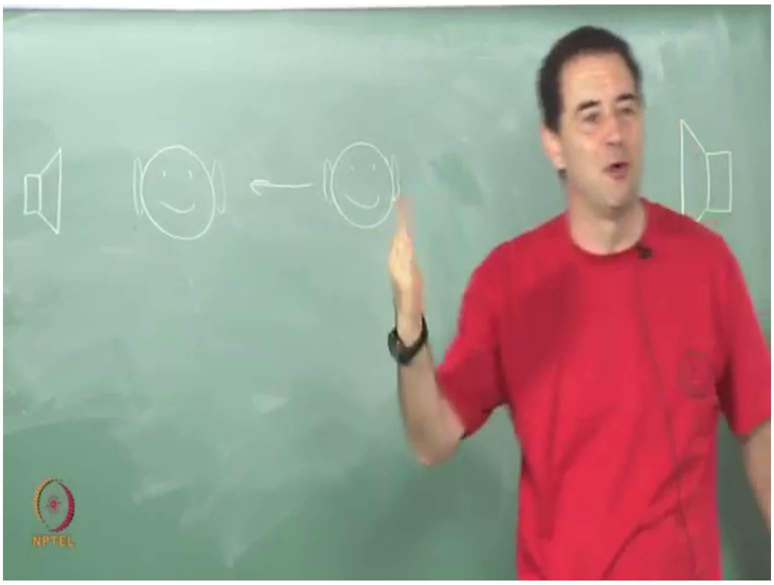
And finally, there is direct versus reverberation energy. And so in the case of reflecting waves right reverberation energy. So, as I speak in this room, my voice is bouncing off of the walls tables off of you, there is reverberation energy so that is causing phase shifts in these waves. So, you are hearing multiple versions of me at different times.

Do I seem to be echoing to you in the room very strongly? Not really, not too much right. You can hear echo in many cases right, but I do not think does not seem like unless we are in some enormous church hall for example, you may hear some echoes because at a large amount of distance that the sound waves have to travel before coming back you may perceive this kind of temporal displacement. So, you may hear second, third echoes coming back as I talk.

This leads me to give you an a interesting example. So, remember that we had optical illusions right for the case of the visual sense. Should not there be audio illusions as well all right, if you have heard of any audio illusions? Let me give you a simple audio illusion that is related to reverberation it helps you to understand this cue.

So, let us put up some a stereo speaker system. So, I have a speaker here; and I have a speaker here. They are supposed to look like the same size. And let us say we put our head here in the middle right. So, we are listening two sounds from the speakers. We do this all the time right. So, I get some kind of stereo sound.

(Refer Slide Time: 14:19)



Let us suppose I just transmit the same sound to both speakers just no worry about stereo separation. You are listening to some kind of music it is really mono music separated and I am just putting the same audio track out to both channels. So, I have a left and a right track, I guess maybe that is the left track and that is the right track. So, everything seems fine.

Now, what I going to do is move my head over here right, and I ask you do you hear the speaker at all. If I go over here, it should be the case that I hear the sound from this speaker and then this one comes in significantly later and there should be a time shift right, because this one is travelling further away. But I do not hear that, do I?

I just hear the sound from this one speaker; I do not hear them both. Try it sometime. Arrange two speakers walk back and forth usually hear the sound. You have done this before; some of you have done this before I think right by just making this up. You try this before you ever notice this you get really close to one speaker you hear only that you get out into the perfect place and you hear both of them you like, wow, this is perfect this is where I am supposed to be. Then you get over to the other speaker and you only hear that one.

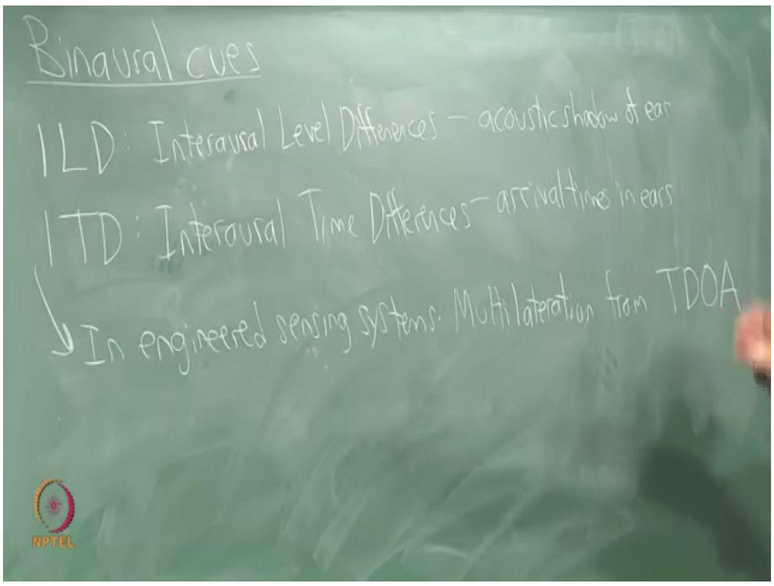
Of course your ears are taking in the sound from both, but your brain is masking away the reverberation from the other one. Because it is a secondary effect, it is essentially it is perceived as the same audio, but it is time shifted and it is lower amplitude. So, it just gets massed away it is an audio or auditory illusion.

