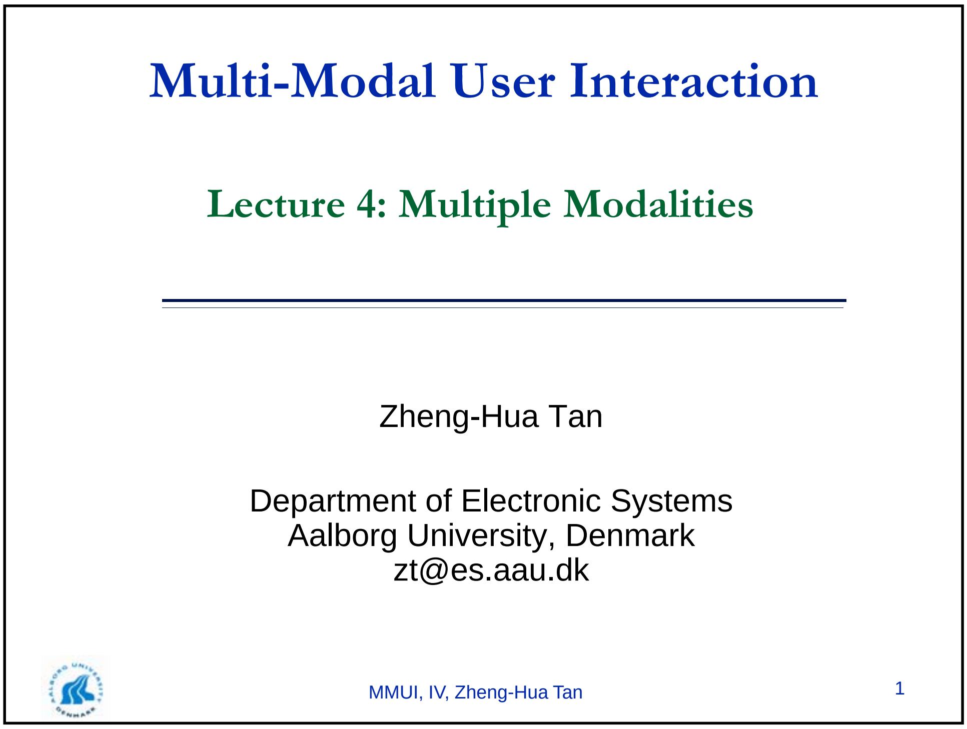
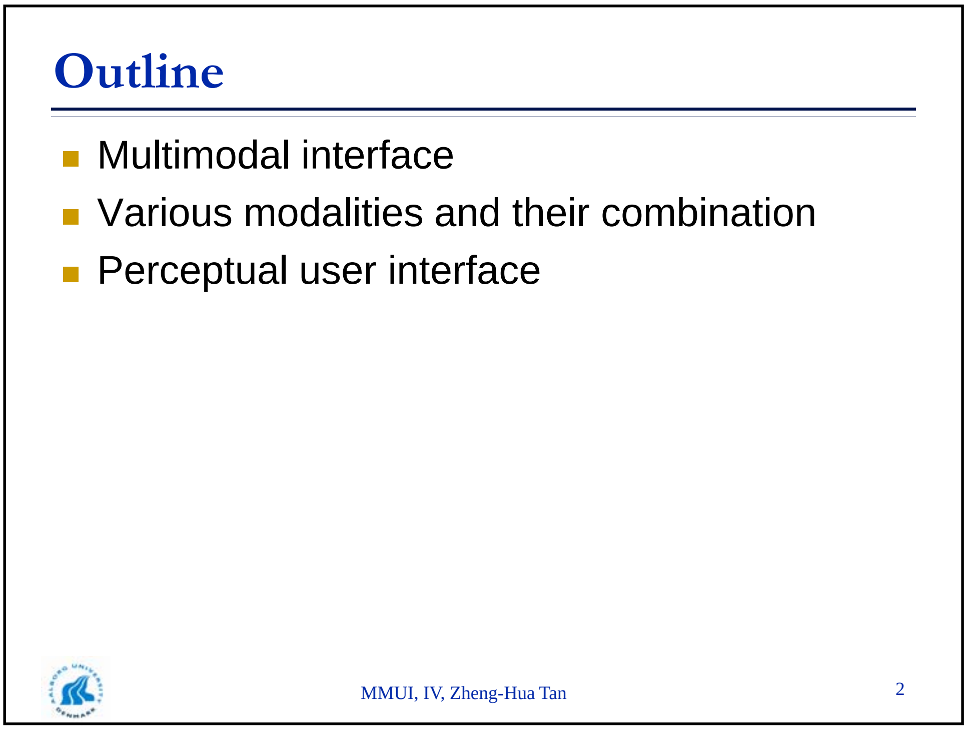


