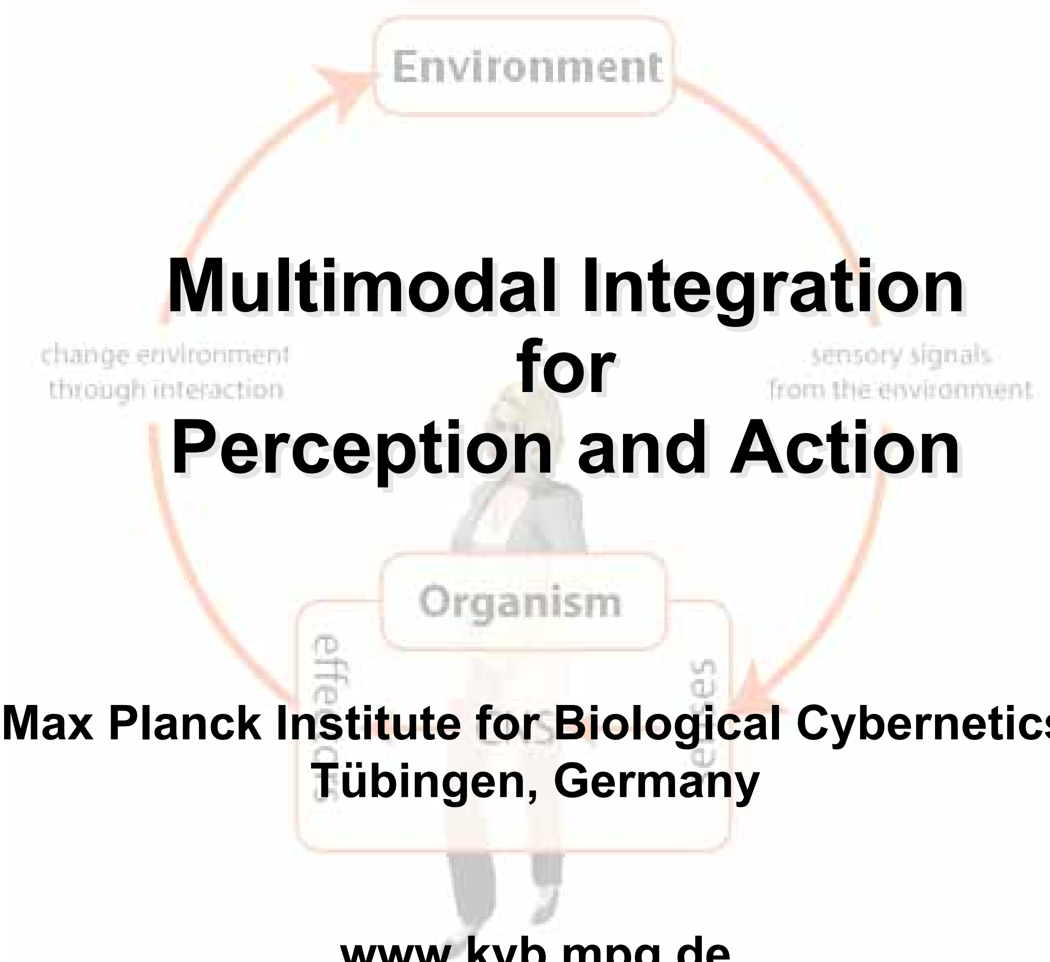
MAX-PLANCK-GESELLSCHAFT

MPI FOR BIOLOGICAL CYBERNETICS

Heinrich H. Bülthoff



Multimodal Integration for Perception and Action

Max Planck Institute for Biological Cybernetics
Tübingen, Germany

www.kyb.mpg.de

Outline

- Sensor fusion at an Early Level
  - Uni-modal Integration (Vision)
    - Shape-from-X (stereo, shading, texture, motion)
  - Multi-modal Integration
    - Visual-Haptic
    - Visual-Vestibular
    - Visual-Auditory
- Sensor fusion at a Higher Level
  - Effects of Attention and Awareness on Integration
- Integration for Control Tasks
  - Multimodal integration for novel ego-motion simulators

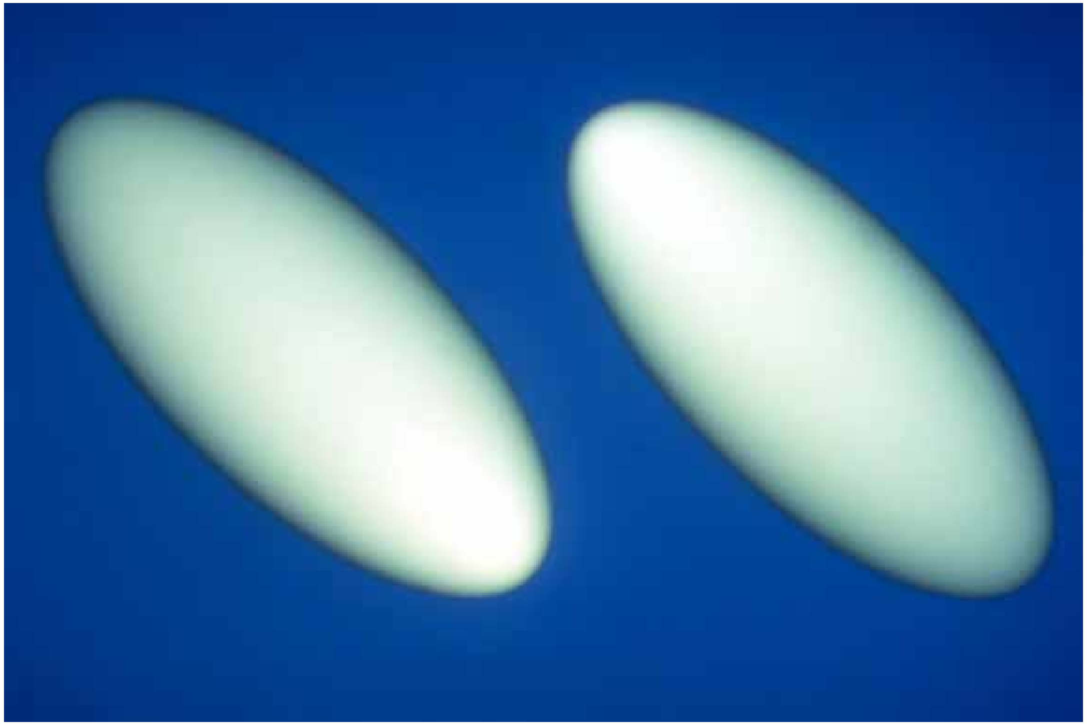
Early Sensor Fusion

examples from vision studies

Bülthoff, Mallot, Blake, Yuille,…(1985-1993)

