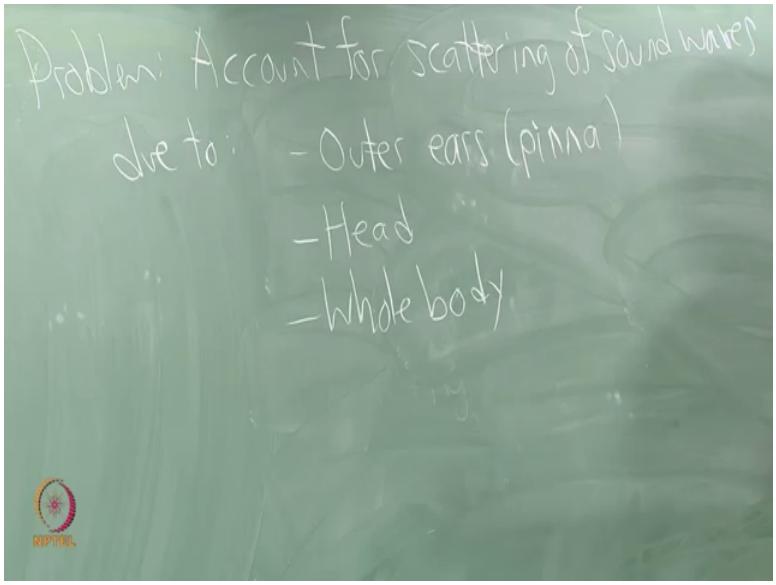
Virtual Reality
Prof. Steve Lavalle
Department of Multidisciplinary
Indian Institute of Technology, Madras

Lecture - 18-1
Audio (spatialization and display)

All right. So, beyond that there is an additional part and the value of this depends on a nber of things one is how much localization do you want to do, and another is where exactly is the display which I needed I think someone to talk about displays as the fourth component. Let me just mention it a bit here is the speaker inside of your ear canal or is the speaker out here somewhere. Another question is, is it a closed system like do I put a kind of closed headphone or earphone on you.

So, blocks out the outside sound and you only get what comes from the speaker or does the speaker just add to the outside sound coming in. So, there is different kinds of choices, but let us suppose that these stimulus is getting very close right up against your tympanic membrane right. So, if that is the case I am not touching, but very very close to your tympanic membrane.

(Refer Slide Time: 01:18)

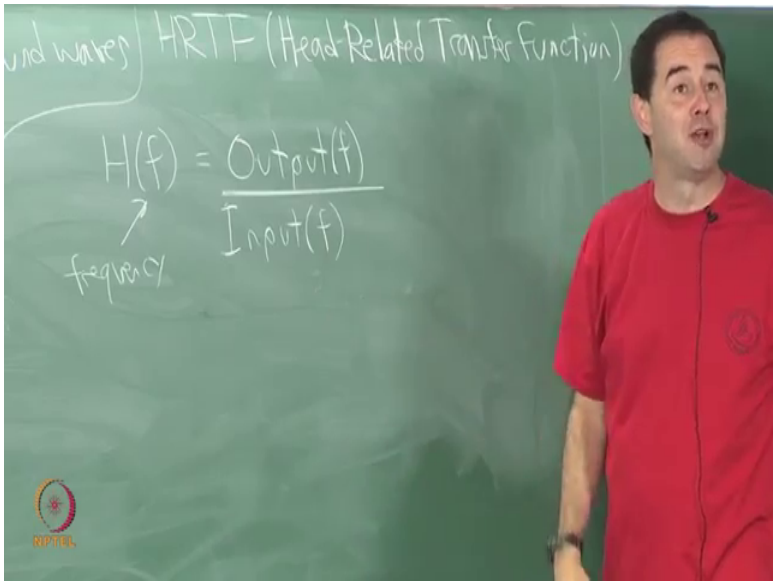


Well we have to account then for some additional scattering of sound waves. So, problem becomes account for a scattering of sound waves due to the outer ears where the part which I called the pinna, and the and the canal is well that goes into your inner ear, what

about the shape of your head, and your whole body. It could even be scattering a sound based on what clothes I am wearing today and maybe it is different from day to day based on what I am wearing could be different based on whether I am wearing a hat all right.

So, the sound will propagate differently that ultimately comes into your ears, you have to simulate the rest of that right remember when we talked about how close is the stimulus generator to the sense. So, if you make it very close, you end up having to simulate the rest that is that is around it right because on the back side of it you end up having to simulate it. So, how do you do that, how do you deal with that well people have studied this very carefully and they have come up with what is called an HRTF.