Multi-Modal User Interaction

Lecture 4: Multiple Modalities

Zheng-Hua Tan

Department of Electronic Systems
Aalborg University, Denmark
zt@es.aau.dk



Outline

- Multimodal interface
- Various modalities and their combination
- Perceptual user interface

Multimodal system characteristics

- Recognition of simultaneous or alternative individual modes
- Type and size of
  - gesture vocabulary
  - speech vocabulary (and grammar)
  - …
- Type of signal fusion
- Type of platform and applications

Challenges

- Development of cognitive theories to guide multimodal system design
- Development of effective natural language processing
- Dialogue processing
- Error-handling techniques
- Function robustly and adaptively
- Support for collaborative multiperson use

http://www.geekstir.com/project-natal-milo-**xbox-360**

What is multimodal interface

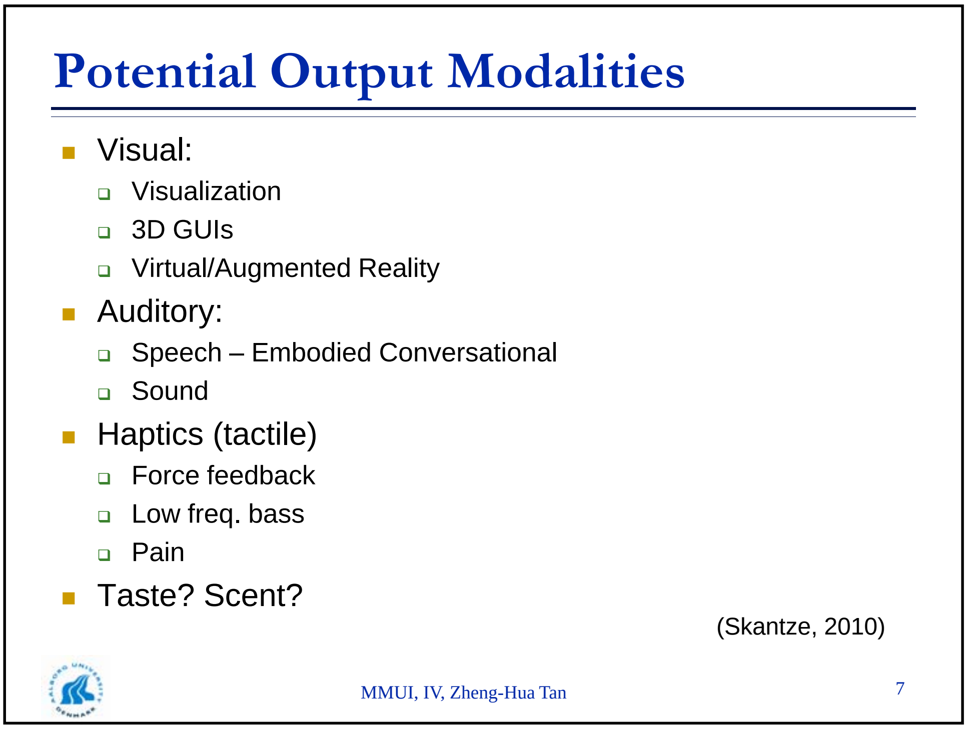
- Humans perceive the world through senses.
  - Touch, Smell, Sight, Hearing, and Taste
  - A mode = Communication through one sense
- Computers process information through modes
  - Keyboard, Microphone, Camera etc.
- Multimodal Interfaces try to combine several different modes of communicating: Speech, gesture, sketch …
  - Provide user with multiple modalities (communication skills)
  - Multiple styles of interaction
  - Simultaneous or not
- Fine-grained distinctions:
  - Visual: Graphics, Text, Simulation
  - Auditory: Speech, Non-verbal sounds

(Skantze, 2010)

Multimedia vs multimodal

- Multimedia – more than one mode of communication is output to the user
  - E.g. a sound clip attached to a presentation.
  - Media channels: Text, graphics, animation, video: all visual media
- Multimodal – computer processes more than one mode of communication.
  - E.g. the combined input of speech and touch in new mobile phones
  - Sensory modalities: Visual, auditory, tactile, …

