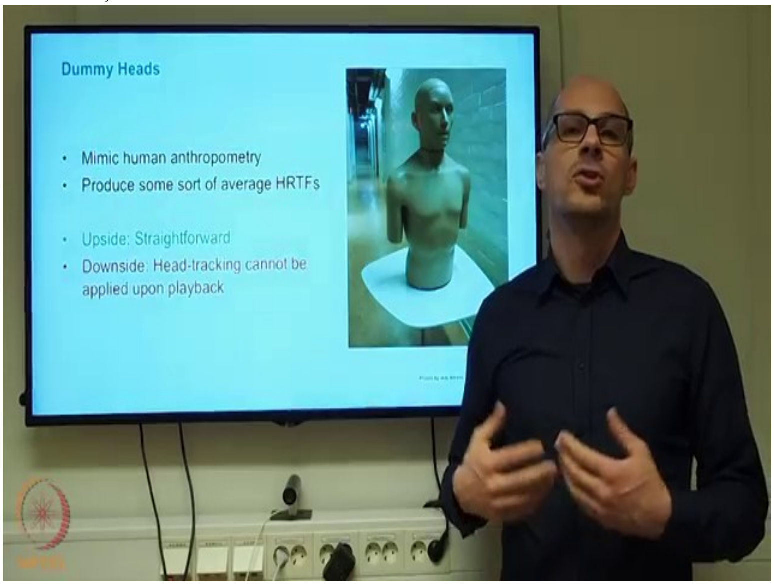
Audio for Virtual Reality
Professor Jens Ahrens
Division of Applied Acoustics
Chalmers Institute of Technology
Dummy Heads

(Refer Slide Time: 00:17)

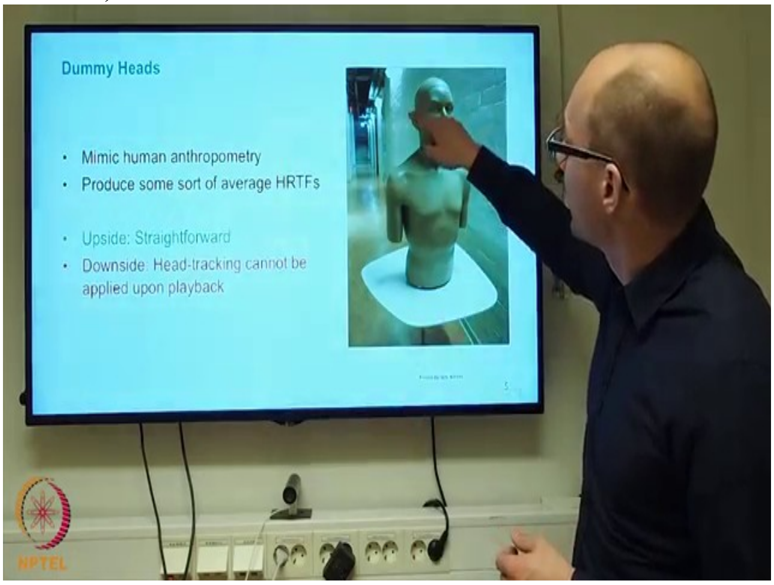


So in order to convey spatial auditory information, spatial acoustic information to a user, we need to create interaural time differences and we need to impose spectral cues, elevation cues onto a signal that we play to the user.

So why not do this straight ahead and record the signals that arrive at the ear of the user. That is of course, a little inconvenient because the user would have to wear microphones in the ears. But it is, there is convenient, a compromise which is called dummy head or mannequin.

These are basically anthropometric copies of the human body. So they are life-size. Here is the picture of one of them. And they include

