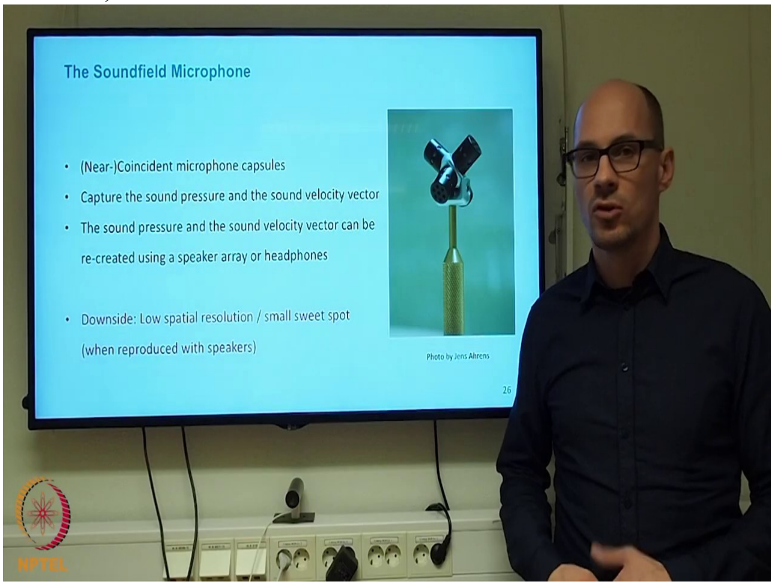
Audio for Virtual Reality
Professor Jens Ahrens
Division of Applied Acoustics
Chalmers Institute of Technology
Capturing of Sound Scenes

(Refer Slide Time: 00:16)



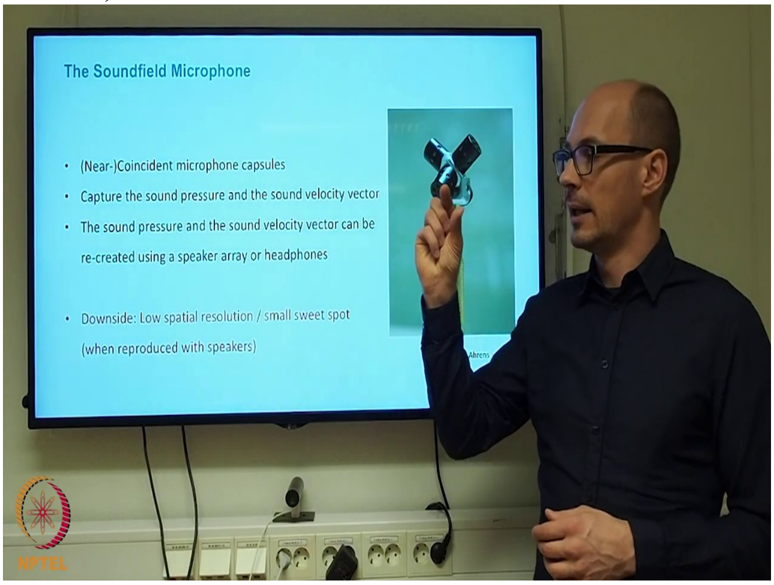
To close this chapter, I would like to briefly talk about methods to capture a special sound scene including the spatial information.

This is, this cannot be done by a single microphone because the output of a single microphone would never allow to draw conclusions on from what direction the signal impinged. Rather the so-called microphone array has to be used.

That is simply any arrangement of more than 1, 2 or more microphones. For example in a fashion, arranged in a fashion like this picture this is a very small microphone composed of small 4 different capsules.

So the diameter of that arrangement is something in the order of

(Refer Slide Time: 01:00)