(Refer Slide Time: 02:48)



Well, sounds very similar to the BRDF when we talked about reflectance models as a similar kind of function, but not exactly the same; it is head related transfer function. And this is actually this extra amount of information the scattering that is due to your outer ear the pinna, the head and your whole body this is the extra information that we use in order to resolve the source of sound inside of this cone of confusion that I talked about last time. So, it is extra amount of scattering. There is a transformation that is happening based on where the sound is coming from and it is a transformation in the frequency domain.

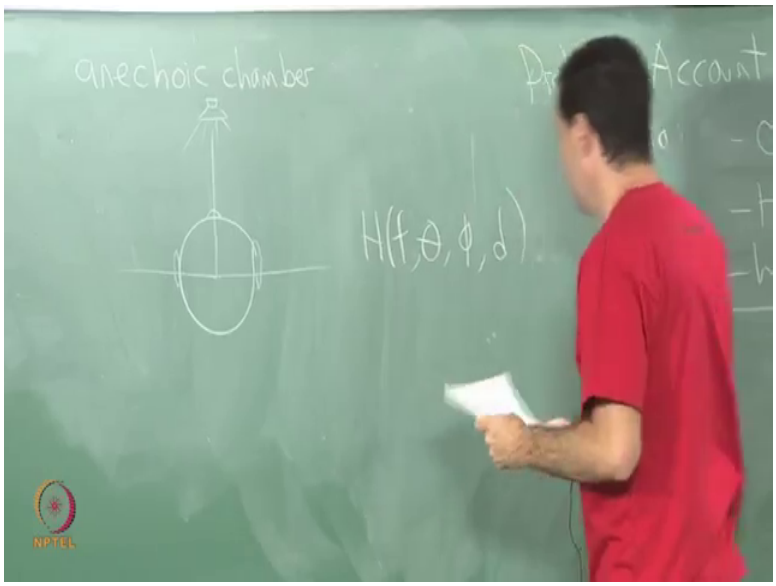
So, the way this function is represented H of f this is a function of frequency, and it looks like this. There is output as a function of frequency over input as a function of frequency. So, this is a corresponds to a linear filter. So, if you have signal processing background or electrical engineering kind of background, then this is just a simple well let us say complex example of a linear filter. It depends on individual ears. So, each one of us will have a different HRTF.

So, if I were to if you were to take off your pinna and swap it with one of your friends, then you may have for at least some short amount of time a lessened ability to localize sounds, especially inside of the cone of confusion. However, our brains due to the kind of plasticity and adaptability that we have, we may be able to adapt. Because if you all of a sudden put on suppose I put on some big Texas cowboy hat that will definitely affect the scattering of sound as it comes into the ears, but you can adjust to that. And after maybe who knows how long I will just make up a time, after 20 minutes of wearing the big Texas cowboy hat, you may be able to then localize sounds just as effective as you could when the hat were off.

So, so they are very interesting questions here about if I want to simulate the scattering of sounds, do you have to learn a particular HRTF that corresponds to your body or at least as close as possible or can you deal with just some generic one and will your brain adapt to it? I think there are a lot of open questions in here and express in audio rendering or studying these in over the last few years. So, it is very current area of researchers.

Certainly people at industry would like to know do you need to have customized HRTF per person or maybe just need three different generic ones and you just select the one that seems to be the best and it is good enough or maybe it does not matter very much as long as there is at least one reasonable representative HRTF that covers most cases and then your brain will just adapt to that after using it for a short amount of time. So, these are the kinds of questions that exist for this.

(Refer Slide Time: 06:20)

