DT2112 Speech Technology

Jens Edlund, edlund@speech.kth.se

ROYAL INSTITUTE OF TECHNOLOGY

Jens Edlund

- Masters in Computational Linguistics, SU
  - Phonetics general linguistics
- PhD Speech technology, KTH
- 1996-1998 Spoken dialogue system researcher at Telia Research and SRI Cambridge
- 1999- Spoken dialogue system researcher at KTH

ROYAL INSTITUTE OF TECHNOLOGY

Speech technology: why and what?

- A broad, multidisciplinary area: any technology in which speech is central, and its applications
- Core technologies include
  - Speech synthesis and generation
  - Speech recognition and understanding
  - Dialogue mangement
- Aorund speech R&D:
  - Data collection
  - Experimentation
  - Evaluation

ROYAL INSTITUTE OF TECHNOLOGY

Speech hype?

Bill Gates 1997:
"In this 10-year time frame, I believe that we'll not only be using the keyboard and the mouse to interact, but during that time we will have perfected speech recognition and speech output well enough that those will become a standard part of the interface."

Bill Gates 2011:
"The next big thing is definitely speech and voice recognition [...]"

ROYAL INSTITUTE OF TECHNOLOGY

Speech as a hardware selling point

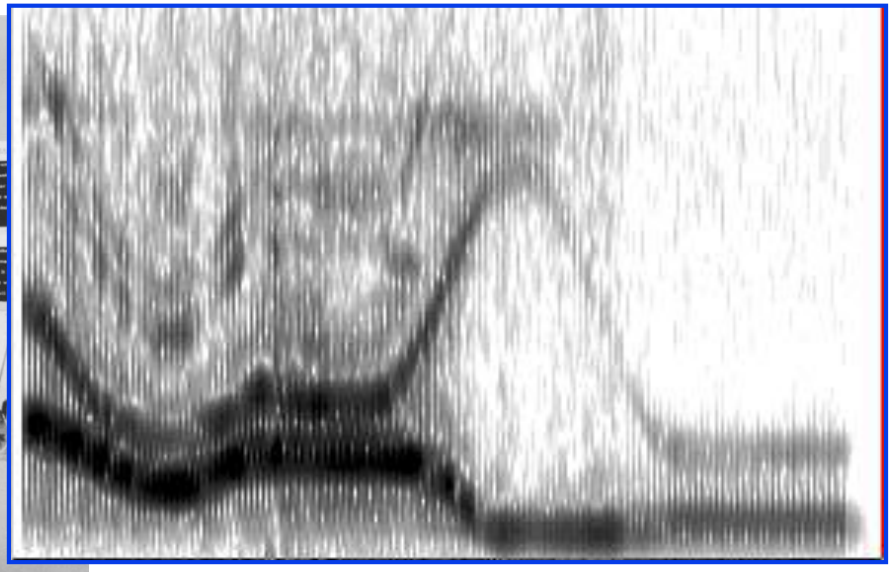
Mobile

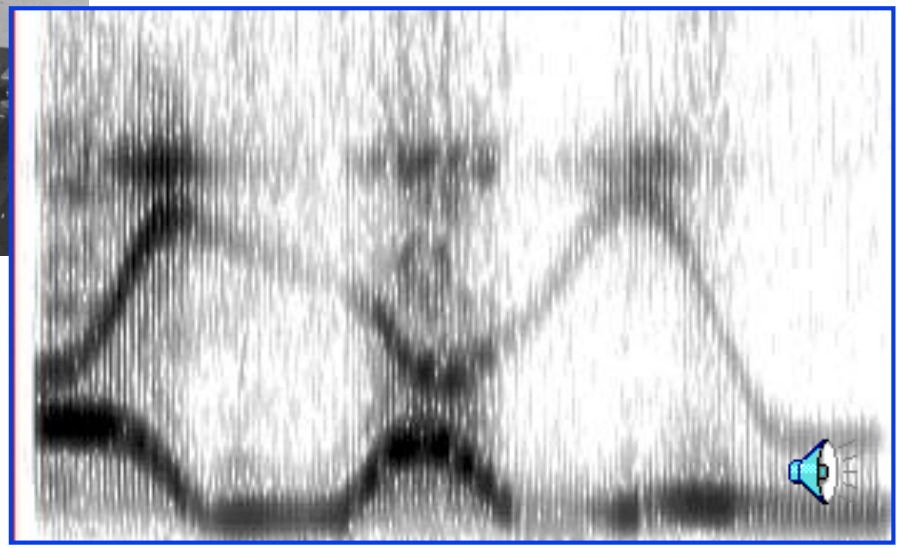
Cars

Ford

The KTH Speech group, early days

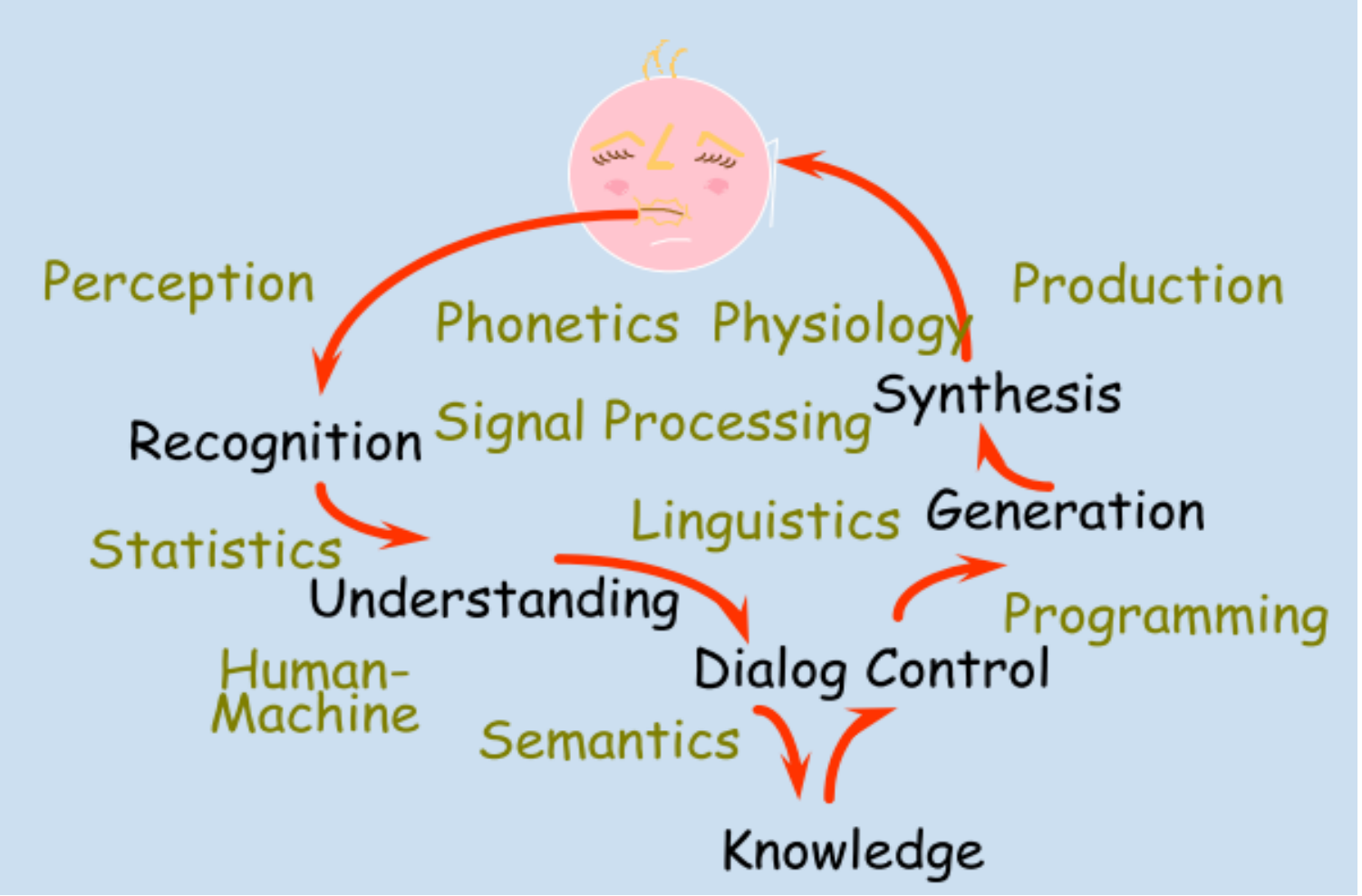
Gunnar Fant and
OVE I 1953





ROYAL INSTITUTE OF TECHNOLOGY

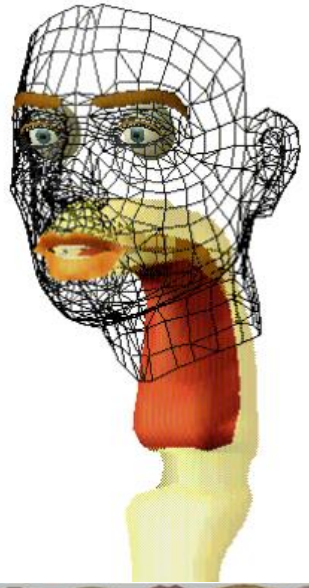
A multi-disciplinary field

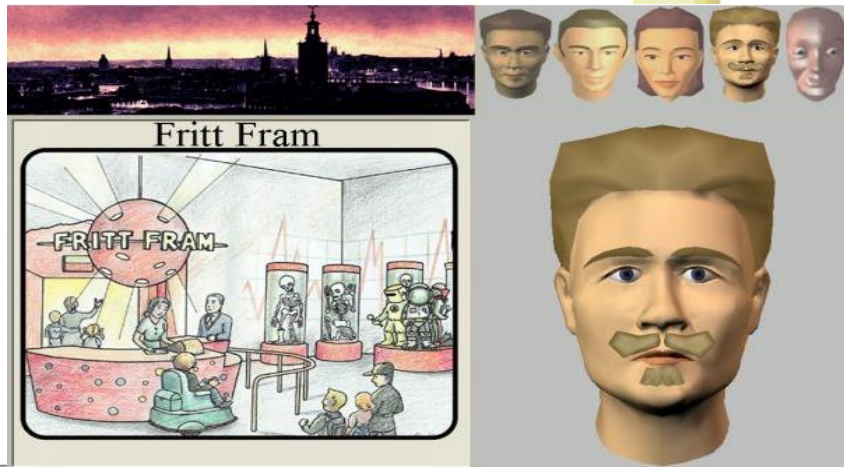


CTT – centrum för talteknologi

Research areas

- Speech production
- Speech perception
- Communication aids
- Multimodal speech synthesis
- Speech recognition
- Conversational speech
- Interactive dialogue systems





ROYAL INSTITUTE OF TECHNOLOGY

Course info

- All lectures held here (Fantum)
- Communication mainly through email: edlund@speech.kth.se

- https://www.kth.se/social/course/DT2112/

ROYAL INSTITUTE OF TECHNOLOGY

Course overview

In-lecture exercises

Lectures:

Human perception and verbal communication
- Speech production by humans
- Speech recognition by humans

Human-Computer Interaction
- Speech production by computers
- Speech & speaker recognition by computers
- Visual speech synthesis and applications
- Dialogue systems
- Data collection
- Evaluation

Participation in research experiments

3 Lab exercises:
1. Speech recognition
2. Dialogue systems
3. Speech synthesis/eval

No written exam!

Home assignments

Project:
You choose, implement, report, present

ROYAL INSTITUTE OF TECHNOLOGY

DT2112 Schedule

- Thursdays!
- Full days, start 10:00 (schedule change!)
  - Today
  - Jan 21th
  - Jan 28th
  - Feb 4th
  - Feb 11th
  - Feb 25th (project seminar)
  - Mar 4th (date open - project seminar)
- Calendar on Social is incorrect!

