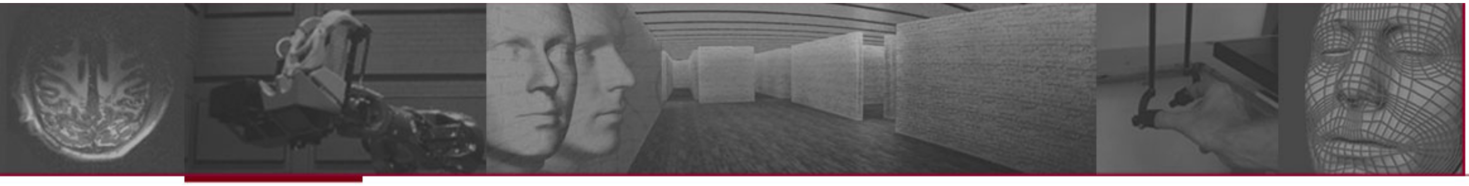
Towards Artificial Systems:
What can we learn from Human Perception?

Heinrich H. Bülthoff

Biological Cybernetics Research at the
Max Planck Institute and Korea University

MAX-PLANCK-GESELLSCHAFT



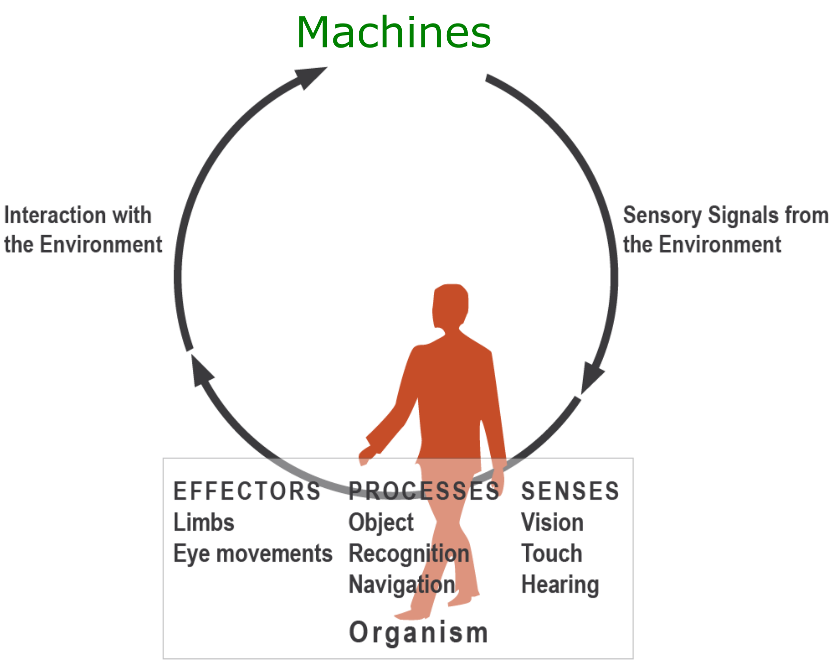
Dept. of Brain and Cognitive Engineering
Korea University

KOREA UNIVERSITY
1905

National Research Foundation of Korea
R31-2008-000-10008-0

My goal for today

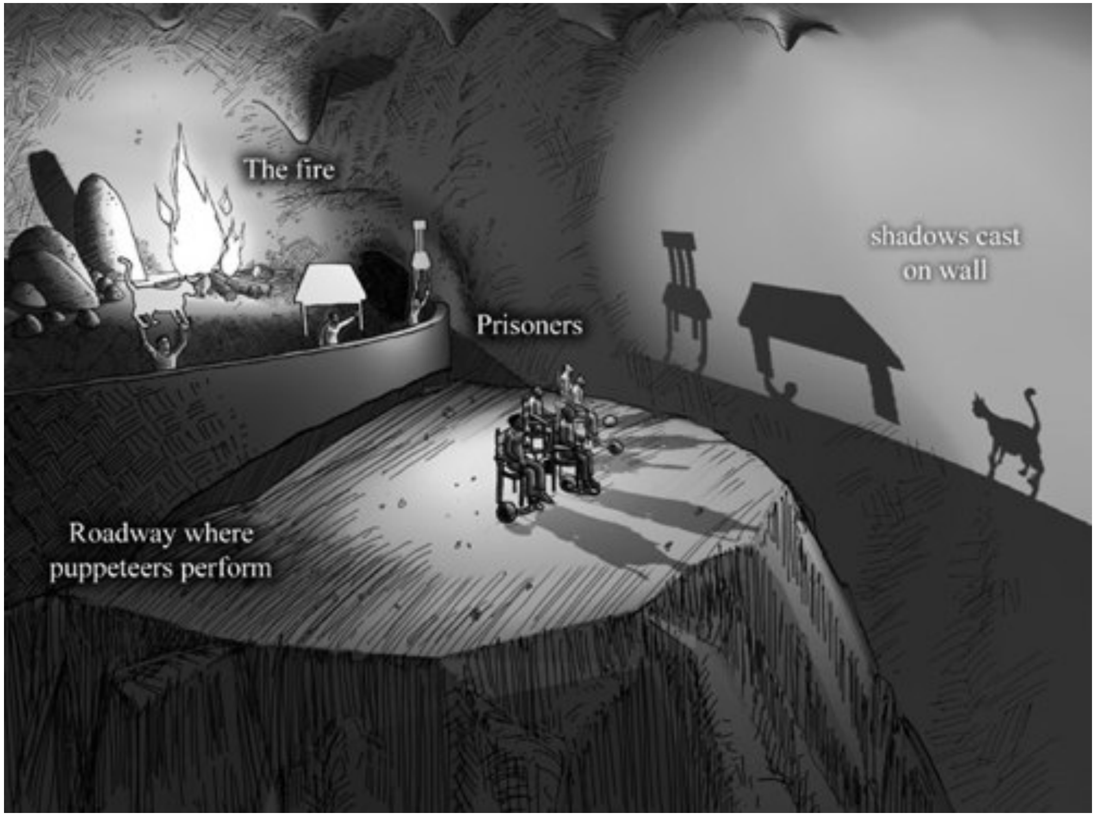
- present interesting findings about human perception which might be useful for building better artificial systems
- show new interfaces for effective and natural interactions
- which integrate humans into the loop
- in order to build better Human-Machine-Interfaces



Talk Outline

1. diagnostic features
2. dynamic information
3. active perception
4. human-in-the-loop

Research Philosophy



 edges

 corners

 curves

Understand human behavior with “realistic” stimuli and scenarios and not the “*shadows of the shadows of reality*”

Talk Outline

1. **diagnostic features**
2. dynamic information
3. active perception
4. human-in-the-loop

What can we learn from human face recognition ?

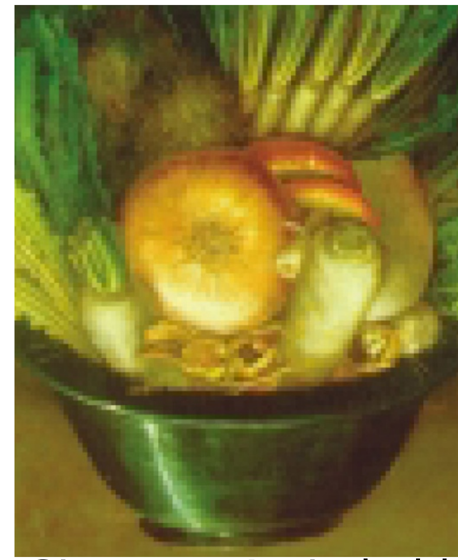
What are important features for face recognition ?

Pawan Sinha (MIT)

suggests 19 **insights from human face recognition**
every computer vision researcher should know about
Proceedings of the IEEE | Vol. 94, No. 11, November 2006

Eyebrows are important features

Thatcher Illusion (Thompson, 1980)



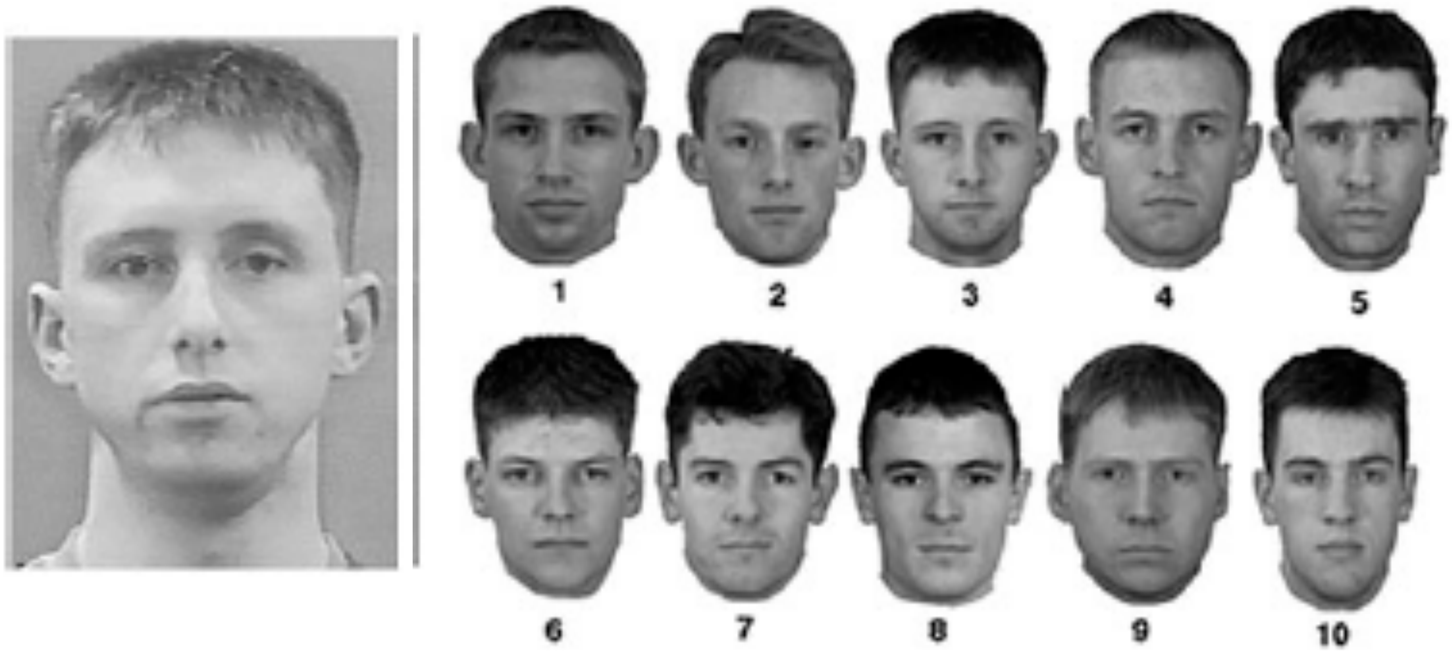
Giuseppe Arcimboldo

When faces are viewed upside-down, our ability to process the configuration is disrupted

How good we are at recognizing **familiar** faces?

1. Michael Jordan
2. Woody Allen
3. Goldie Hawn
4. Bill Clinton
5. Tom Hanks
6. Saddam Hussein
7. Elvis Presley
8. Jay Leno
9. Dustin Hoffman
10. Prince Charles
11. Cher
12. Richard Nixon

How bad are we with unfamiliar faces?

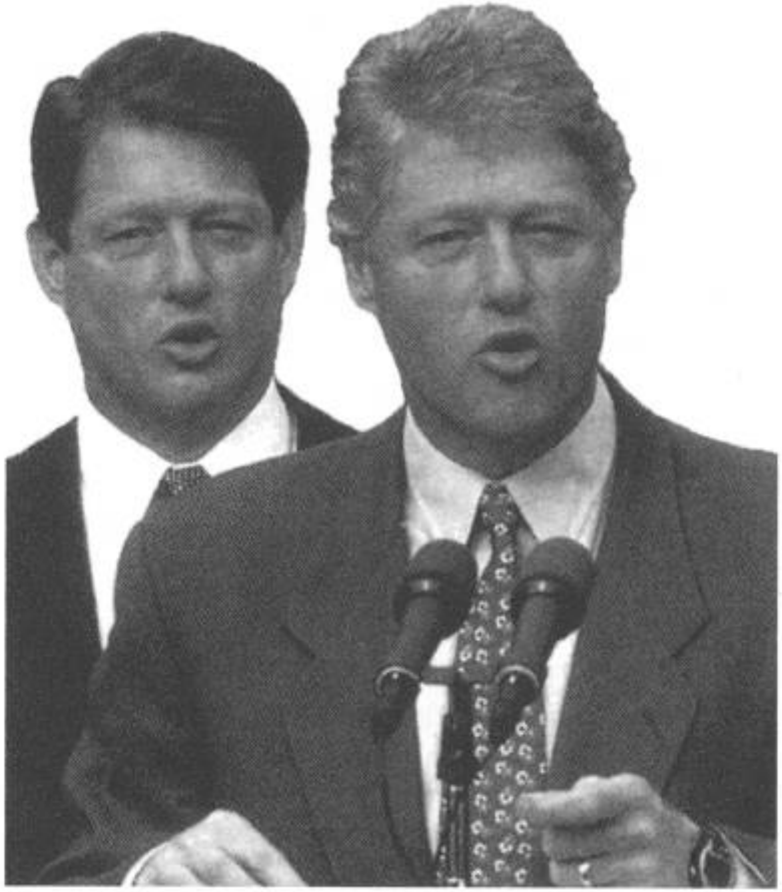


Bruce, V. et al. (1999). J. Exp. Psychol. Appl. 5, 339–360

Caveat...

