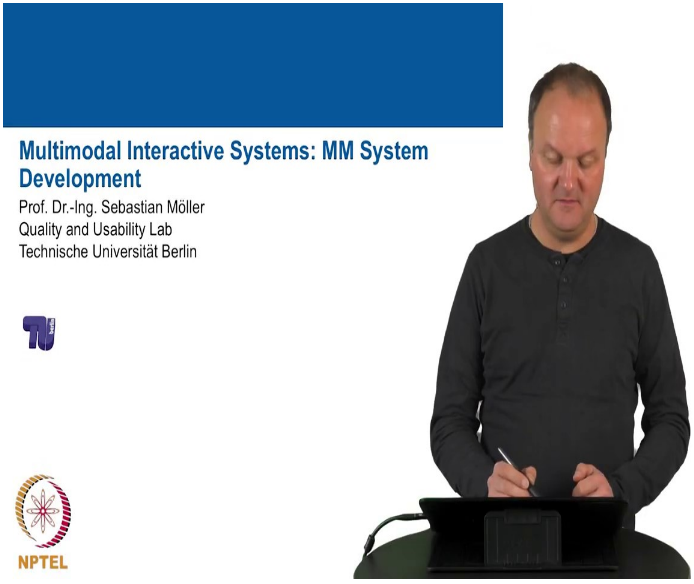
Multimodal Interactive Systems
Professor Doctor Sebastian Moller
Quality and Usability Lab
Technische Universitat Berlin
MM System Development

(Refer Slide Time: 00:15)



Multimodal Interactive Systems: MM System Development

Prof. Dr.-Ing. Sebastian Möller
Quality and Usability Lab
Technische Universität Berlin

In this video I would like

(Refer Slide Time: 00:17)

Processes and tools for system development. Empirical development possibilities:

- Collect **data** in **human-to-human interactions** for a comparable task

to illustrate some principles of multimodal system development. I will first concentrate on empirical development first the possibilities and then on model based development possibilities.

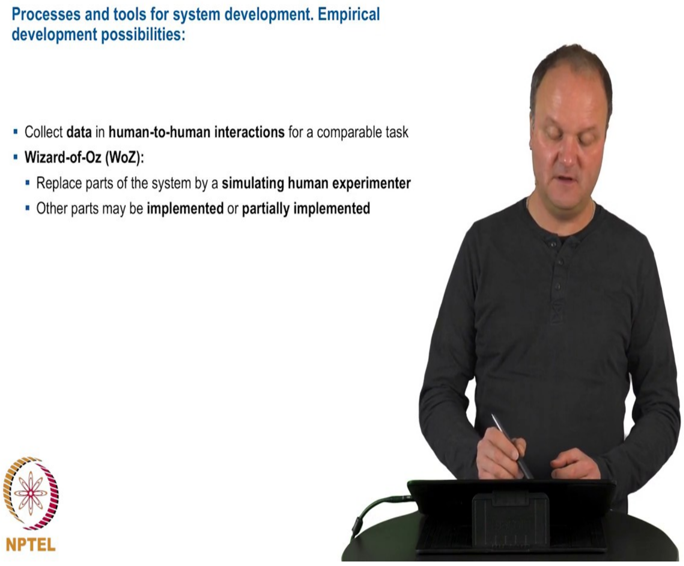
In order to develop a multimodal interactive system on the basic of data we need to collect data. And the first way to collect data is to observe human-human interactions for comparable tasks; the multimodal interactive system has been built for.

This has some disadvantages because the user behavior is usually not the same in human-human interaction than it is in the human-machine interaction so users behave differently when being confronted to a machine than they would behave with human.

There are also in human-human interaction no or very few recognition and understanding errors and in many tasks there is actually no comparable human-human interaction situation which the human-machine interaction could be based on.

This is why the people have developed

(Refer Slide Time: 01:22)



Processes and tools for system development. Empirical development possibilities:

- Collect **data** in **human-to-human interactions** for a comparable task
- **Wizard-of-Oz (WoZ):**
  - Replace parts of the system by a **simulating human experimenter**
  - Other parts may be **implemented** or **partially implemented**

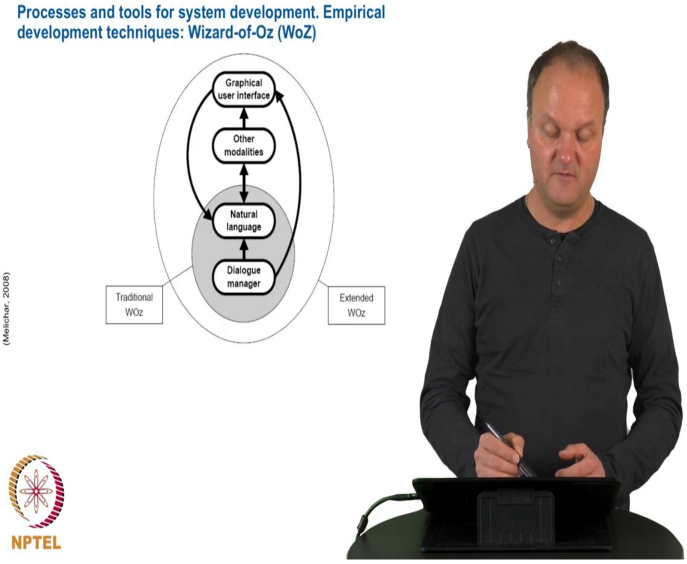
the second technique which is called the Wizard-of-Oz technique. The Wizard-of-Oz is actually an expert which replaces parts of the system by simulating a human experimenter. Of course the term comes from the famous novel from Baum.