You do not hear the extra echo from that or the time shifted version of that that it falling onto ears. If I were going to suddenly while I am over here, turn off the speaker of course, I would hear that one right, but you do not perceive it at all when you turn on this one on. So, auditory illusions any questions so far?

I want to talk about binaural cues now right. So, just like we have stereo with our eyes, we should have some kind of stereo with our ears. Another interesting part if we want to compare two eyes is we had a vestibular ocular reflex do not we should not we have something like a vestibular oral reflux or not? Why do not we have that? This some animals can rotate their ears right. And so I think horses can do that, for example, cats can do that.

So, if you could orient your ears then you should be able to also have your vestibular signal connected to that from your vestibular sense, so that you could orient your ears to keep them pointed at some audio source right. We do not have that. We cannot even reorient our ears, but some animals can do that. So, I am just pointing out some of the interesting differences I think it is nice to compare the reason and understand the differences. So, we do not have a vestibular oral reflex could have happened, but did not seem so important for our survival I suppose all right.

(Refer Slide Time: 17:55)



So, let us go to binaural cues. So, there is two concepts here. One is called ILD, which is a inter interaural level differences. One thing that becomes very important this was called the acoustic shadow of the ear all right.

So, if I am facing this way, and if one of you are to speak it is much louder for this eras than the ear that is in the acoustic shadow right. So, just as if it were light, the sound waves as well due to some diffraction I may get some bending around the corner due to reverberation of the board I will hear more, but generally speaking the sound should be louder for this ear than for this ear so that is interaural level difference so that is one very important binaural cue. These are just like the back faces when we talked about rendering right.

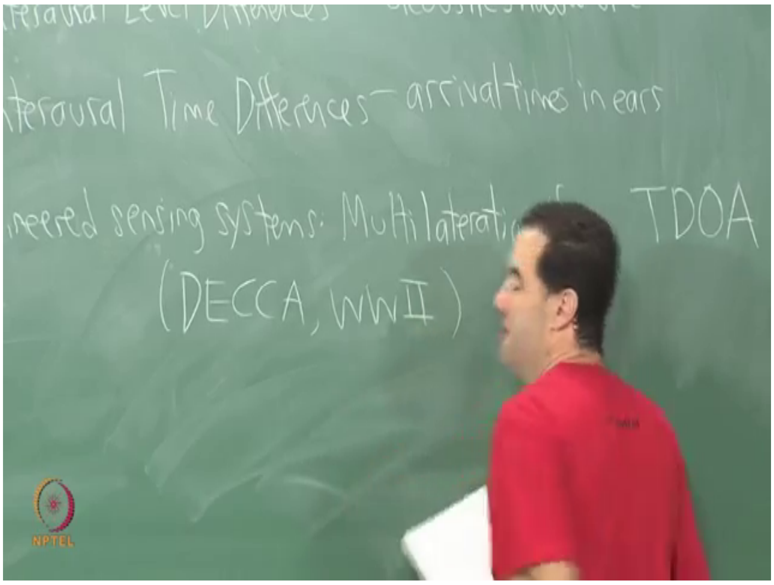
And the other one is ITD which is interaural time differences. So, this is based on different arrival times in the ears. Just for reference the distance between the ears is about 14 centimetres depends on your head size of course. But so there is that amount of distance maximum, and you think about a sound source right maybe coming from 45 degrees away from centre, and it hits this ear first and then this ear all right.

So, based on that time difference, believe it or not our brains our neural structure is resolving that temporal difference and it is using it is measuring that temporal difference or the phase shift between these waves that are coming in. And it is using that information to determine where the sound is coming from.

What I find really interesting about that is that the same thing has been done in engineering for a long time. So, if you have studied sensing systems in engineering, this is called so in engineered sensing systems. This is called multi lateration using what is called time difference of arrival or TDOA.