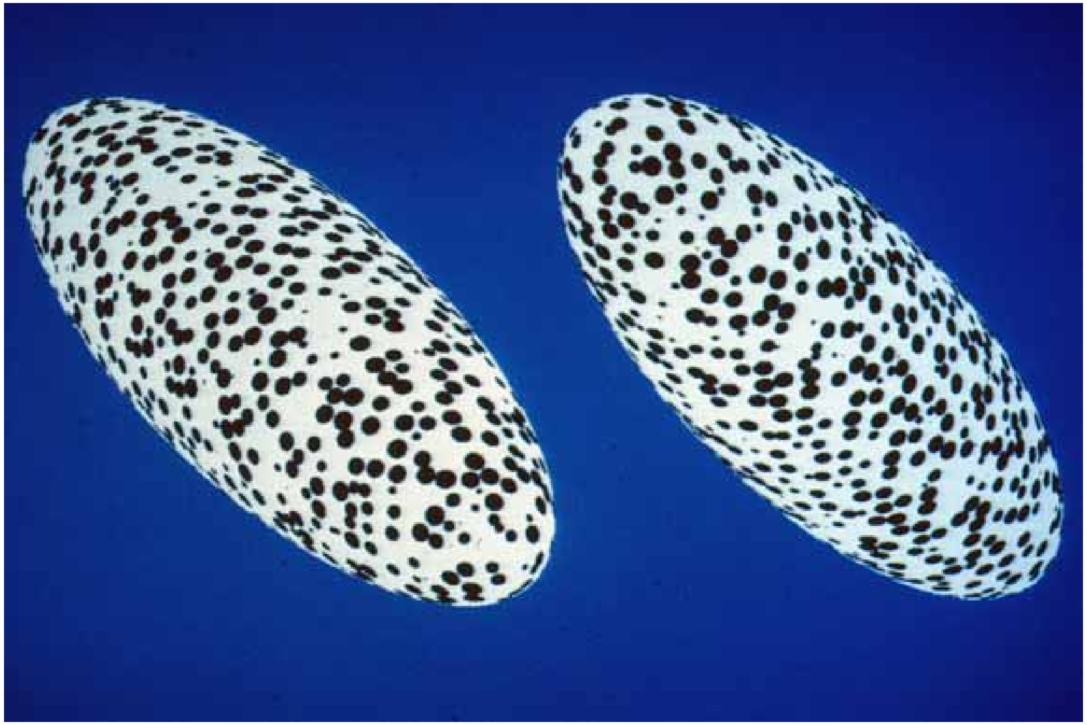
- **accumulation** (the-more-the-better policy)
  - linear combination of shape-from-x modules
  - joint regularization (with cost function)
- **cooperation or strong fusion**
  - likelihood functions for individual cues are often not independent
  - shape-from-shading and shape-from-texture are weak cues
  - combined they provide almost perfect shape perception

Shape-from-Shading



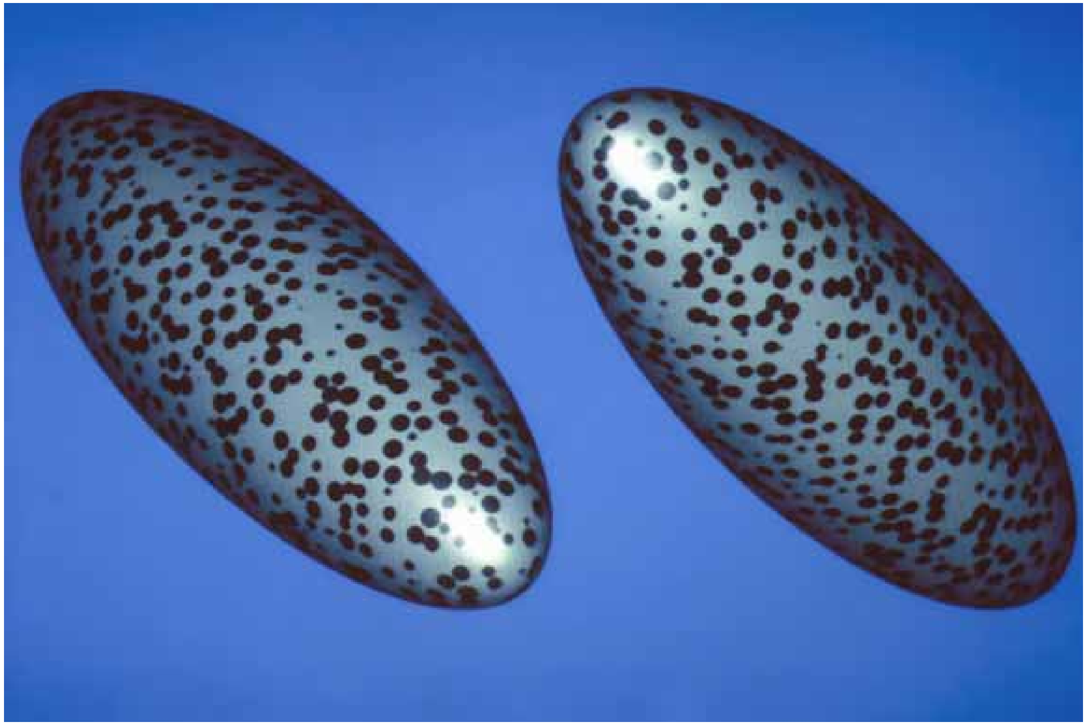
Shape-from-shading is a weak cue
to the perception of orientation.

Shape-from-Texture



Shape-from-texture is a weak cue
to form and orientation.

Shading + Texture



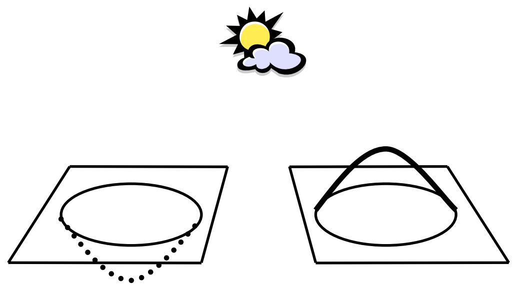
Integration of all cues provides good perception of form and orientation.

Early Sensor Fusion

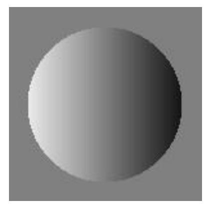
- **accumulation** (the-more-the-better policy)
  - linear combination of shape-from-x modules
  - joint regularization (with cost function)
- **cooperation or strong fusion**
  - likelihood functions for individual cues are often not independent
  - shape-from-shading and shape-from-texture are weak cues
  - **combined they provide almost perfect shape perception**
- **disambiguation**
  - stereo can disambiguate shading (convex / concave)

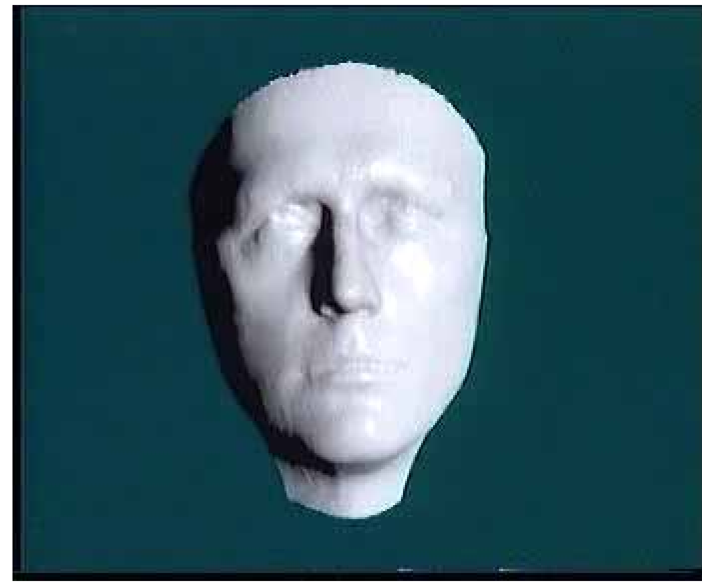
Disambiguation

shape-from-shading is ambiguous



valley hill

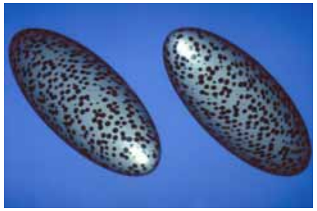




- **convexity prior** (familiarity) dominates ambiguous interpretation
  - ***Langer, M.S. and H.H. Bülthoff:*** A prior for global convexity in local shape-from-shading. Perception **30**, 403-410 (2001)
  - ***Langer, M.S. and H.H. Bülthoff:*** Depth discrimination from shading under diffuse lighting. Perception **29**, 649-660 (2000)
- **stereo disambiguates** shape-from-shading