This wizard has to behave exactly like the future system or the system component the experimenter is intended to replace would behave. So it could not behave as a normal human but it should behave actually as the future machine will behave.

It may also replace only parts of the system so modules of the system and other parts may be implemented already then we talk about a bionic wizard.

The parts which are replaced by

(Refer Slide Time: 02:15)



Processes and tools for system development. Empirical development techniques: Wizard-of-Oz (WoZ)

Graphical user interface

Other modalities

Natural language

Dialogue manager

Traditional WOz

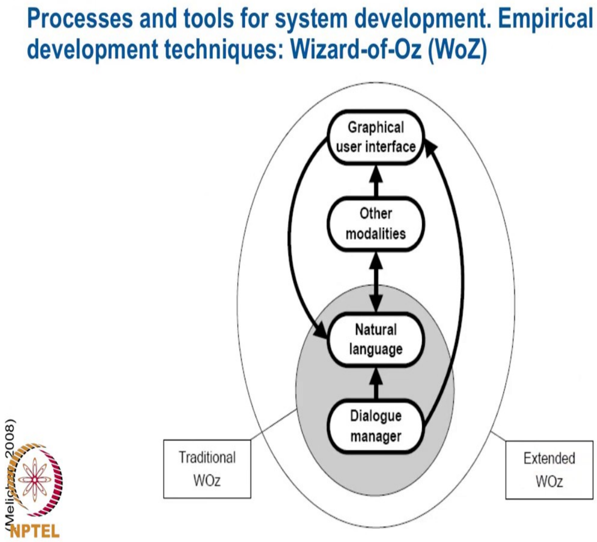
Extended WOz

(Melichar, 2008)

NPTEL

the wizard can be illustrated in the picture behind me.

(Refer Slide Time: 02:19)



Processes and tools for system development. Empirical development techniques: Wizard-of-Oz (WoZ)

Graphical user interface

Other modalities

Natural language

Dialogue manager

Traditional WOz

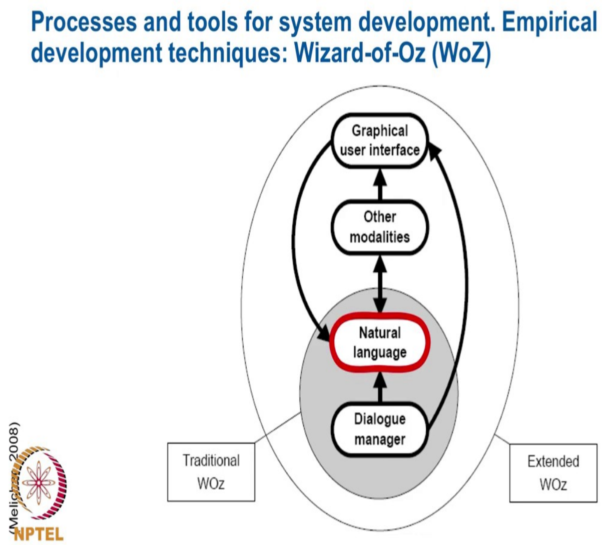
Extended WOz

(Melichar, 2008)

NPTEL

In a classical, for example spoken dialog system you would have a wizard replacing a natural language

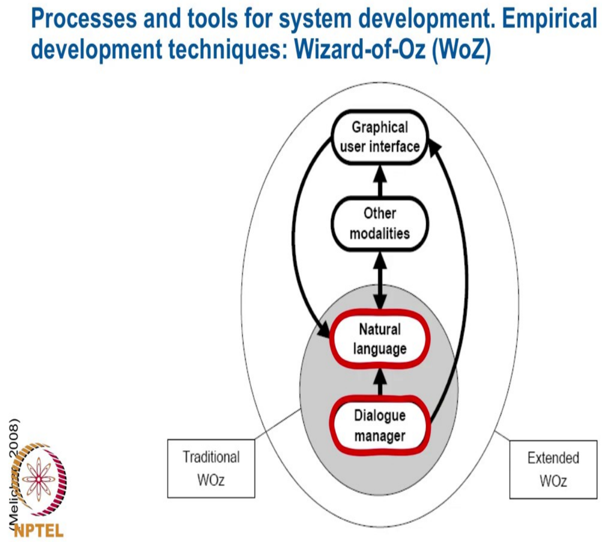
(Refer Slide Time: 02:25)



Processes and tools for system development. Empirical development techniques: Wizard-of-Oz (WoZ)

generation component and

(Refer Slide Time: 02:27)



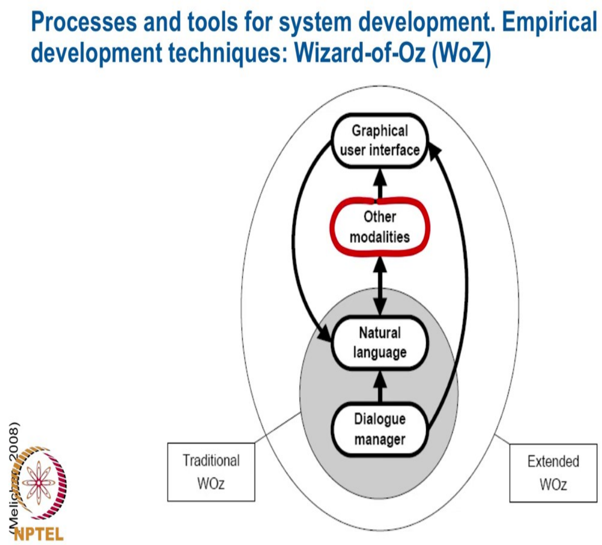
Processes and tools for system development. Empirical development techniques: Wizard-of-Oz (WoZ)

a dialog manager.

