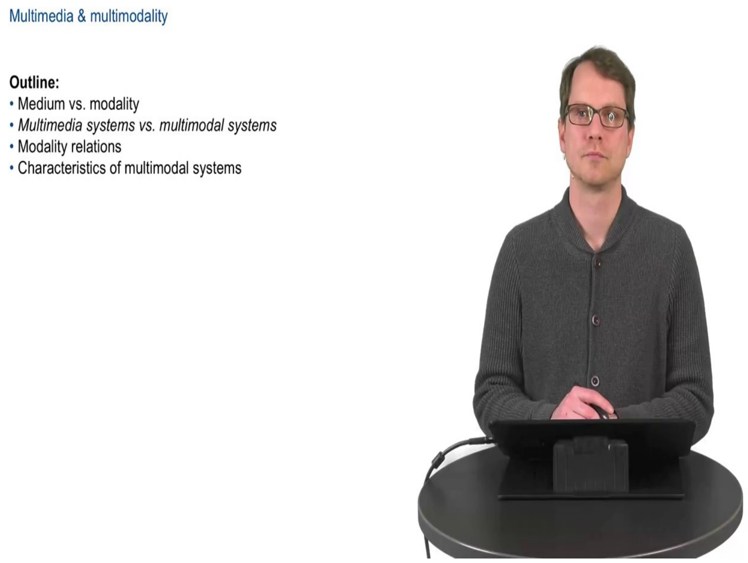
Multimodal Interaction
Multimedia and Multimodality
Professor Benjamin Weiss
Quality and Usability Lab
Technische Universitat Berlin
Multimedia and Multimodality

(Refer Slide Time: 00:16)

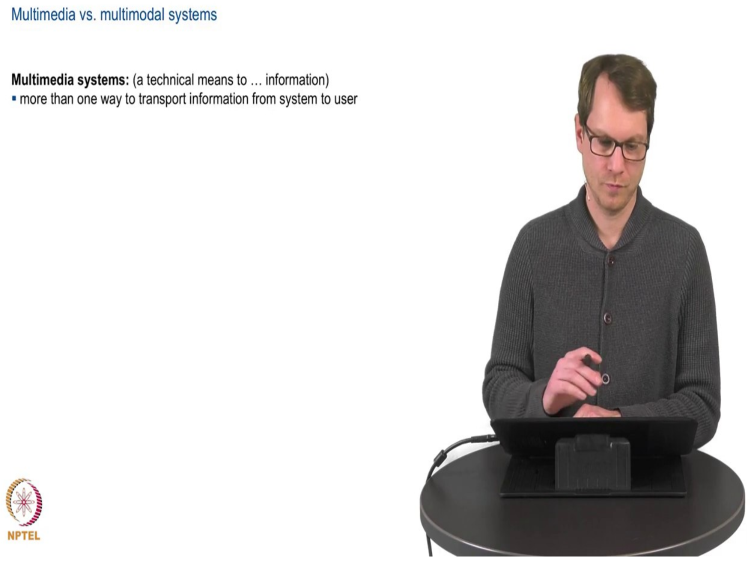


Multimedia & multimodality

Outline:
- Medium vs. modality
- *Multimedia systems vs. multimodal systems*
- Modality relations
- Characteristics of multimodal systems

Let us continue with the actual systems.

(Refer Slide Time: 00:20)

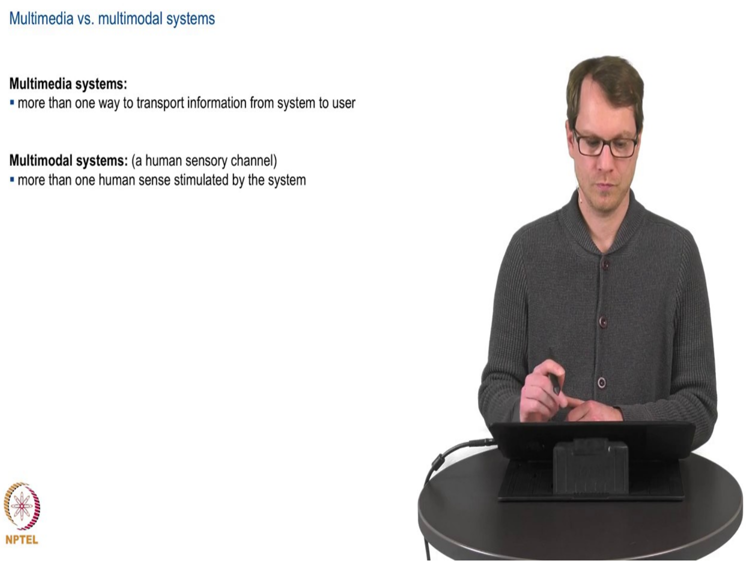


Multimedia vs. multimodal systems

Multimedia systems: (a technical means to … information)
- more than one way to transport information from system to user