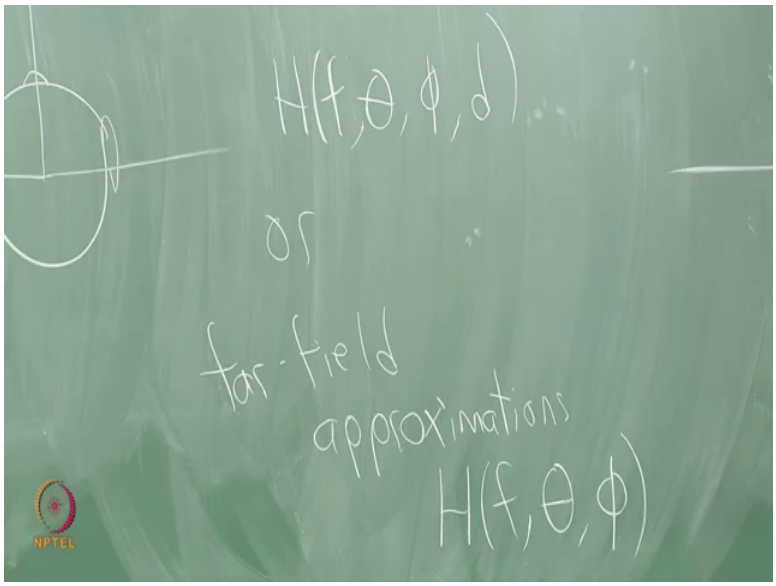


Let me show you how these are measured in a studio. So, you have you have the head, I will put some ears here, and suppose it is looking upward or just have a top down view. And we place the subject in what is called an anechoic chamber that means that the walls all around the subject are fully absorbing the sound, and so there is no echo back. So, it is absorption for the materials all around.

And then inside of this chamber, you place speakers you have the ability to place a speaker that generates a sound source, and you put it at different locations perhaps every 15 degrees, you put the speaker, and then you generate an impulse, just a single impulse and you look at the impulse response. So, this is standard way of designing a filters. So, you look at the impulse response from that particular location, and you record it across the frequency spectr.

And then you move to another location and you do the same. You can do this at various angles both in you know across the the horizontal direction and the vertical direction. You can also look at different distances or you can just make a simplifying assption that these sounds are coming from sufficiently far away then I am not going to worry about distance. So, so it could be that we measure this HRTF function of frequency, we look at these two angles horizontal and vertical and some distance d. So, as I said we could put these speakers further and further away.

(Refer Slide Time: 08:18)



Or we use a far-field approximation. This is like the assption of parallel wave fronts for a light which it does correspond to this case, in which case it just simplifies down to only taking measurements based on we characterizing this transfer function here in terms of frequency, but it only depends on beyond that theta and phi right, all right.

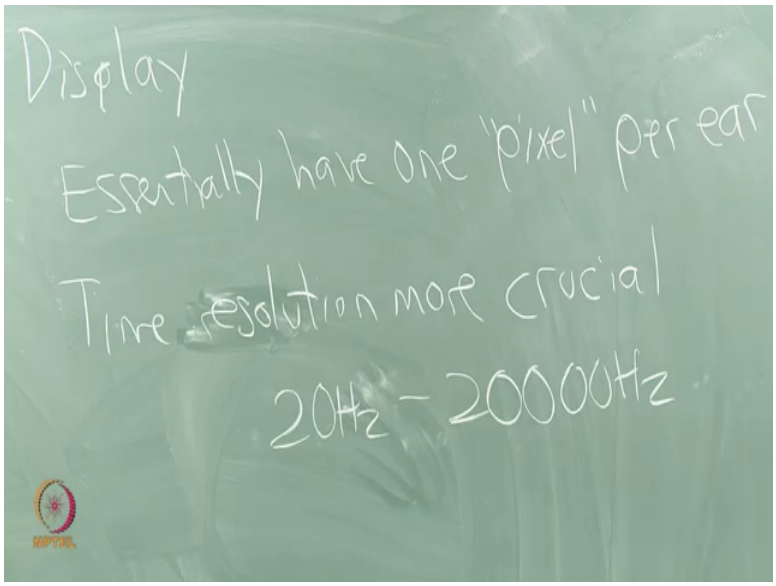
And where I have drawn the center here this makes me wonder. I put at the center of the head, it may make more sense to study the HRTF for each particular ear and just move this over. So, I could take this picture and center it on the ear, and then move all the way around at a fixed distance from the ear. In fact, that might make more sense than this. I have drawn it based on the center of the head, I could do it once for the right ear and once for the left ear pretty sure are the pinna for each ear is not exactly the same for ear to ear from left and right for an individual person all right.

So, it is still going to be you know different based on whether the person is sitting in a chair or standing up. And again if you put a hat on, you will get a different result, but that is the idea you generate these perfect impulse signals from a speaker placed at different locations, you gather all that information and you can reconstruct this HRTF function of frequency, but also function of the angle of that the sound is coming from two angles that it is coming from and distance in the more general case all right.

So, you take that into account and then you apply that as a kind of distortion at the end. So, when you have figured out, how you are going to render the audio signal, at the end

of that you can apply this filter as a kind of distortion. Just as we have optical distortions in the video rendering case, we have this audio distortion that we apply to take into account the additional scattering of sound before it goes into our the inside of our ear, the inner ear part all right.