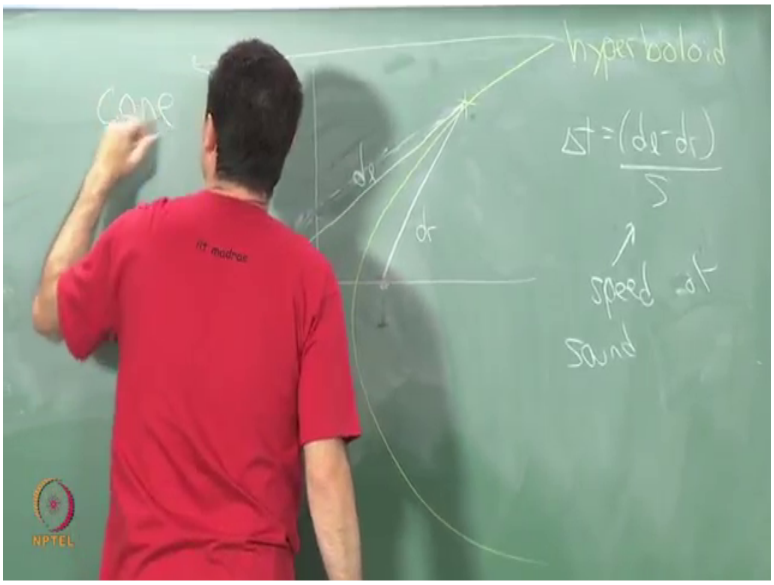
So, if you want to read about the engineering of these systems, you can go ahead and explore it this way. Just look up multilateration and time difference of arrival, so if I have some transmitter and it is transmitting sound and there receivers out in a field somewhere, you can figure out where the transmitted sound is coming from. Interestingly enough this was even used in World War II it is called the DECCA system for submarine localization from World War II. So, you can look that up if you like as well.

(Refer Slide Time: 21:24)



Let me say something about the geometry of that, and then we can take a break.

(Refer Slide Time: 21:42)



So, I have receivers what say I mean two locations may be here and here. So, they have some distance between them as I said it is about 14 centimetre for human ears. And we make that a little bit better they should look like they are and kind of through the centre here, so and body here all right.

And now there is a sound source somewhere in the space. And then we want to look at we have this distance and we have this distance should be straight lines you know it look particularly bad, but I make there and so straight lines here. And so we have distance to the left and distance to the right.

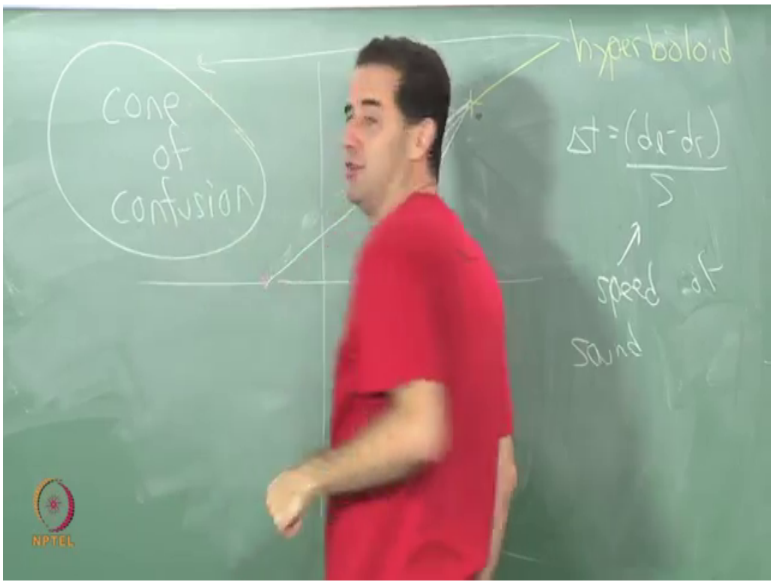
And now based on the difference in time of arrivals here right, there is some delta t that I get right some difference in time that should be equal to the dl minus dr divided by s which is the speed in the medium, speed of sound in this particular case right to shape speed or sound.

So, if I make some calculation like this, I now need to think about what is the set of all places right. If I work backward, so I started with the sound source and I said ok, we need to look at the difference between these two distances and that will give us a difference in time based on the propagation speed of the ways.

So, now let us work backwards. Suppose I have two ears they hear a sound source and at difference in time has been detected what is the set of possible places, where the sound could be coming from. And if you work through the algebra for that, it turns out to be a hyperboloid.

And generally we may remember from basic connect sections and analytic geometry, hyperboloids come in two sheets. The two sheets will be you will get one sheet if one signal came first, and you will get the other sheet if the other signal came for, so it depends on the actual order as far as which sheet you get. And I am drawing it in 2D, but you actually get a hyperboloid, it should be peaking right on the axis here.

(Refer Slide Time: 24:56)



And the hyperboloid is referred to by perceptual psychologists by the great name or the cone of confusion all right. So, this is the cone of confusion. So, there is a cone shaped region hyperbolic cone over which you cannot localize any further using only interaural time differences.

Now, one thing I find fascinating about that is that we can in fact determine where sound is coming from inside of the cone of confusion. Now, part of that is because we are using interaural level differences, but part of it is because of some more information that is coming later, but to give you just a hint of it, it has to do with the pinna. So, we can do more information, but if you are only looking at interaural time differences, you have a cone of confusion in region within which you cannot distinguish any further where the sound is coming from based only on this time difference of arrival of the sound waves questions about that?