Sensor fusion of visual modules

- **accumulation** (the-more-the-better policy)
  - linear combination of shape-from-x modules
  - joint regularization (with cost function)
- **cooperation or strong fusion**
  - likelihood functions for individual cues are often not independent
  - shape-from-shading and shape-from-texture are weak cues
  - combined they provide almost perfect shape perception
- **disambiguation**
  - stereo can disambiguate shading (convex / concave)
- **veto**
  - very strong cues should not be challenged by others
  - edge-based stereo vetoes intensity-based stereo



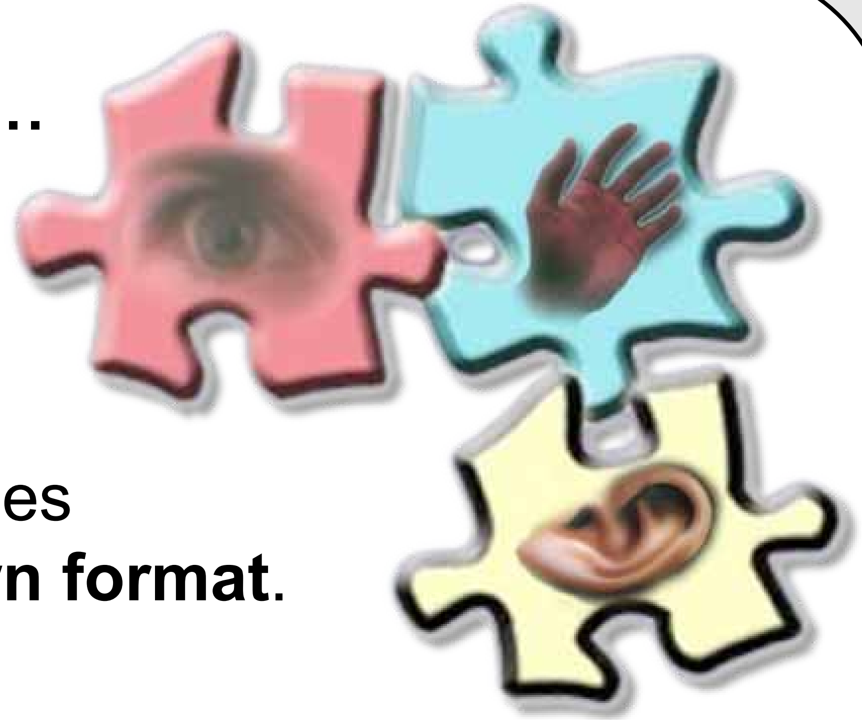
Bülthoff, H.H. and H.A. Mallot: Integration of depth modules: stereo and shading. Journal of the Optical Society of America 5, 1749-1758 (1988)

Blake, A. and H.H. Bülthoff: Does the brain know the physics of specular reflection? Nature 343, no. 6254, 165-168 (1990)

Bülthoff, H.H.: Shape from X: Psychophysics and Computation. Computational Models of Visual Processing, M. Landy and A. Movshon Eds., M.I.T. Press, 305-330 (1991)

Multimodal Sensor Fusion
The Puzzle of the Senses

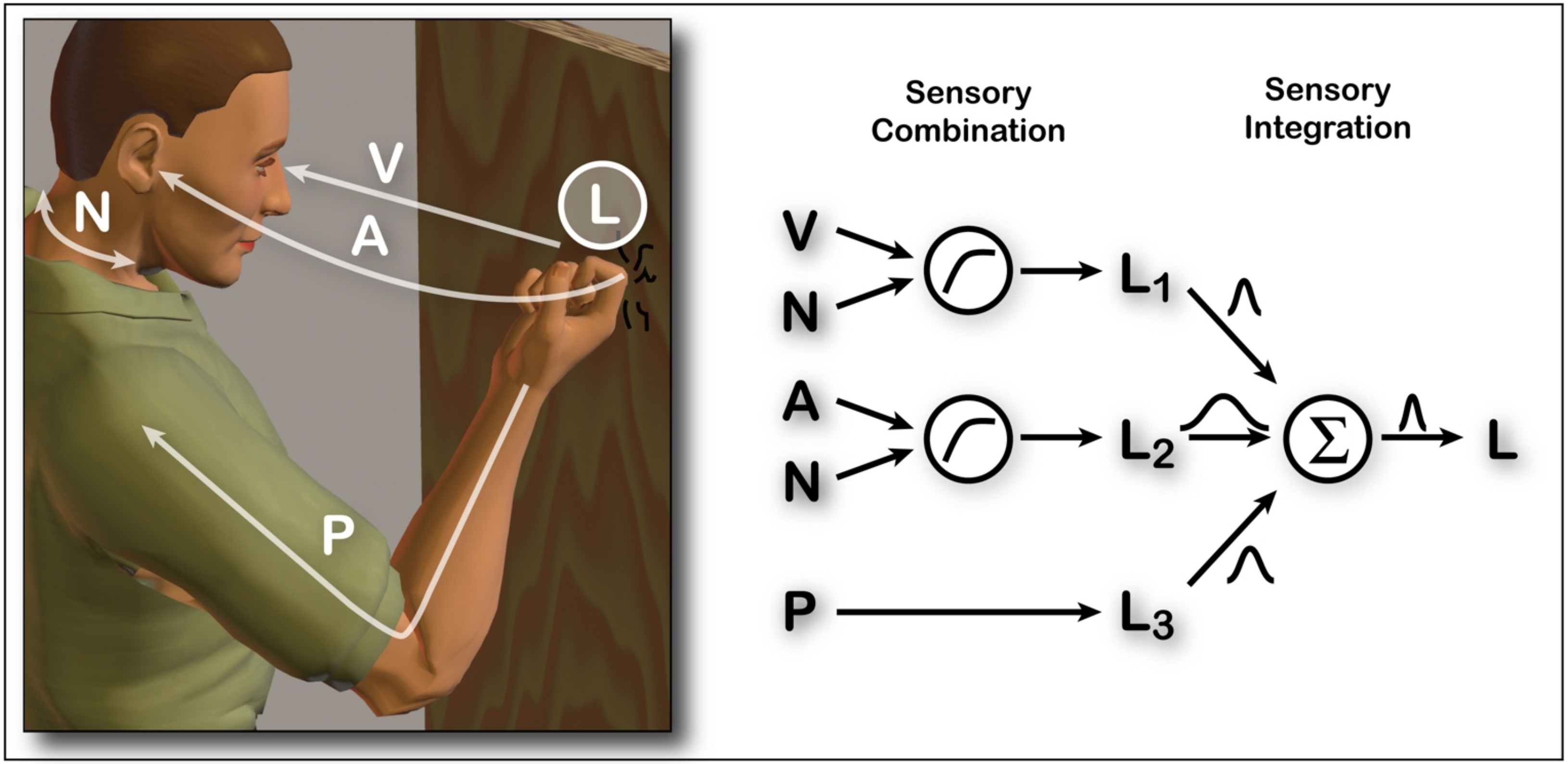
Vision, Touch, Audition, ...



Each sensory modality provides **unique information** in its **own format**.

How is all this sensory information put together to form a coherent percept?