- Clinton and Gore
- True ????
- Inner parts of both faces are Clinton

- It is not always good to model human vision

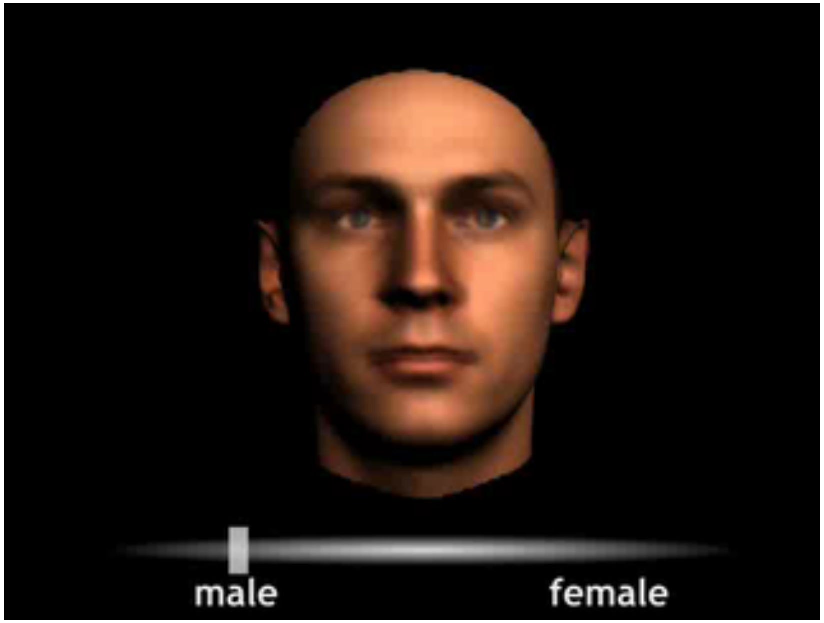


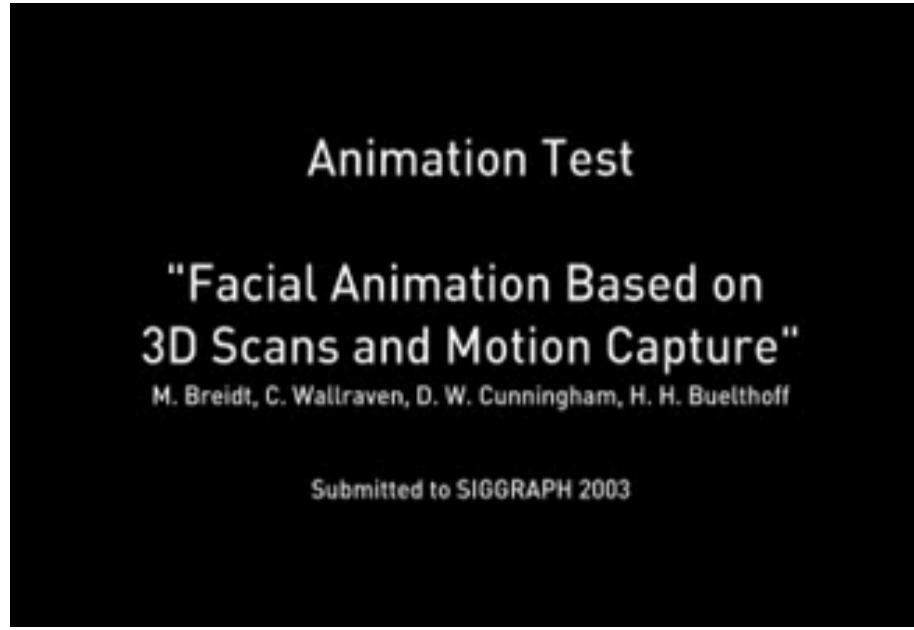
Sinha & Poggio 1996

British Version by Ian Thornton

DAVID CAMERON & NICK CLEGG *

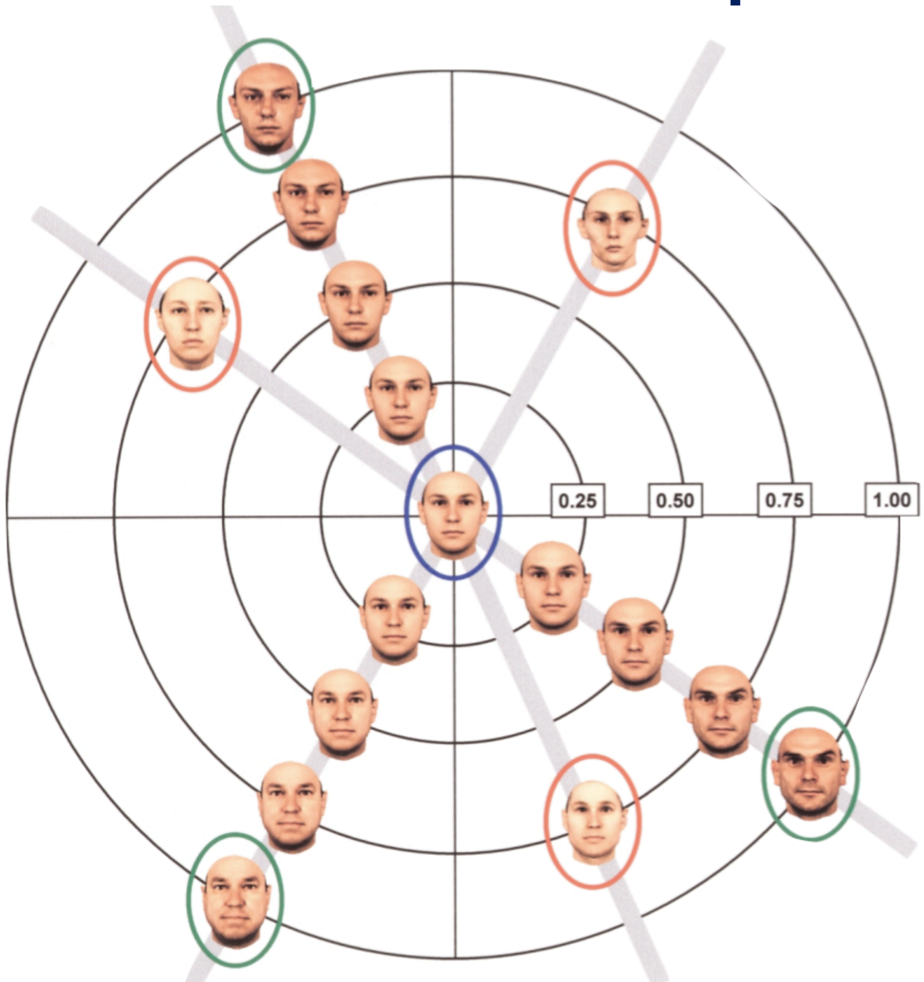
Our approach to study recognition





- high resolution 3D head models all in correspondence
- interpolation in high dimensional face space
- precise control of physical and perceptual dimensions

Organization of Facial Representations in a High-dimensional Face Space



green: original face
blue: average face
red: anti-face

Leopold, O'Toole, Vetter & Blanz (2001)
Armann & I. Bülthoff (2009)

Subjective Perceptual Modeling

More feminine

More masculine

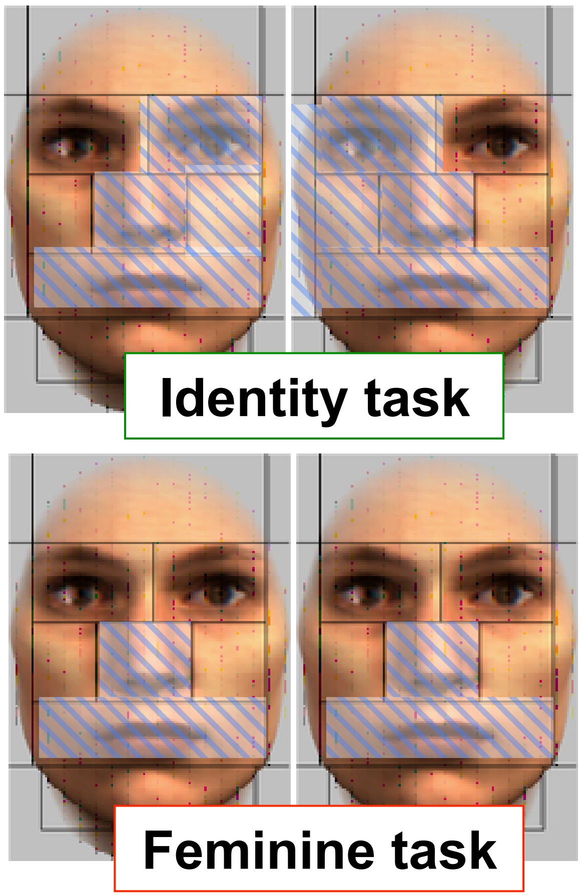
Friendlier

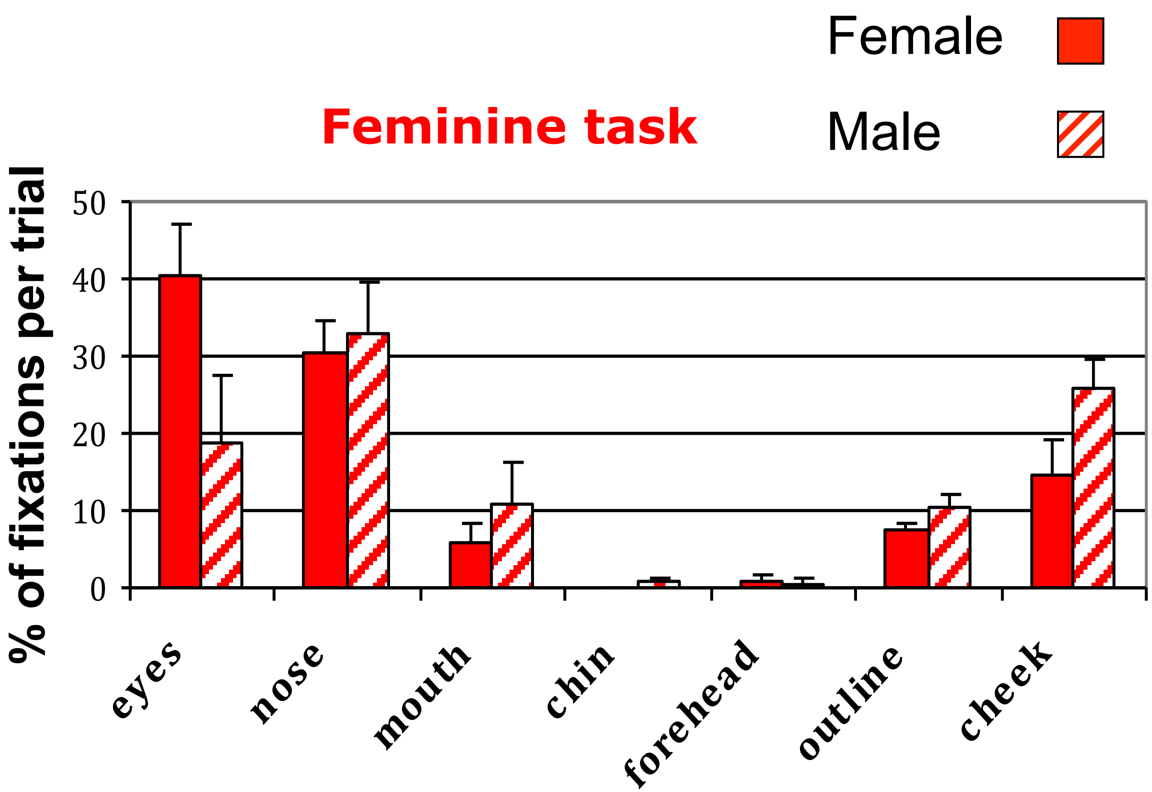
More attractive

Task dependent eye movements

- Eye movement recordings for different tasks:
  - identity judgment
  - feminine judgment
  - eye movements recorded for male and female participants

Diagnostic features vary according to task and observer

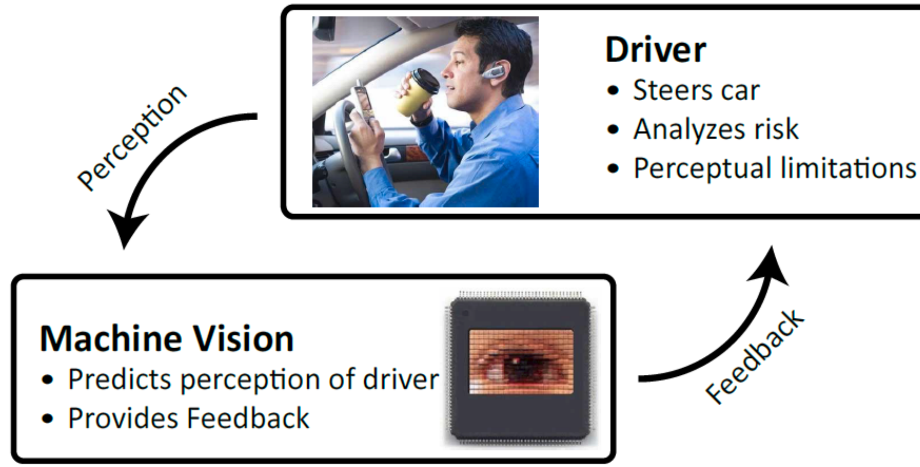




Armann & I. Bülthoff, 2009

Driver assistance systems (David Engel and Cris Curio)

Using computer vision to predict human perception performance

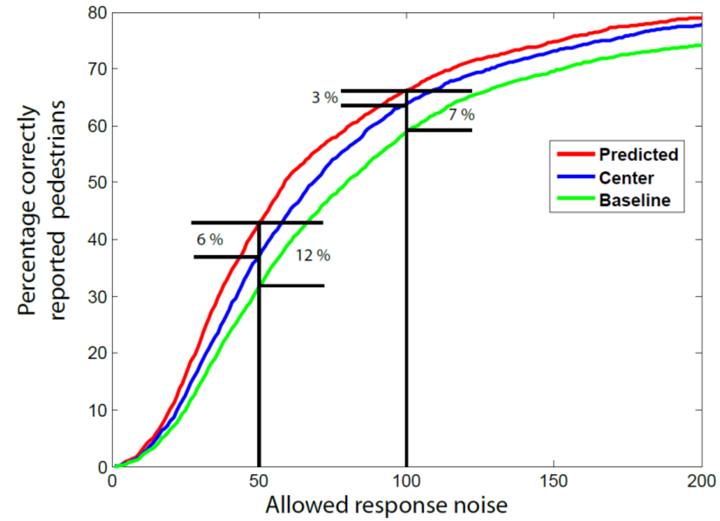


- “Detectability” measures the probability that a pedestrian will be noticed at a glance
- Pedestrians with low detectability are in more danger since they do not factor in the risk estimation of the driver
- If we can predict the detectability we can optimize the perception of the driver

- The driver can usually do risk estimation very well even in hazard situation with short reaction times
- But if he didn‘t perceive a pedestrian he will miscalculate the risk
- Idea: Provide feedback that will help the driver notice all pedestrians

Detectability:

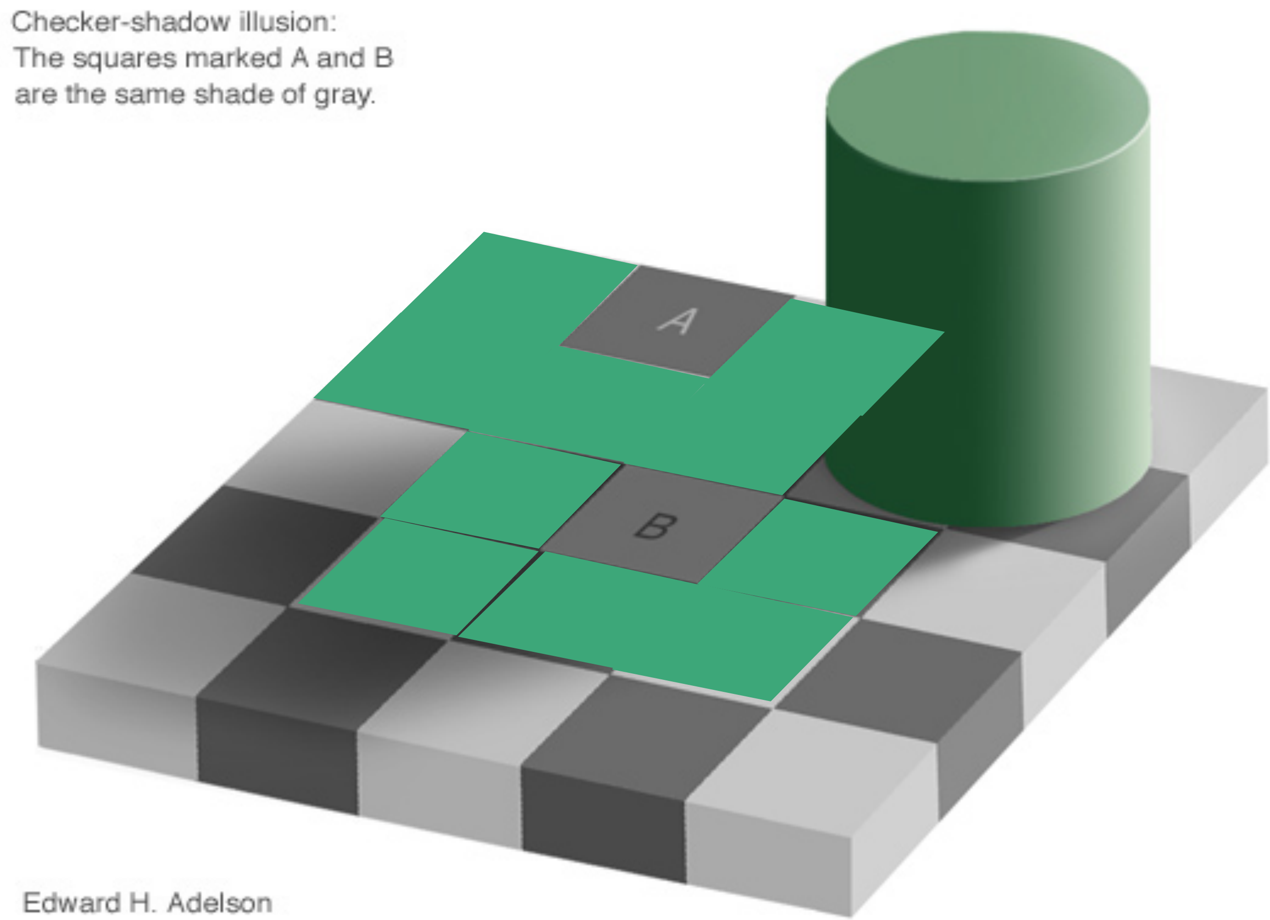
Optimizing the perception of drivers



- Trained a regressor to predict the detectability of all pedestrians in an image given a fixation cross (for short image presentation times)
- Predict the optimal position of the fixation cross

- Performance increases of 30% over random fixation crosses (12% is absolute value)
- 30% pedestrians that would not have been noticed earlier
- Currently extending the method to dynamic stimuli

What are the imaged-based features ?
Image descriptors not pixels are important



The brain does not act like a photometer, it is not interested in the light intensity but more what it can infer from it.