And as soon as a text-to-speech synthesizer becomes available then perhaps also this component here can be taken over by the machine.

If we have a multimodal interactive system we might want to also address other modalities,

(Refer Slide Time: 02:43)

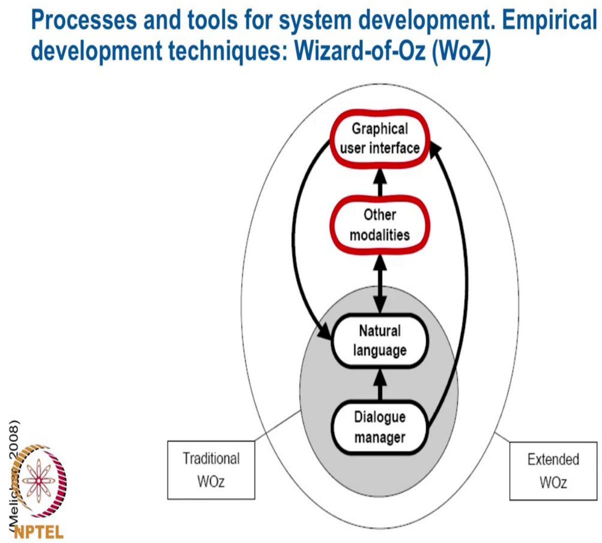


Processes and tools for system development. Empirical development techniques: Wizard-of-Oz (WoZ)

(Melichar, 2008)

for example gesture recognition, face recognition and things like that. It might also provide some output, additional output capabilities like a graphical user interface

(Refer Slide Time: 02:53)



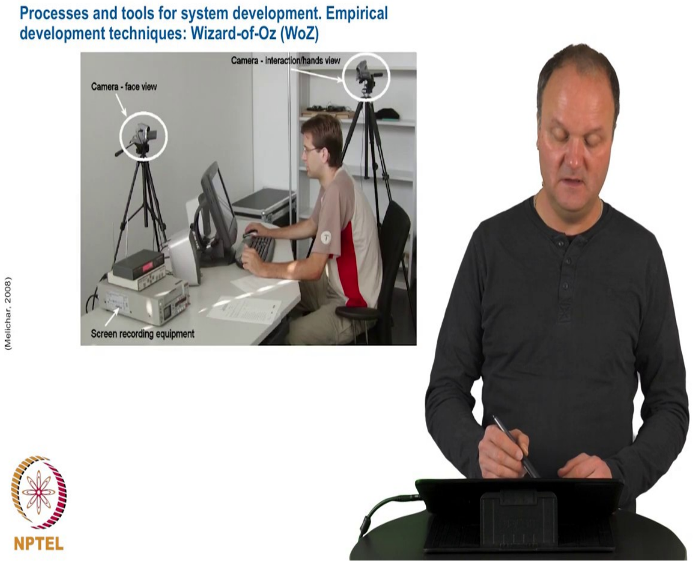
Processes and tools for system development. Empirical development techniques: Wizard-of-Oz (WoZ)

(Melichar, 2008)

when we are talking about extended Wizard-of-Oz scenario.

Such a Wizard-of-Oz setup can be quite

(Refer Slide Time: 03:01)



Processes and tools for system development. Empirical development techniques: Wizard-of-Oz (WoZ)

complicated. Here you see a view of Wizard-of-Oz setup from the human participants' point of view. So you see that the human is sitting in front of a computer, multimodal computer interaction.

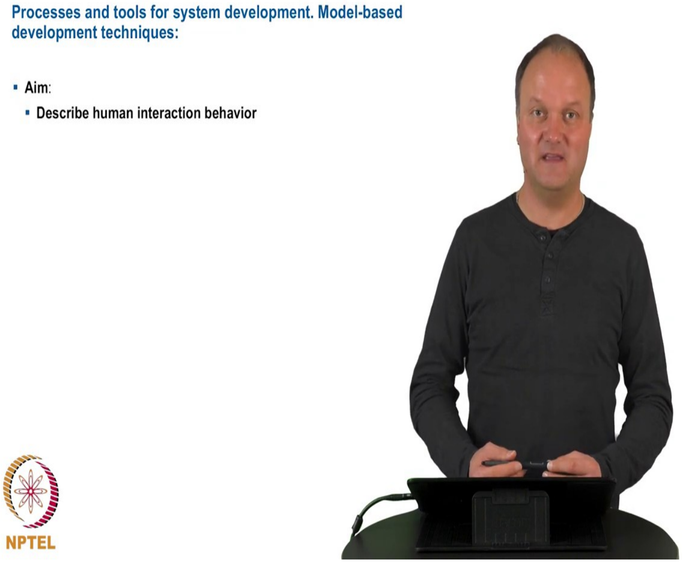
It is observed by a couple of cameras, may have a camera which shows the face view and the camera which showes the interaction and the enhanced view. We also have a screen recording equipment and so on, of course and a microphone put on a table so that we can follow what the test participant is doing.

And on the other side, there might be one or in this case even two wizards which observe and which simulate parts of the system.