(Refer Slide Time: 10:49)



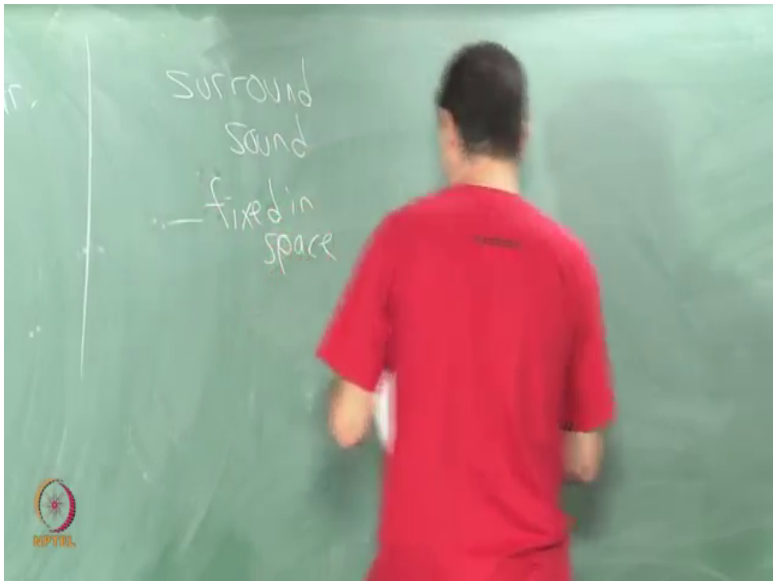
Let me get to the last part which is last part of the four points that I am making here. And just simply mention the display choices. So, display we essentially have one pixel per ear right, just generating a sound pressure wave per ear. It is interesting when I compare again I am always trying to do compare to a visual.

So, for visual displays, we worry a lot about the spatial resolution, but as I said we do not have an imaging system here do it right. So, it is just a matter of generating what looks like a single scalar kind of pressure wave right it does not have some kind of spatial resolution. However, note that the time resolution is much more important. Remember when we talked about frames per second, and we got up to a 75 and 90 hertz maybe even got as high as a 1000 hertz when I talked about this problem of if you take a blinking led, and you are not tracking it, you may perceive it as separate pulses hitting the retina separate images on the retina.

So, so we did get up that high, but in general for audio the frames per second even though it appears to be sort of like a one pixel for per ear the temporal resolution is significantly higher right. So, we normally deal with 60 frames a second on a standard

display, here we have time more time resolution being more important or crucial right. So, you have up to or we go from let us say 20 hertz to 20,000 hertz, so that is just something to think about right. It is the temporal resolution that ends up being crucial for audio not some kind of spatial arrangement which we do not have because we do not have an optical system.

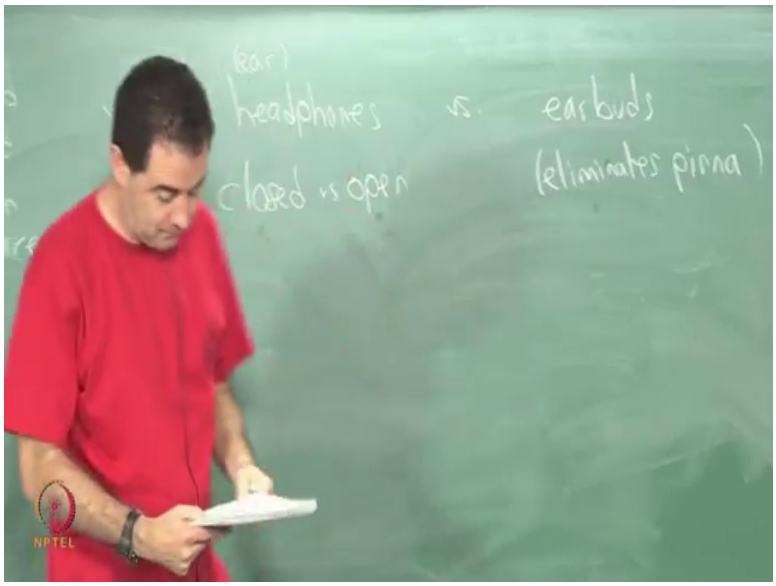
(Refer Slide Time: 12:52)



And then if we think about the choices that we have, we have a surround sound which I gave an example of that in the first lecture. We could have a surround sound system that is fixed in space. I suppose I could mix things right; I could have a head mounted display for the visual part, but I could still have audio being surrounding in the room right. So, it could be half cave like and half head mounted display like right, we could do that we could separate them out that is probably not too convenient.