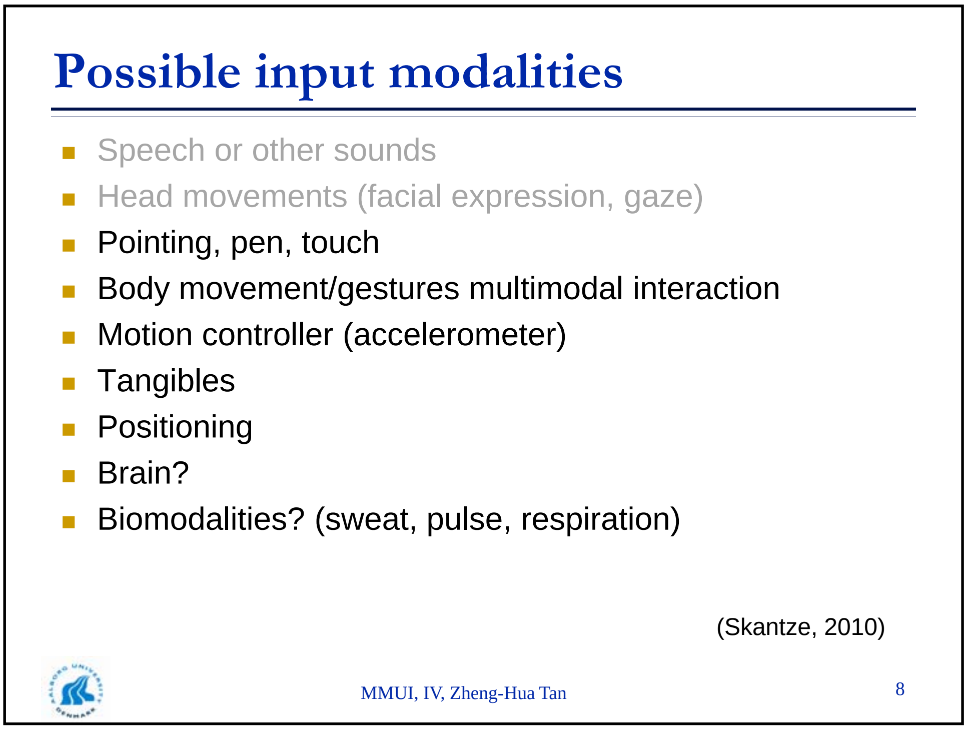
(Skantze, 2010)



Potential Output Modalities

- Visual:
  - Visualization
  - 3D GUIs
  - Virtual/Augmented Reality
- Auditory:
  - Speech – Embodied Conversational
  - Sound
- Haptics (tactile)
  - Force feedback
  - Low freq. bass
  - Pain
- Taste? Scent?

(Skantze, 2010)



Possible input modalities

- Speech or other sounds
- Head movements (facial expression, gaze)
- Pointing, pen, touch
- Body movement/gestures multimodal interaction
- Motion controller (accelerometer)
- Tangibles
- Positioning
- Brain?
- Biomodalities? (sweat, pulse, respiration)

(Skantze, 2010)

Outline

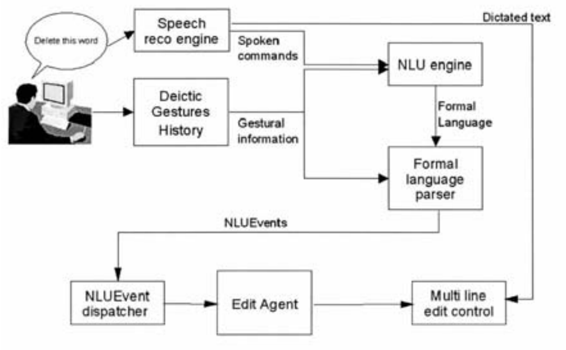
- Multimodal interface
- Various modalities and their combination
- Perceptual user interface

Multimodal speech and pen-gesture applications

- Interpret speech and pen-based gestureal input in a robust manner
- Bolt’s ”Put That There” concept

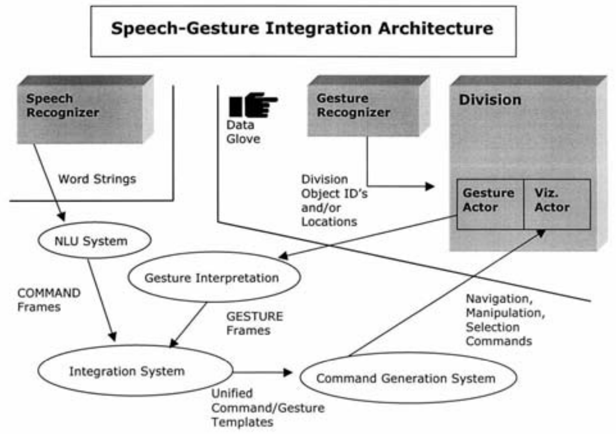
IBM's human-centric word processor

- Architectural flow of signal and language processing



"Delete this word <points to word>."

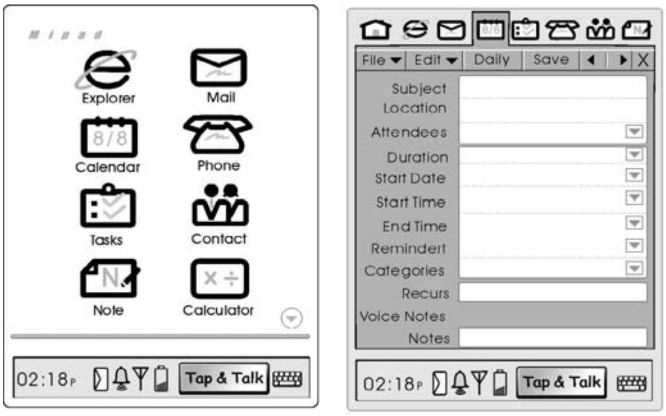
Boeing's speech and gesture system


Speech-Gesture Integration Architecture
Speech Recognizer
Data Glove
Gesture Recognizer
Division
Word Strings
Division Object ID's and/or Locations
Gesture Actor
Viz. Actor
NLU System
Gesture Interpretation
COMMAND Frames
GESTURE Frames
Navigation, Manipulation, Selection Commands
Integration System
Command Generation System
Unified Command/Gesture Templates

"Give me that <points to an object>."