Simulation may, for example be split into one experimenter doing just the transcription of what the user said, so replacing this speech recognition component and another one generating the behavior of the interactive system so replacing the dialog manager component.

But there might be different setup according to which components of the system have already been implemented and how well they work.

(Refer Slide Time: 04:13)



Processes and tools for system development. Model-based development techniques:

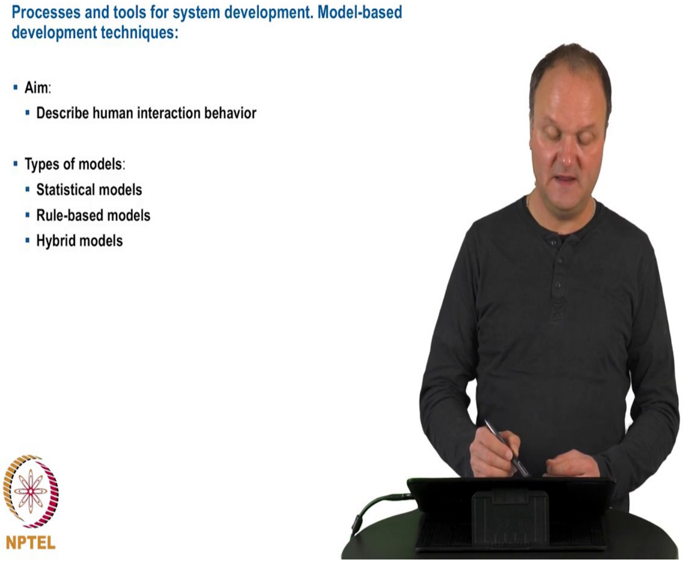
- Aim:
  - Describe human interaction behavior

The second type of technique for developing multimodal interactive system is based on models; that is algorithms which intend to describe human behavior.

Such a model can be used for making assumptions about the next interaction step but they can also be used for training for example a statistical dialog manager or for evaluating without relying on human test participants.

So you see that the purpose and the application areas of such models are multifold. There are different types of models with different

(Refer Slide Time: 04:47)



Processes and tools for system development. Model-based development techniques:

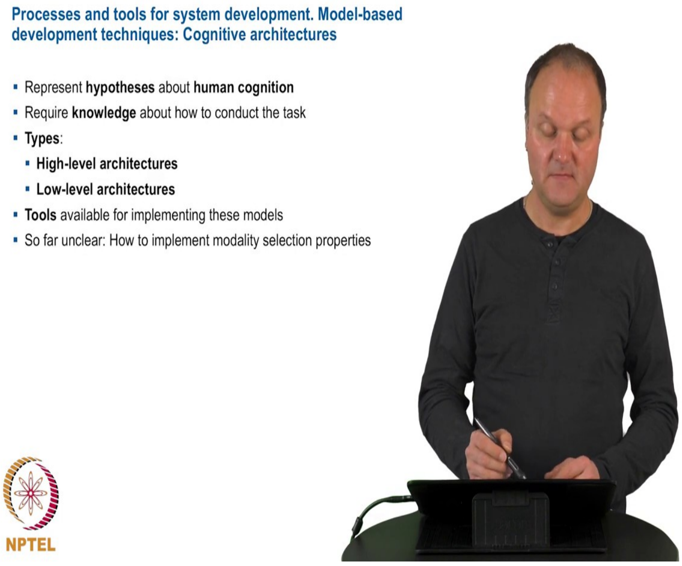
- Aim:
  - Describe human interaction behavior
- Types of models:
  - Statistical models
  - Rule-based models
  - Hybrid models

aims. Usually we distinguish between statistical models that are models learnt from data.

For example, using reinforcement learning, partially observable Markov decision processes or POMDPs would be an example of this type of model.

And on the other hand, rule-based models which describe human interaction behavior by rules. There are also hybrid models which combine statistical and rule-based approaches as we will see with the MeMo workbench at the end of this video.

(Refer Slide Time: 05:18)

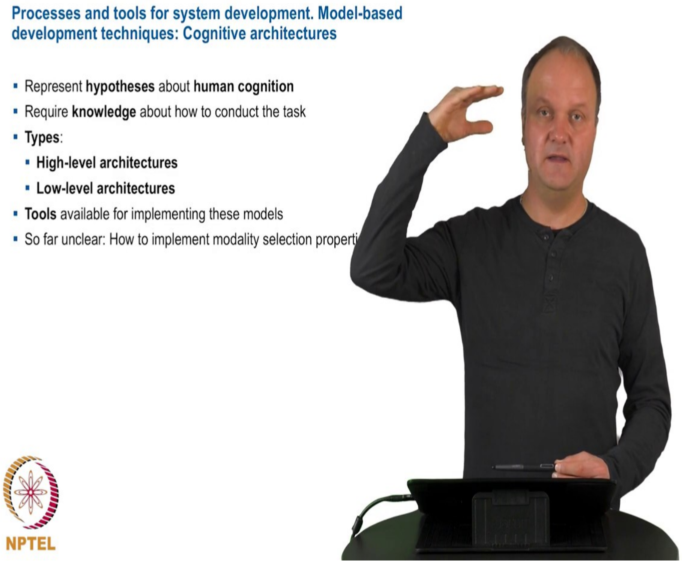


Processes and tools for system development. Model-based development techniques: Cognitive architectures

- Represent **hypotheses** about **human cognition**
- Require **knowledge** about how to conduct the task
- **Types**:
  - **High-level architectures**
  - **Low-level architectures**
- **Tools** available for implementing these models
- So far unclear: How to implement modality selection properties