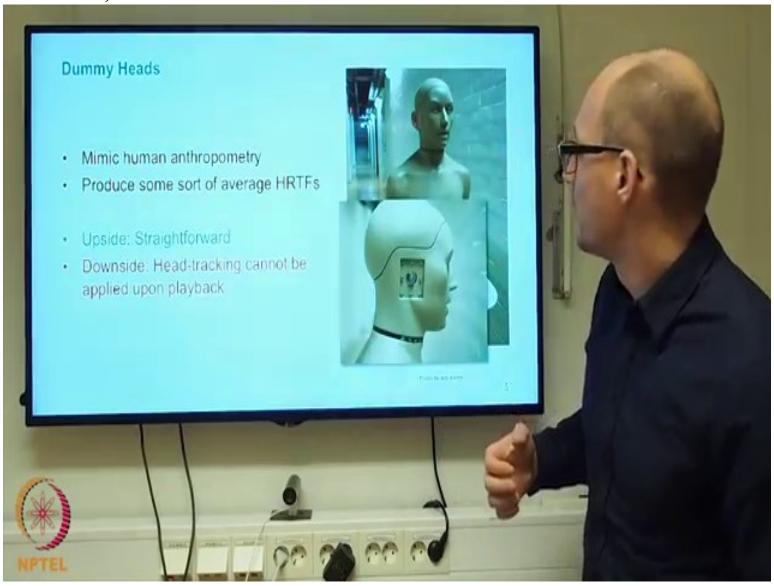
(Refer Slide Time: 00:59)



all the details that are important for a signal that reaches the eardrum of the listener.

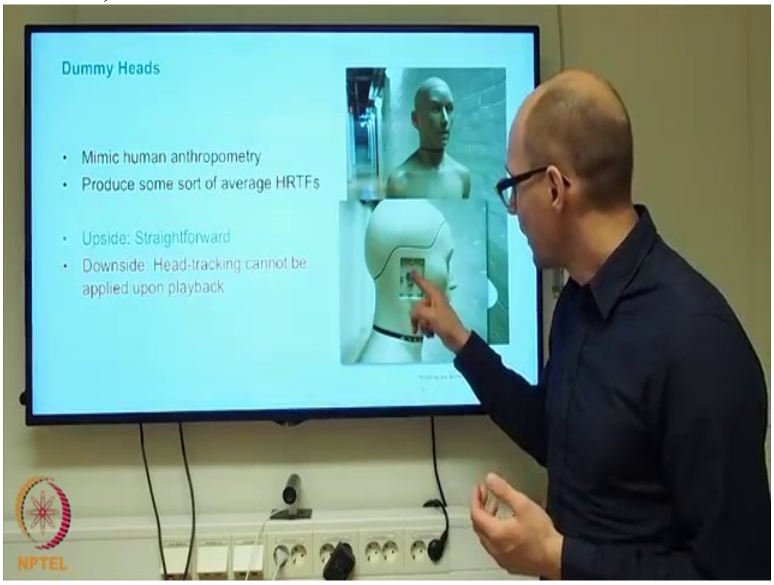
And of course these

(Refer Slide Time: 01:08)



mannequins, they do not have ears but they have

(Refer Slide Time: 01:11)



microphones at those locations where human, humans have the entrances of the ear canal.

So these basically record the ear signals directly and due to sound sink that can be any complex thing, does not to be just a single sound source, it can be anything and there can be dynamic changes, the sound sources may move, etc. And these record the ear signals directly.

And if we then playback these signals through headphones directly into the ear canals of the user then the user is provided with all these acoustic cues that were created by the anthropometry of the dummy head.

And if the anthropometry of the dummy head is similar enough to the anthropometry of the user then the user, well the user's auditory system will recognize all these acoustic cues and will hear the spatial sound scene.

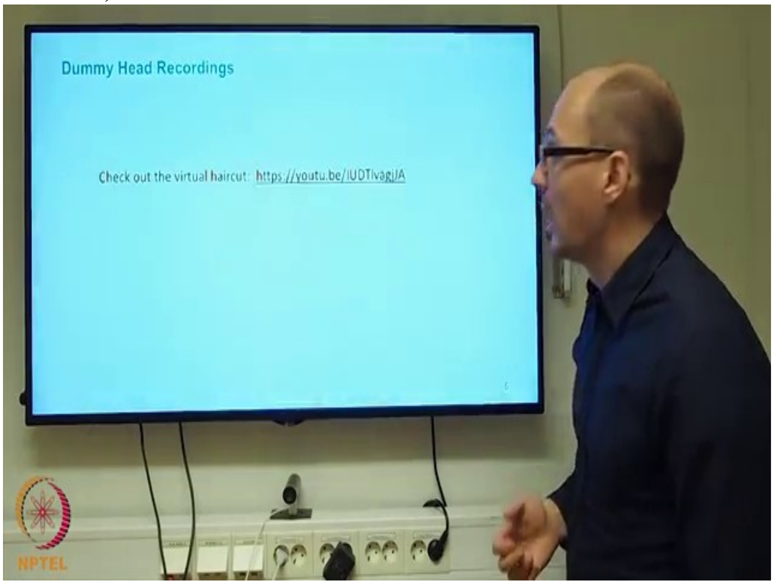
So for example if the dummy head recorded the sound, the signals that arose at its ears due to a sound source slightly elevated, slightly lateral then a user who listens to the recorded signals through headphones will localize a sound source over here, a sound source at this same location.