"Fly forward", or a flying gesture.

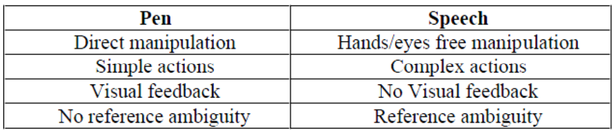
Microsoft MIPAD: A multimodal interaction prototype



(Huang et al. 2001)

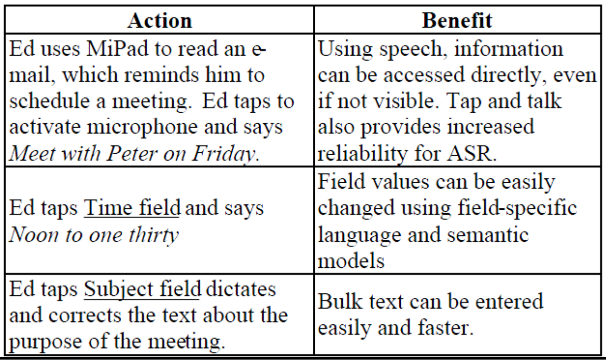
Microsoft MIPAD: A multimodal interaction prototype

Table 1 Complementary strengths of pen and speech as input modalities



Pen	Speech
Direct manipulation	Hands/eyes free manipulation
Simple actions	Complex actions
Visual feedback	No Visual feedback
No reference ambiguity	Reference ambiguity

Table 2 Benefits to have speech and pen for MiPad



Action	Benefit
Ed uses MiPad to read an e-mail, which reminds him to schedule a meeting. Ed taps to activate microphone and says *Meet with Peter on Friday.*	Using speech, information can be accessed directly, even if not visible. Tap and talk also provides increased reliability for ASR.
Ed taps Time field and says *Noon to one thirty*	Field values can be easily changed using field-specific language and semantic models
Ed taps Subject field dictates and corrects the text about the purpose of the meeting.	Bulk text can be entered easily and faster.

(Huang et al. 2001)

Microsoft MIPAD: A multimodal interaction prototype

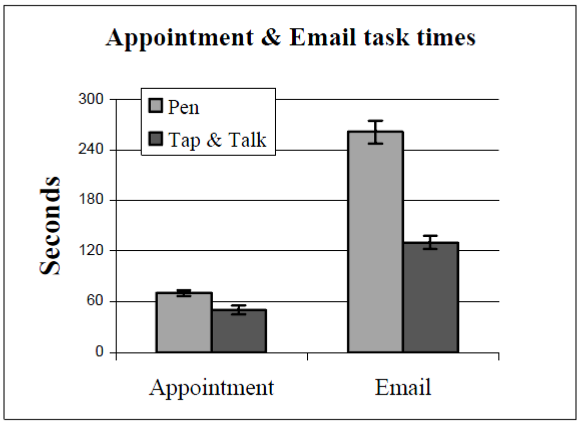
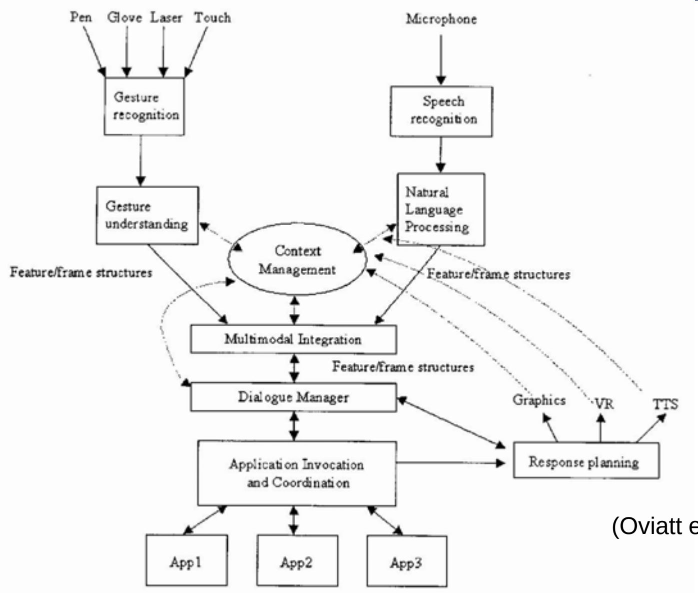


Figure 5 Task completion time of email transcription between the pen -only interface and *Tap and Talk* interface. The standard deviation is also shown above the bar of each performed task.