A very popular class of model is the so-called cognitive architecture. Actually cognitive architecture tries to represent hypothesis about human cognition that is they are bias based quite

(Refer Slide Time: 05:34)



Processes and tools for system development. Model-based development techniques: Cognitive architectures

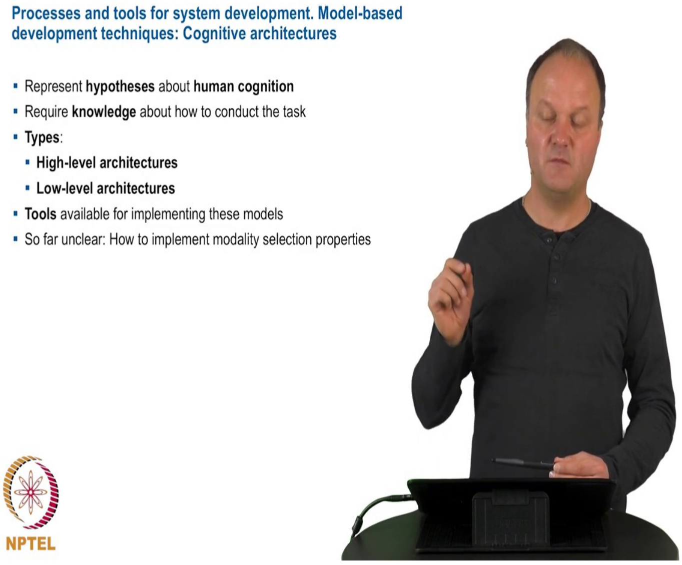
- Represent **hypotheses** about **human cognition**
- Require **knowledge** about how to conduct the task
- **Types**:
  - **High-level architectures**
  - **Low-level architectures**
- **Tools** available for implementing these models
- So far unclear: How to implement modality selection properti

upwards in the human decision making process.

These models require knowledge about how to conduct the task and this is knowledge which usually an expert needs to provide to the model in order to be functional. We can distinguish between so-called high level architectures with a relatively coarse temporal resolution. GOMS would an example of this class.

And low level architectures which describe interaction behavior in terms of

(Refer Slide Time: 06:03)



Processes and tools for system development. Model-based development techniques: Cognitive architectures

- Represent **hypotheses** about **human cognition**
- Require **knowledge** about how to conduct the task
- **Types**:
  - **High-level architectures**
  - **Low-level architectures**
- **Tools** available for implementing these models
- So far unclear: How to implement modality selection properties

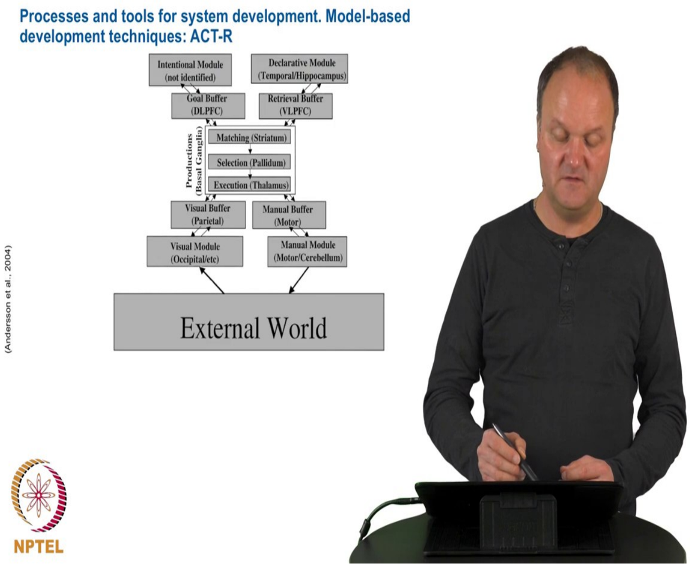
atomic events, EPIC or ACT-R or Soar would be examples of these low level architectures.

In order to implement these architectures and to make them useful for simulating human machine interaction, tools are available for example GLEAN or CogTool.

Unfortunately, it is so far unclear how modality specific tools for example rules for modality selection can be included inside such models, the very first approach is by Schaffer which try to model quantitative decision taking on modality selection.

A very popular

(Refer Slide Time: 06:44)



Processes and tools for system development. Model-based development techniques: ACT-R

Intentional Module (not identified)

Declarative Module (Temporal/Hippocampus)

Goal Buffer (DLPFC)

Retrieval Buffer (VLPFC)

Productions (Basal Ganglia)

Matching (Striatum)

Selection (Pallidum)

Execution (Thalamus)

Visual Buffer (Parietal)

Manual Buffer (Motor)

Visual Module (Occipital/etc)

Manual Module (Motor/Cerebellum)

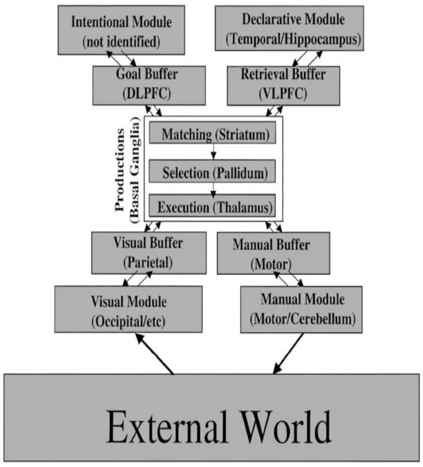
External World

(Andersson et al., 2004)

such cognitive architecture is ACT-R which you see in the picture behind me.