The upside of this approach is that it is straight forward so we do not need any complicated arrangement. Of course it is a little bulky but there is no processing required or the like. The downside is that there is no dynamic cues contained in these signals.

So that means if the user turns the head up on playback the entire scene will turn with the user. So you are losing a lot of spatial fidelity through this, and of course it is not as easy for the user to resolve front and rear for example because the head orientations are not taken into account.

Still it can be very impressive and

(Refer Slide Time: 03:03)



very well-known example of dummy head recording is termed the virtual haircut. You can find it for example on YouTube through this link above. Please be sure, this is a recording of a dummy head and it is neatly done.

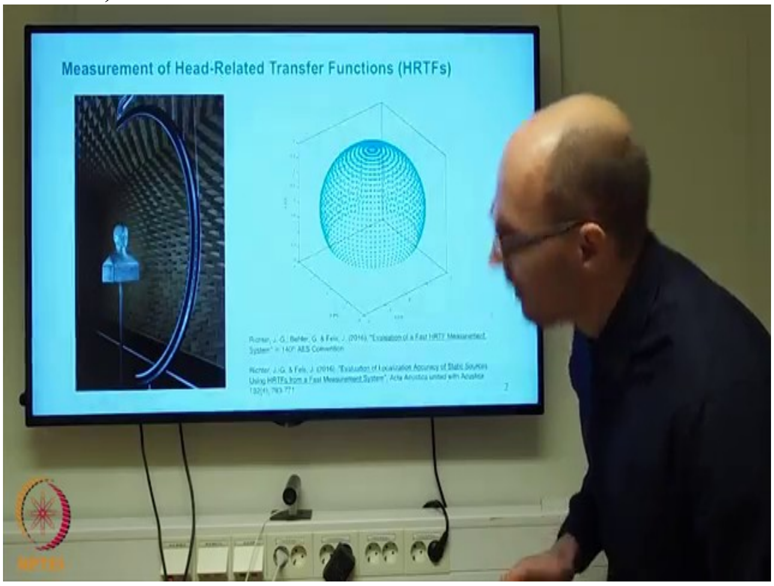
It is a virtual barber shop that you are experiencing, a haircut you are getting in a virtual barber shop and there will be narration, there is music, there is background noises, noise so we can all, well a lot of information but please be sure that you are wearing headphones when you are listening to it.

Because if you listen to these signals through, for example loudspeakers in your room, the loudspeakers will not play the signal directly into your ear but the signal will reach both ears so that there will be crosstalk apparent.

And this changes the interaural cues so that you will, you might perceive something totally different compared to the case when you are consuming the content over headphones.

So what can you do in order to integrate, to take the head rotations of the user into account? Well, first of all

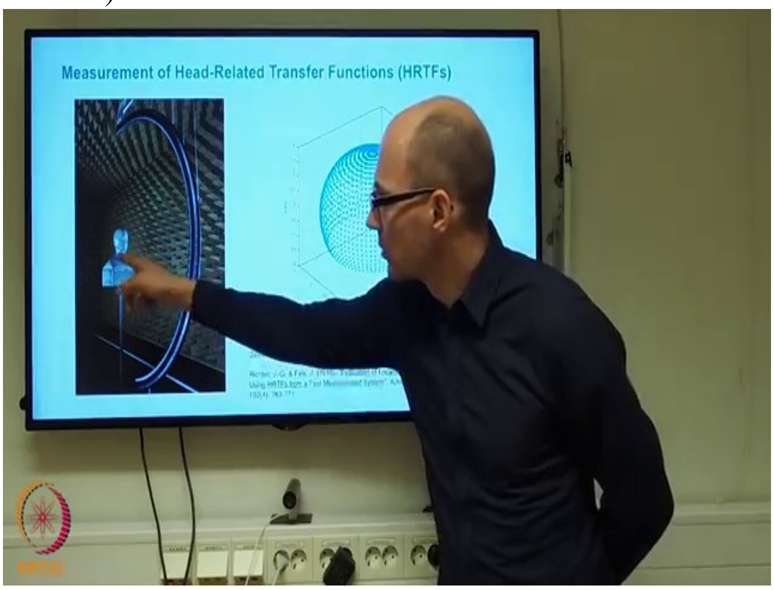
(Refer Slide Time: 04:13)



we can measure these head related transfer functions.