Following the definition of a medium, which is technical means to store, transfer and to convey information what is actually a multimedia system? It would be a system with more than just one way to transport information from the system to the user.

(Refer Slide Time: 00:37)



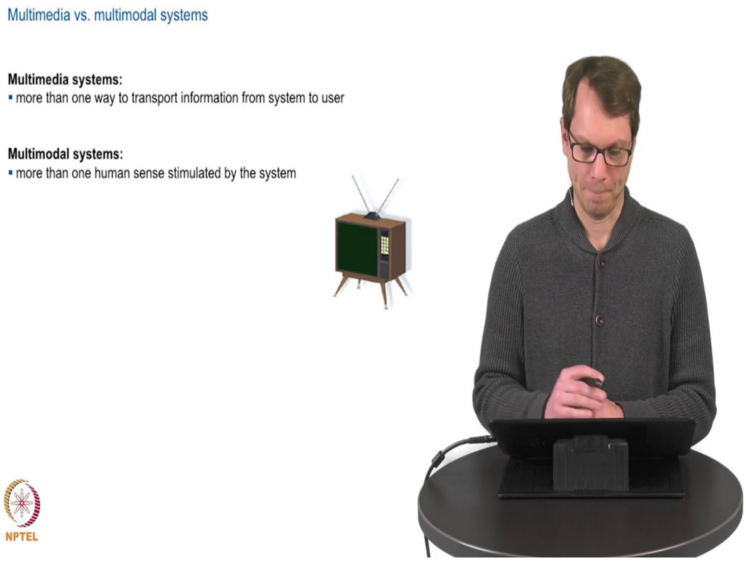
Multimedia vs. multimodal systems

Multimedia systems:
- more than one way to transport information from system to user

Multimodal systems: (a human sensory channel)
- more than one human sense stimulated by the system

Following the definition of modality which is a human sensory channel, a Multimodal system would be a system that can stimulate more than one human sensory channel. The problem here is that it is actually quite similar despite the differences I laid out between modality and media.

(Refer Slide Time: 00:58)



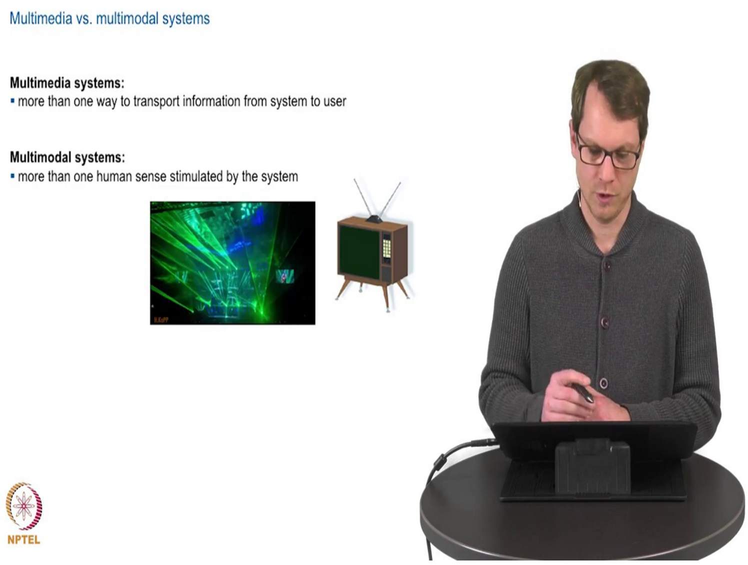
Multimedia vs. multimodal systems

Multimedia systems:
- more than one way to transport information from system to user

Multimodal systems:
- more than one human sense stimulated by the system

The problem is for example, that any kind of TV, which offers a screen and loud speakers, could be considered as a multimedia and a multimodal system, following this definition,