(Refer Slide Time: 06:50)

Processes and tools for system development. Model-based development techniques: ACT-R



Intentional Module (not identified)

Declarative Module (Temporal/Hippocampus)

Goal Buffer (DLPFC)

Retrieval Buffer (VLPFC)

Productions (Basal Ganglia)

Matching (Striatum)

Selection (Pallidum)

Execution (Thalamus)

Visual Buffer (Parietal)

Manual Buffer (Motor)

Visual Module (Occipital/etc)

Manual Module (Motor/Cerebellum)

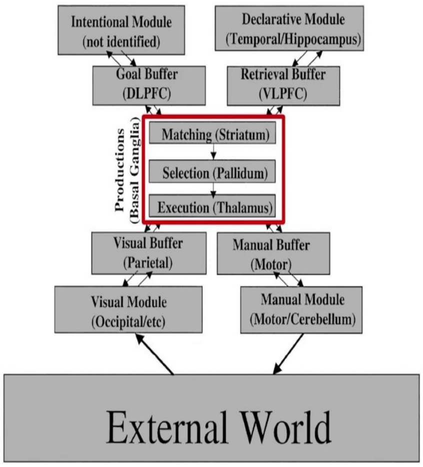
External World

(Anderson et al., 2004)

ACT-R is based on production engine which is at the center of the picture

(Refer Slide Time: 06:54)

Processes and tools for system development. Model-based development techniques: ACT-R



Intentional Module (not identified)

Declarative Module (Temporal/Hippocampus)

Goal Buffer (DLPFC)

Retrieval Buffer (VLPFC)

Productions (Basal Ganglia)

Matching (Striatum)

Selection (Pallidum)

Execution (Thalamus)

Visual Buffer (Parietal)

Manual Buffer (Motor)

Visual Module (Occipital/etc)

Manual Module (Motor/Cerebellum)

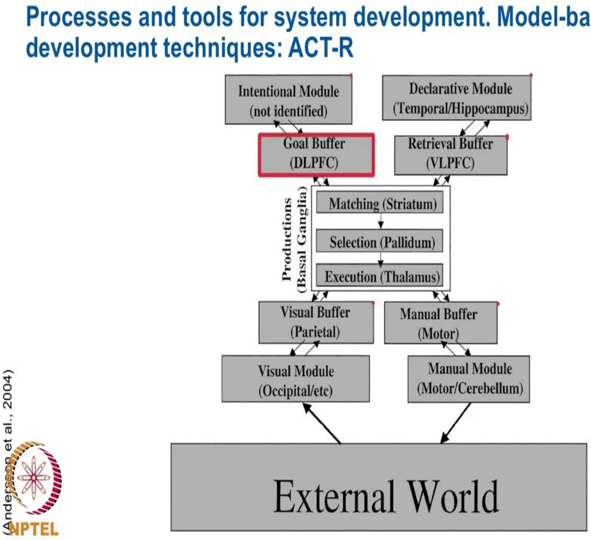
External World

(Anderson et al., 2004)

here, which is matching, selecting and execution, executing certain productions in terms of interaction steps.

This machine interacts with buffers for the goals,

(Refer Slide Time: 07:07)



Processes and tools for system development. Model-based development techniques: ACT-R

Intentional Module (not identified)

Declarative Module (Temporal/Hippocampus)

Goal Buffer (DLPFC)

Retrieval Buffer (VLPFC)

Productions (Basal Ganglia)

Matching (Striatum)

Selection (Pallidum)

Execution (Thalamus)

Visual Buffer (Parietal)

Manual Buffer (Motor)

Visual Module (Occipital/etc)

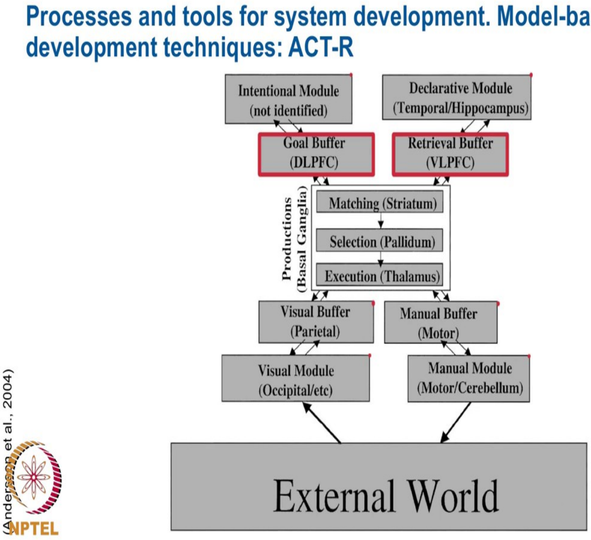
Manual Module (Motor/Cerebellum)

External World

(Anderson et al., 2004)

for the retrieval

(Refer Slide Time: 07:09)



Processes and tools for system development. Model-based development techniques: ACT-R

Intentional Module (not identified)

Declarative Module (Temporal/Hippocampus)

Goal Buffer (DLPFC)

Retrieval Buffer (VLPFC)

Productions (Basal Ganglia)

Matching (Striatum)

Selection (Pallidum)

Execution (Thalamus)

Visual Buffer (Parietal)

Manual Buffer (Motor)

Visual Module (Occipital/etc)