ROYAL INSTITUTE OF TECHNOLOGY

Lectures

- Appear!
- We take breaks…

ROYAL INSTITUTE OF TECHNOLOGY

Project work, home exam, labs…

- On-site work spaces and “red sofas”
- Coffee machine ☺

ROYAL INSTITUTE OF TECHNOLOGY

E-mail addresses

- Your e-mail two ways
  - Paper list
  - Send to edlund@speech.kth.se with DT2112 in subject
- Always put DT2112 in subject line when writing to me

ROYAL INSTITUTE OF TECHNOLOGY

Names!

- Make signs, place in front of you
- Re-use, unless you really like making signs

Lectures

Speech perception/production by humans

(David House)

Speech synthesis / Dialogue systems:
(Joakim Gustafson)

Speech recognition

(Giampiero Salvi)

Visual speech synthesis

(Jonas Beskow)

ROYAL INSTITUTE OF TECHNOLOGY

Lab exercises

- Lab1 speech recognition
- Giampiero Salvi

- Lab2 speech synthesis
- Jens Edlund

- Lab 3 dialogue systems
- Gabriel Skantze

ROYAL INSTITUTE OF TECHNOLOGY

Requirements and grades

Home exam:
Satisfactory solutions for each assignment submitted on time.

Project:
Choose, Implement, Report, Present

3 laboratory exercises, with short written reports

Grades from A-F on *each of the home assignments* and on the *project report* is the main basis for the final grade.

Extra grading point will be assigned for attending research experiment (perception studies)

Spoken dialogue systems

ROYAL INSTITUTE OF TECHNOLOGY

The spoken dialogue system vision

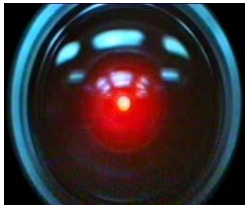
An interface that allows speakers to interact with a computer using spontaneous, unconstrained speech