(Refer Slide Time: 01:10)



Multimedia vs. multimodal systems

Multimedia systems:

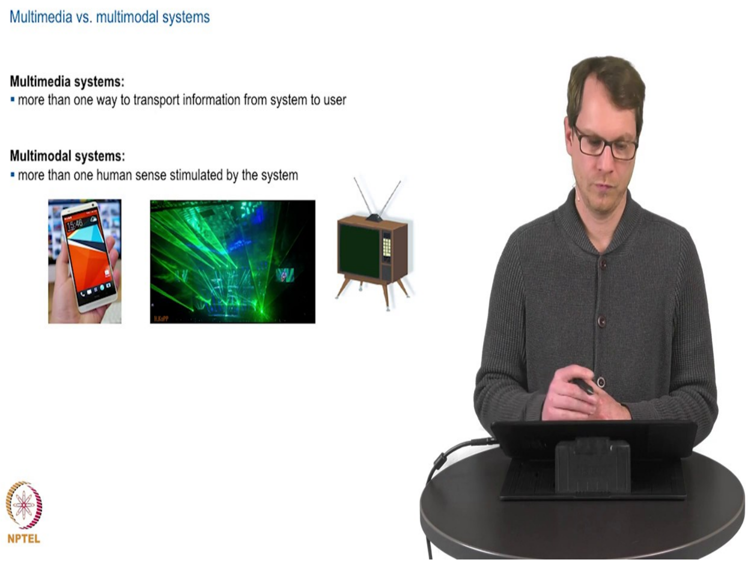
- more than one way to transport information from system to user

Multimodal systems:

- more than one human sense stimulated by the system

but also any kind of concert hall which provides sound and light.

(Refer Slide Time: 01:17)



Multimedia vs. multimodal systems

Multimedia systems:

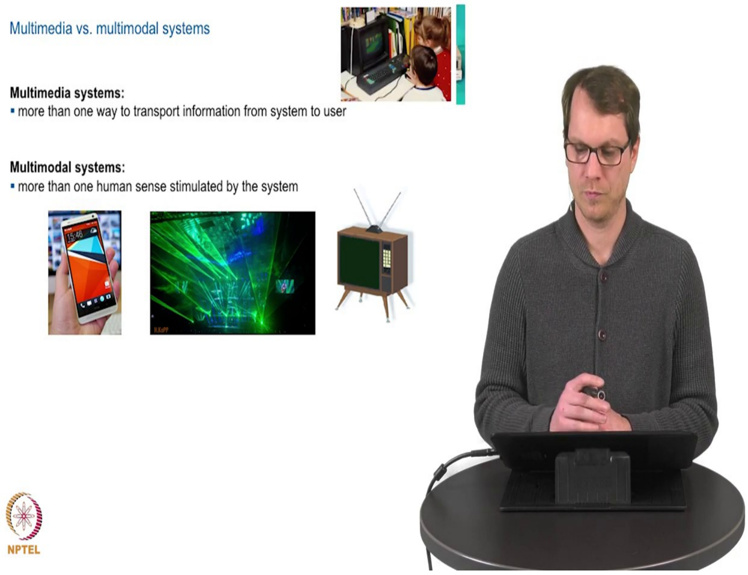
- more than one way to transport information from system to user

Multimodal systems:

- more than one human sense stimulated by the system

Same holds for smart phone with its screen and vibration, and the loudspeakers.

(Refer Slide Time: 01:23)



Multimedia vs. multimodal systems

Multimedia systems:
- more than one way to transport information from system to user