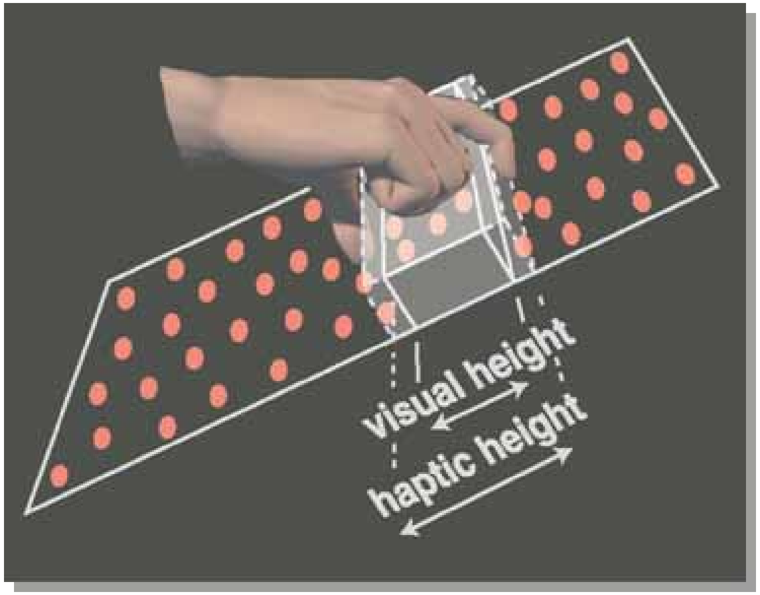
Forming a Unique Percept



M. Ernst & H. Bülthoff, TICS 2004

The problem of sensory integration

Size estimation with vision and touch



Physics

Vision: photons striking retinas

Haptics: changing pressure on fingers

Sensor bias & noise

Systematic bias: between senses arises from consistent distortion (e.g., glasses or gloves)

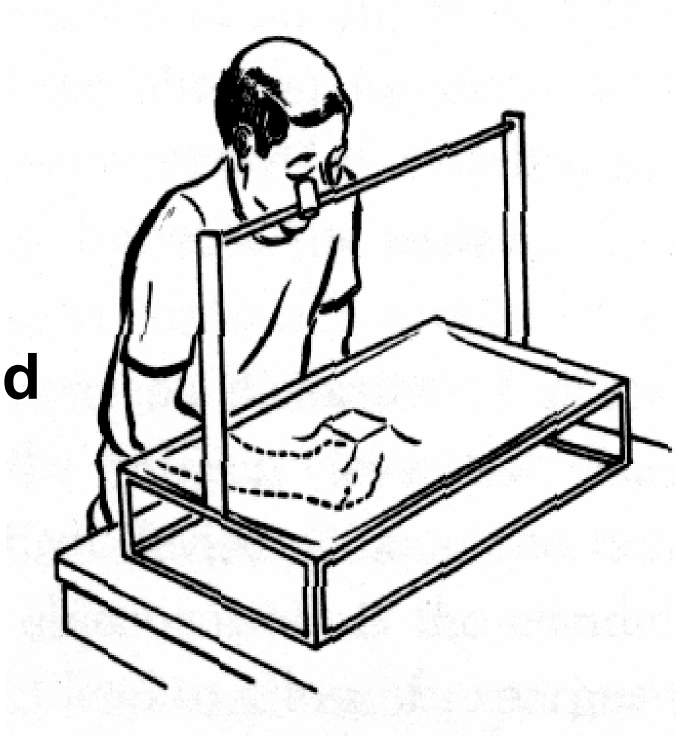
Unsystematic bias: arises from measurement noise

What is the optimal way to integrate sensory information?

Multimodal Cue Integration
Rock & Victor (1964)

Irv Rock

View cube through distorting lens while exploring object haptically.

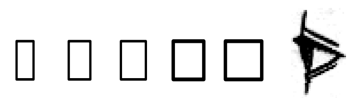
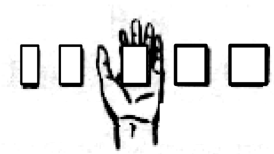


Visually and haptically specified shapes differ.

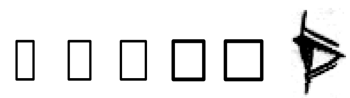
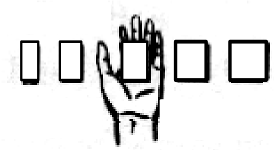
What shape is perceived?

Cube or elongated box ?

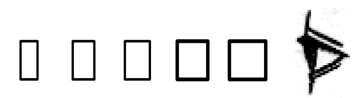
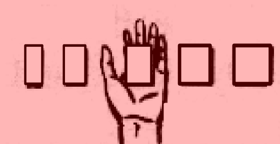
Rock & Victor (1964) Experimental Design

Stimulus Presentation			Response Method
Vision alone	Haptic alone	Conflict	
V	*H*	*V* *H*	Drawing 
V	*H*	*V* *H*	Vision alone 
V	*H*	*V* *H*	Haptic alone 

Rock & Victor (1964) Results

Stimulus Presentation			Response Method
Vision alone	Haptic alone	Conflict	
V 1.90	*H* 0.98	*V* *H* 1.85	Drawing 
V 13.4	*H* 23.1	*V* *H* 14.1 mm	Vision alone 
V 14.1	*H* 20.5	*V* *H* 14.5 mm	Haptic alone 

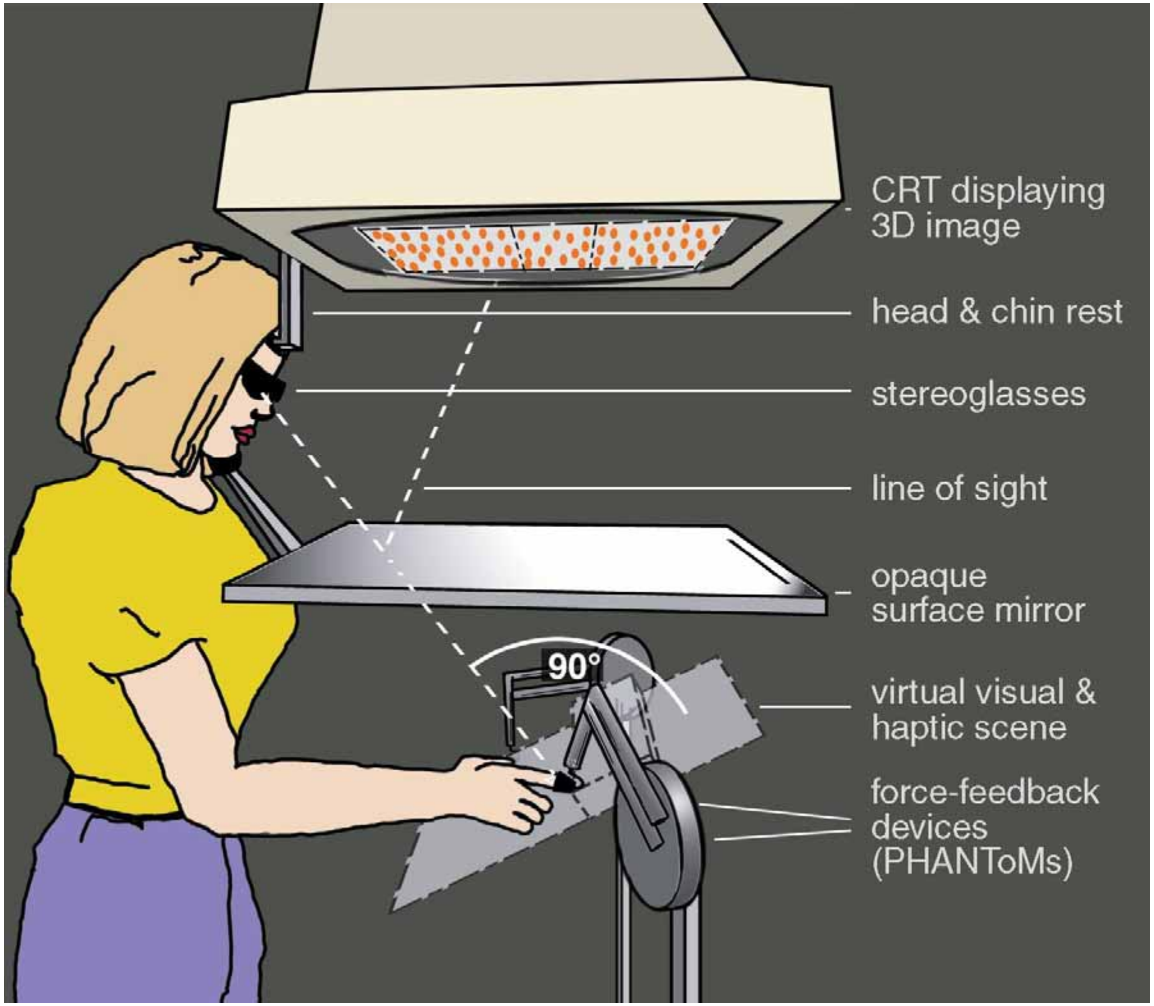
Rock & Victor (1964)
Results

Stimulus Presentation			Response Method
Vision alone	Haptic alone	Conflict	
V 1.90	*H* 0.98	*V* *H* 1.85	Drawing 
V 13.4	*H* 23.1	*V* *H* 14.1 mm	Vision alone 
V 14.1	*H* 20.5	*V* *H* 14.5 mm	Haptic alone 