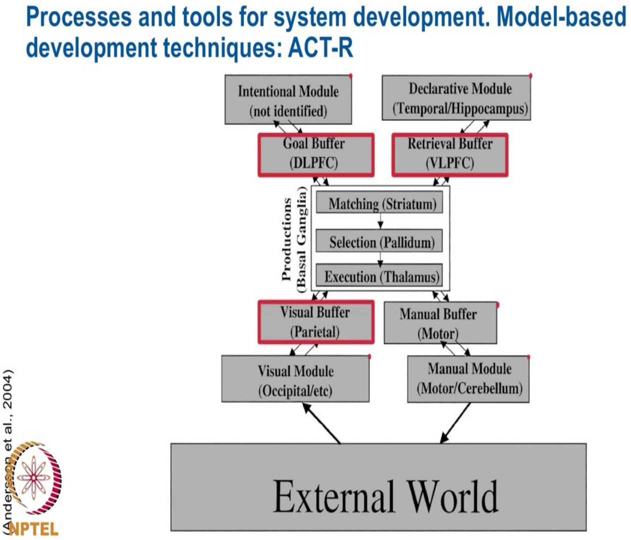
Manual Module (Motor/Cerebellum)

External World

(Anderson et al., 2004)

of information from memory, for the visual information

(Refer Slide Time: 07:12)

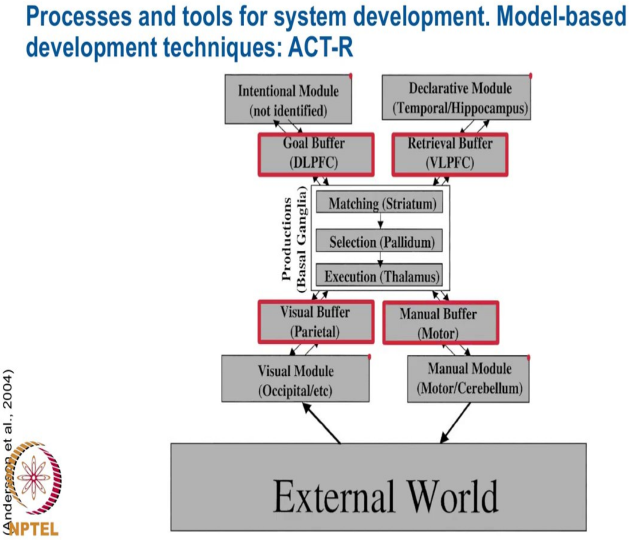


Processes and tools for system development. Model-based development techniques: ACT-R

(Anderson et al., 2004)

and for the manual information.

(Refer Slide Time: 07:14)

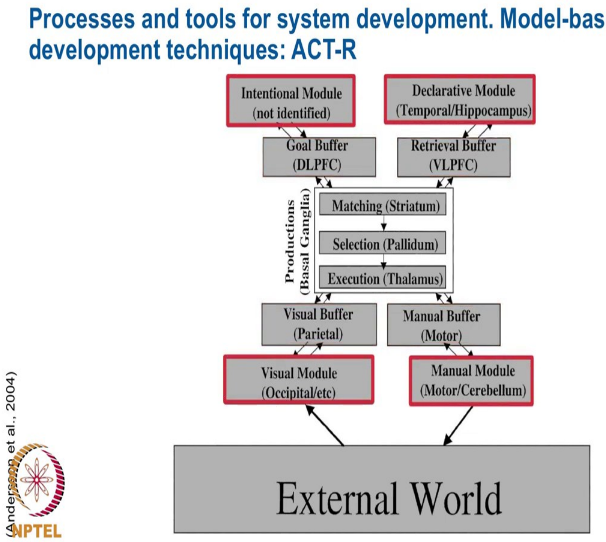


Processes and tools for system development. Model-based development techniques: ACT-R

(Anderson et al., 2004)

And for each of those Basal, buffers, there are modules available

(Refer Slide Time: 07:19)



Processes and tools for system development. Model-based development techniques: ACT-R

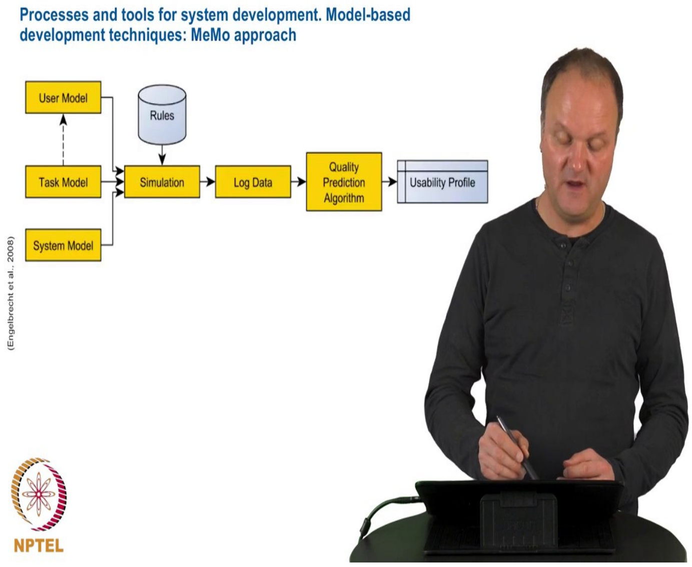
(Anderson et al., 2004)

which describe how the interaction actually would work. And some of those modules then have a connection to the external world.

These models have been used to describe interaction behavior but they are not very popular so far in practically building multimodal interactive systems.