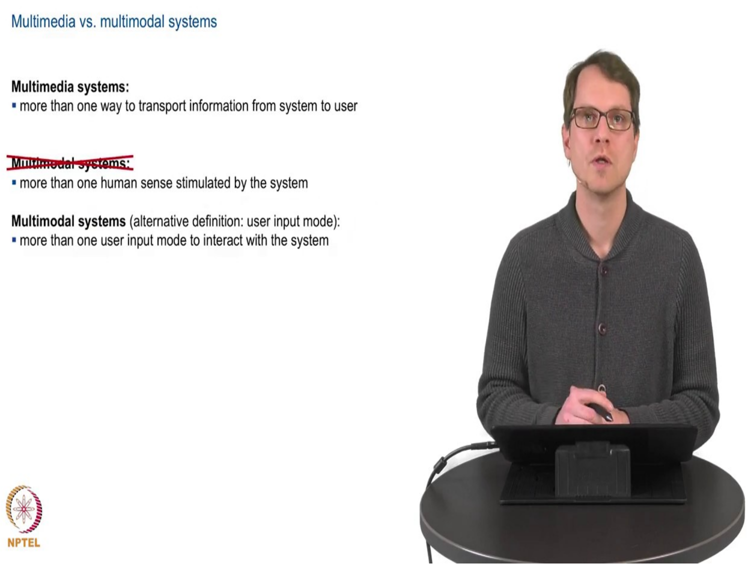
Multimodal systems:
- more than one human sense stimulated by the system

And of course, any kind of old computer, desktop computer which follows the WIMP principle, which means Windows, Icons, Menus and Pointers.

This could also all be Multimodal systems and Multimedia systems. This is not what we want

Therefore, we follow human-computer interaction, a different definition, the alternative definition. All these systems are presented will be considered as multimedia systems, not multimodal systems.

(Refer Slide Time: 01:56)



Multimedia vs. multimodal systems

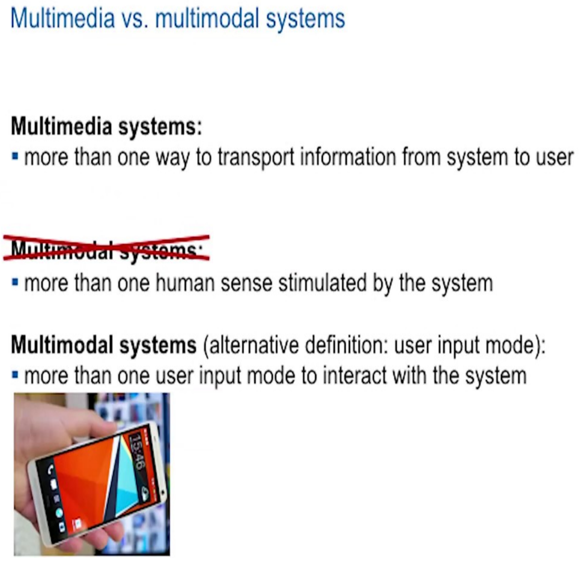
Multimedia systems:
- more than one way to transport information from system to user

~~**Multimodal systems:**~~
- more than one human sense stimulated by the system

Multimodal systems (alternative definition: user input mode):
- more than one user input mode to interact with the system

So let us follow the alternative definition. Modality is a user input mode to interact with the system. Following this, a multimodal system would be a system that provides more than just one user mode to interact with it.

(Refer Slide Time: 02:13)



Multimedia vs. multimodal systems

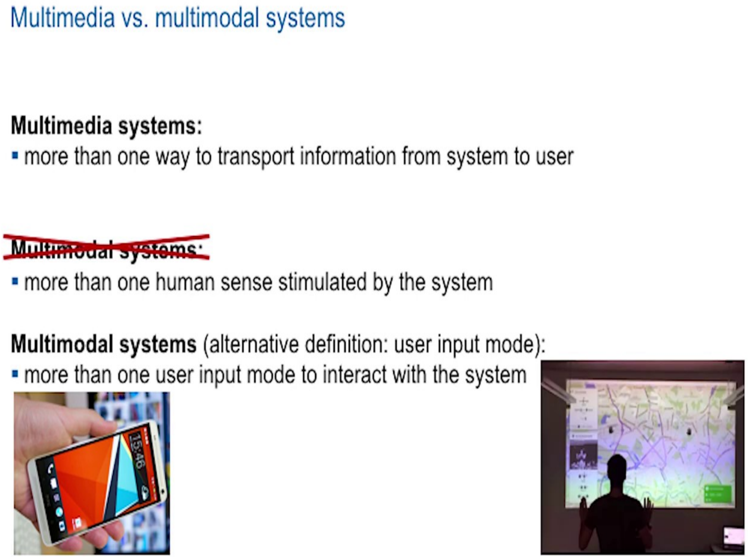
Multimedia systems:
- more than one way to transport information from system to user

~~**Multimodal systems:**~~
- more than one human sense stimulated by the system

Multimodal systems (alternative definition: user input mode):
- more than one user input mode to interact with the system

Again, a smart phone would be a multimodal system, just because it offers a touch screen for touch but also sensors for 3D, three dimensional gestures and also microphones for speech for example.