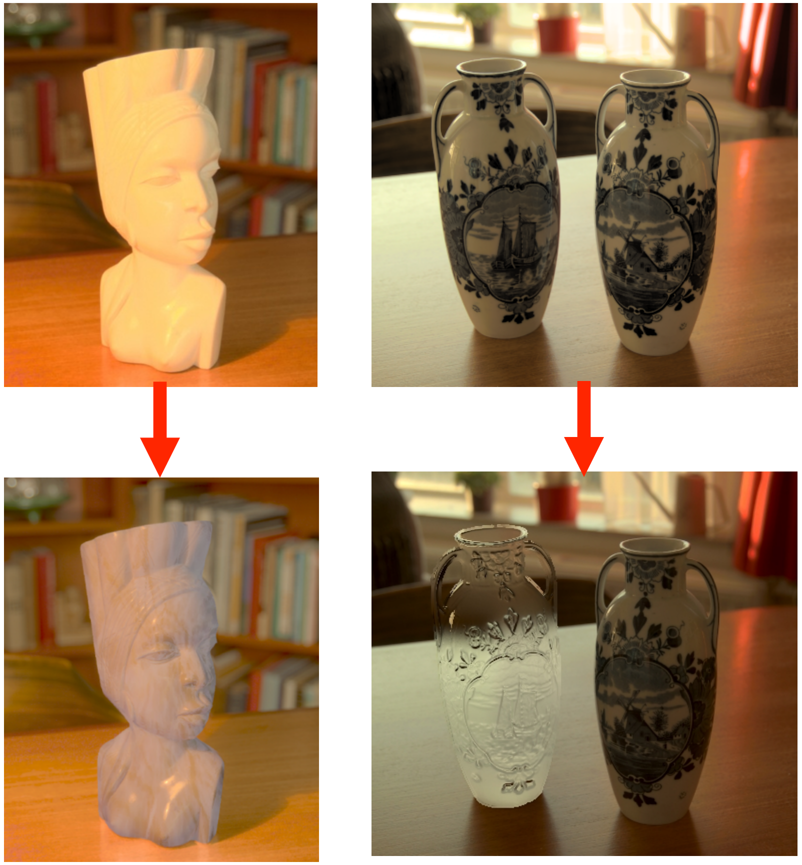
Image-based material editing

Kahn, Reinhard, Fleming, Bülthoff [*SIGGRAPH*, 2006]

- Question:
  Can we exploit perceptual tricks to change materials in a photograph (without a 3D-model)

- Method:
  Crude 3D shape reconstruction using image statistics

  - Simple background-inpainting for transparency



re-textured   transparency

Image-based material editing

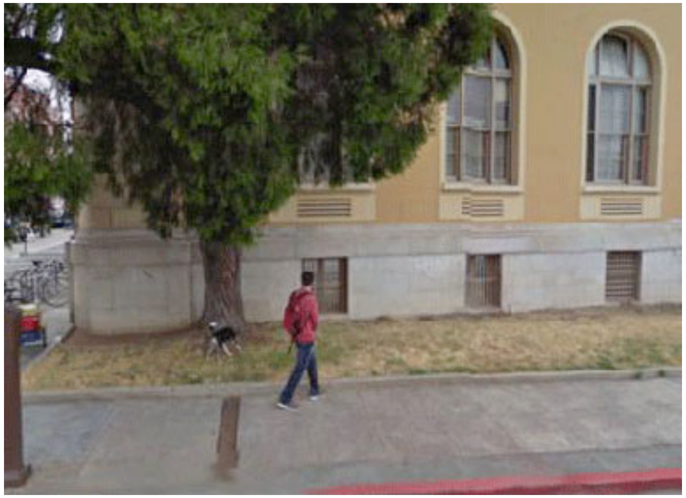
Kahn, Reinhard, Fleming, Bülthoff, 2006

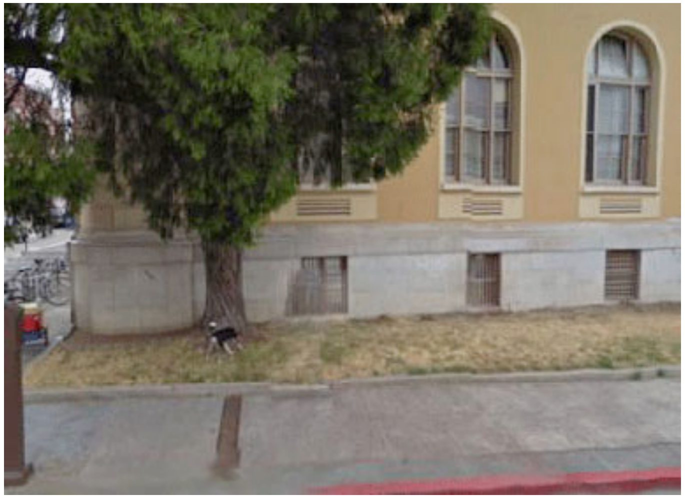
Image-based material editing

Kahn, Reinhard, Fleming, Bülthoff, SIGGRAPH 2006

Google Streetview Privacy

A. Flores & S. Belongie, University of California, San Diego





Short summary: Diagnostic Features

- The brain does **not** seem to apply an inverse physics approach to perception
- The brain does not work like a photometer
- The brain does not work like an architect (no 3D models)
- Rather, the brain uses:
  - view-based features for object recognition
  - heuristics to estimate material properties
  - subjective properties (friendliness, attractiveness, ---) for building multi-dimensional face representations

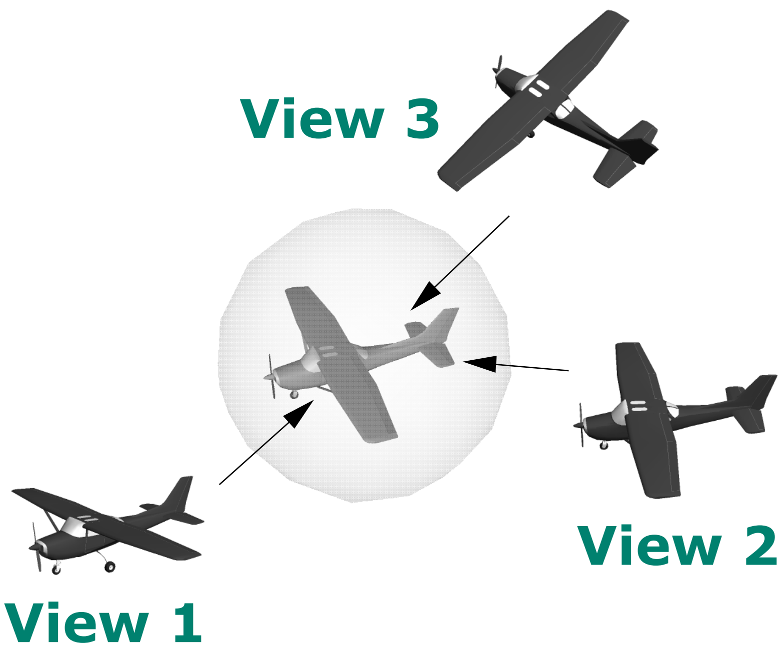
What can artificial systems learn from perception?

Develop models for representing diagnostic **view-based** features and **perceptual dimensions**

Talk Outline

1. diagnostic features
2. dynamic information
3. active perception
4. human-in-the-loop

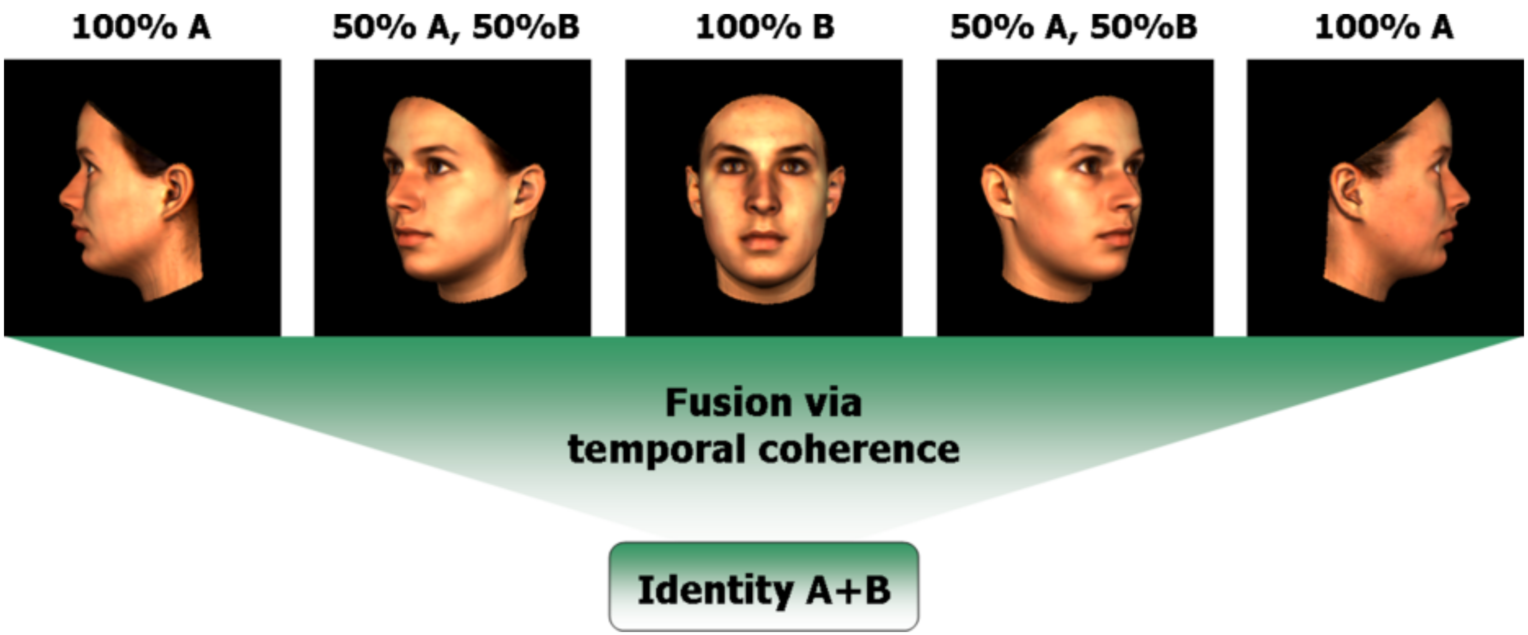
The binding problem of View-based recognition



How does the brain know that different views of an object belong together?

The role of time in object learning

- The **temporal association hypothesis** predicts that morph sequences of a rotating head which changes identity from A to B should bind all images to one single person.



Wallis & Bülthoff, PNAS (2001)

MPI Face Database

- Face Database with 200 3D head models in correspondence
  - faces.kyb.tuebingen.mpg.de (open access)
    - Troje & Bülthoff(1996) Vision Research **36**, 1761-1771
    - Blanz & Vetter (1999) SIGGRAPH'99 Conference Proceedings, 187-194
- Face Video Database with facial action units from 6 viewpoints
  - vdb.kyb.tuebingen.mpg.de (open access)

- Facial Expression Database
- edb.kyb.tuebingen.mpg.de (open access)

Semantic motion coding

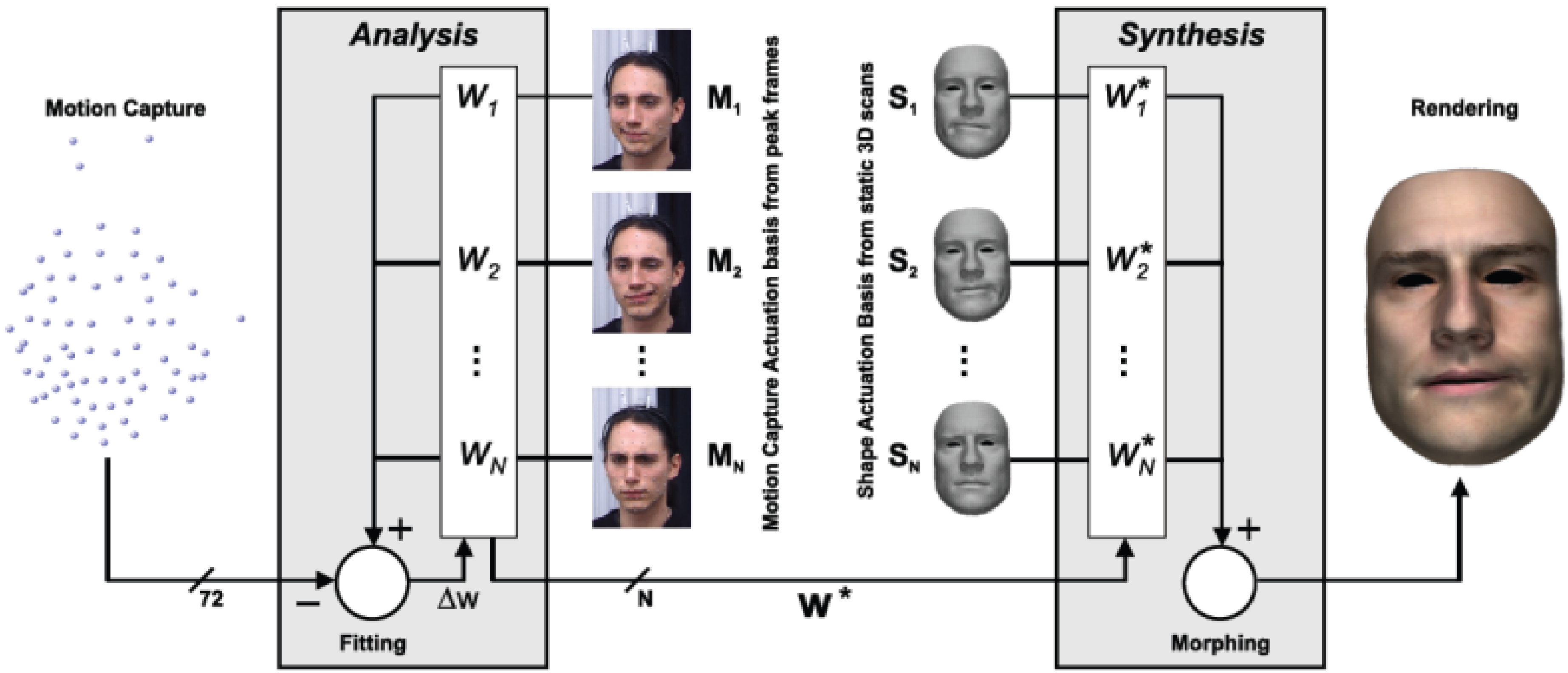
Facial Action Unit Coding System: FACS (Ekman & Friesen)

Examples of Action Units (MPI Video-Database, 6 views, synchronized video)

Coding system for facial expressions based on Action Units (AU)

- Complex facial expressions can be disassembled into AUs
- Most AUs correspond to anatomical facial muscle activations
- AUs have a semantical meaning