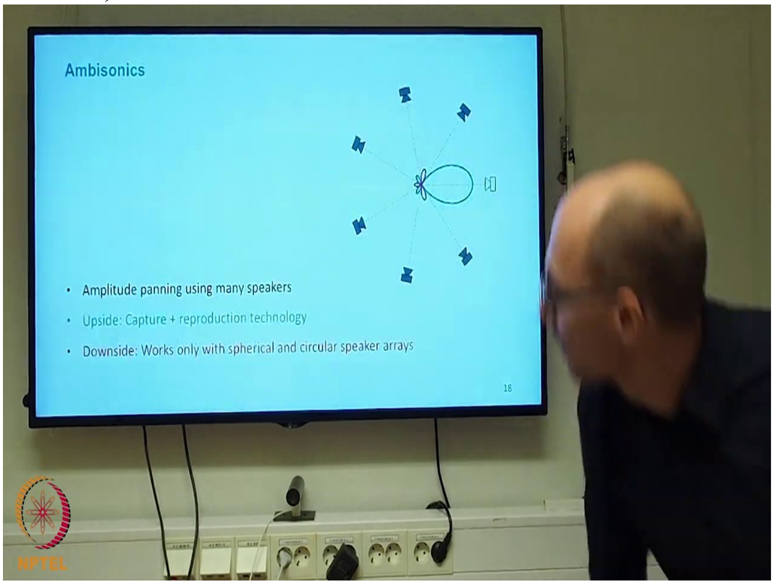


a few centimeters. These are in this particular case, cardioid-like capsules and they are near co-incident meaning that for most of the frequency range they can be considered so close to each other that there is no timing difference between the signals that they capture.

And from these, from the signals that are captured by the individual microphone capsules one can compute the sound pressure and the sound velocity of the signal meaning that, you know we can know what the signal is, that is the sound pressure and velocity will tell you in what direction the sound field is propagating.

This microphone arrangement is termed sound field microphone which is both a method and a commercial product and this was proposed in the ambisonics community. So if you recall ambisonics, quickly switch to that slide, this is

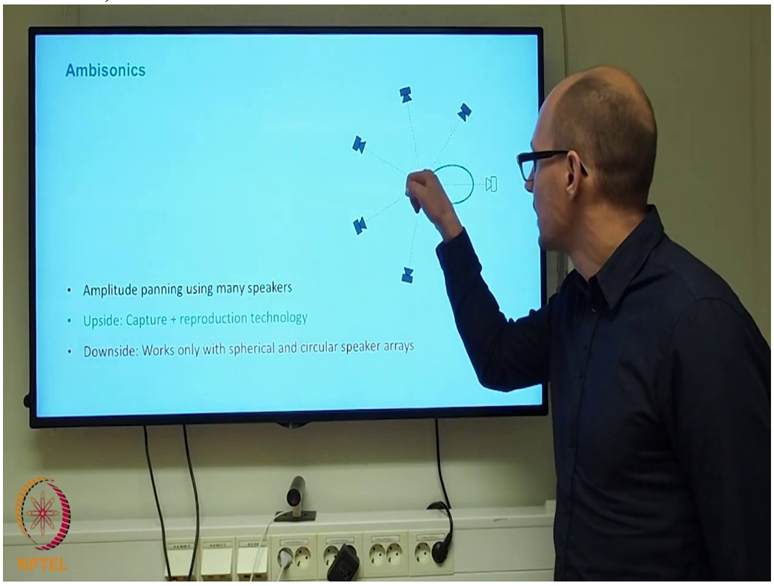
(Refer Slide Time: 02:02)



slide number 18.

It is possible to compute the signals that individual loudspeakers of a suitable arrangement would need to radiate to recreate the sound field that was captured by the microphone array at the center of

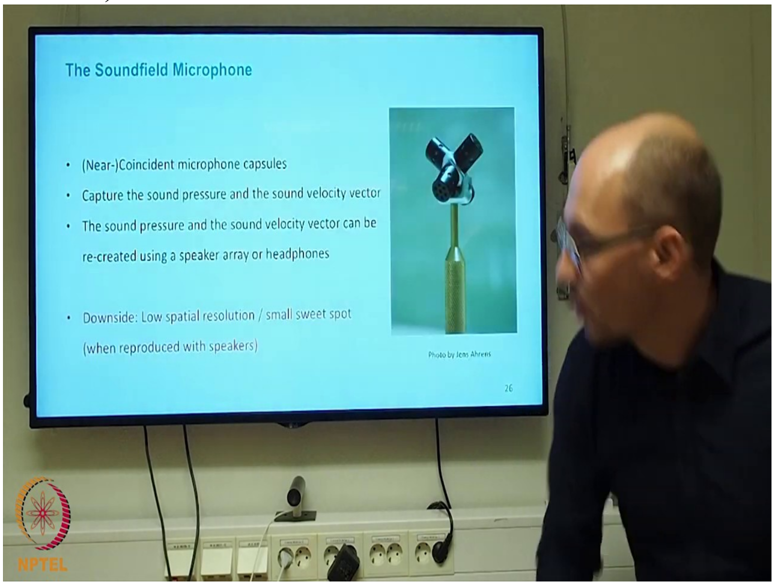
(Refer Slide Time: 02:21)



the arrangement.