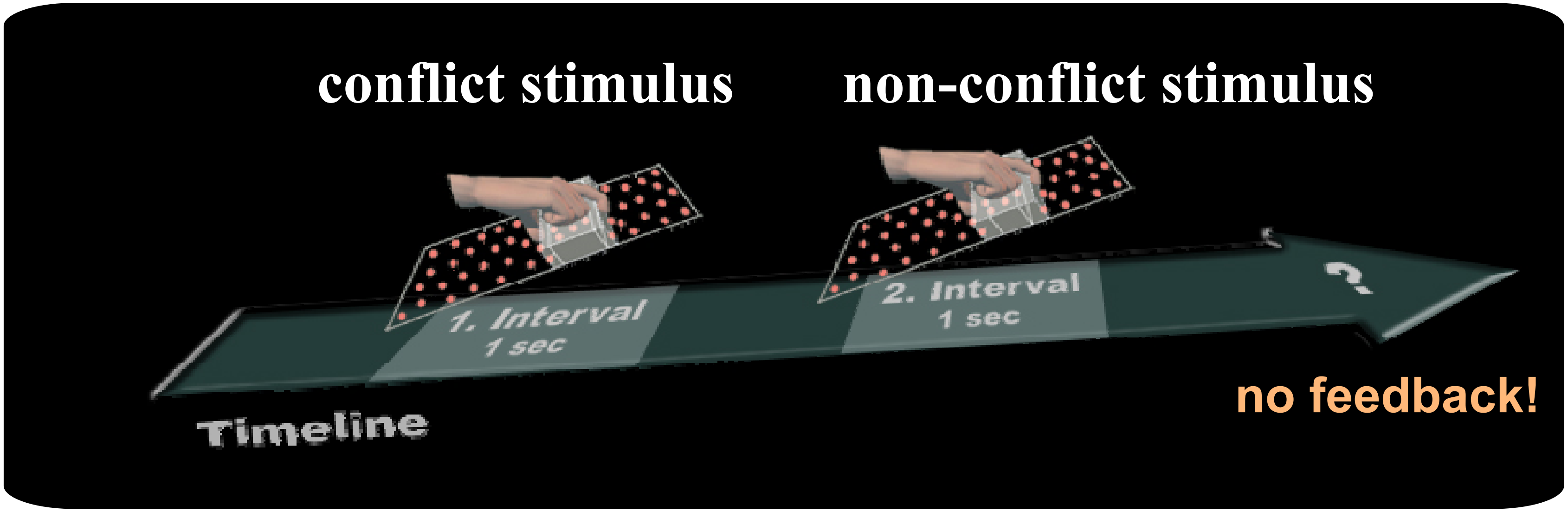
“Visual Capture”