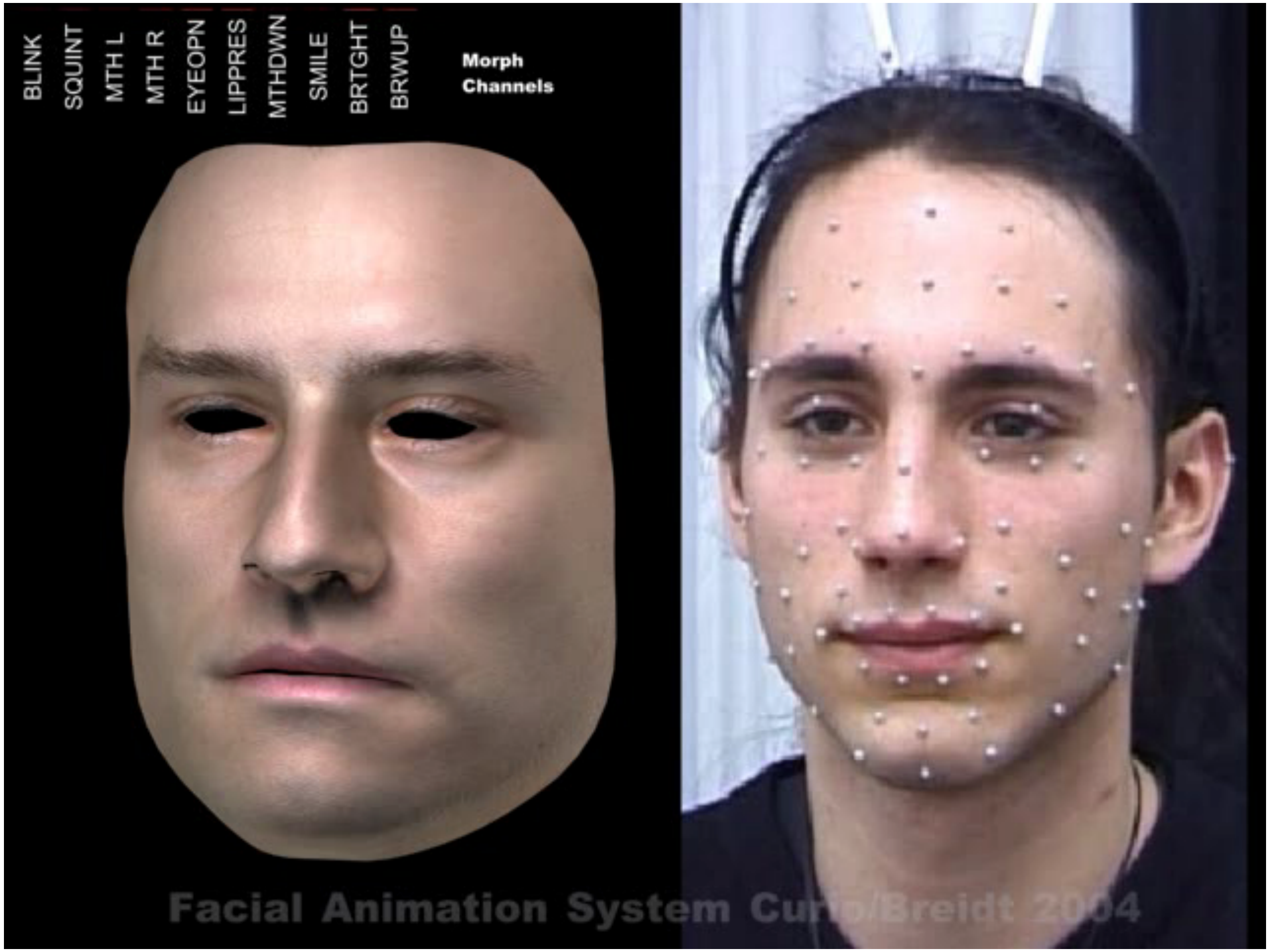
Semantic facial animation pipeline



Curio et al 2006 APGV

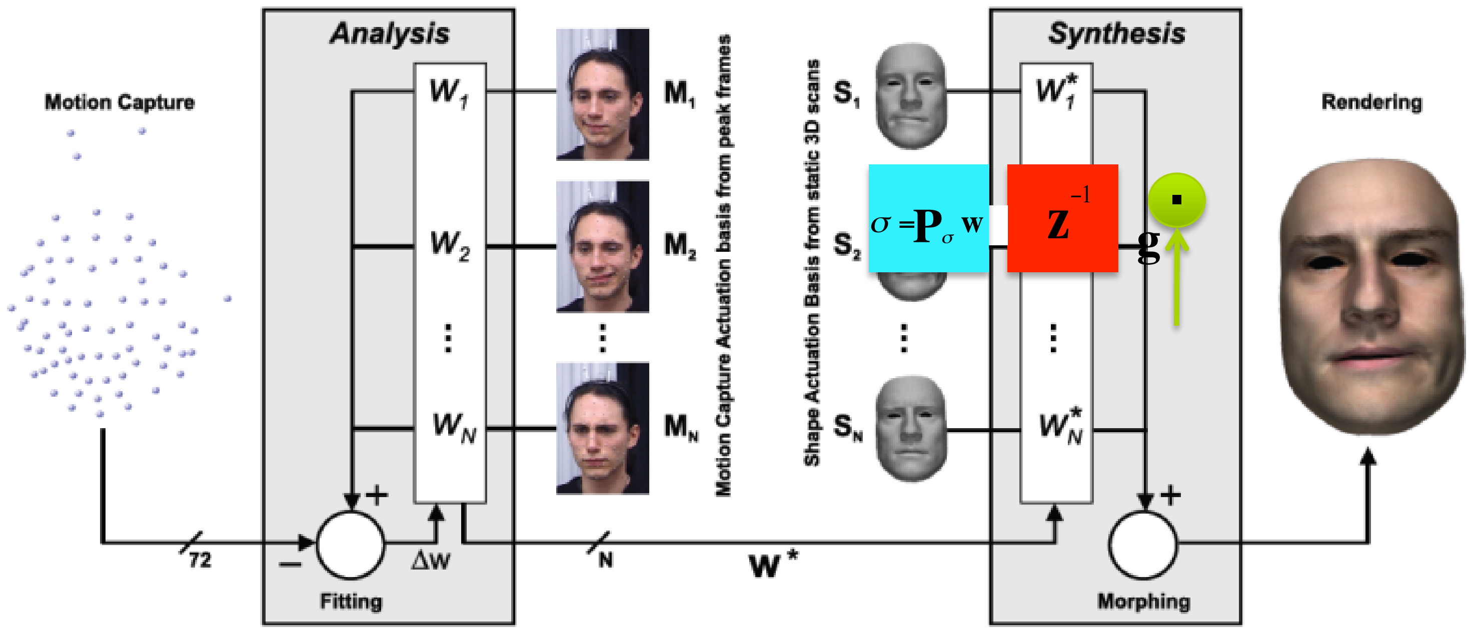
Study based on animation

Morph weights **w**



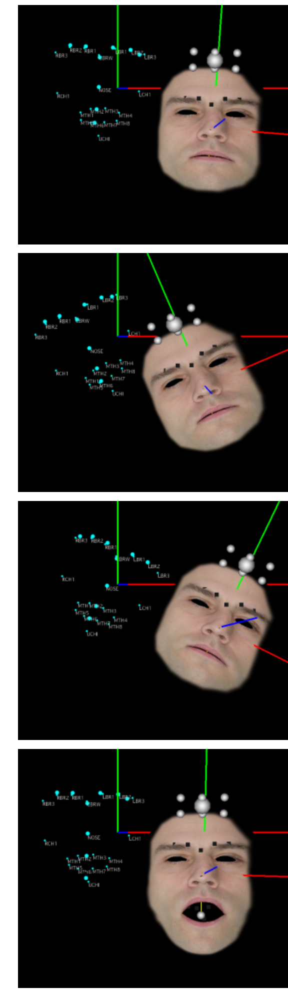
Semantic Facial Motion Analysis & Synthesis (Curio et al 2006)
based on Facial Action Unit Coding System (Ekman & Friesen)

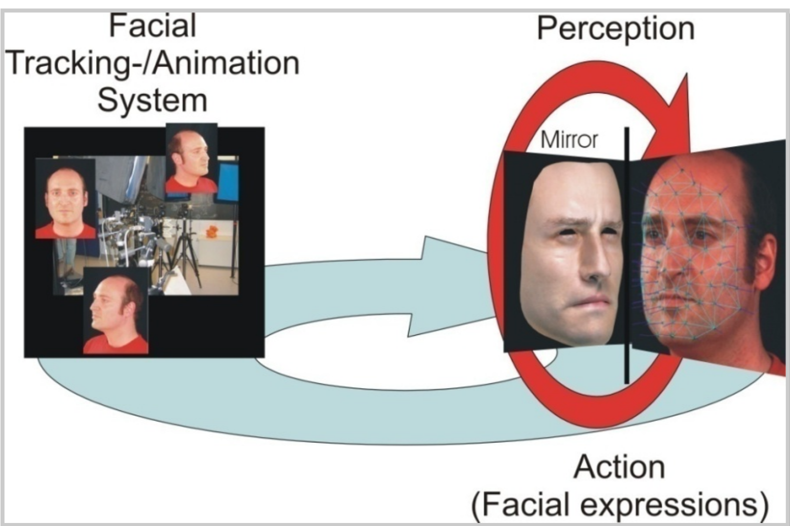
Virtual Mirror (Cris Curio)

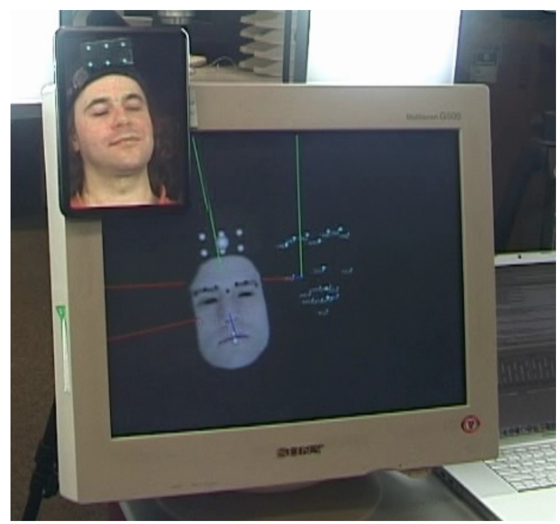


Max Planck 'Virtual Mirror' for closed-loop facial expression studies

- Psychiatry
  - self-perception
  - Autism, Schizophrenia
- Rehabilitation/Healthcare
  - Clinical interventions, bio-feedback
  - About 40-50% of all stroke patients suffer from partial facial movement palsies
- Communication training







Short summary: Spatio-temporal Representations

- The temporal properties of objects play an important role during learning and recognition
- Object representations are spatio-temporal
- Can be exploited in many applications:
  - medical diagnosis and therapy
  - video conferencing
  - game design
  - sports coaching

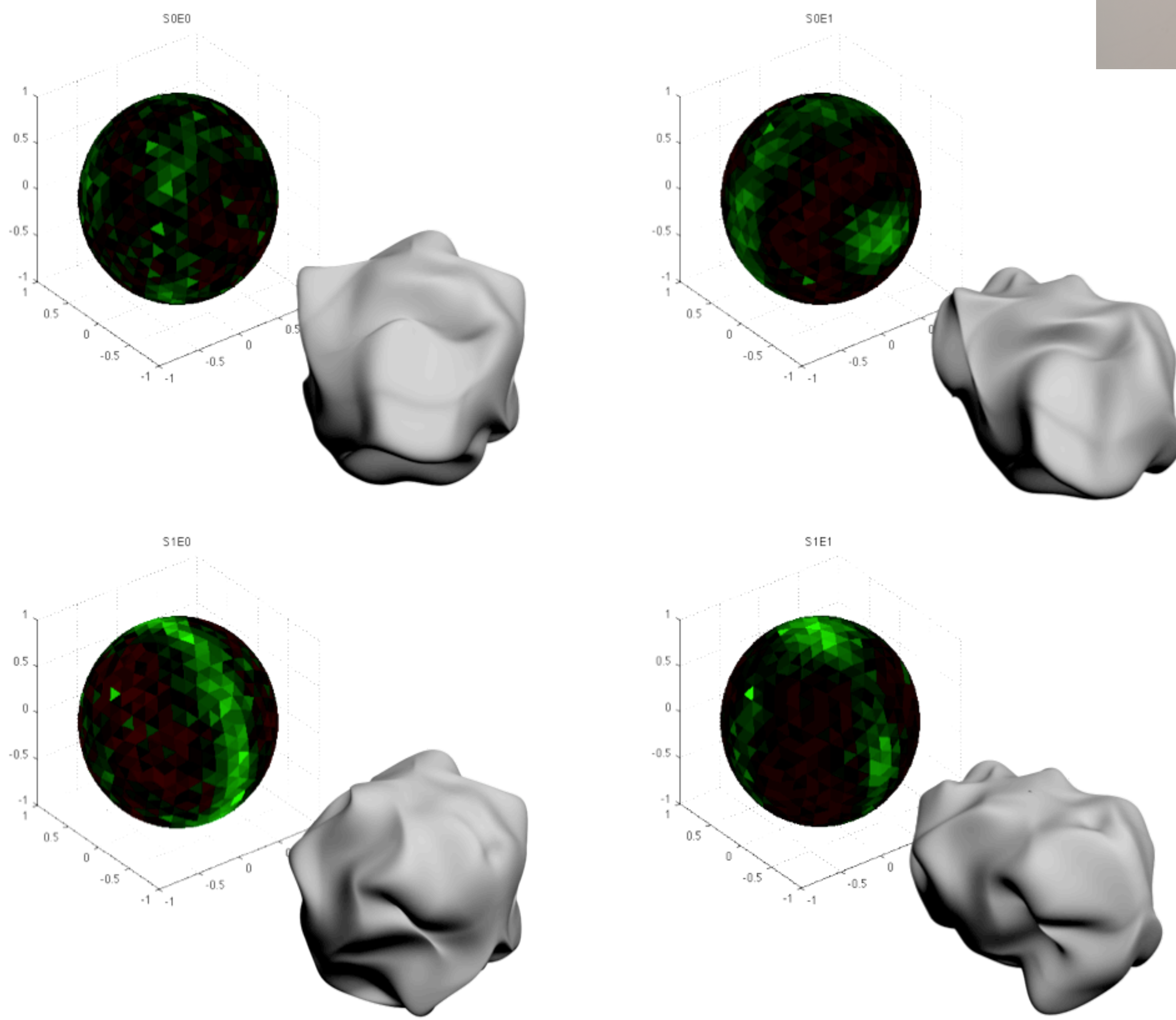
What can artificial systems learn from perception?

Make use of the time dimension of the visual input
(motion statistics, feature tracking, view integration, incremental learning…)

Talk Outline

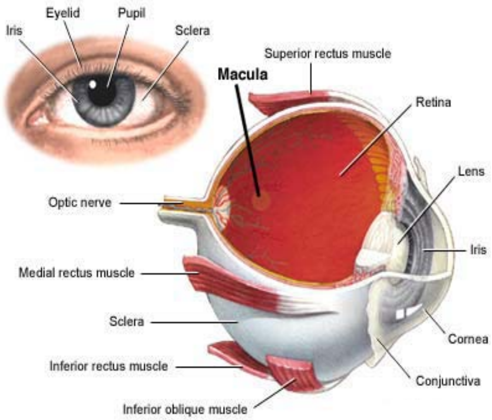
1. diagnostic features
2. dynamic information
3. active perception
4. human-in-the-loop