For example, for many, many locations around the user, so this is done as follows.

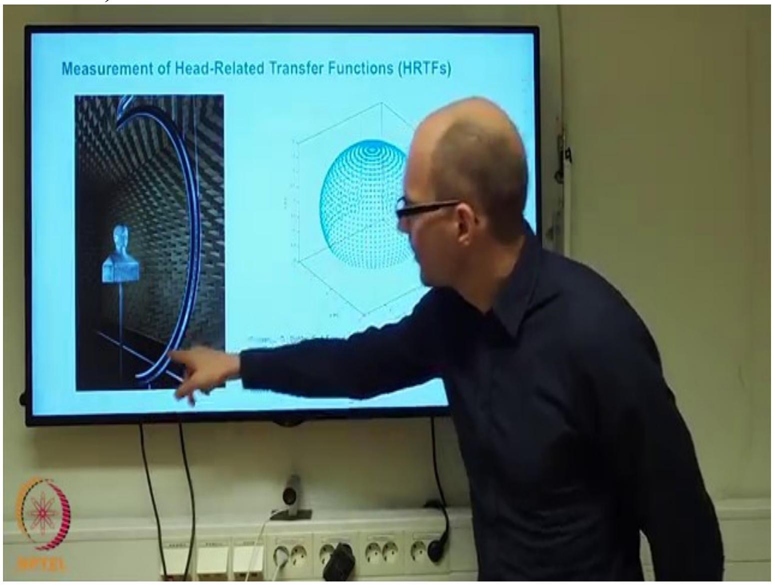
(Refer Slide Time: 04:21)



The user or in this case a dummy head is sat into anechoic chamber and because one would usually not want to measure the acoustic use of a room, of a specific room one wants to do this in free field to capture only the acoustic properties of the body.

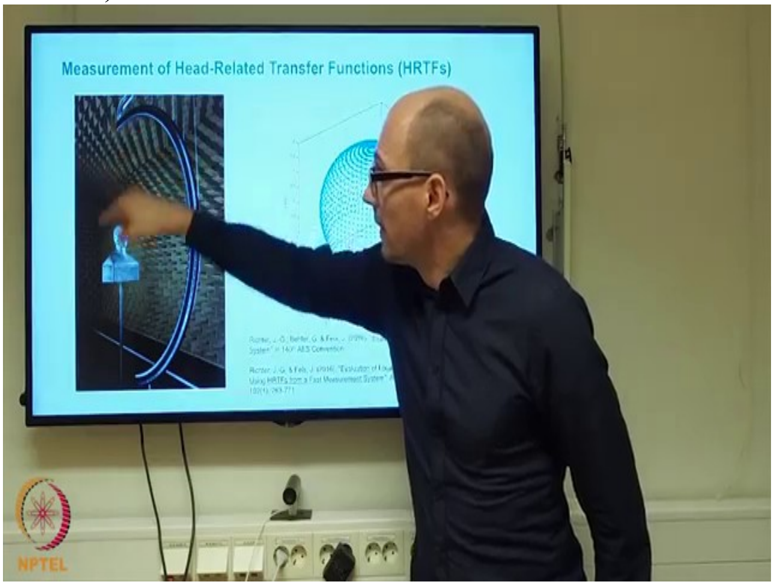
And then there would be loudspeakers,

(Refer Slide Time: 04:42)



for example, in this case there are loudspeakers distributed over this arc and they will play sounds, specific measurement signal that will be