ROYAL INSTITUTE OF TECHNOLOGY

Talking machines as we know them from movies…

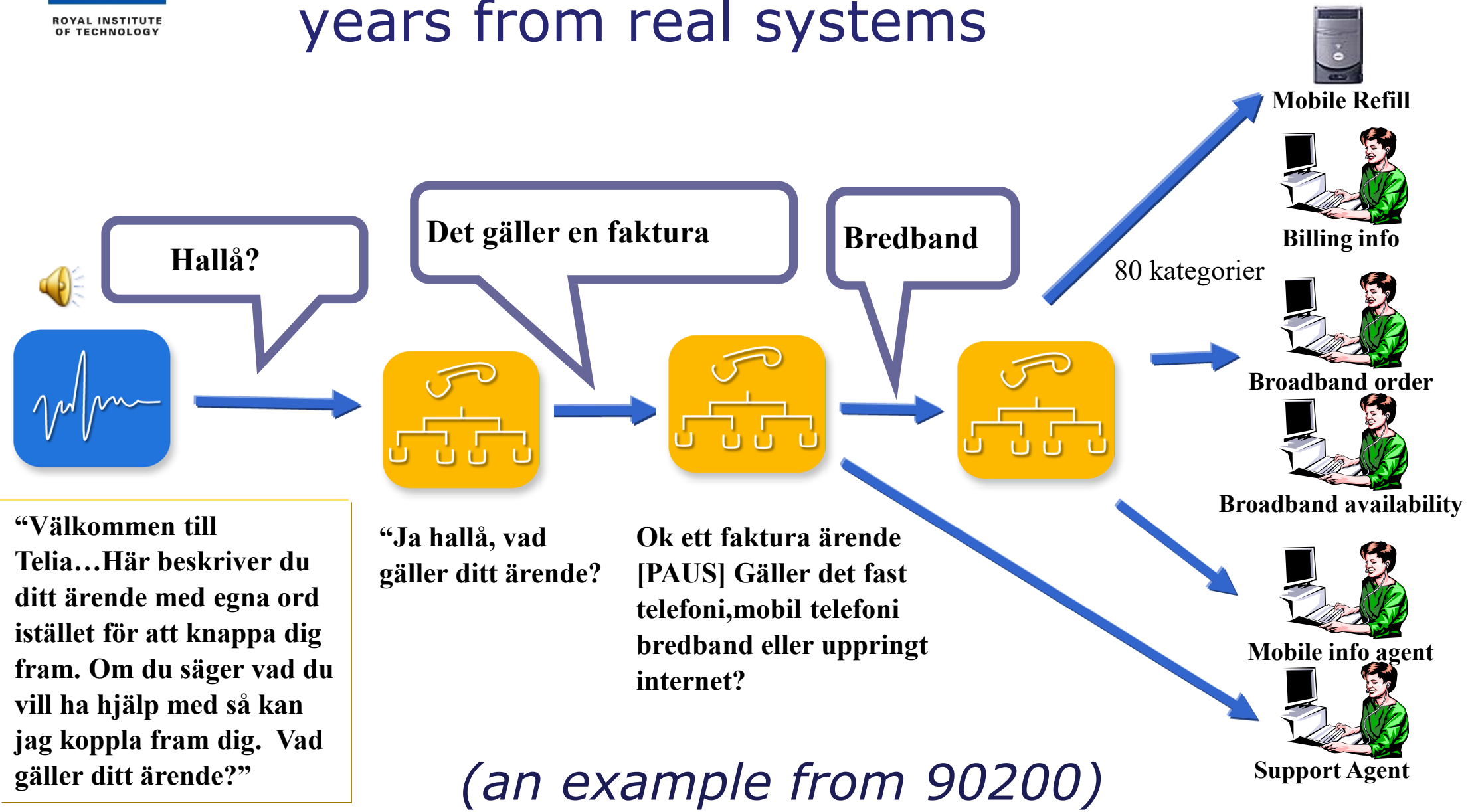
- Star Trek


- HAL 9000 (2001)



- C3P0 (Star Wars)

...and since a couple of years from real systems



(an example from 90200)

ROYAL INSTITUTE OF TECHNOLOGY

The beauty of speech

1. Works in hands free situations
2. Works in eyes free situations
3. Works when other interfaces are inconvenient:
4. Works where disabilities make other interfaces useless
5. Works with common hardware, e.g. a telephone
6. Efficient information transfer (as far as humans are concerned).

ROYAL INSTITUTE OF TECHNOLOGY

Speech application Domains

- Information retrieval systems
  - e.g. train time table information or directory inquiries
- Ordering
  - ticket booking, buying music
- Command control systems
  - home control (“turn the radio off”) or voice command shortcuts (“save”)
- Dictation
- Navigation
- Mobile assistants (siri, google search)

ROYAL INSTITUTE OF TECHNOLOGY

The beauty of speech (cont)

7. Reasoning.
8. Problem solving.
9. Naturalness
10. Easy-of-use
11. Flexibility
12. Error handling and hedging
13. Mutual adaptation enables error handling and more efficient information transfer
14. Social, bond-building.

ROYAL INSTITUTE OF TECHNOLOGY

Application Domains (cont)

- Games and entertainment
  - Games can take advantage of the more social aspects of spoken dialogue
- Co-ordinated collaboration
  - The task of controlling or over-viewing complex situations requires flexibility and efficiency.
- Expert systems
  - Diagnose and help systems that need to reason about facts and goals and may benefit from natural dialogue
- Learning and training
  - Naturalness, flexibility, and robustness are attractive features in training environments

ROYAL INSTITUTE OF TECHNOLOGY

Speech enabled computer games and interactive movies

Tom Clancy's EndWar

13th Street's Last Call

ROYAL INSTITUTE OF TECHNOLOGY

How do we want speaking machines to behave?

ROYAL INSTITUTE OF TECHNOLOGY

Interface metaphors

- Exploit specific knowledge that users already have from other domains
- Give the user instantaneous knowledge about how to interact
- Examples include
  - The desktop – the computer is a workplace
  - File system – the computer is a filing cabinet
  - Games – the computer is a toy