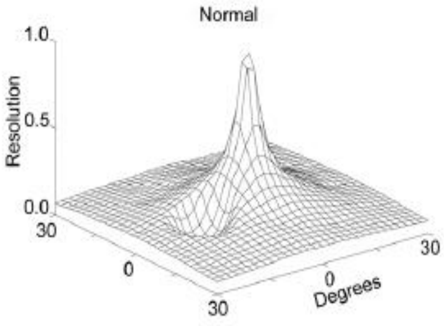
Active Object Exploration



Chuang, Bülthoff & Wallraven, 2011

Human perception is 'forced' to be active

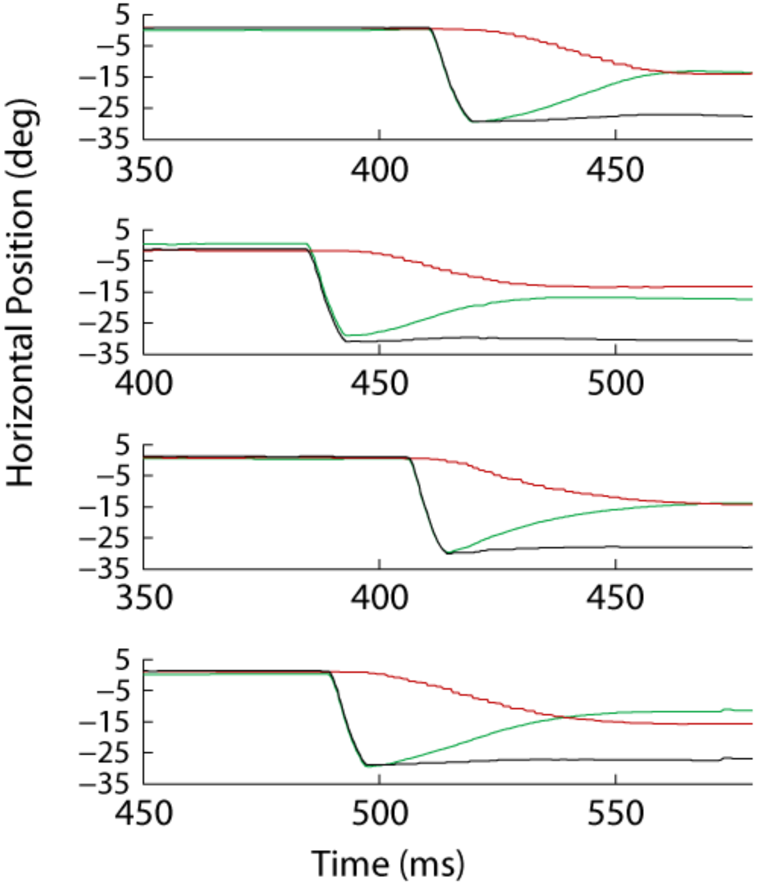


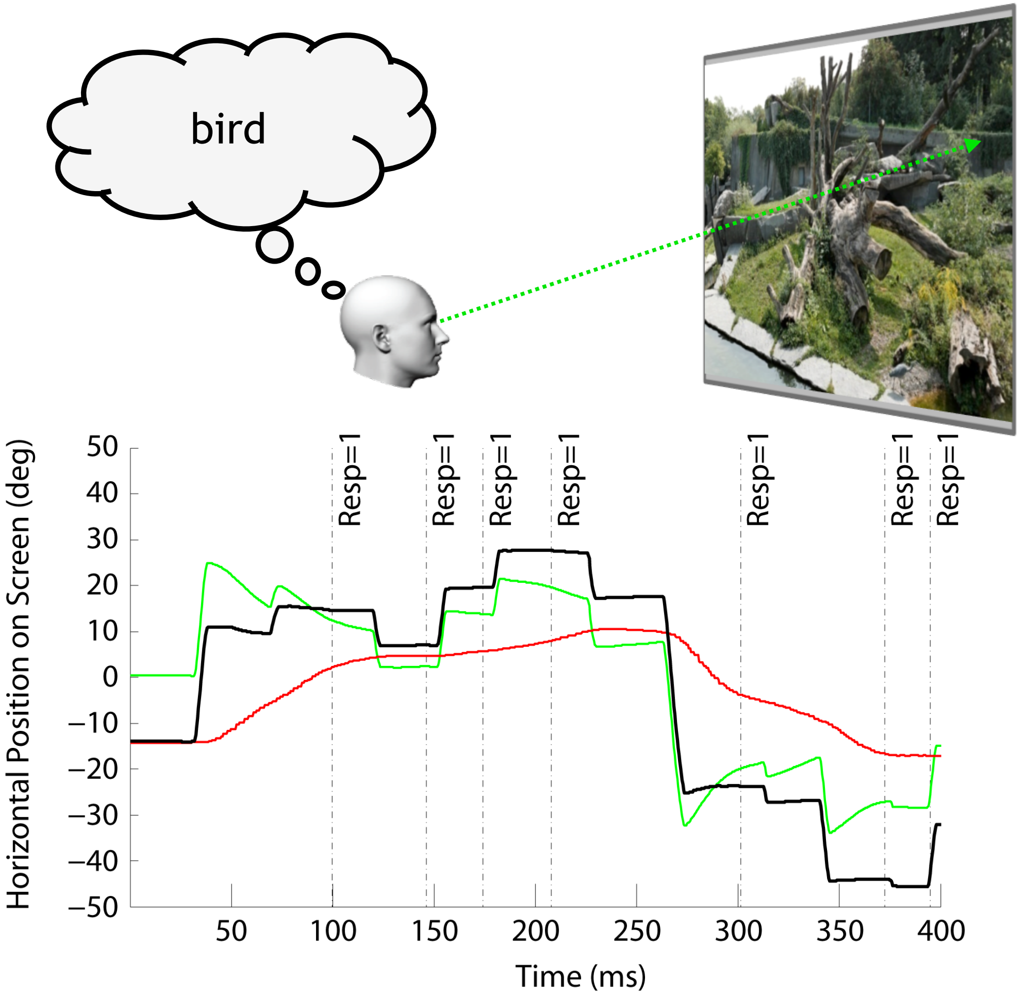


Gaze-assisted interfaces

GigaPixel project: Deussen et al, 2010

From single to multiple gaze shifts



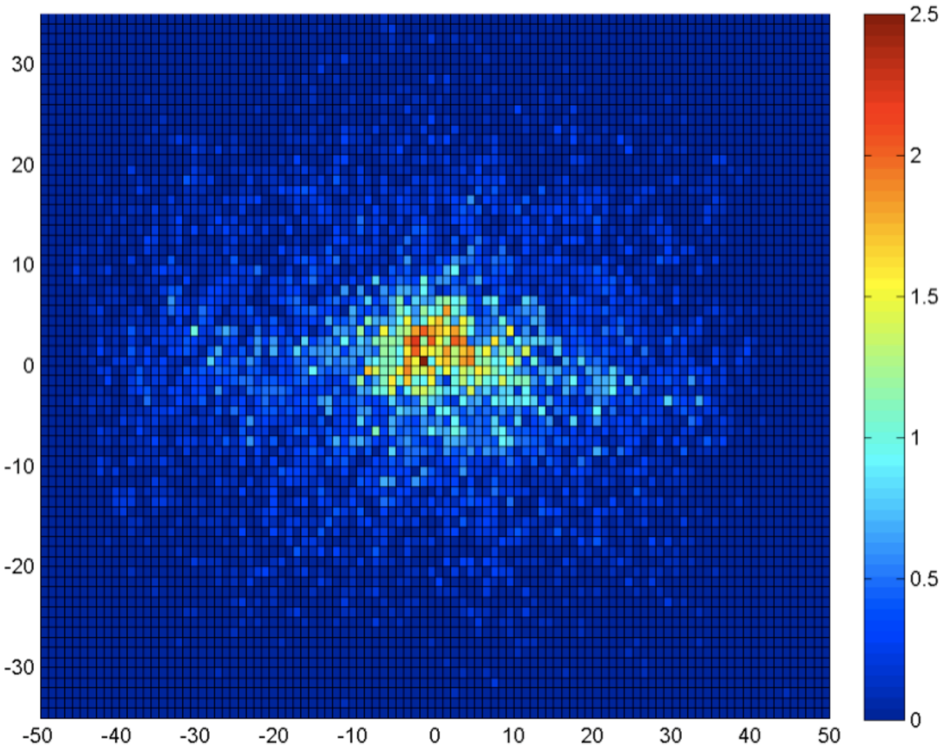


Herholz, S., L. Chuang, T. Tanner, H. H. Bülthoff and R.W. Fleming, 2008;
Chuang, L. Bieg, H-J., H.H. Bülthoff, and R.W. Fleming, 2010

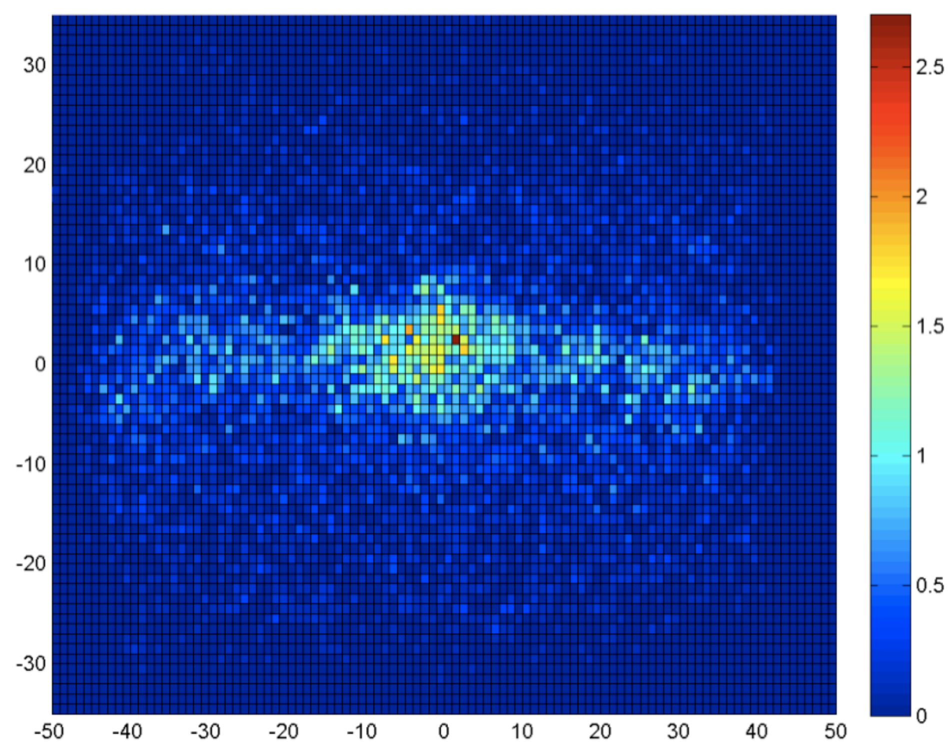
Image-based properties are not sufficient to predict gaze behavior

Saliency regions drawn using Itti & Koch, 2001

Motor biases and Utility constrain gaze behavior



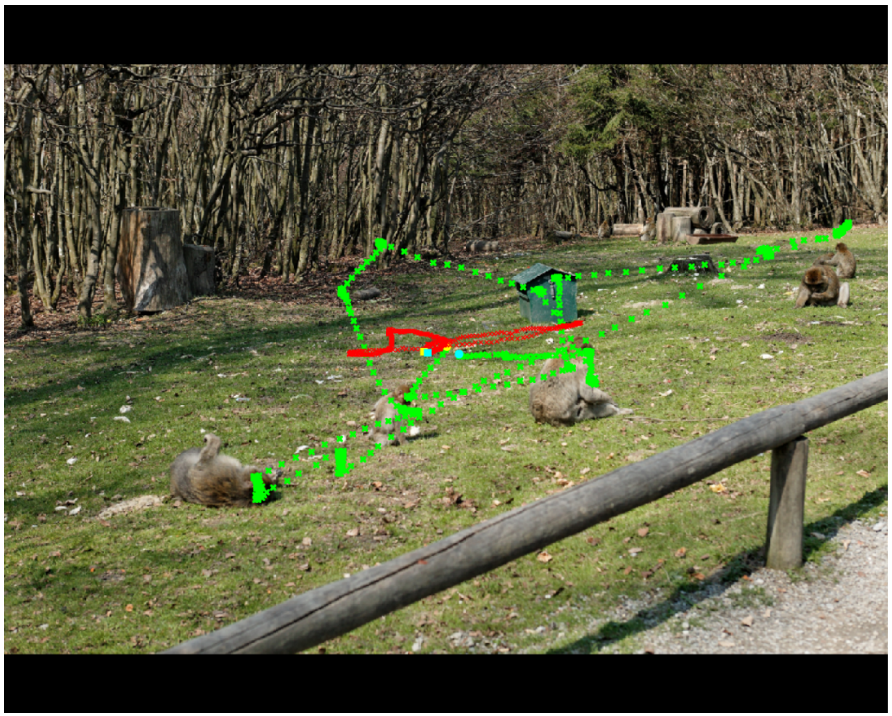
Attractiveness Rating



Count Animals

Motor biases and Task demands constrain gaze behavior

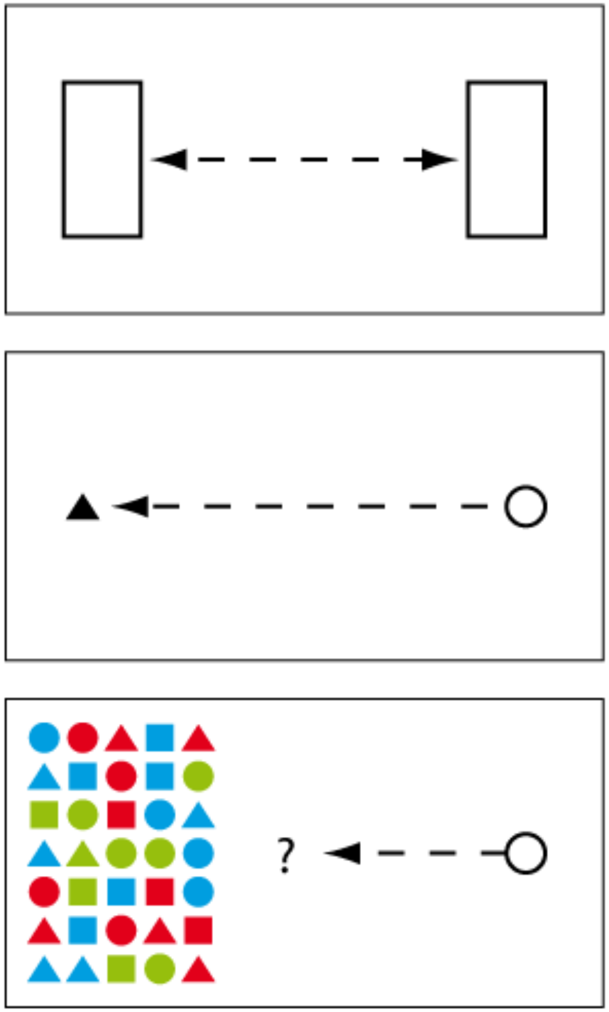
Head movements
Gaze movements



Rate Attractiveness

Count Animals

Gaze-assisted interfaces



Short summary: Active Perception

- Perception can be considered as an action modality
  - we are not simply receiving information but actively planning the selection of information
  - image saliency is not sufficient in explaining how we choose targets
- Human Gaze is a perfect example of a control system for information acquisition via eye and head movements

What can artificial systems learn from perception?

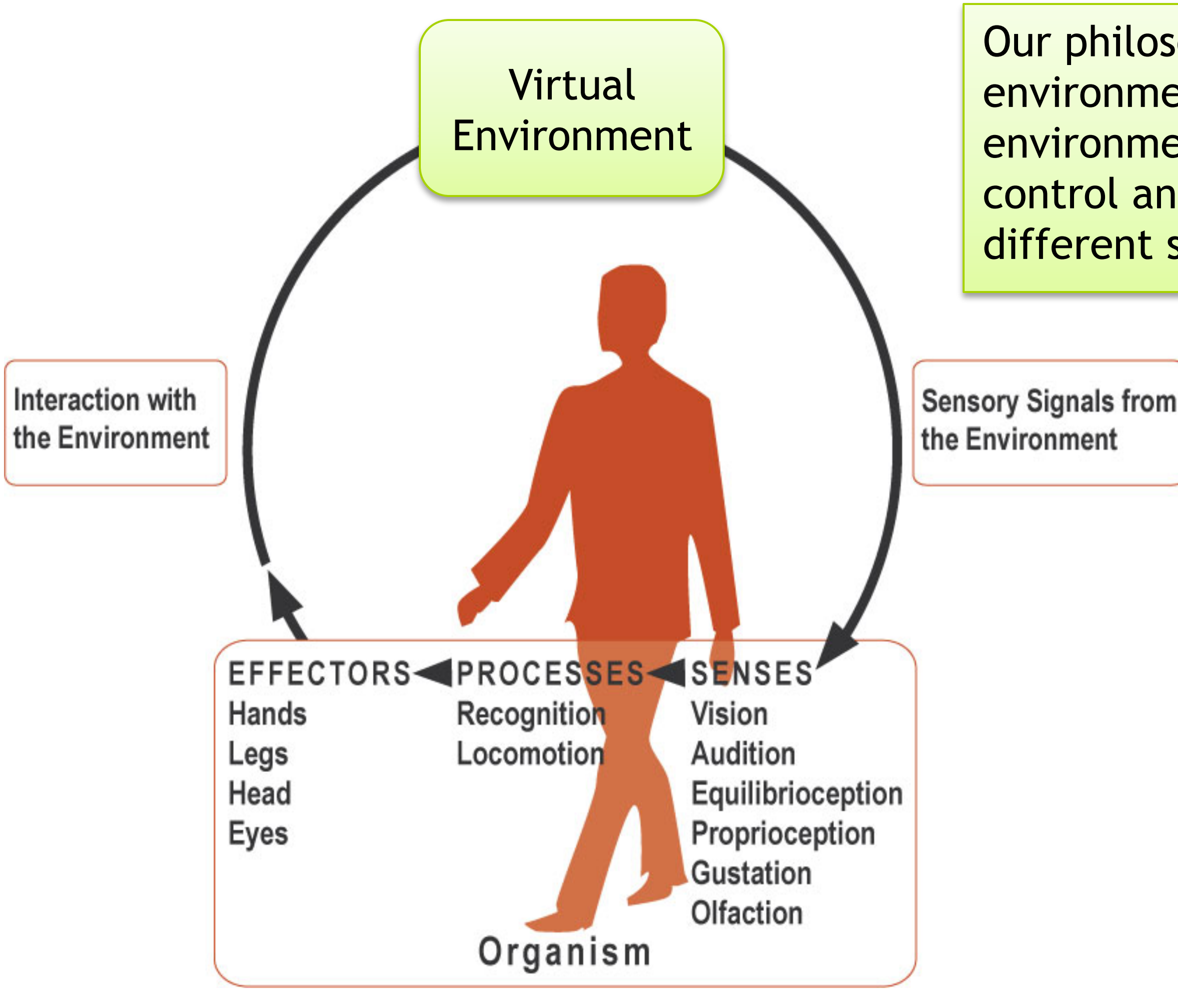
by integration the control system for eye, head and hand coordination we can build better devices(e.g., robots) and human machine interfaces.

Talk Outline

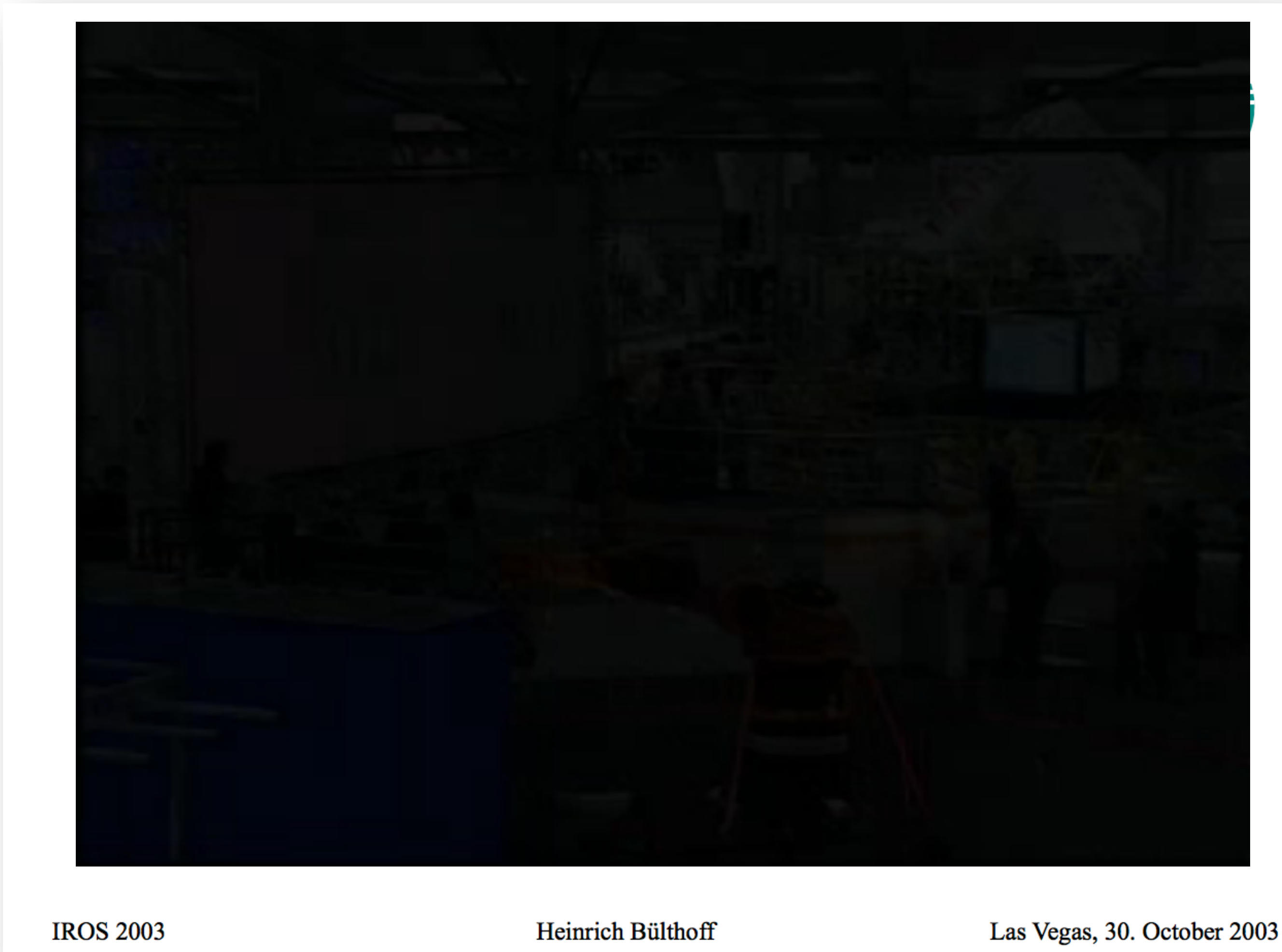
1. diagnostic features
2. dynamic information
3. active perception
4. human-in-the-loop

The Human: a complex cybernetic system

Our philosophy is to replace the environment with a virtual environment for better experimental control and to decouple the different sensory channels



A new tool (toy) for the investigation of multi-sensory closed-loop control



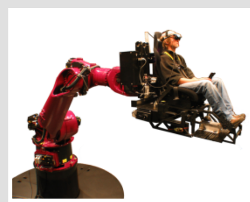
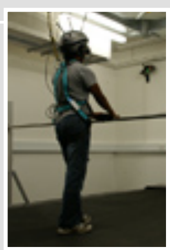
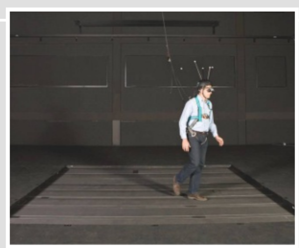
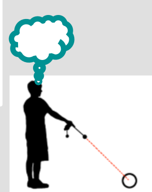
Cyberneum

The CYBERNEUM is a new facility for basic research in the area of multi-sensory perception and human-machine interaction.