(Huang et al. 2001)

Typical info flow in a multimodal architecture

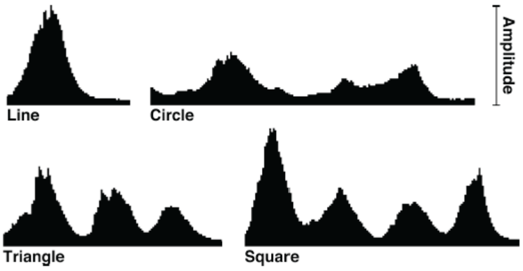


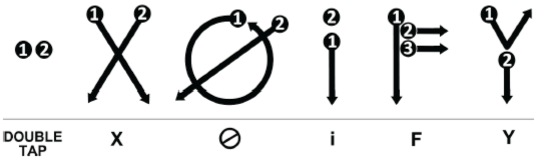
(Oviatt et al. 2000)

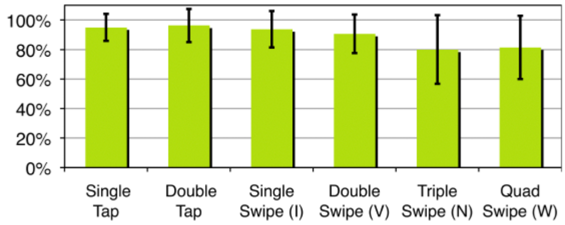
Interactive surface

- Scratch input
- Touch table

Scratch input

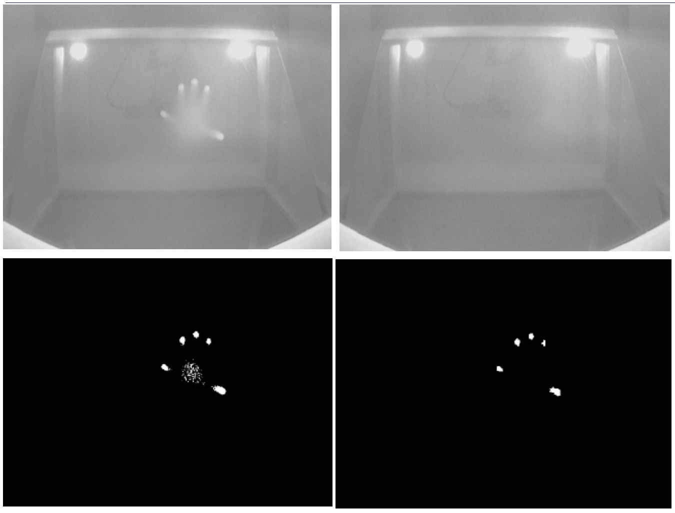




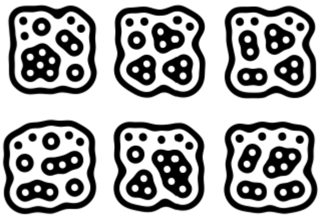


(C. Harrison, 2008)

Touch table

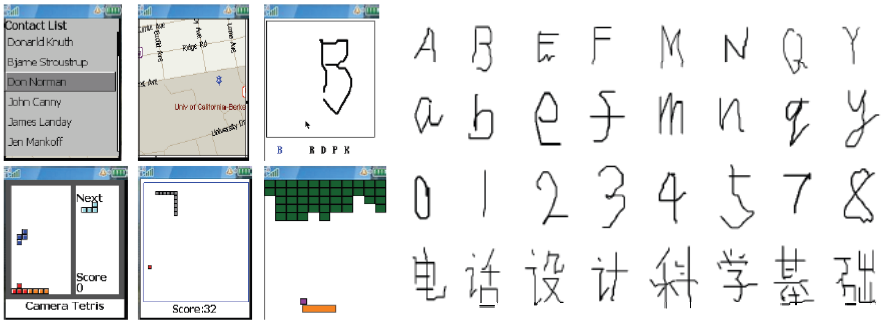


Reactable



Fiducial markers used in reacTIVision

Camera Phone Based Motion Sensing



(Wang et al. 2006)

Egocentric interaction

- Exploits the spatial relation between user and device and uses changes in this relation as input commands.