But this is what we would have in a cave system or if we wanted to do that for a head mounted display, we could put it surrounding. So, in this case, surround sounds the display or the speakers are fixed in space right versus I could be wearing what we normally refer to as headphones of course they are only for the ears what I am talking about I was calling them earphones.

(Refer Slide Time: 13:41)



And in this particular case, we have speakers that are placed on the outside of the ear. There is an interesting question of does that compress your pinna when you put them on right. And very often for closed headphones what are called closed headphones, you are just compressing the pinna blocking off all of the outside sound and then generating a sound for your ear. So, this part gets the part that ends up being important here this extra scattering may get lost.

So, do you try to compensate for that with an HRTF, you may have to write versus open headphones which may leave more room for the sound to propagate in through the pinna and also combined with outside sounds all right. So, open headset open headphones this means that the sound from outside of the headphones is being added.

Now, if you are not in a quiet place, and you want to use your virtual reality system then you have to deal with the sound from the outside it is just like the difference here between having a head mounted display that was mainly designed for virtual reality which blocks out the outside light like the (Refer Time: 15:03) is like that or take something such as Microsoft hollow lens, it is designed for augmented reality in which case its combining the light from the outside with the light that it is generating. So, same two choices exist for standard stereo headphones that you can buy.

And then you know presenting the stimulation as close as possible, you could put ear buds into your ear canal, this is eliminating the pinna altogether. So, eliminate pinna and

the outer ear canal altogether right so or at least some portion of the outer ear canal. And in this case, you had better account for the scattering of sound using the HRTF if you want to have highly local highly accurate localization ability right, because you have eliminated all that by bypassing the pinna all together by using earbuds.

Student: Sir.

Yes.

Student: (Refer Time: 16:07).

That is a good question. I guess it depends on whether the earbuds themselves are more closed or open all right, so that is a good question. So, you have the earbud it is blocking sound coming in directly, maybe there is still some vibration from the pinna could be the case. I am thinking of the effect of the pinna as mainly being reverberation of sound as it goes into the canal while still going through the air, but you could also think about mechanical vibration right through the pinna itself that may transmit some sound.

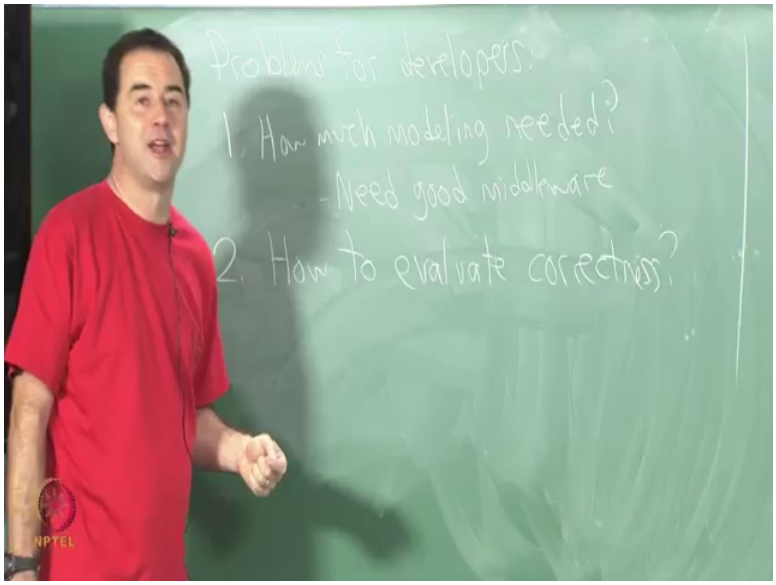
I am not I am certainly not an audio expert. So, I do not, I do not know, but it they could be still some effect from that. And I am thinking of the earbuds as being closed as I says so that its blocking vibrations coming through the ear canal, but the whole earbud could be vibrating I suppose and transmitting some sound from the outside. You know how quiet does it sound when you put earbuds in, and there is and there is their sound coming in from the outside, it does block a fraction of it, but you still do hear sounds from the outside right.

So, I guess there is still some combination, but that is not the virtual part that you wanted right. So, you will be able to local. So, if there is some sound seeping through from the physical world, you may be able to localize that, but it is not the sounds from the virtual environment that you care about all right. So, it was good question, did not think of.