Two large experiment halls (15x12x9m) with Vicon tracking and MPI CyberMotion simulator.

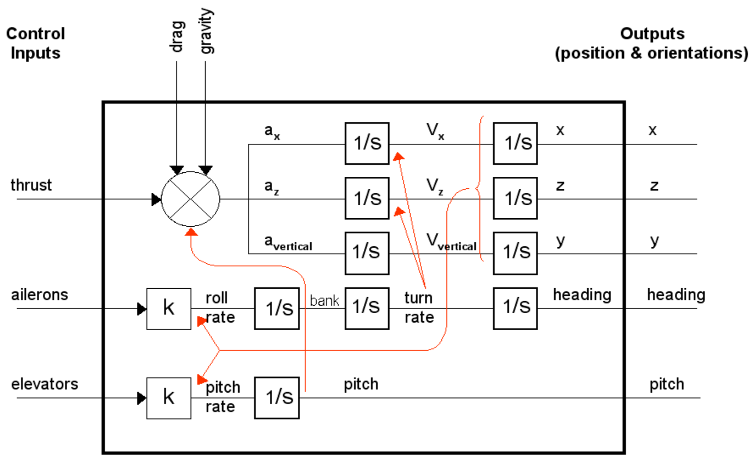
www.cyberneum.de

Isolating sources of non-visual information during locomotor tasks

	Proprioceptive	Vestibular
Full-Scale Walking 	✓	✓
Passive Transport  	✗	✓
Treadmill Walking  	✓	(✗)
Visual Self-Motion  Imagined Self-Motion 	✗	✗

Flight Control

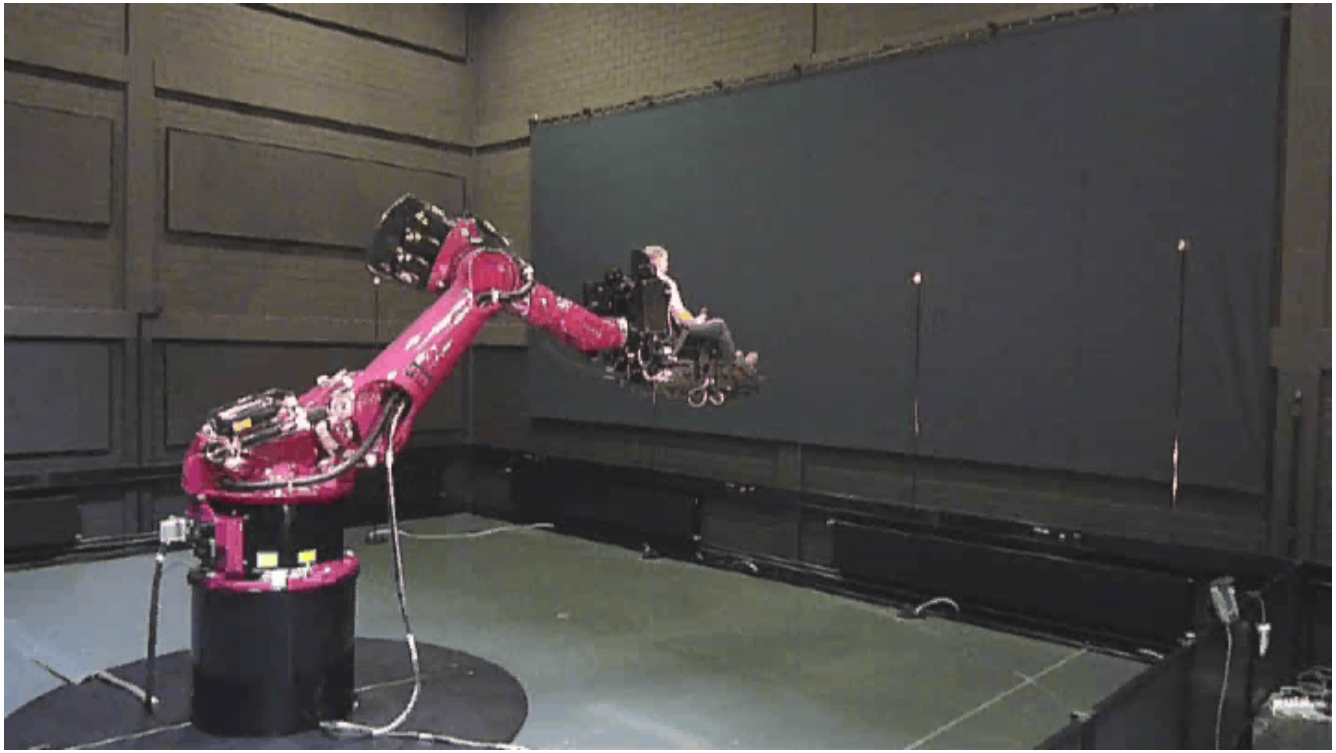
- is an interesting problem for understanding how the **brain** makes **assumptions** in order to close the perception-action loop
- is based on
  - an “**internal model**” of the dynamics of the aircraft



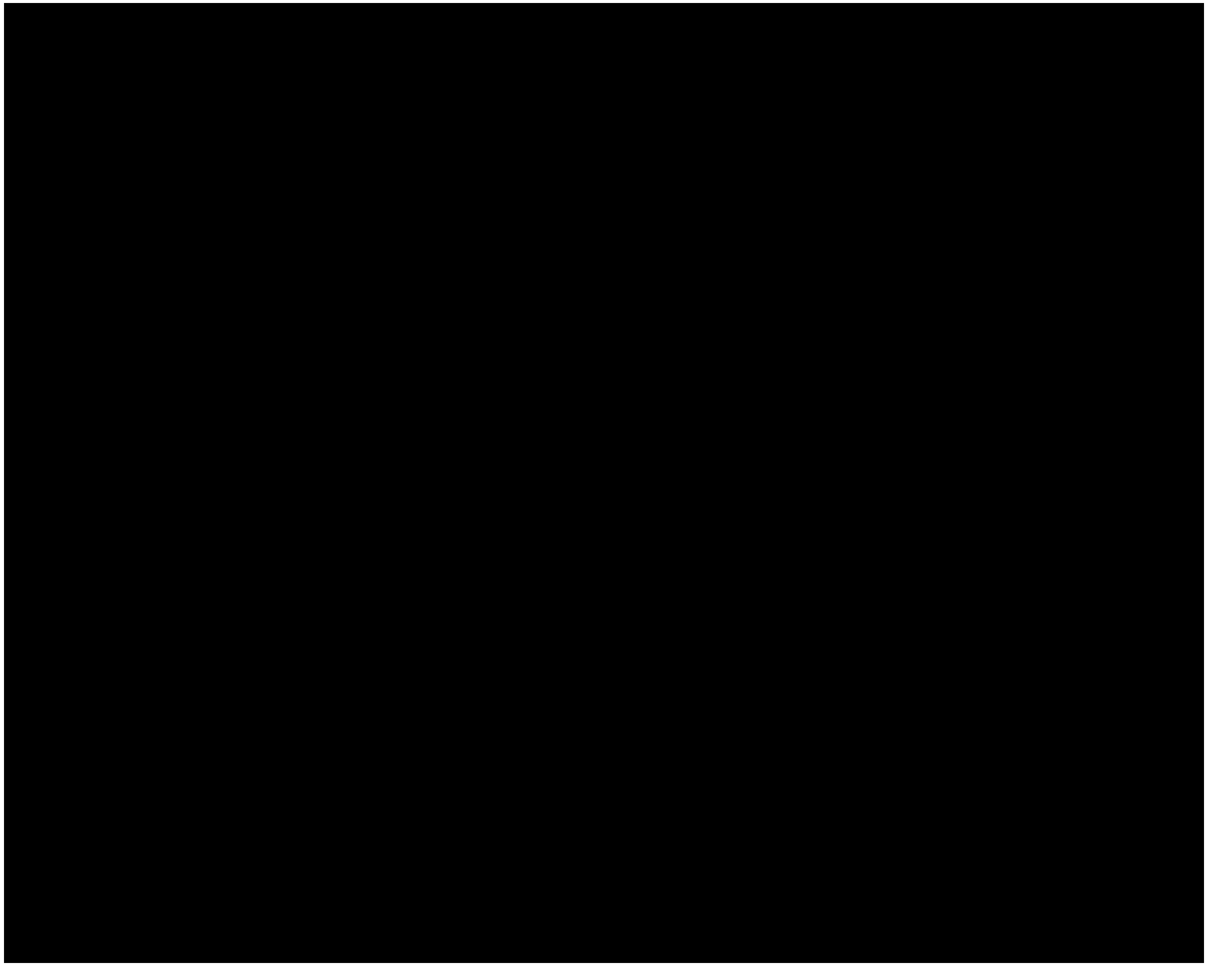
  - a “**perceptual model**” of the 3-dimensional environment
  - **sensing of self-motion**
    - visual information (optical flow)
    - vestibular and other inertial information (“seat-of-the-pants”)

First Steps to Helicopter Simulation

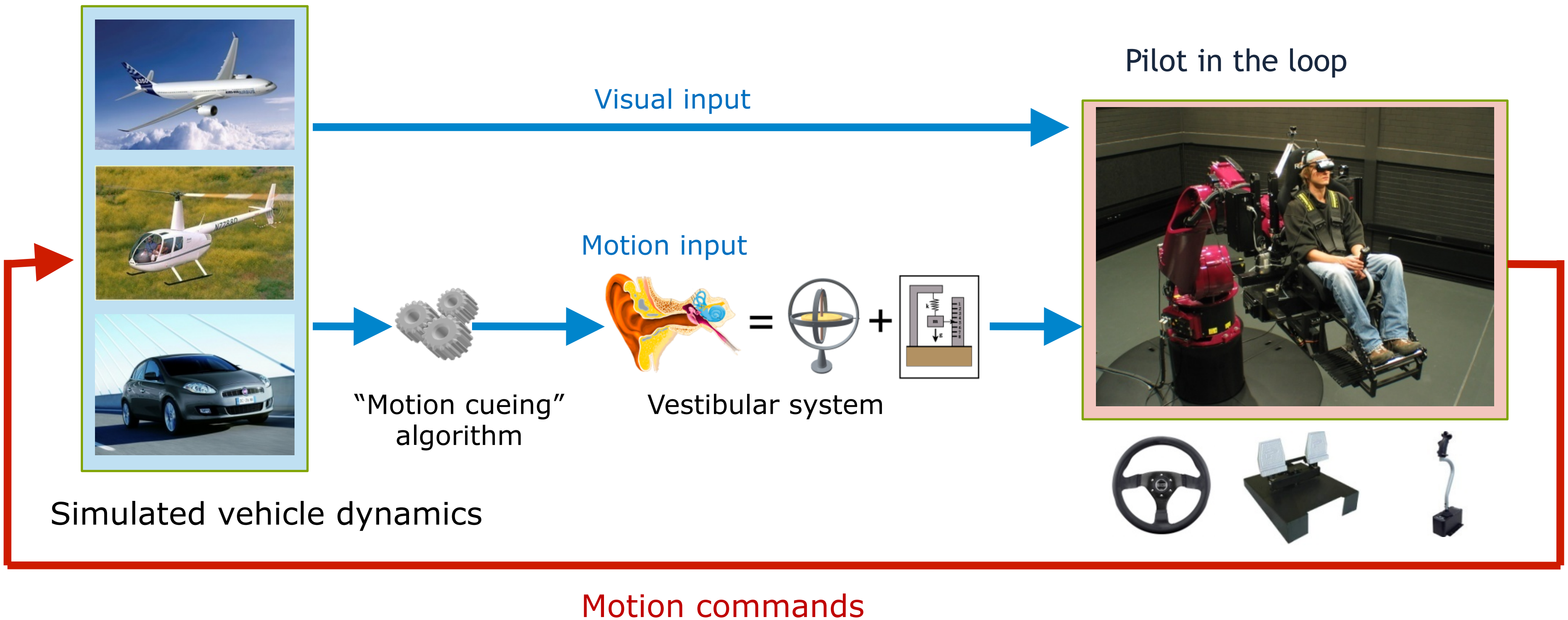
- The MPI CyberMotion Simulator can be used with realistic helicopter maneuvers in the real world
  - Nusseck, H.-G., H. J. Teufel, F. M. Nieuwenhuizen and H. H. Bülthoff: Learning System Dynamics:Transfer of Training in a Helicopter Hover Simulator. Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit (AIAA 2008), 1-11, AIAA, Reston, VA, USA (08 2008)



Virtual environment (HMD pilot view) visual-vestibular integration studies



The CyberMotion Simulator
The perfect tool for multi-sensory control studies



- visual simulation via HMD or projection system
- vestibular and somatosensory simulation (seat-of-the-pants)
- haptic simulation with control loading (cyclic stick and pedals)

From basic research to applied research

- Helicopter flying is difficult and training is expensive
- safer and inexpensive training
  - for hover maneuvers
  - emergency procedures
    - autorotation
    - quickstops
- familiar environment
  - inside and outside
  - no visualization problems
    - delays, field of view
- lower cost than Hexapod sims