Visual-Haptic Integration

Irv Rock revisited by Marc Ernst

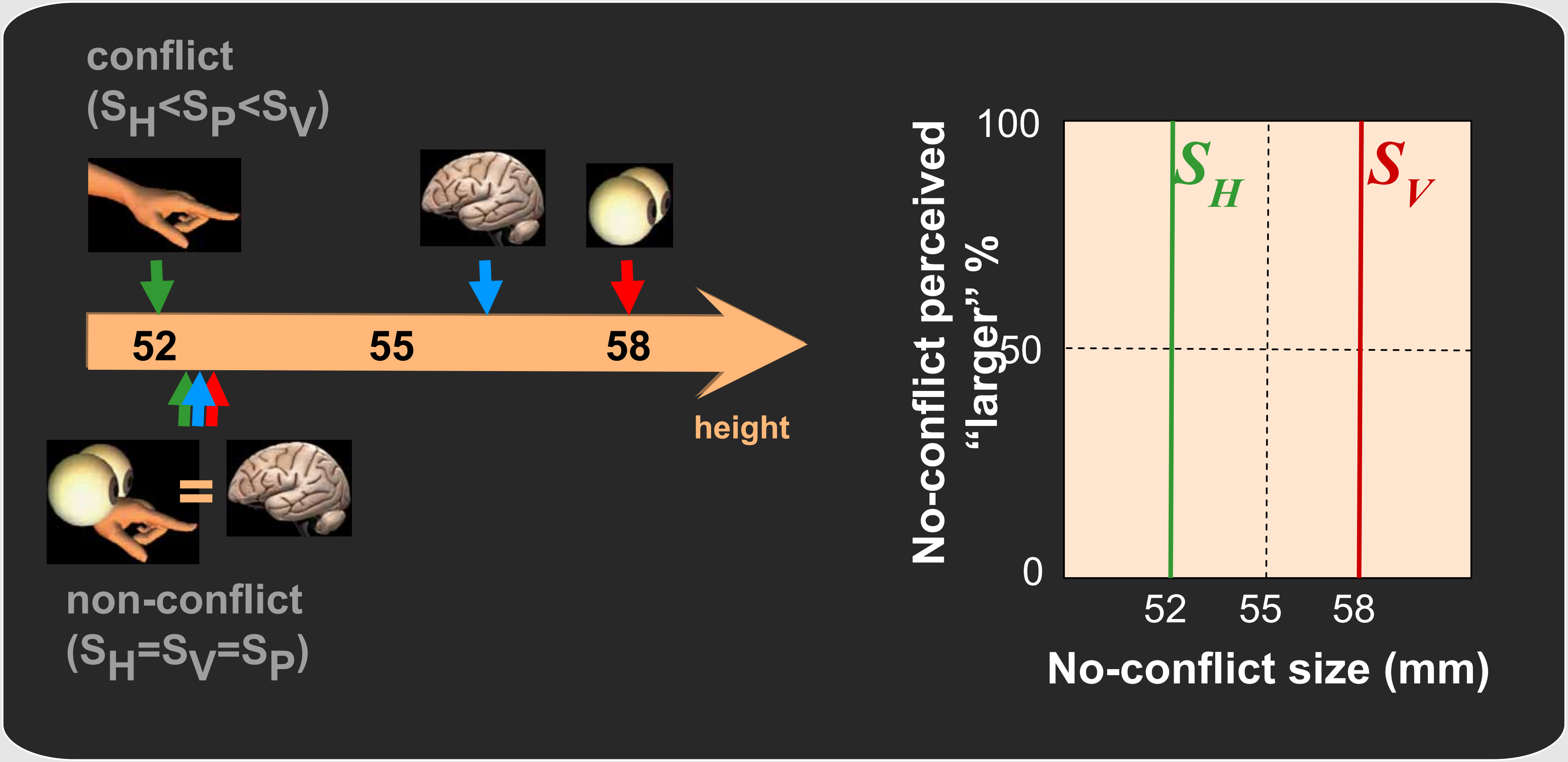


Size Comparison
with visual and haptic cues

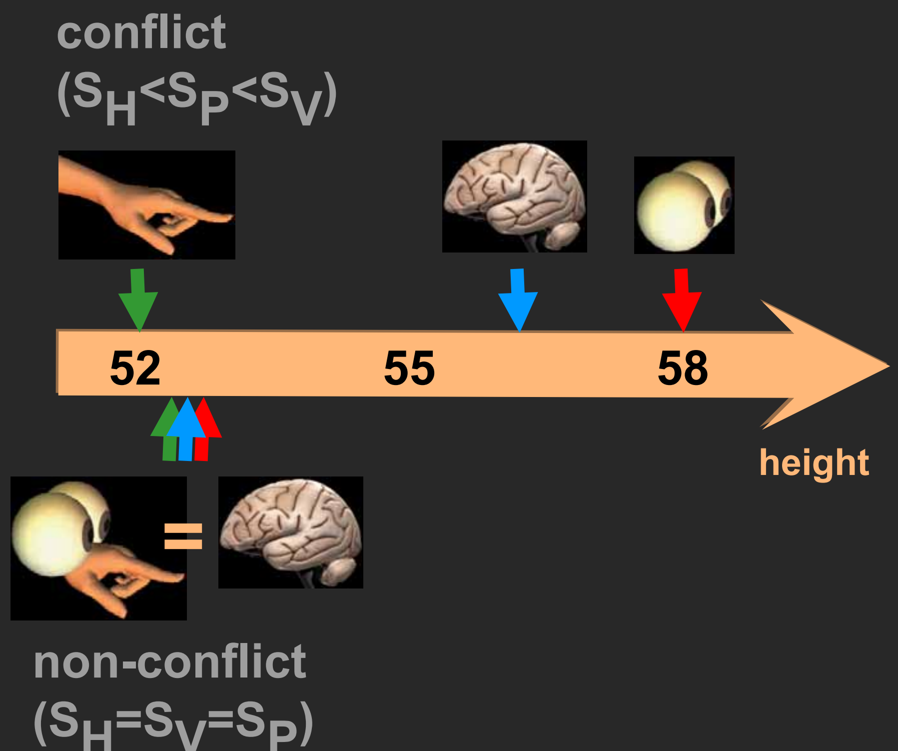


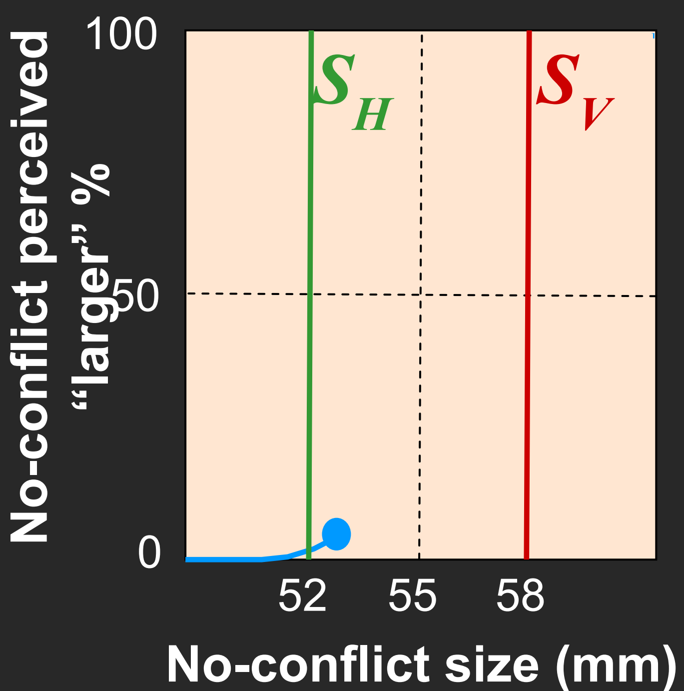
- Two Interval Forced Choice Task (2-IFC)
  - which one is bigger?
- which size do we need for the non-conflict stimulus
  to be perceptual the same size as the conflict stimulus?
- conflicts are below perceptual threshold