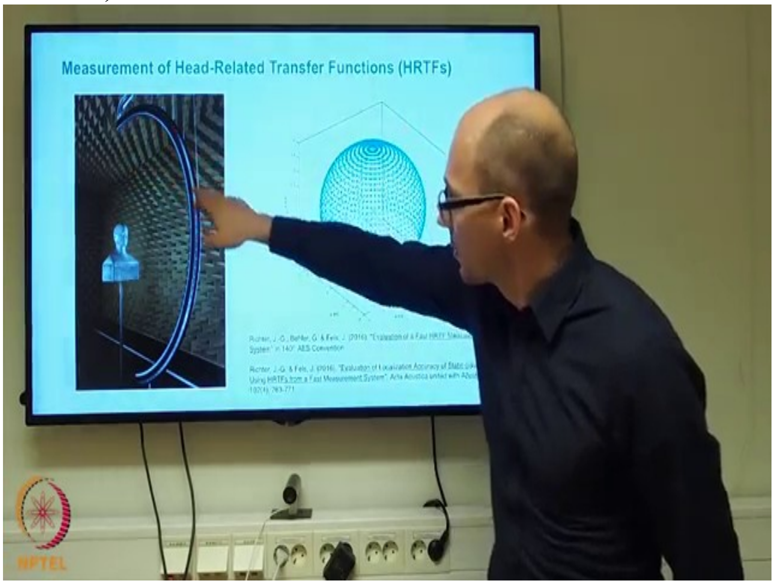
(Refer Slide Time: 04:51)



captured by the ears either inside the ears of this, microphones inside the ears of a person or microphones inside the ears of the dummy head that allow to compute the transfer function from that loudspeaker position to the, to the ear of the mannequin or the person.

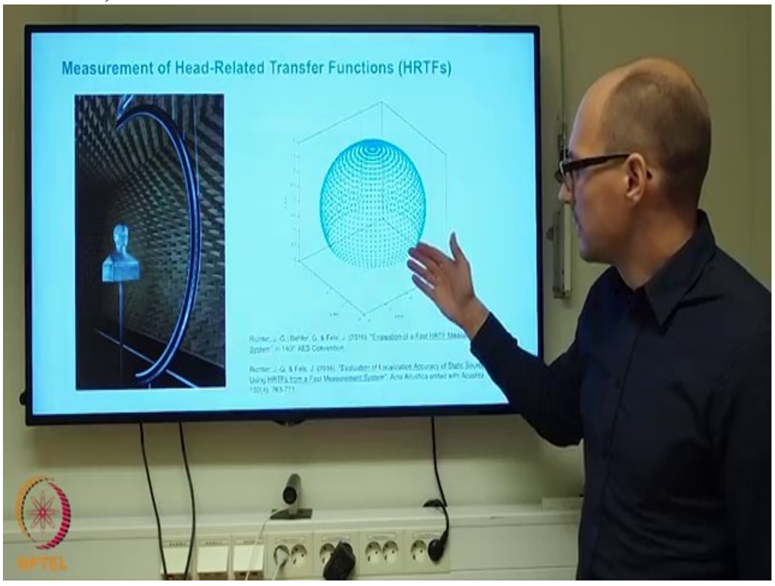
And this will be performed for each of those loudspeakers.

(Refer Slide Time: 05:14)



In this case it is several dozens of loudspeakers and then either the arc will rotate to the new angle or the user or the dummy head will rotate slightly and then the same procedure is repeated to cover all possible angles around the user as it is indicated

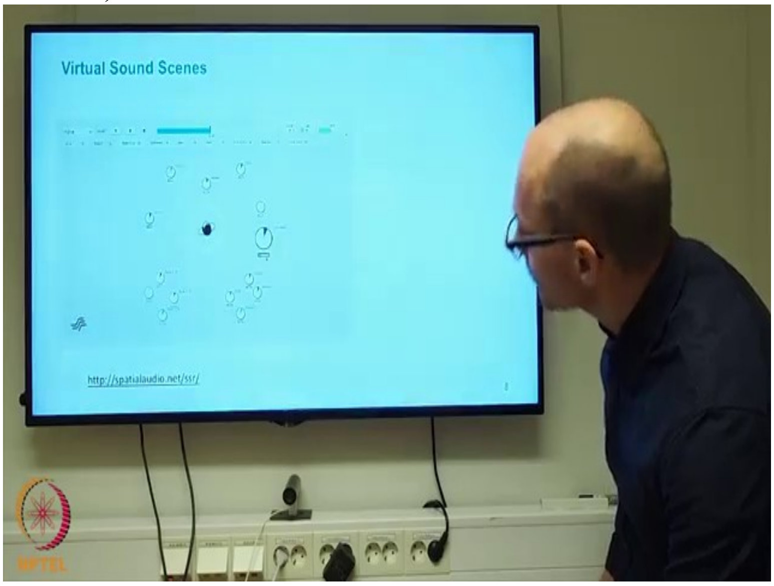
(Refer Slide Time: 05:32)



in this graph.