HELITRAINER

HELIAVIATION services · school · technology

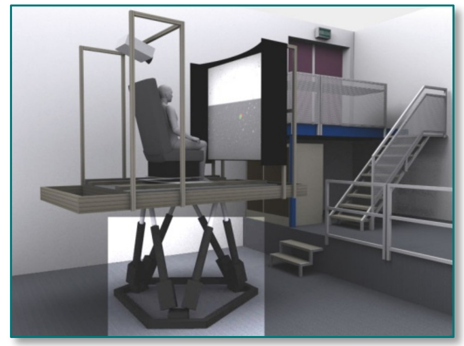
KUKA

Heli-Trainer

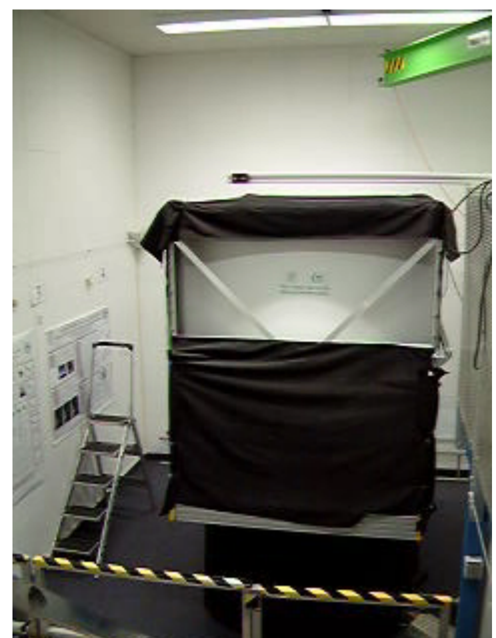
ILA, Berlin 2010

Today’s Flight Simulators

- are all based on a Hexapod (Stewart) platform

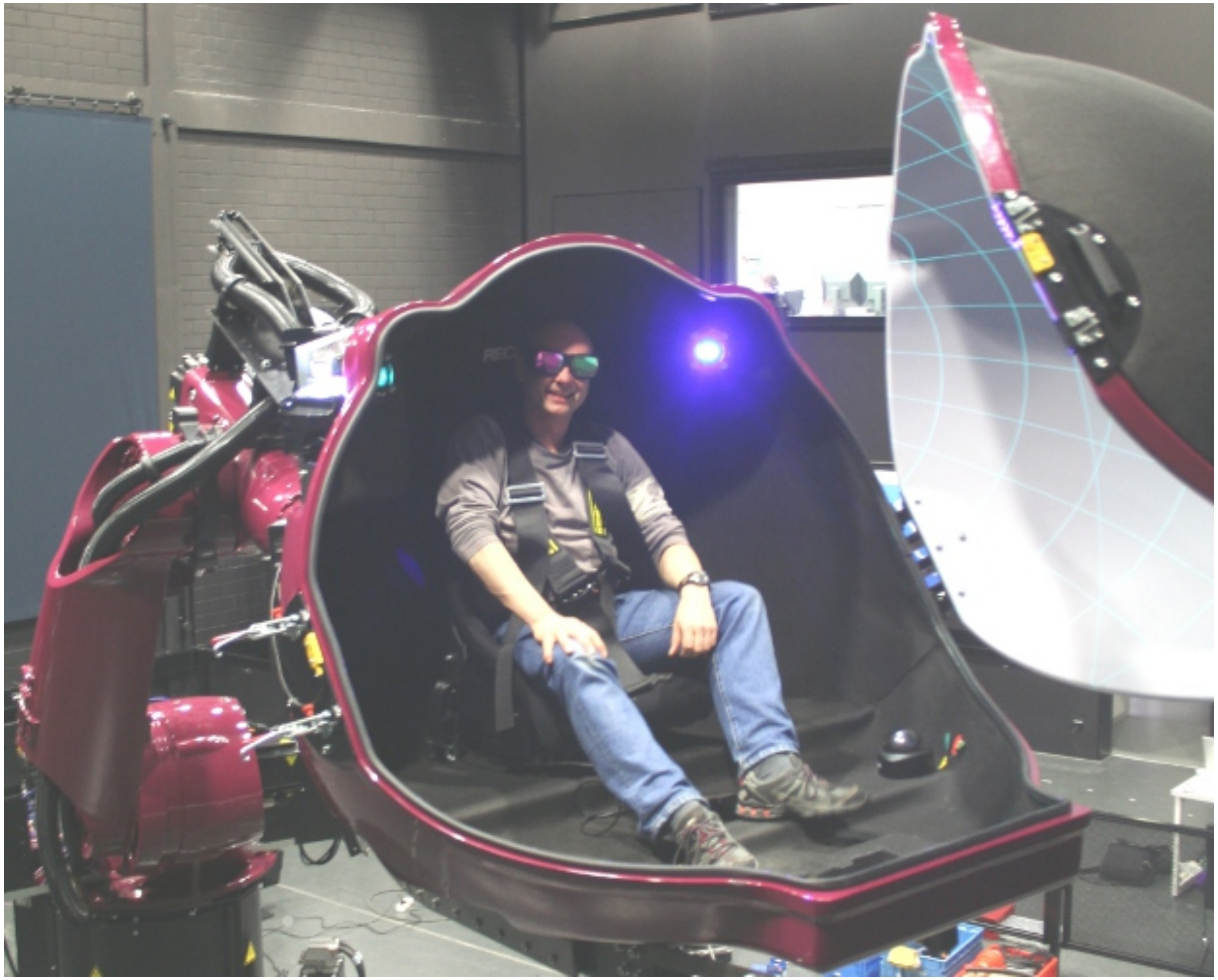


  - six degrees of freedom



- Advantage:
  - large payload
- Disadvantage:
  - limited workspace
  - roll-translation coupling

CyberMotion Simulator with seventh axis and stereo projection



BUCK ENGINEERING & CONSULTING

eyevis PERFECT VISUAL SOLUTIONS

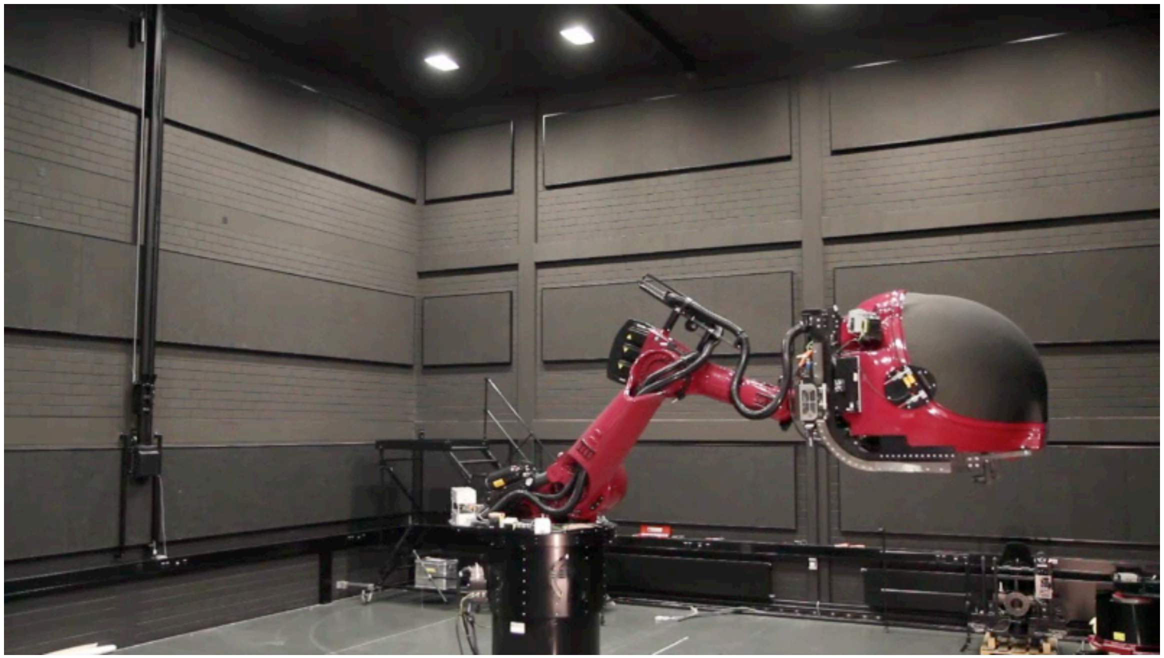
KUKA

- **Empty weight: ~ 390kg**
- **7.-Achse: 90°**
- **3D-Stereoprojection**
- **FOV > 110° horizontal**
  > 90° vertical

Future Developments

- Linear Axis (12m) Summer 2012
  - extend the linear workspace by 10m
  - for lane change manoeuvres in driving simulation
  - for autorotation manoeuvres in helicopter training

SUPRA– Upset recovery training



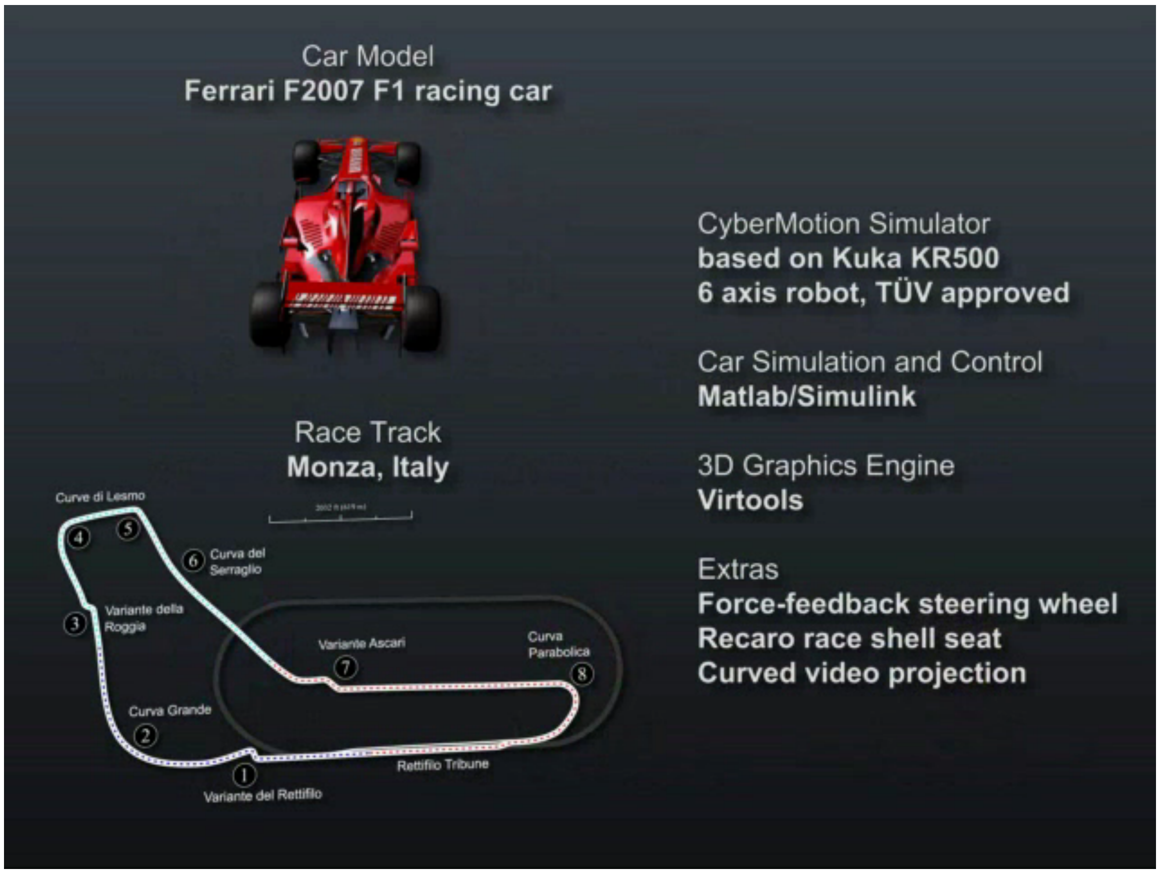
SIMULATION OF UPSET RECOVERY IN AVIATION

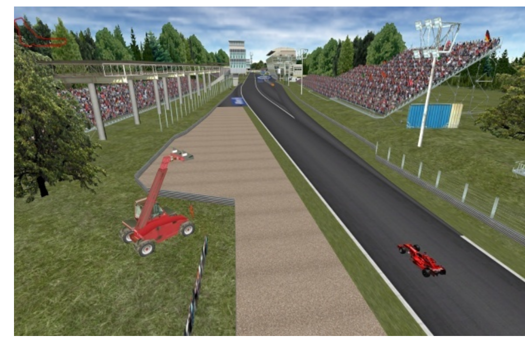
supra

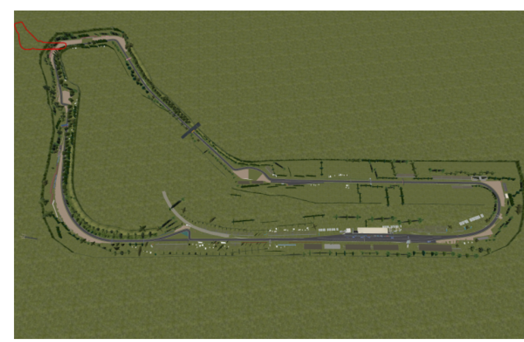
TNO · AMST AMST-SYSTEMTECHNIK GMBH · dmu.ac.uk DE MONTFORT UNIVERSITY LEICESTER · BOEING · Max-Planck-Institut für biologische Kybernetik · DINAMIKA aviation simulators · ЦАГИ

- Loss of Control in Flight (LOC-I) resulting from unsuccessful upset recovery during flight is one of the primary causes of fatal accidents in civil aviation
- Nine established research organisations from six different countries collectively aim at enhancing flight simulator technology beyond its current capabilities to allow for effective upset recovery training.

The next generation of multisensory games



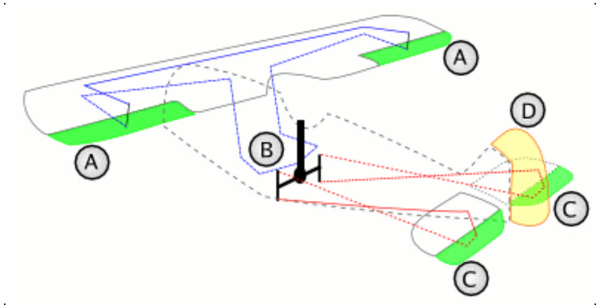




Haptic feedback

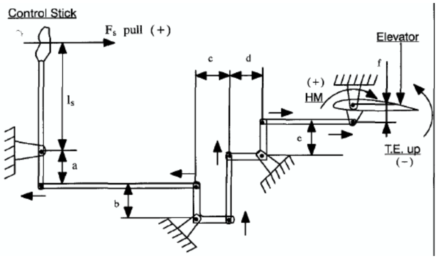
- Feeling force cues can be crucial for piloting a vehicle

- The force felt on the steering wheel informs car pilots on the amount of grip between tires and road



- An airplane pilot can judge the aerodynamic load or occurrence of wind gusts
  - He can “feel” the state of the aircraft

- Assume a fly-by-wire system or a remote control scenario



- What is the “best” haptic feedback to help pilots in his task?

Haptic Control for Airplanes and Teleoperation of Remotely Piloted Vehicles

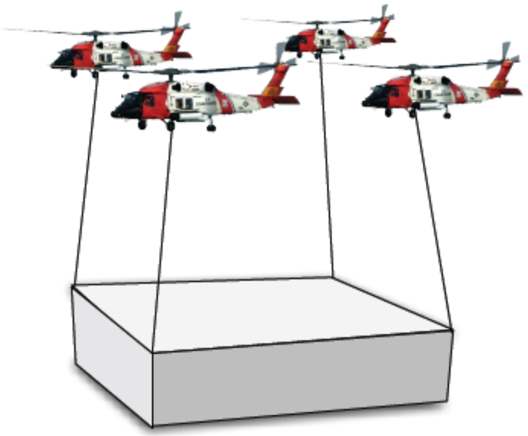
omega 6 force dimension™

- Alaimo SMC , Pollini L and Bülthoff HH (2011) **Admittance-based bilateral teleoperation with time delay for an Unmanned Aerial Vehicle involved in an obstacle avoidance task** AIAA Modeling and Simulation Technologies Conference 2011 (MST-2011), American Institute of Aeronautics and Astronautics, -. accepted
- Alaimo, S. M.C., L. Pollini, A. Magazzù, J.-P. P. Bresciani, P. R. Giordano, M. Innocenti and H. H. Bülthoff: **Preliminary Evaluation of a Haptic Aiding Concept for Remotely Piloted Vehicles**. EuroHaptics 2010, 418-425 (07 2010)
- Alaimo SMC , Pollini L , Bresciani J-P and Bülthoff HH (2010) **A Comparison of Direct and Indirect Haptic Aiding for Remotely Piloted Vehicles** 19th IEEE International Symposium in Robot and Human Interactive Communication (IEEE Ro-Man 2010), IEEE, Piscataway, NJ, USA, 506-512.
- Alaimo SMC , Pollini L , Bresciani J-P and Bülthoff HH (2011) **Evaluation of Direct and Indirect Haptic Aiding in an Obstacle Avoidance Task for Tele-Operated Systems** 18th World Congress of the International Federation of Automatic Control (IFAC 2011), 1-6. accepted

Visual/ Haptic control of a team of flying robots

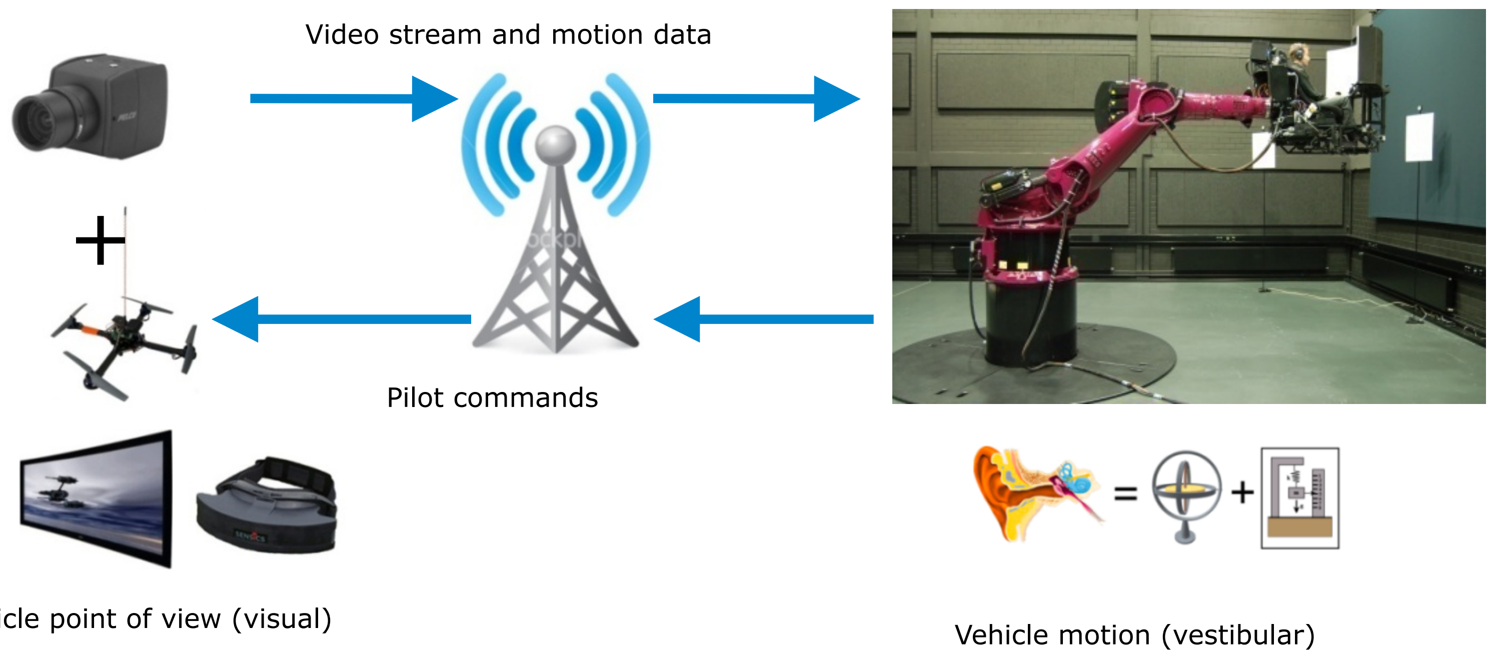
- Human commands the collective motion
- Robots must have autonomy:
  - Keep the formation
  - Avoid obstacles
  - Gather a map of the environment
  - Pick and place operations