Combined Visual-Haptic Experiment

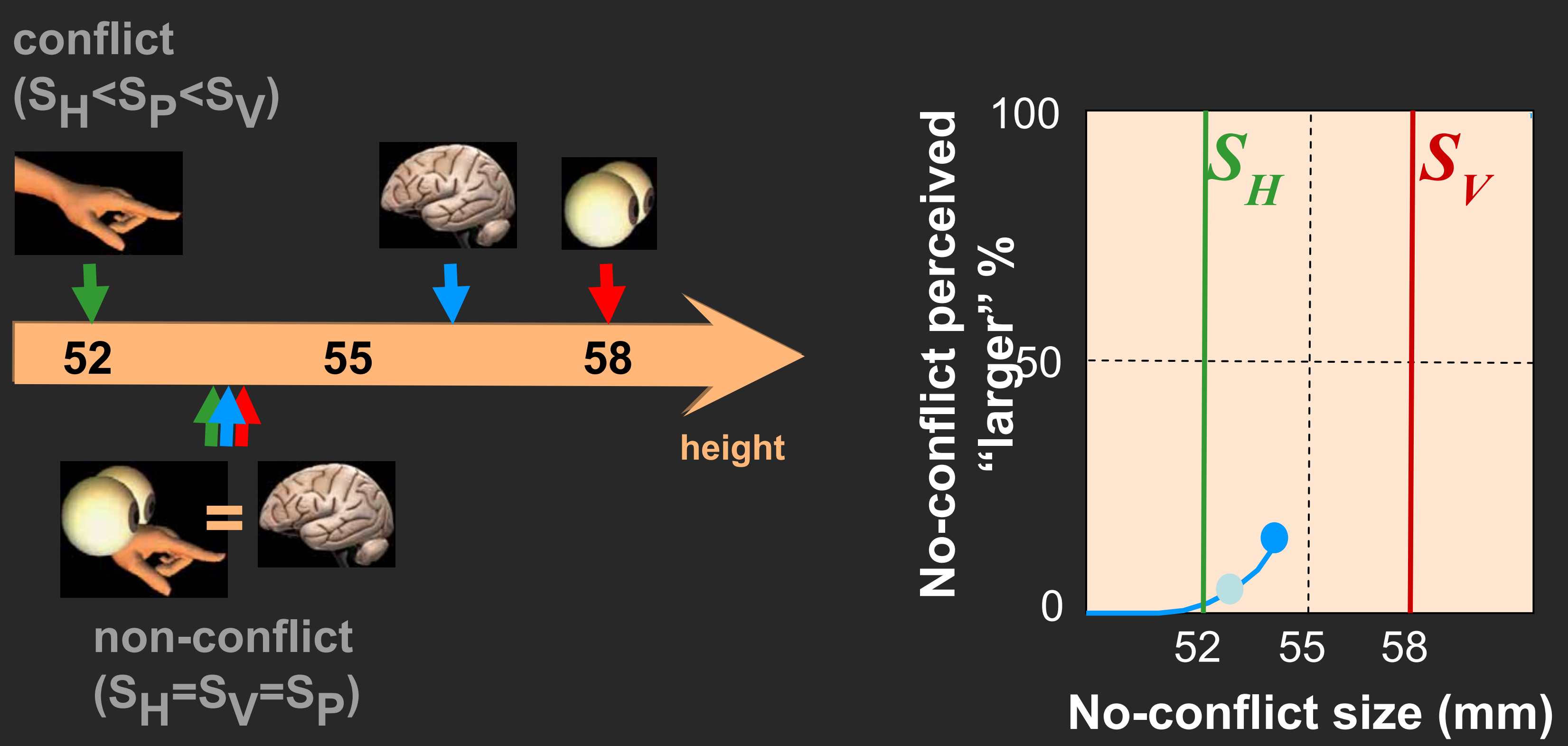


Combined Visual-Haptic Experiment

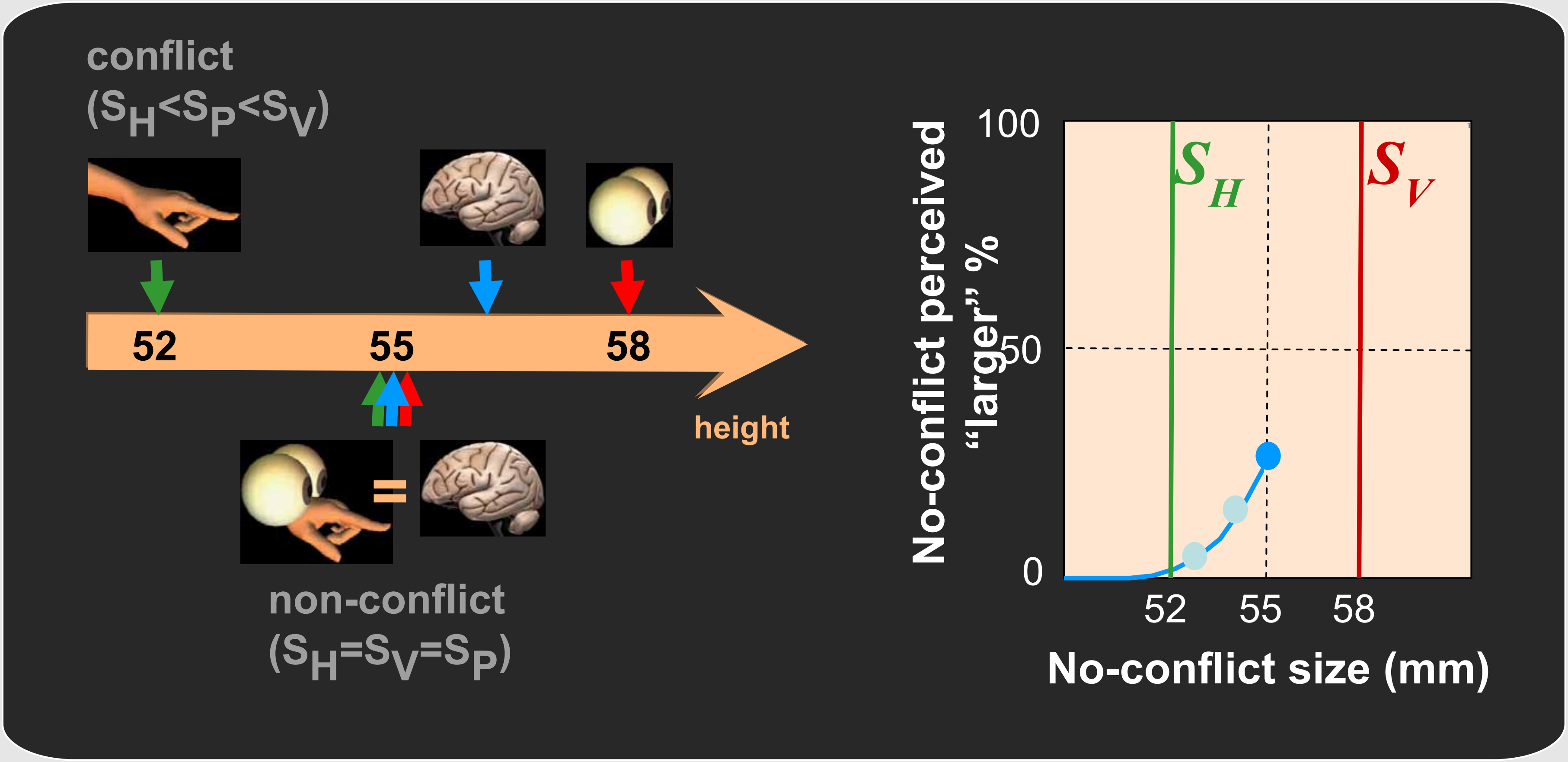




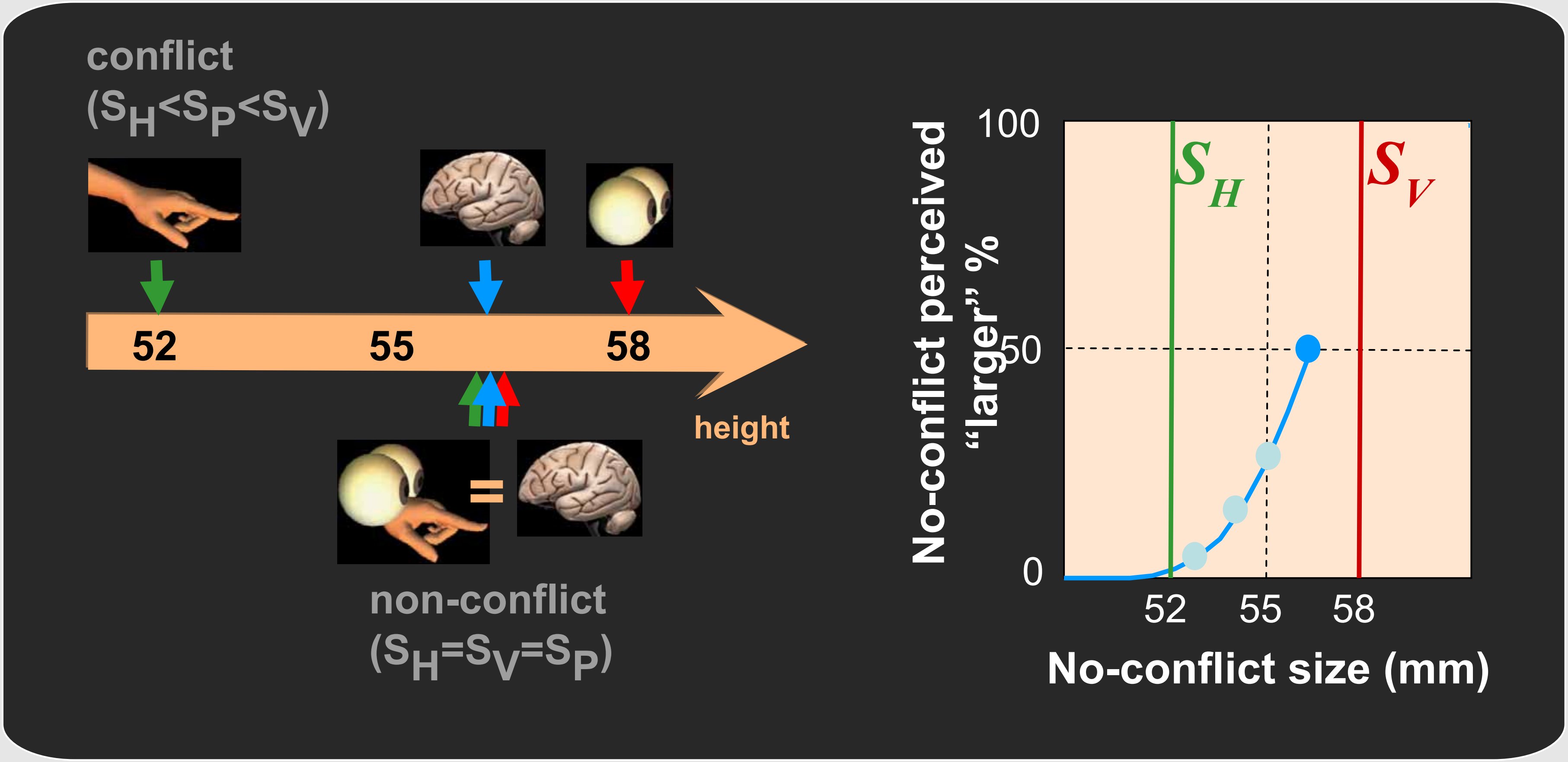
Combined Visual-Haptic Experiment