A model which has the aim

(Refer Slide Time: 07:41)



Processes and tools for system development. Model-based development techniques: MeMo approach

(Engelbrecht et al., 2008)

to be practically usable is the so-called MeMo workbench which is an example for the hybrid approaches I depicted at the beginning of this video.

The MeMo workbench makes use of a user model, a task model and a system model in order to simulate human machine interaction behavior.

The system model is very easy. It can be deduced from the machine description, the task model as well because we know what the multimodal interactive system has been built for.

And then there is a user model. And the user model then tries to use the task model in order to turn this into a probable sequence of steps a user would perform with the help of the system.

I am talking about a probable sequence because such a user model may generate many, many different interaction paths with different probabilities and these interaction paths are manipulated by rules.

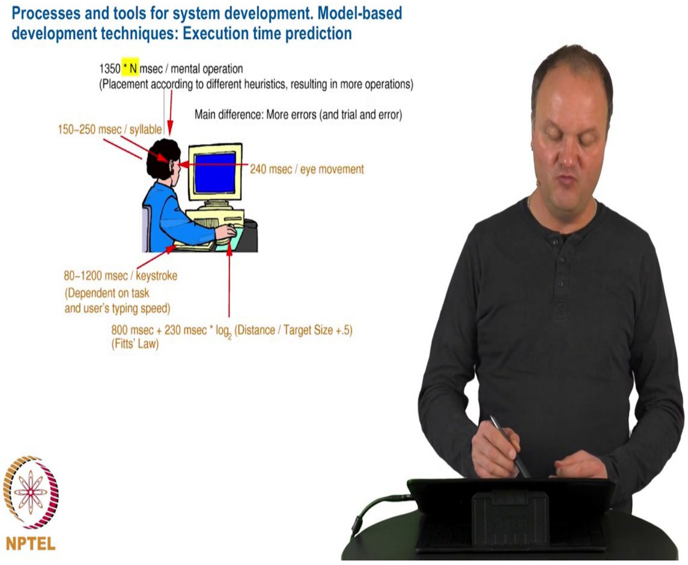
These rules are governed by either user characteristics, for example a user who has certain preferences for certain interactive modalities, who has certain knowledge, who is used to, who has experience with certain interaction techniques, but also rules which are governed by the characteristics of the system, that is whether there are one or two or many buttons on a web page or alike.

Then using these models, typical user machine interactions would be simulated. They would be logged and on the basis of the log files we can evaluate how well these interactions work, that is we can, for example determine the task success, we can determine the efficiency of the interaction

But perhaps also if the user has to step back several times in the flow of the interaction, then this might be judged as bad. So we might be able to provide a full usability profile of the simulated interaction between the user and the system.

In order to describe some specific characteristics of the user, there might be

(Refer Slide Time: 09:52)



Processes and tools for system development. Model-based development techniques: Execution time prediction

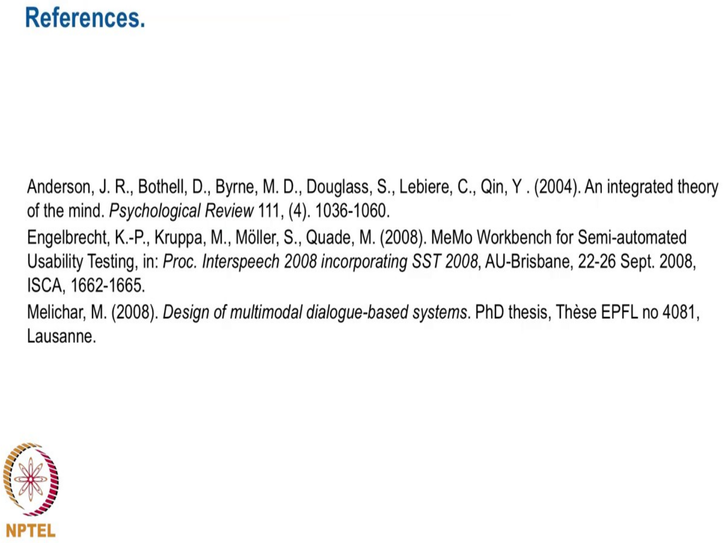
1350 * N msec / mental operation
(Placement according to different heuristics, resulting in more operations)
Main difference: More errors (and trial and error)
150–250 msec / syllable
240 msec / eye movement
80–1200 msec / keystroke
(Dependent on task
and user's typing speed)
800 msec + 230 msec * $\log_2$ (Distance / Target Size +.5)
(Fitts' Law)
NPTEL

rules being implemented in such an approach

For example a rule for predicting the execution time which you see here, which is actually based on Fitt's Law and which illustrates how long user will take in order to find a certain icon on a screen and then click that icon on the screen.

These rules may be used in the MeMo approach but they might also be attached to a cognitive architecture like ACT-R.

(Refer Slide Time: 10:22)



References.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., Qin, Y . (2004). An integrated theory of the mind. *Psychological Review* 111, (4). 1036-1060.

Engelbrecht, K.-P., Kruppa, M., Möller, S., Quade, M. (2008). MeMo Workbench for Semi-automated Usability Testing, in: *Proc. Interspeech 2008 incorporating SST 2008*, AU-Brisbane, 22-26 Sept. 2008, ISCA, 1662-1665.

Melichar, M. (2008). *Design of multimodal dialogue-based systems*. PhD thesis, Thèse EPFL no 4081, Lausanne.