- Human receives a “suitable” feedback:
  - Inertia
  - Forbidden directions (e.g., obstacles)
  - External disturbances (wind)
  - Goal location



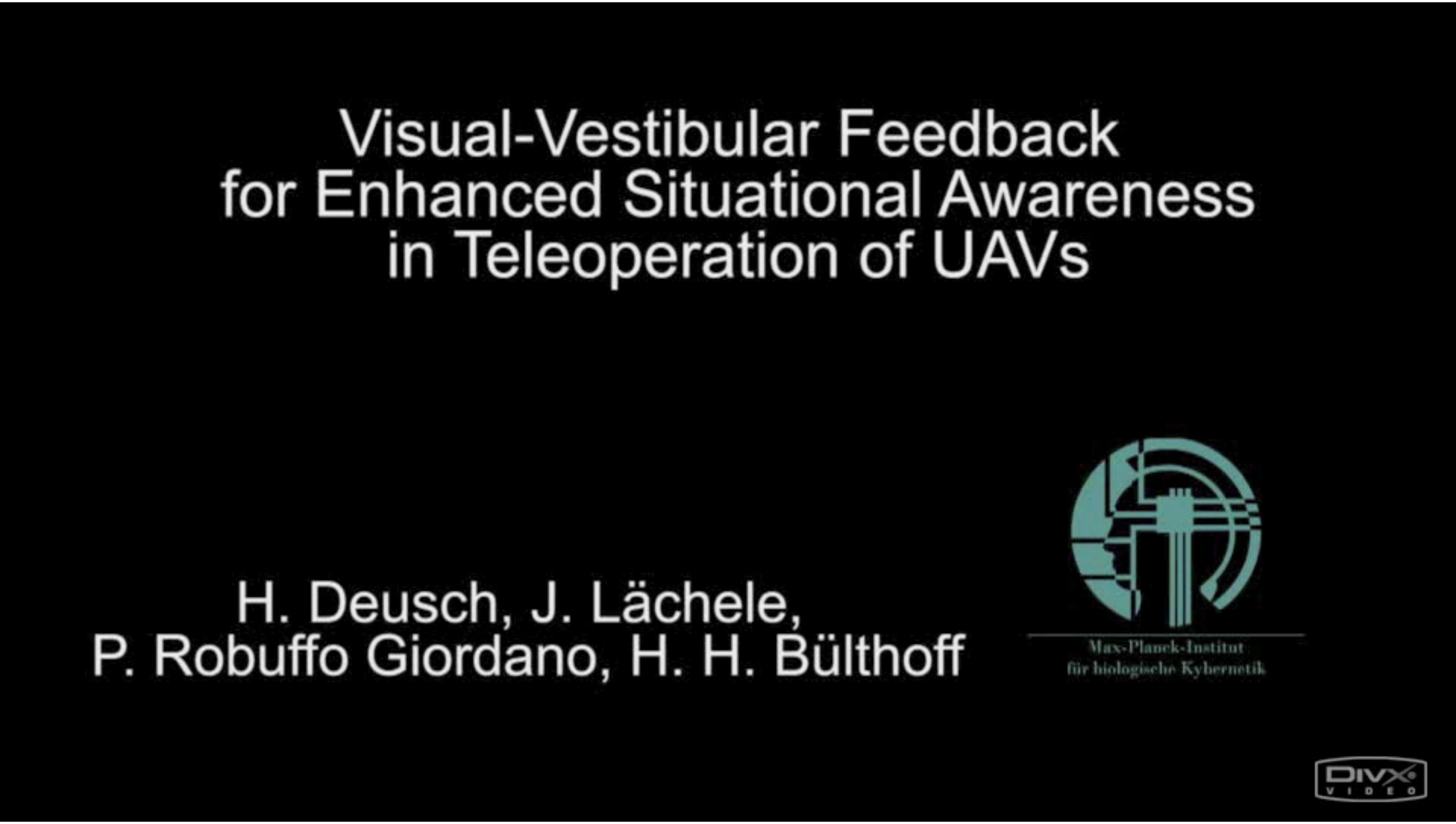
Multi-sensory control of Unmanned Aerial Vehicles (UAVs)

- Add vestibular feedback to enhance situational awareness
  - Scenario: remote teleoperation of a flying vehicle (in our case a quadcopter)
  - Hypothesis: vestibular feedback improves situational awareness for the pilot (and thus facilitates task execution)



Teleoperation of Unmanned Aerial Vehicles

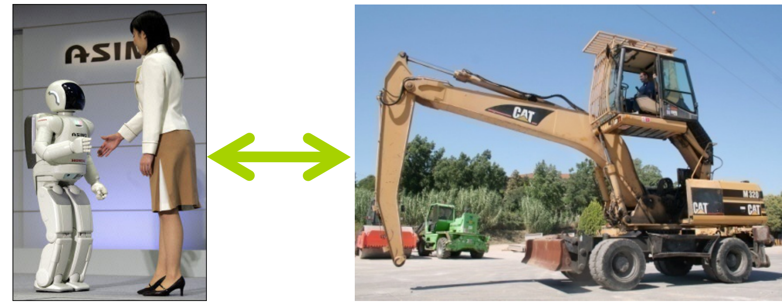
AHS 66th (2010)



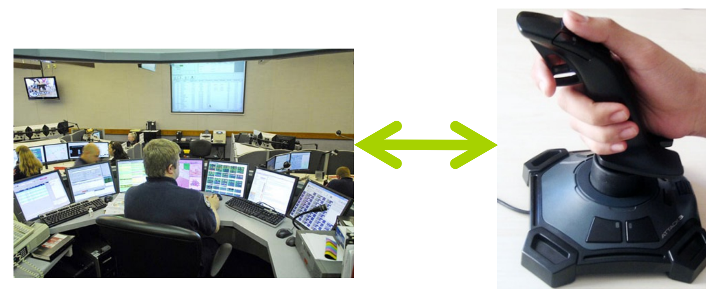
Conclusion and further questions

- By understanding better the human perception and action control system we can design novel ways to interface humans and (<u>autonomous)</u> machines

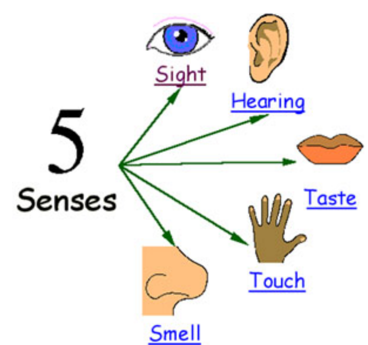
- Many possible scenarios, e.g.,
  - human-robot interaction
  - Autonomous or "passive" machines?
  - Physical or remote/virtual interaction?



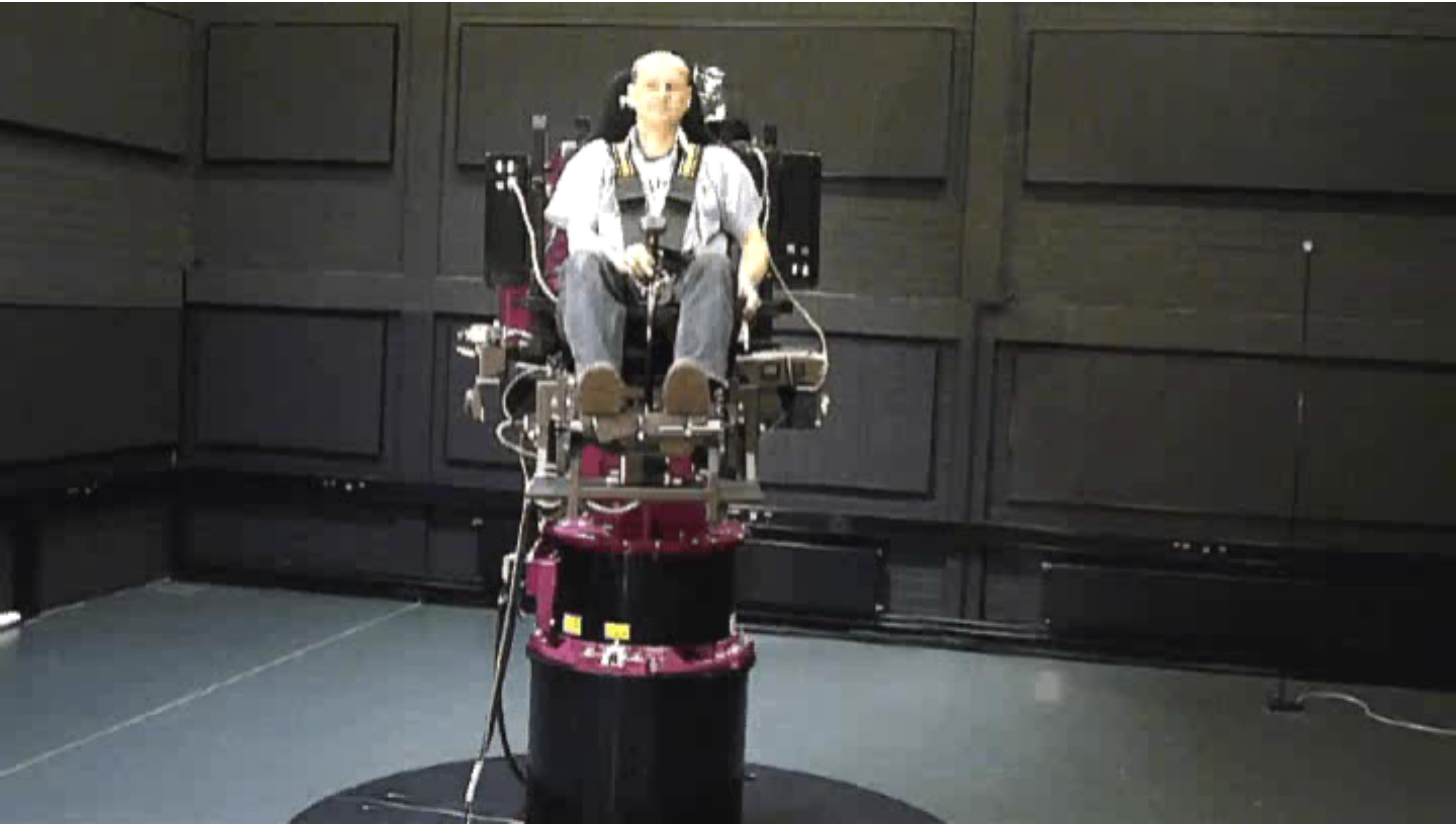
- Human role:
  pure supervisor or full control?



- What feedback is best suited for the humans?
  - Which sensory channel?
  - What information (quantity and quality)?
  - What is the best way to integrate information (Bayes)?



Thank you very much for your attention



Interdisciplinary team

- Biology

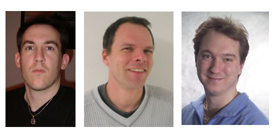
- Neuroscience

- CS/Eng

- Medicine

- Mathematics

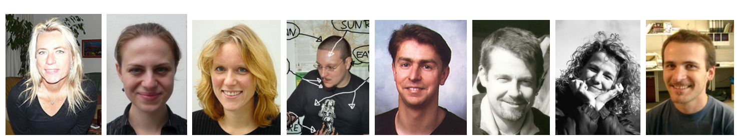
- Physics



- Optometry

- Psychol.

- Support

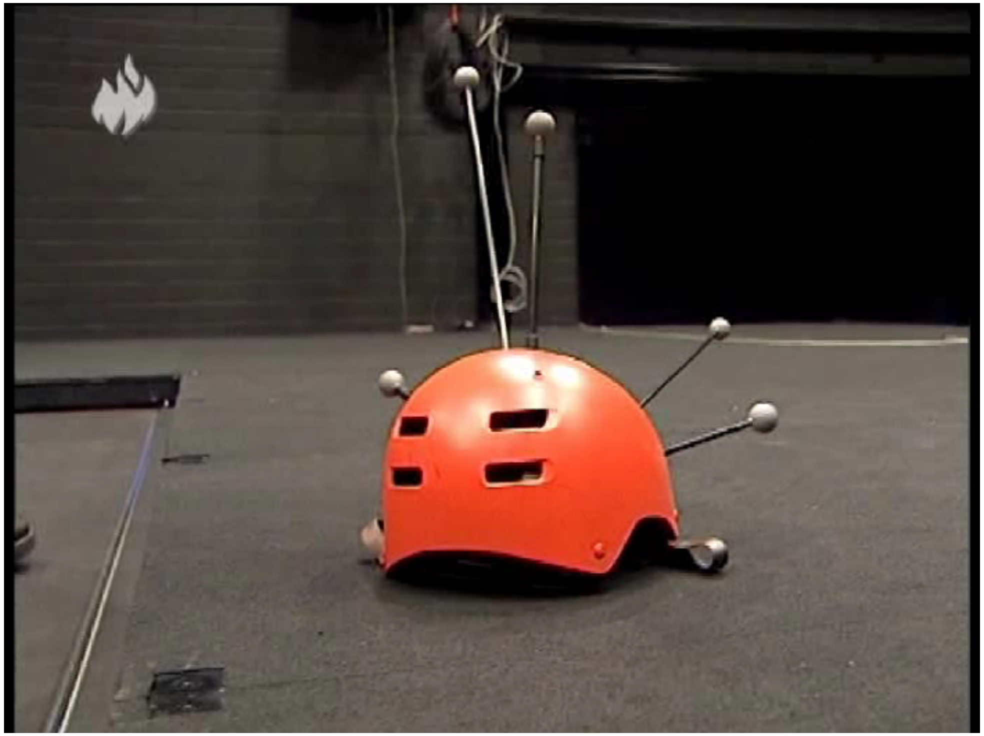


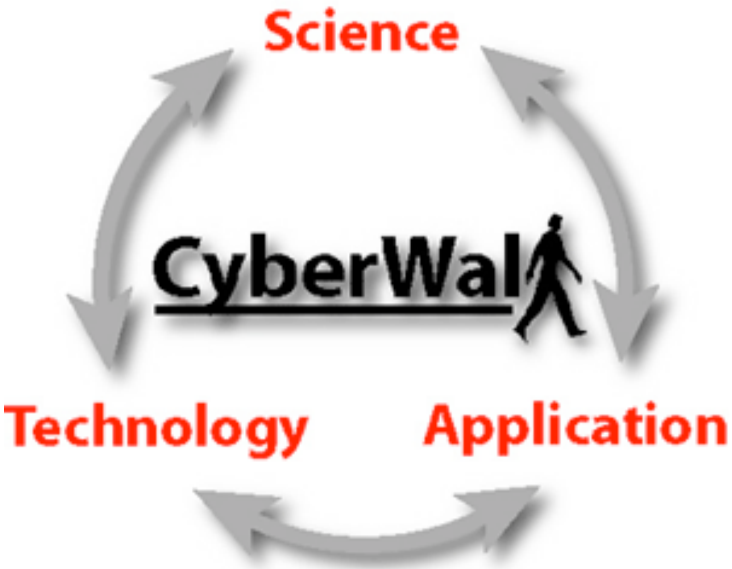
Open questions

- ICT questions / Computer Science
  - How can we make better and controllable animations?
  - How can we build a realistic and real-time virtual human?
  - Can we train recognition systems providing fast generalization capabilities in real-world?

- Perception/Cognition / Neuroscience
  - How do we perceive biological motion in interactive situations?
  - How is the production and perception of actions coupled?
  - How innate is biological movement perception, and
  - What are the mechanisms of learning them?

Omni-Directional Treadmill for Spatial Cognition Research



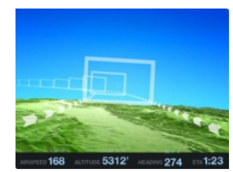


- Unconstrained walking in all directions (2D), creating a truly immersive locomotion interface for VR

Passive machines

- Example: a human piloting a vehicle, either directly or remotely
  - The human is in full control: need of enough information to complete the task (drive a car, fly an airplane, stabilize a helicopter, etc.)

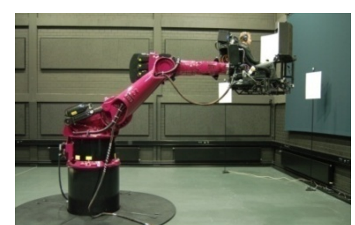
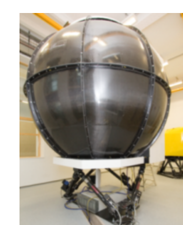
- How to present the needed information?
  - Increase situational awareness (esp. in remote control tasks)
  - Facilitate task execution
  - Develop better/faster training procedures



- Visual cues: tunnel-in-the-sky, glass cockpit




- Haptic/force cues: force-feedback devices, tactile vests

- Vestibular (self-motion) cues: motion platform, motion simulators

Haptic feedback

- The sense of touch carries rich and “fast” information
- Widely exploited in teleoperation applications (e.g., telesurgery)