(T. Luel and F. Mazzone, 2009)

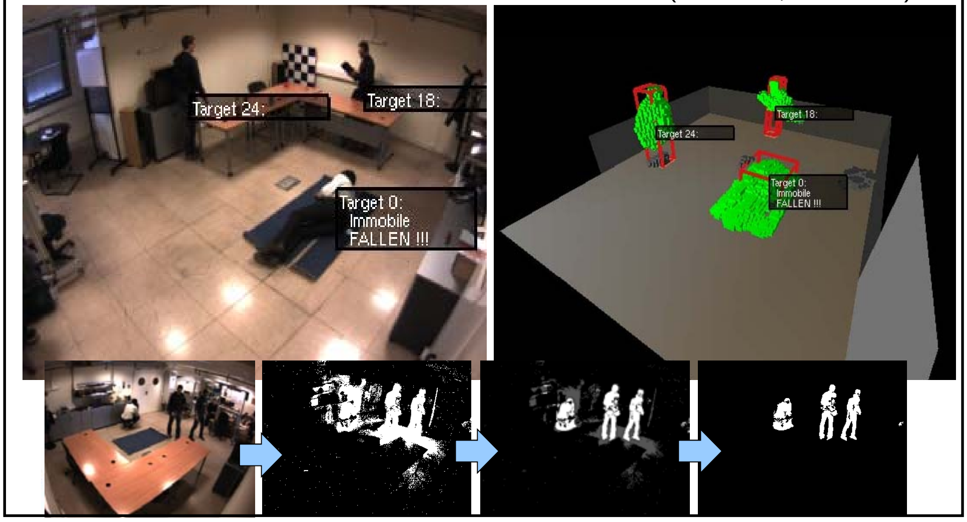
(M.H. Justesen, et al. 2010)

3D sensing

a) an image with persons and information overlay

b) detected foreground and information

Elderly care, survelience
(Andersen, et al. 2010)



Finding information

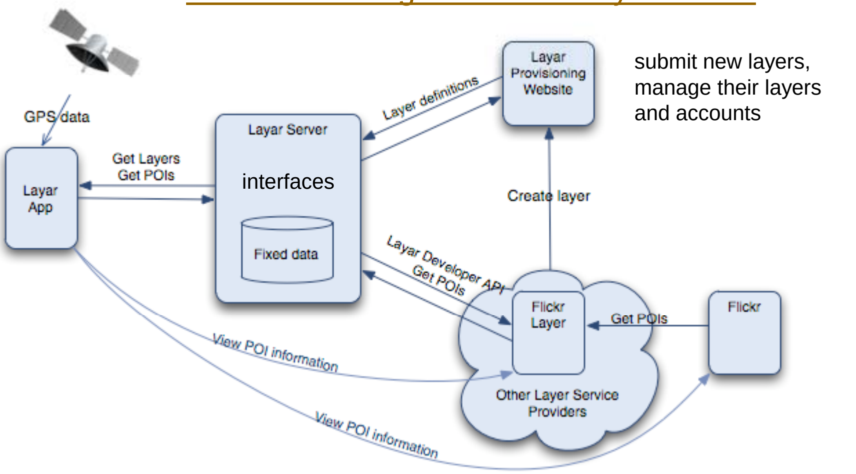
Google it! ☺ ☹

Layar First Mobile Augmented Reality Browser

The world is the interface!

Layar

First Mobile Augmented Reality Browser



submit new layers, manage their layers and accounts

http://layar.com/ : founded in June 2009 in Amsterdam

Layar – cont.

- Superposition of multiple layers: Reality, Design Layout, Point of Interest (POI)
- Layar browses ressources on the server to display the POI.
- Layar uses the HTTP GET request (Requests a representation of the specified resource. A simple action of retrieval.)
- Layar allows the creation of layers by developers. No license costs involved creating a layer.

Layar – cont.

- 500 layers developed by from individuals to small enterprises to large companies; 2000 layers in development.
- Over 1 million active end-users.
- Applications for marketing.
- Support all Android devices and the iPhone 3GS. A Symbian version is in development. Need internet connection, camera, GPS and compass.

The world is my interface

- Mobile devices can be used to interact with the "Internet of Things”.

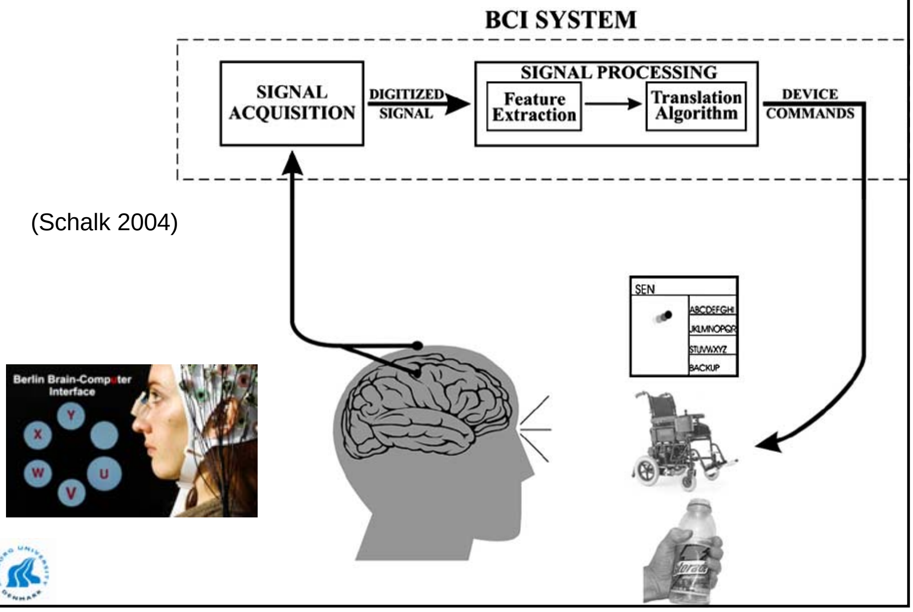
Information
Bluetooth
LUCID
Photos
Maps & GPS
Multimedia
Tickets
Email & IM
Mobile Barcodes
Social Networking