So, let me finish this topic of audio by mentioning challenges for developers. Let me just erase actually, it is going to draw it there, but let me erase here enough. So, what are the problems for developers in this space? So, audio rendering for virtual reality is not as far developed as it is for graphical rendering, you know for the visual part we leverage all of the techniques from computer graphics, many of them are useful, many of them do not

work well in VR, we talked about that in previous lectures. In a in the audio space, we have a lot less work that is been done in terms of audio rendering, it is been considered much less important. And I think now that we have the ability to do head tracking and combined audio and visual senses in virtual reality; it is now getting gaining a lot of moment and a lot of interest.

(Refer Slide Time: 18:15)



So, what are the problems or challenges for developers? One is how much modeling a detail or accuracy is needed all right. So, how much do we have to worry about that right, can we make very, very coarse crude simplified models and that will be sufficient or do we really have to pay attention to every bit of detail right? Do I have to worry about the fabric in all of your clothing, if I want to model the acoustics of my lecture today right? No, I do not know how much if it and if so how to accomplish this? What we would like to have is good middleware to facilitate this right.

We have geometric modeling tools, we have game engines; we have all kinds of things out there for the visual part. We need good tools for the audio part. It is going to take many years to develop these to make it easier for the developers of virtual reality content, and so right now they have to be there sort of in the in the early ages of this right where you have to do all the work yourself, all the hard work yourself.

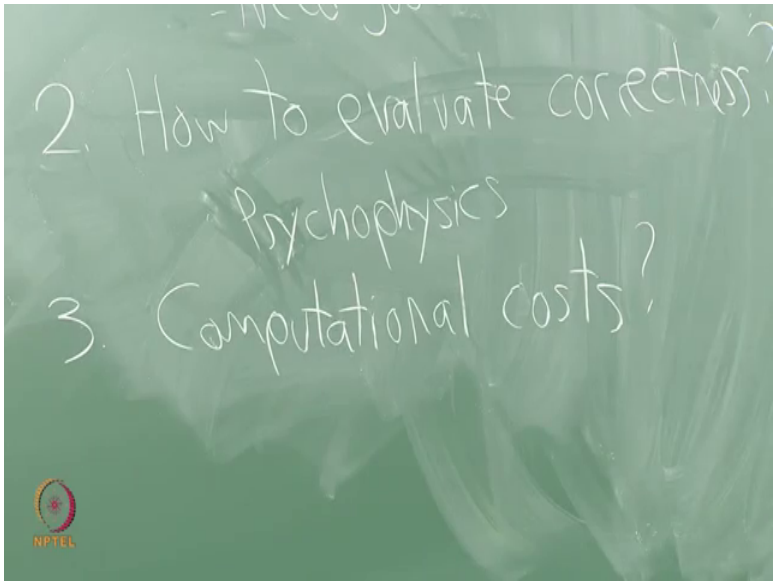
So, the person the developer is trying to be creative has to get into a lot of the technical details and do the implementations themselves. Whereas, if it is the visual part you could

go to something like unity or unreal engine, and very quickly make beautiful visual content without worrying about the technical implementation aspects in in most cases.

Another challenge is how to evaluate correctness or sufficiency for your task right have you accomplished your goal or not if its visual you may look at it and say it looks fine right. If it is audio, again do are you do you care about the fact that you might have lost some localization capability, and if you if it is important for your task to have that localization capability, how do you know that you have maintained it? I mean how do you know that you have reproduced the sounds well enough to maintain that.

So, the ears are not as sensitive as the eyes in some ways. And you know these HRTFs is not important to get those correct or does your brain just adapt to a different HRTF or if you eliminate it all together how much of your localization capability have you lost and is it critical for your application. So, this gets very difficult.

(Refer Slide Time: 20:57)



And generally you have a problem that I would call a one of psychophysics when we talked about the perception of sound, you have to design experiments to determine whether or not you have succeeded. So, it may become much more complicated, you may have to design experiments on and bring in a han subjects to evaluate whether or not you have gotten it correct or sufficient for your task in terms of the simulation you are performing.

And generally you know what are the computational costs associated with doing these audio simulations. Can you take shortcuts and still be effective with regard to a nber two, are you getting it correct is it sufficient for your task and within your computational budget all right.

So, in computer graphics people struggle with this for a very long time then they design GPUs, are there going to be audio processing units that are going to be handling exactly the most important acoustic aspects or is it going to turn out that it does not have to be so high fidelity as it was for the visual case, so that specialized processing units are not needed all right, how far do we have to go.