(Refer Slide Time: 02:28)



Multimedia vs. multimodal systems

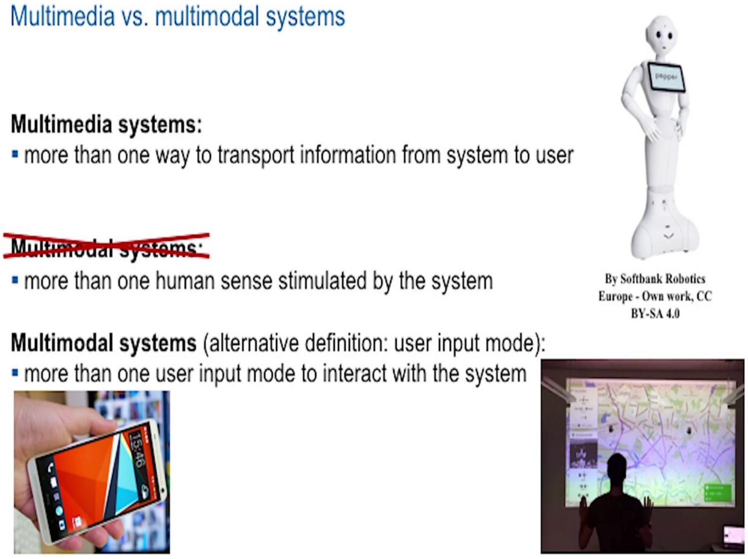
Multimedia systems:
- more than one way to transport information from system to user

~~**Multimodal systems:**~~
- more than one human sense stimulated by the system

Multimodal systems (alternative definition: user input mode):
- more than one user input mode to interact with the system

But also this rather old system which provides voice recognition and a touch screen would also be considered as multimodal system because it offers two different input modes.

(Refer Slide Time: 02:40)



Multimedia vs. multimodal systems

Multimedia systems:

- more than one way to transport information from system to user

~~**Multimodal systems:**~~

- more than one human sense stimulated by the system

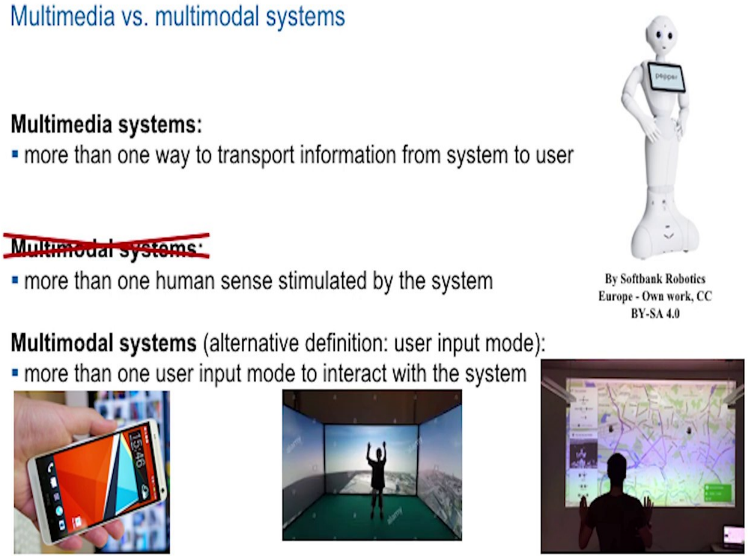
Multimodal systems (alternative definition: user input mode):

- more than one user input mode to interact with the system

By Softbank Robotics Europe - Own work, CC BY-SA 4.0

And of course any kind of social robot or robot that can provide and interact with speech or vision and here even a touch screen would be a multimodal system.