ROYAL INSTITUTE OF TECHNOLOGY

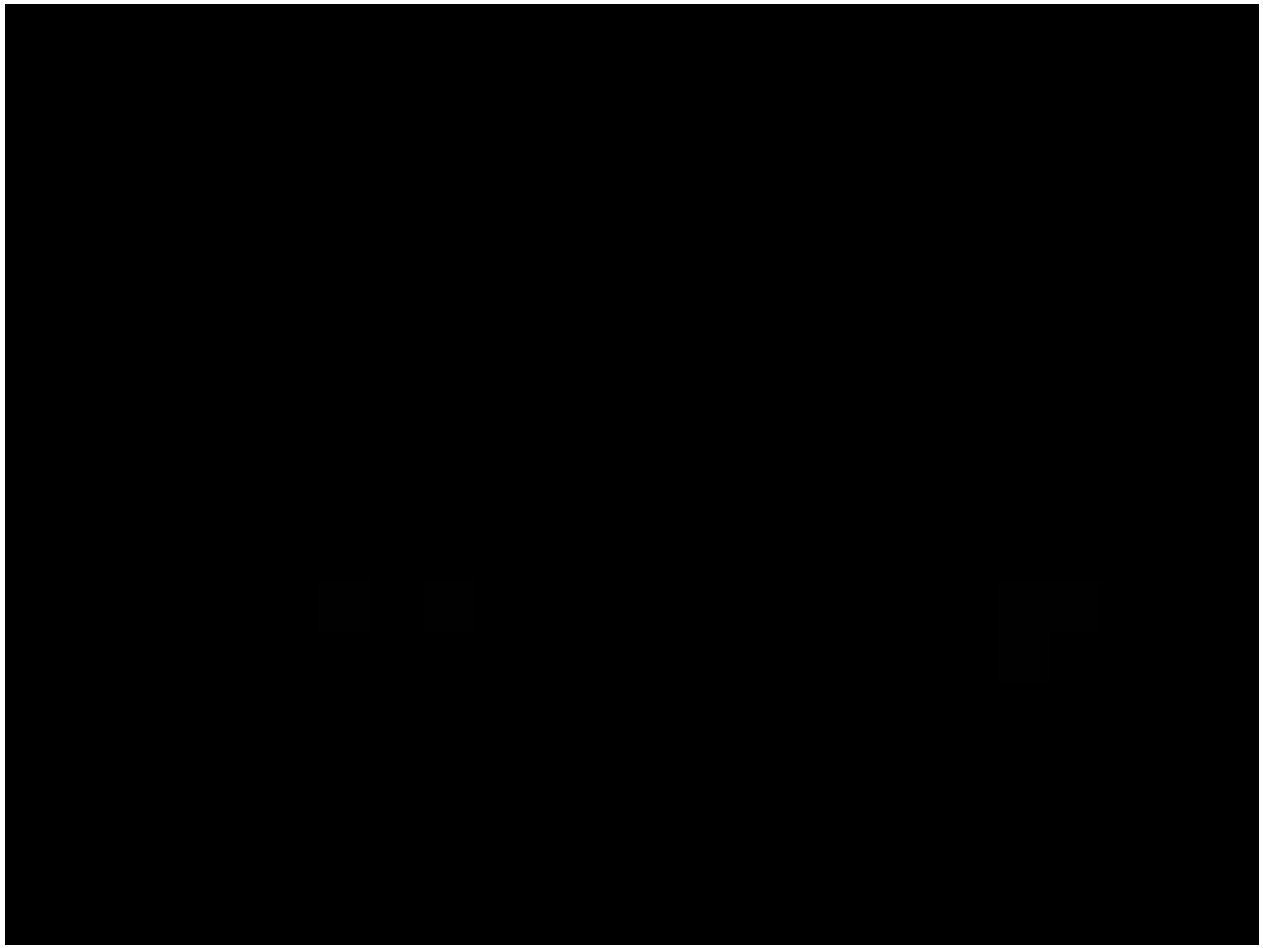
Metaphors for speech interfaces

- The tool/interface metaphor
  - Speech as an alternative interface technology in GUIs (multimodality)
- The servant/human metaphor
  - The computer as an entity with human-like conversational abilities