So each of these marks inside the three dimensional space corresponds to a measured loudspeaker position. So this tells us, this allows us to measure all the transfer paths from source directions to the user, so if we then

(Refer Slide Time: 05:50)



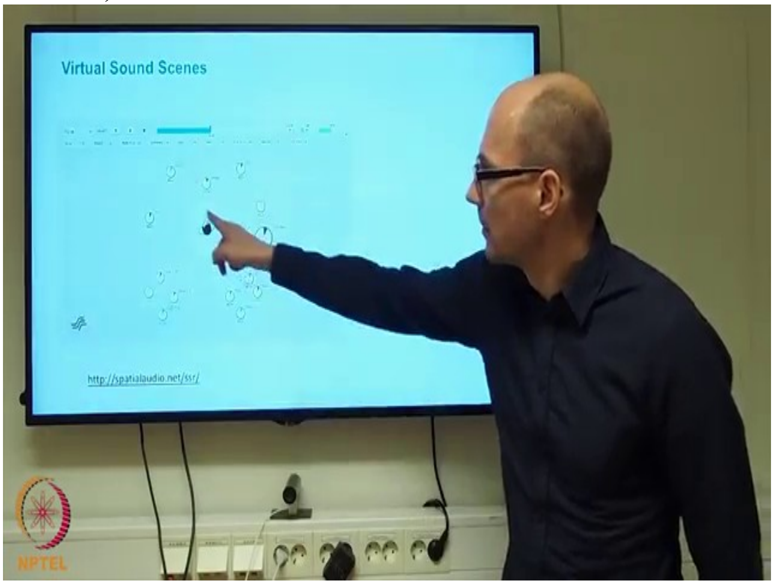
use a software that can impose these transfer functions onto a signal

For example, a signal that will record with the microphone like my speech or a music signal, if we filtering this signal through, with the data that we have measured previously then we are imposing all the acoustic cues of the internal timing differences and the spectral difference, differences and alter that source signal so for example, my voice

And the user will then localize or hear the signal at the location that corresponding to, the whether loudspeaker was in the measurement.

So this would allow us to compose a sound sync of virtual sources like it is the case here

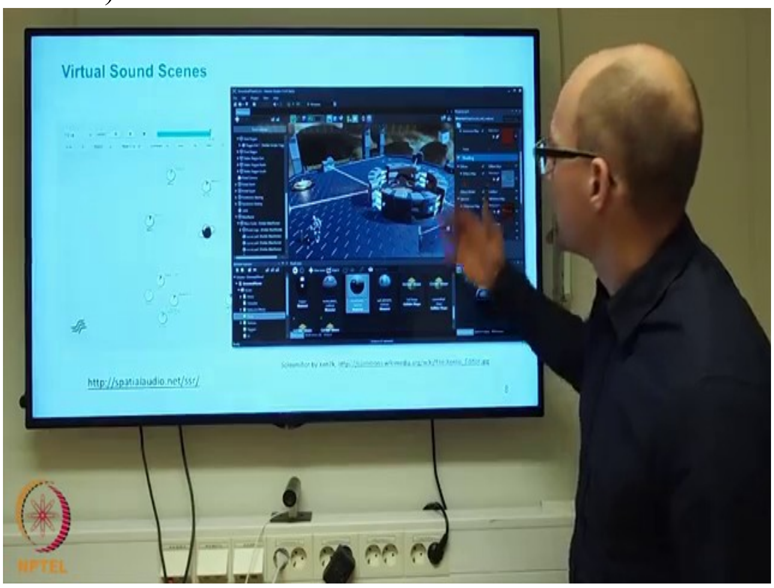
(Refer Slide Time: 06:35)



in the center. This indicates, this little object indicates the user and the other objects are virtual sound sources in this scene so they all have a position associated to it.