(Refer Slide Time: 02:54)



Multimedia vs. multimodal systems

Multimedia systems:

- more than one way to transport information from system to user

~~**Multimodal systems:**~~

- more than one human sense stimulated by the system

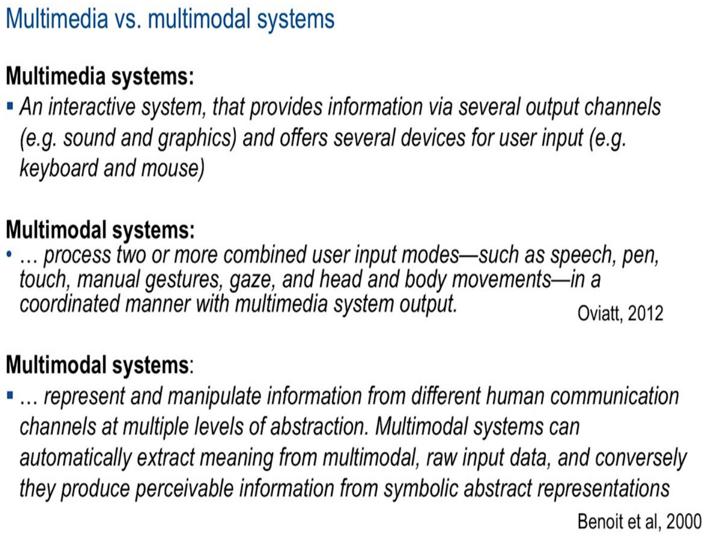
Multimodal systems (alternative definition: user input mode):

- more than one user input mode to interact with the system

By Softbank Robotics Europe - Own work, CC BY-SA 4.0

The last example is a virtual environment, where actually any kind of social signal or information that we can provide as users, could be in principle be used to trigger any kind of system reaction. This would be all multimodal systems.

(Refer Slide Time: 03:12)



Multimedia vs. multimodal systems

Multimedia systems:

- *An interactive system, that provides information via several output channels (e.g. sound and graphics) and offers several devices for user input (e.g. keyboard and mouse)*

Multimodal systems:

- *… process two or more combined user input modes—such as speech, pen, touch, manual gestures, gaze, and head and body movements—in a coordinated manner with multimedia system output.* Oviatt, 2012

Multimodal systems:

- *… represent and manipulate information from different human communication channels at multiple levels of abstraction. Multimodal systems can automatically extract meaning from multimodal, raw input data, and conversely they produce perceivable information from symbolic abstract representations*

Benoit et al, 2000

So, let us start with some real definitions here.

Multimedia systems are interactive systems that provide information via several output channels, for example, sound and graphics, and offer several devices for user input, for example, keyboard and mouse.

In contrast to this we have a nice definition by Sharon Oviatt who is a actually pioneer in multimodal interactions. She says that a multimodal system process two or more combined user input modes such as pen, speech, touch, manual gestures, gaze, and head and body movements in a coordinated manner with multimedia system output."

And here are two important aspects in this definition. The first one is that these are combined input modes. That means for truly multimodal system, she assumes that the information we provide via different input modes are actually related somehow.

Second topic is that a multimedia system is a kind of complement to a multimodal system while multimedia is more focused on the system output, the multimodal part of a system refers to the capabilities of interpreting and processing information from the user.

But we also have a second definition of what a multimodal system is. This is by Benoit et al, and they say- "a multimodal system represents and manipulates information from different human communication channels at multiple levels of abstraction.

Multimodal systems can automatically extract meaning from multimodal raw input data, conversely they produce perceivable information from symbolic abstract representations.”

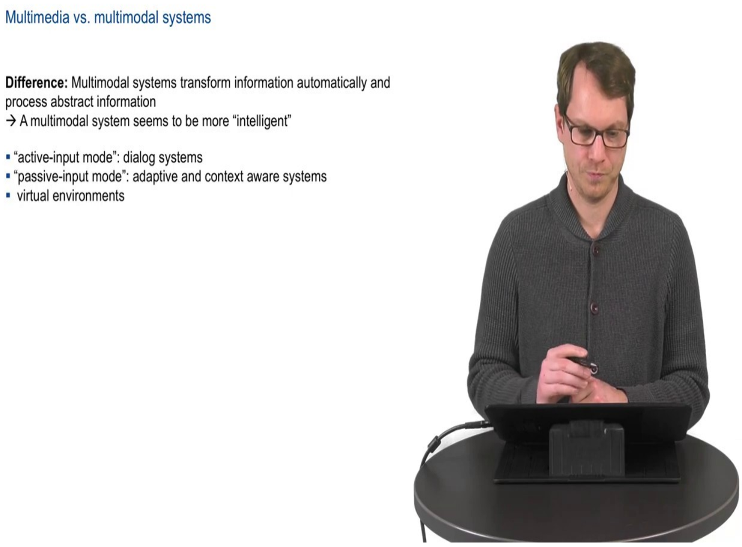
Here is a really different angle, a different view on what a multimodal system is. It is about symbolic processing of data.