Sensors in smartphones to revolutionize the UI:
- microphones
- cameras
- motion sensors
- proximity sensors, and
- location sensors.

Many application examples

http://www.lucidproject.org/

Brain-computer interface (BCI)



BCI SYSTEM

SIGNAL ACQUISITION

DIGITIZED SIGNAL

SIGNAL PROCESSING

Feature Extraction

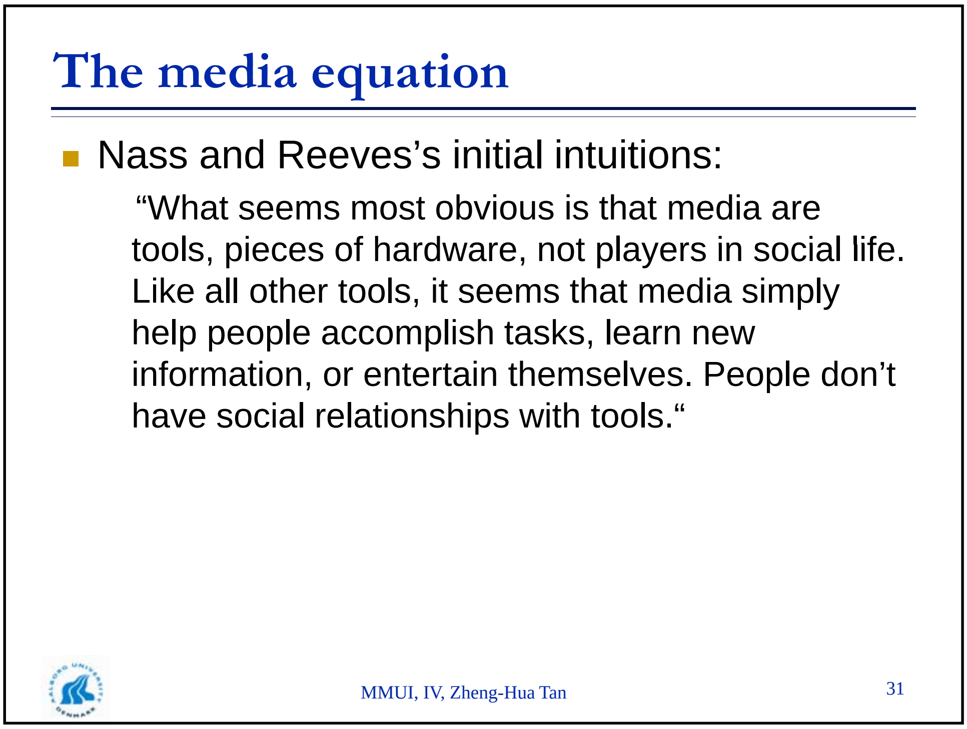
Translation Algorithm

DEVICE COMMANDS

(Schalk 2004)

Outline

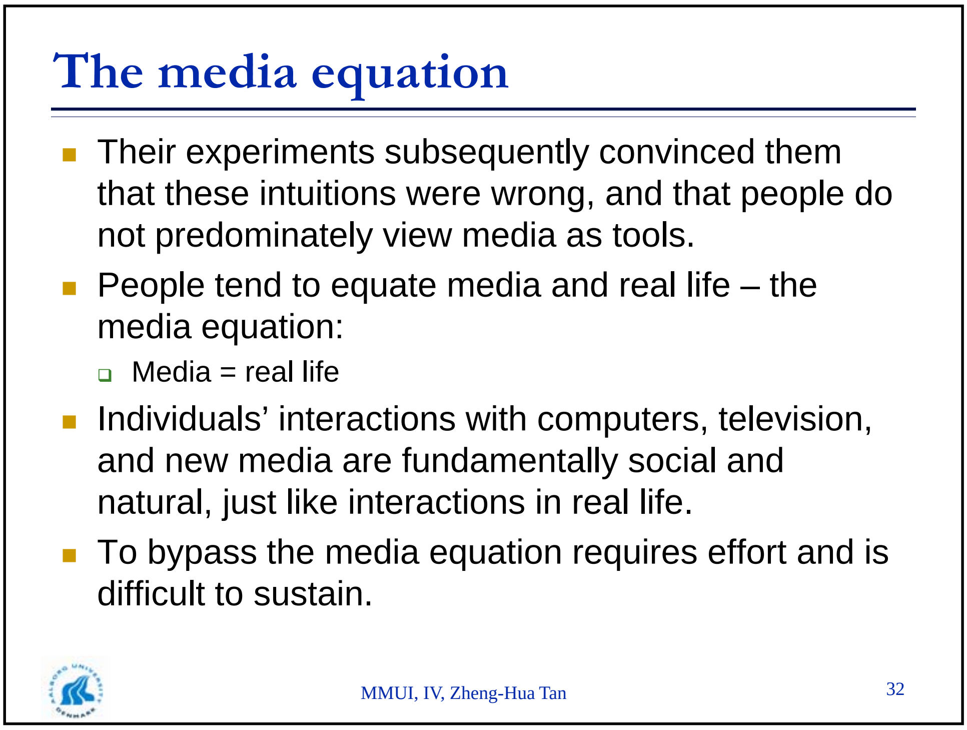
- Multimodal interface
- Various modalities and their combination
- Perceptual user interface