This can also look differently like in a

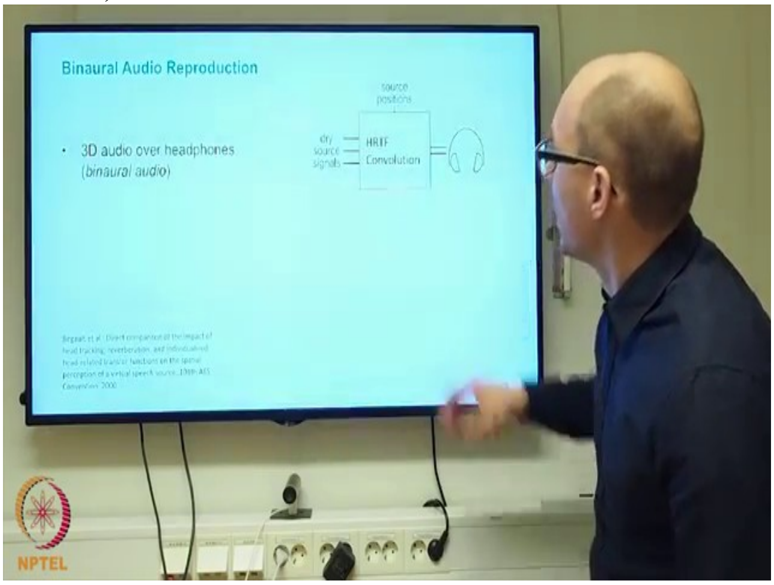
(Refer Slide Time: 06:49)



game engine for example when you are composing three dimensional content and you have this kind of an editor where you can add objects to the scenario

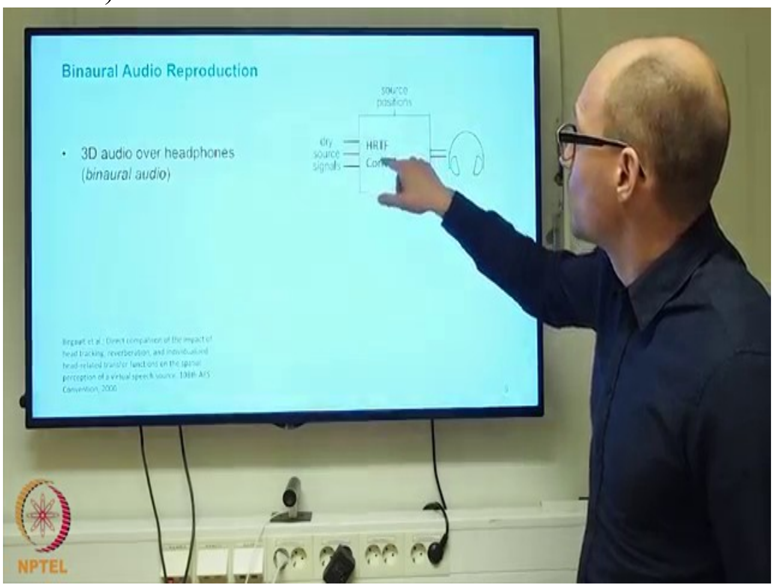
And each object of course has a location and you can assign a sound to each of these objects and then you need an audio engine that picks the right set of measured head related transfer functions for a corresponding sound source position relative to the instantaneous orientation of the listener

(Refer Slide Time: 07:21)



And then it needs, it imposes

(Refer Slide Time: 07:23)



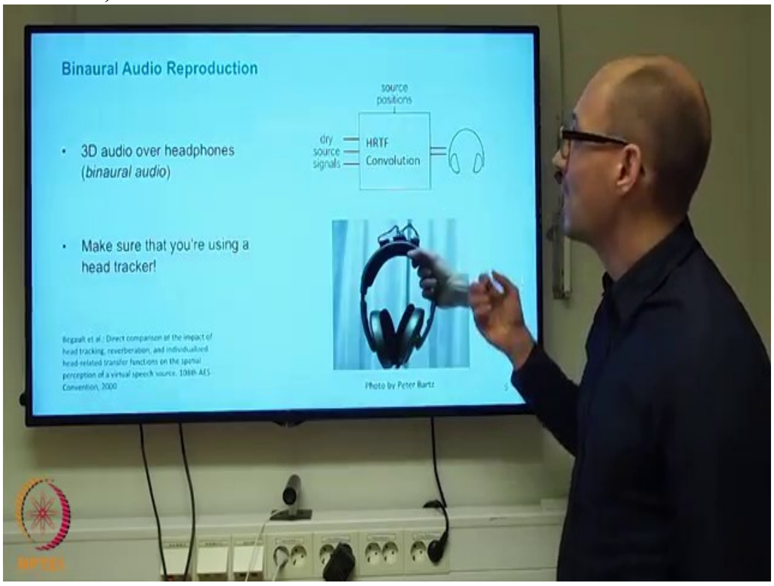
and it produces the source signals, microphone signals what are the recorded signals and with the corresponding set of head related transfer function of HRTFs.