<u>One is not necessarily better than the other!</u>

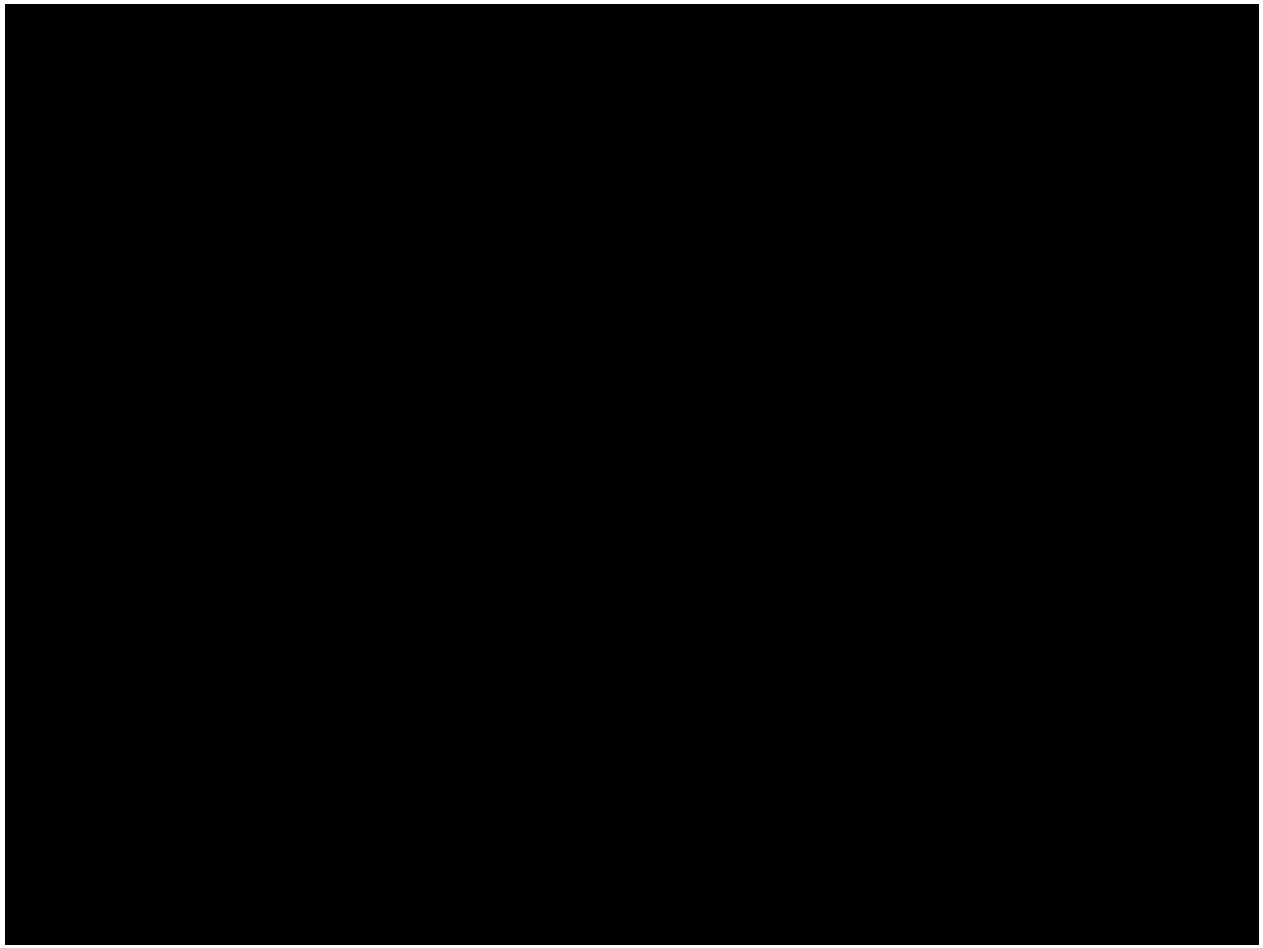
ROYAL INSTITUTE OF TECHNOLOGY

Interface metaphor



ROYAL INSTITUTE OF TECHNOLOGY

Human-like metaphor



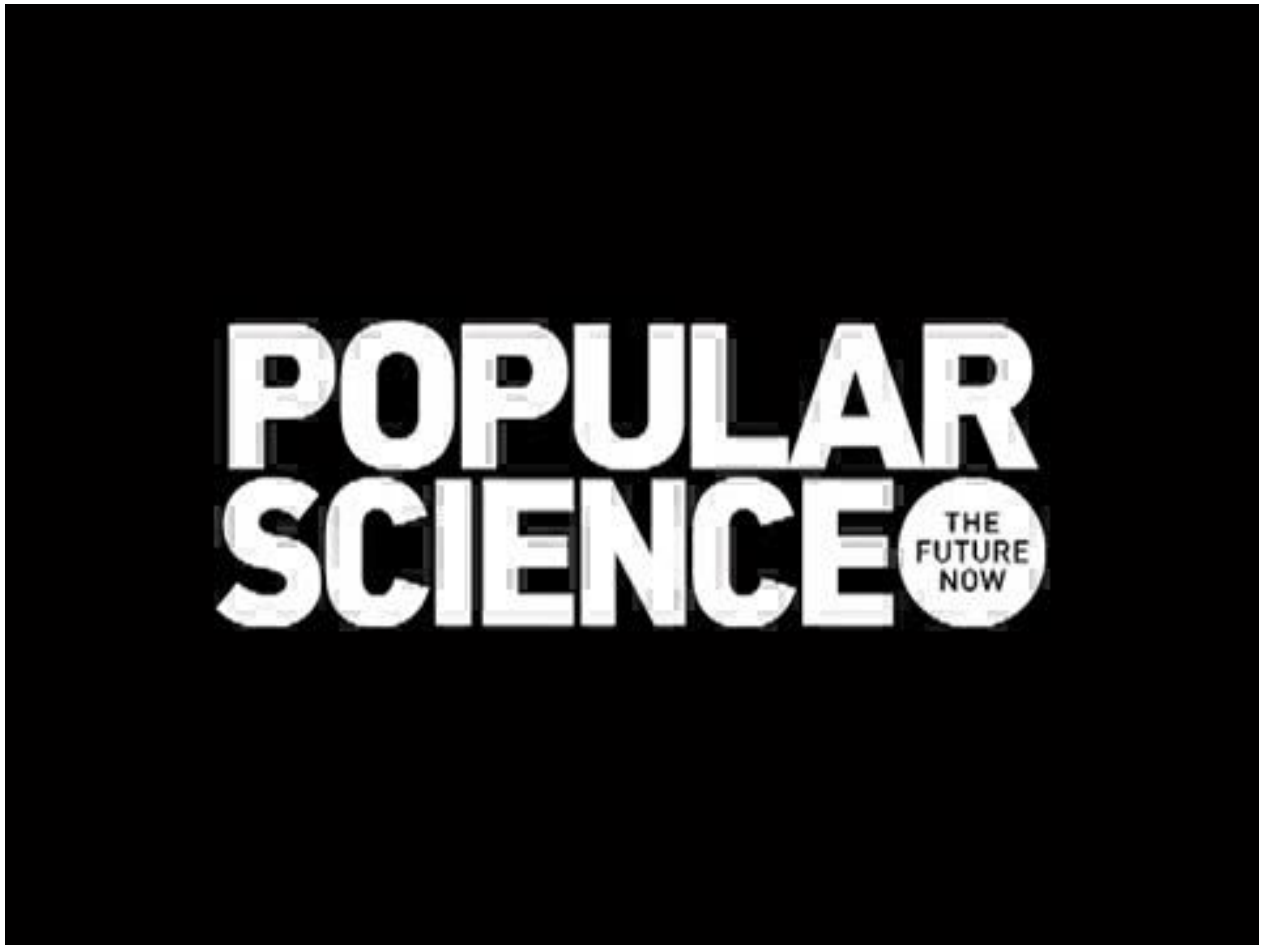
ROYAL INSTITUTE OF TECHNOLOGY

Applications of machines with human metaphor interaction

- Research tool
  - Cassel: "a machine that acts human enough that we respond to it as we respond to another human"
- Second Language Learning
- Social Robots
- Computer Games

ROYAL INSTITUTE OF TECHNOLOGY

Is more human-like always better?



ROYAL INSTITUTE
OF TECHNOLOGY

Human-like interaction control

ROYAL INSTITUTE
OF TECHNOLOGY

Thank you!

Questions?