The media equation

- Nass and Reeves's initial intuitions:

  "What seems most obvious is that media are tools, pieces of hardware, not players in social life. Like all other tools, it seems that media simply help people accomplish tasks, learn new information, or entertain themselves. People don't have social relationships with tools."



The media equation

- Their experiments subsequently convinced them that these intuitions were wrong, and that people do not predominately view media as tools.
- People tend to equate media and real life – the media equation:
  - Media = real life
- Individuals' interactions with computers, television, and new media are fundamentally social and natural, just like interactions in real life.
- To bypass the media equation requires effort and is difficult to sustain.

Perceptual user interface

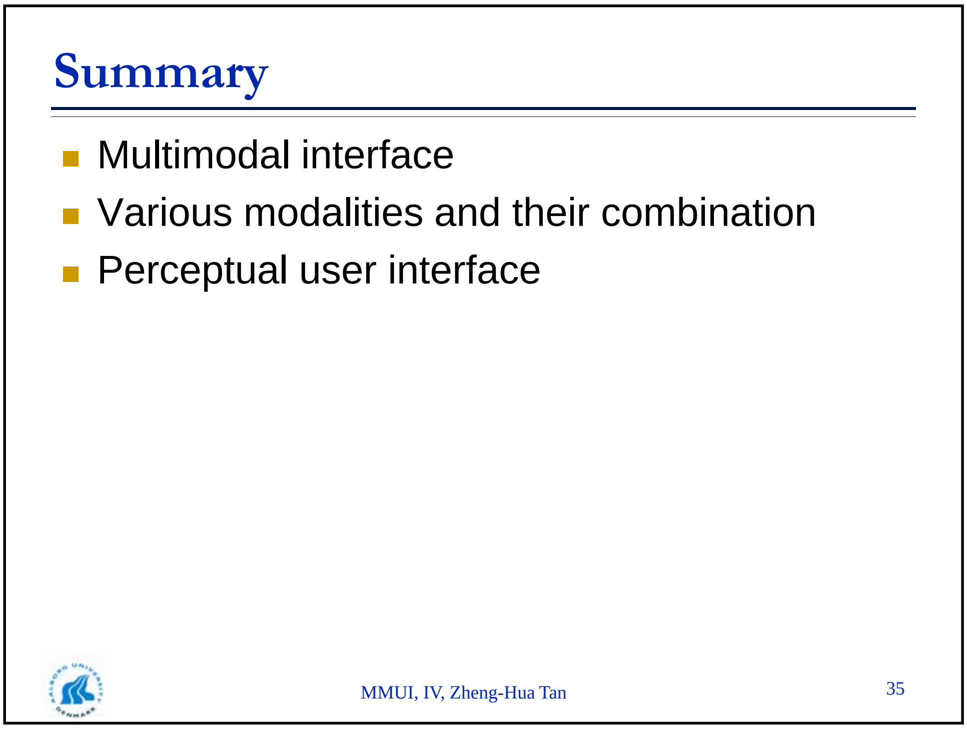
Highly interactive, multimodal interfaces modeled after natural human-to-human interaction, with the goal of enabling people to interact with technology in a similar fashion to how they interact with each other and with the physical world.

(Matthew Turk)

Perceptual user interface

- Vision based interfaces
  - Gesture recognition
  - Full body tracking
  - Head tracking
  - Eye-gaze tracking
- Audio based interfaces
- “Interaction between man and machine should be based on the very same concepts as that between humans, i.e., it should be intuitive, multi-modal and based on emotion.”

  - Reeves and Nass (1996), *The Media Equation.*



Summary

- Multimodal interface
- Various modalities and their combination
- Perceptual user interface