One terms this binaural audio reproduction. The term binaural refers to the fact that we are listening with both ears and it has, it is established term for this kind of reproduction.

Now we still do not have any, we are still not taking into account the movements of the listener upon playback so this technically does not, is not very much different from a dummy head recording

But since we have all the source signals independent

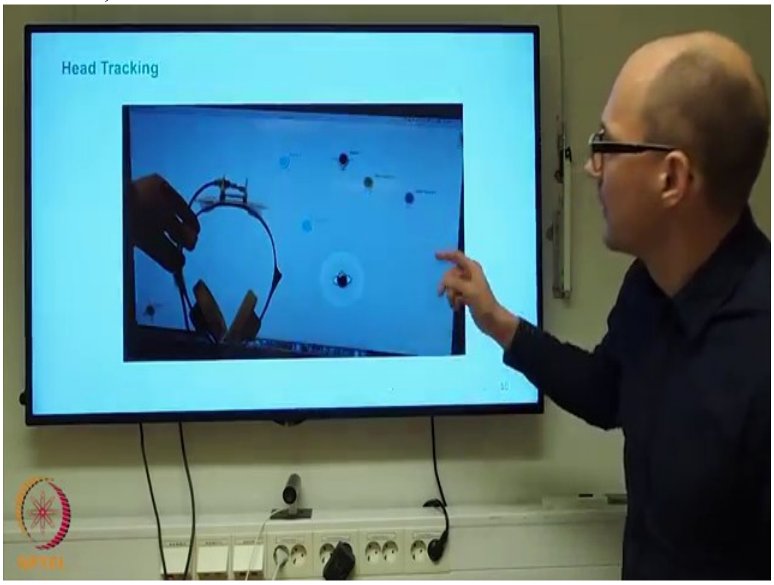
(Refer Slide Time: 08:07)



of each other we can actually mount the equipment, the headphones of the user with the sensor that measures the instantaneous orientation and then we can take the orientation, the instantaneous orientation of the user into account and adopt the processing, the selection of head related transfer functions such that even if the user moves the head the sound sources stays stable in space.

This is indicated in this video.

(Refer Slide Time: 08:35)

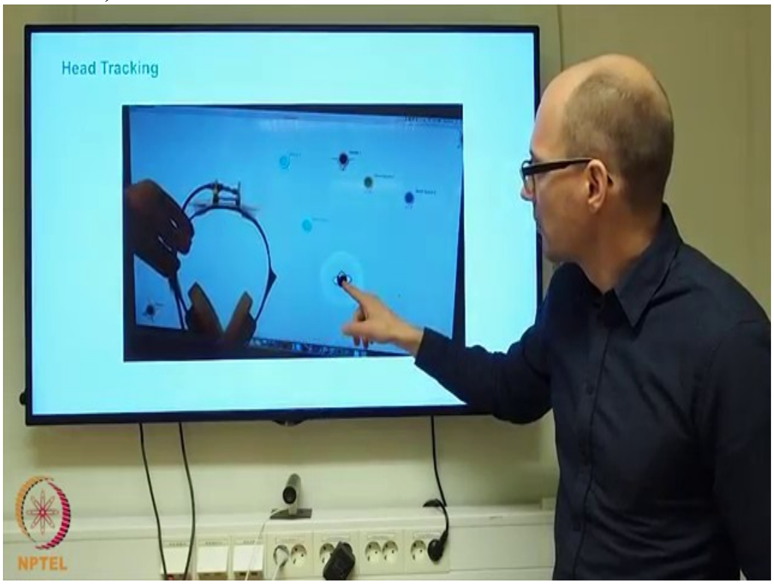


If you see, this is the head tracker

(Refer Slide Time: 08:39)

and you can see the movements of these paired headphones are

(Refer Slide Time: 08:43)



directly translated on to the user in this scenario.

So and the sound sources they are not moving inside the virtual scene and, which means that if the user consumes this content with head tracking then the sound sources are stable in space

And the human auditory system can make use of the, all the dynamic cues inside the signal so that the spatial fidelity is much higher than what you can achieve with the dummy head recording where no head tracking is incorporated.