So, we have different input modes like acoustic information from our speech or voice or some visual information, from our movement or even from some haptic information from our touch gestures and so on.

But this raw data is not only processed in some way, it is also represented in a symbolic way by the system which is trying to find out what our true intentions are. So, some semantics are represented in the system on a symbolic level.

And this is the true aim of this definition which makes the multimodal system by Benoit et al much more intelligent than other systems.

(Refer Slide Time: 05:58)



Multimedia vs. multimodal systems

Difference: Multimodal systems transform information automatically and process abstract information
→ A multimodal system seems to be more “intelligent”

- “active-input mode”: dialog systems
- “passive-input mode”: adaptive and context aware systems
- virtual environments

So, the main difference between the multimedia system and multimodal systems are that these kind of information are processed in an intelligent way and represented in the system.

We have actually three different domains of multimodal systems. One is called by Oviatt the so-called active Input Mode. This means that user can intentionally give information to the system, for example, by commands or by learnt gestures.

This is intentionally and consciously signalled to the system. For example a dialog system would be one such a kind of multimodal system as long as it provides different input modes.

There is also the passive input mode by Oviatt again and this means that all kind of information that users do not produce intentionally, like emotions or signaling the user's interest by gazing somewhere without actually knowing or purposely doing this to trigger the system.

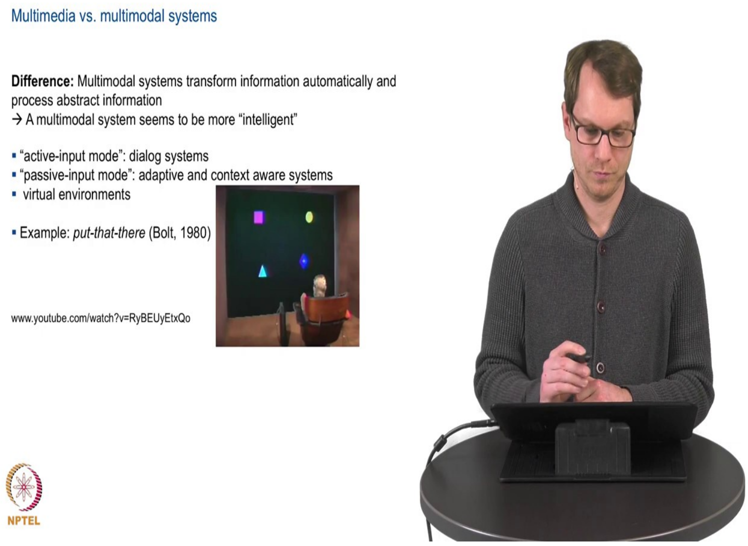
So, all these systems which actually interpret signals which might not be voluntarily or intentionally be given by users are intelligent and adaptive systems which rely on this passive input mode.

Of course, a system, a multimodal system can be processing active or passive input modes.

The last domain is truly virtual environments where users can use all kinds of information behave just naturally as he or she would do in a normal environment.

We are humans and the system would potentially be able to analyse all this kind of information or most of them and then act accordingly.

(Refer Slide Time: 07:52)



Multimedia vs. multimodal systems

Difference: Multimodal systems transform information automatically and process abstract information

→ A multimodal system seems to be more “intelligent”

- “active-input mode”: dialog systems
- “passive-input mode”: adaptive and context aware systems
- virtual environments

- Example: *put-that-there* (Bolt, 1980)

www.youtube.com/watch?v=RyBEUyEtxQo

Here is very old example but a seminal one, by Bolt-it is called as the put-that- there scenario from MIT and I will give you a link on the website for the video where you can have a look at it

It is one of the first or even the first system which could produce and process information in a multimodal way which is commands by voice and pointing gestures, three dimensional gestures in a combined way.

This is called put-that-there scenario because the user could actually sit there in this nice chair, point somewhere and say, say something like “Put the yellow cross in the red cycle”.