But then the recreation is physically only valid at this particular point. Of course we can imagine the head of a person is larger than only that particular point, so the ear signals they differ somewhat from the ear signals that would arise if the user were exposed

(Refer Slide Time: 02:39)

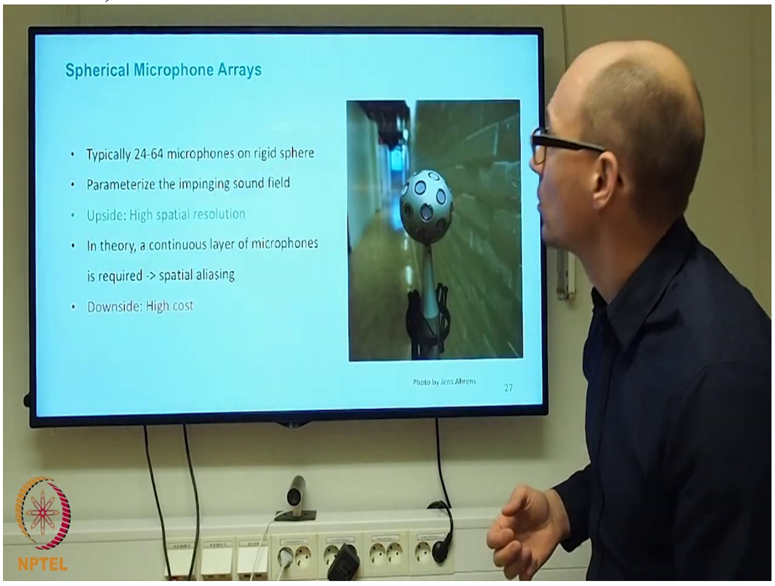


to the sound field directly.

And it also has a little bit of a low spatial, one consequence is that it has a little bit of a low spatial resolution perception-wise so it is not well, possible to hear details in the spatial arrangement, for example differentiate the location of sound sources that are very close to each other and also reverberation recorded with such a microphone is sometimes not very spacious.

This is better with

(Refer Slide Time: 03:13)



so-called spherical microphone arrays which are basically the advanced version of the sound field microphone.

In this case, they use microphones which are usually distributed over the surface of a spherical scattering object so by design this object which hosts the microphone capsules, each of these circles is one capsule. It reflects the sound field.

That is specific reason for why this is desired. The one on the picture has 32 microphones but there is also larger ones, the size of this one is somewhat a bit larger than the tennis ball.

There are also larger arrays with the size of the volley ball which is even more large, even more microphones, something in the order of 64, or sometimes even 128.

There is no commercial product available of those large arrays that is established, this one, the small one is a commercial product for example but most of these arrays that exist, they are currently being used in research community, in research institutions because the methods have not fully matured. So there is no, they are not readily available yet but they will be so in the next few years.

And what these microphone arrays are doing, they are using fairly involved mathematical methods to decompose the impinging sound field into, the sound field that is captured by the microphone array into elementary components and that allows them to undo the effect of that scattering object.

So you can remove the effect of that object, of the microphone array itself, the effect that the microphone array has on the sound field and then obtain basically the acoustic fingerprint of the sound field that was captured by the microphone array.

And this can be then rendered over headphones by means of head related transfer function, head related transfer functions and so conceptually this thing records an acoustic fingerprint of a sound field. And head related transfer functions, they tell us how the ear reacts to sound

So you can guess probably that there is a way to combine this information so that you can use head related transfer functions to virtually put a person into the sound field that was captured by the microphone array so that one can compute the signals that would arise at the ears of that person if that person would have been exposed to the, to this particular sound field.

And then you can rotate the sound field and the microphone array against each other to incorporate head tracking on the rendering side

This is of course fairly involved and computationally fairly expensive but it is the topic that is being actively researched and the current results are very promising. So with a little bit of luck in a few years such a technology would be readily available.

Besides binaural rendering as a, that the one that I have explained, it is also possible to use loudspeaker arrays to recreate the physical structure of that sound field that was captured by the microphone array.

Unfortunately as with loudspeaker arrays the theory requires a continuous layer of microphones on the surface of that sphere meaning again an infinite amount of infinitesimal microphones.

Since this is not the case in practice because we have simply no way of implementing this and we also have to, or we also have spatial aliasing apparent in the sound field that is captured by the microphone array and that is fairly similar to the spatial aliasing that we have seen with loudspeaker array sound field, based on sound field synthesis, and it just happens on the captured side.

The downside is that the costs are high and the resource costs, both in terms of computational effort but also in terms of hardware effort but this is currently the only means to obtain a high resolution capture of the spatial information of a sound scene.