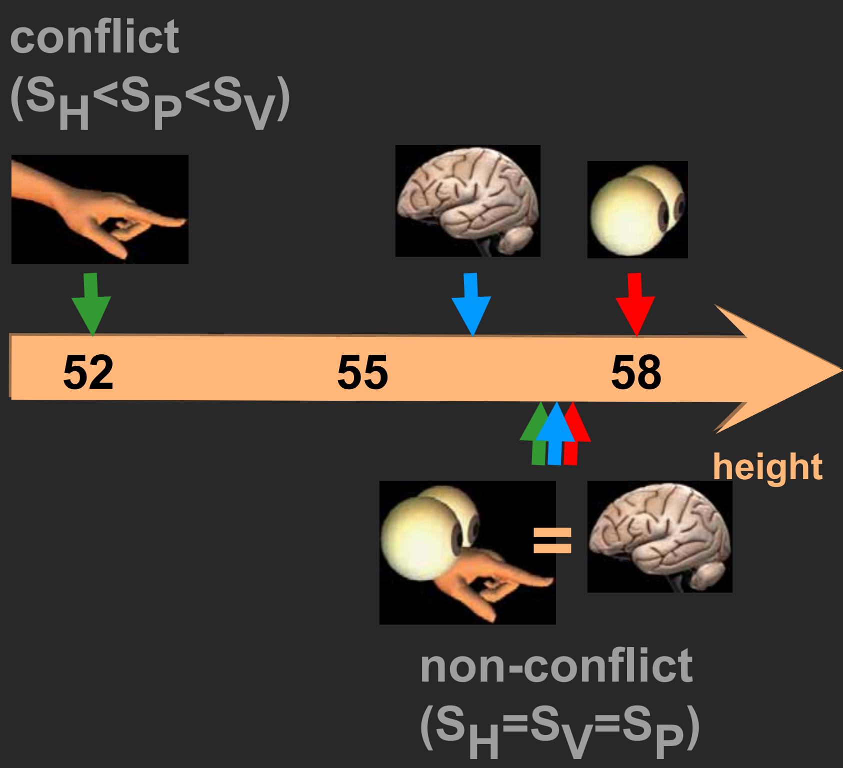
Combined Visual-Haptic Experiment

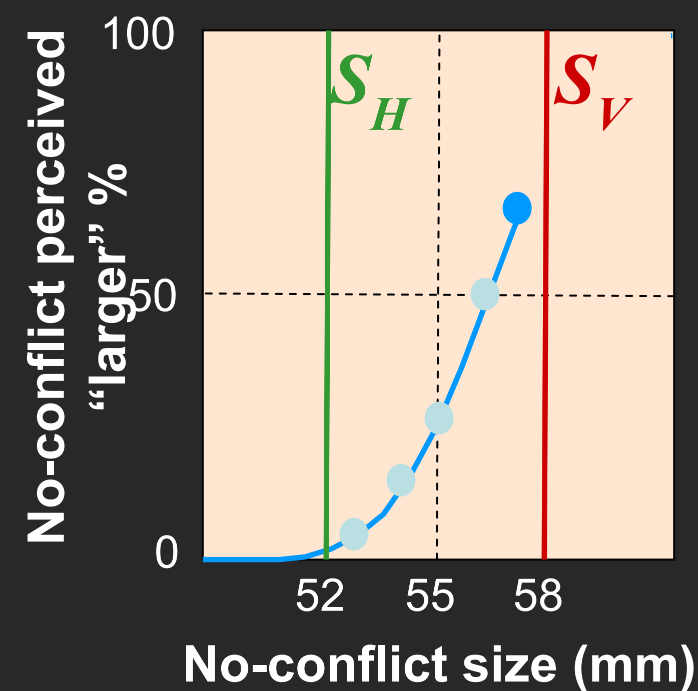


Combined Visual-Haptic Experiment

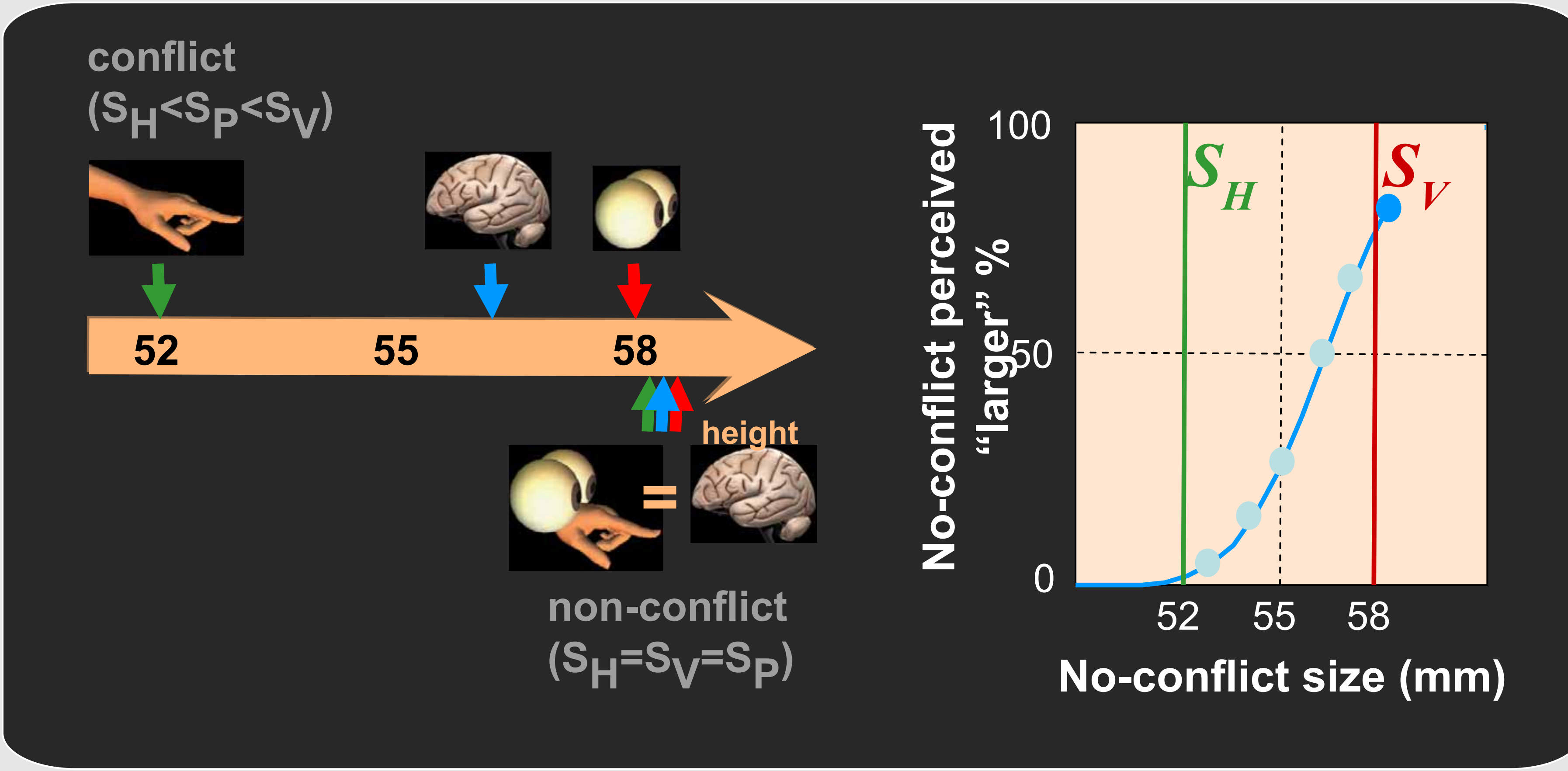


Combined Visual-Haptic Experiment

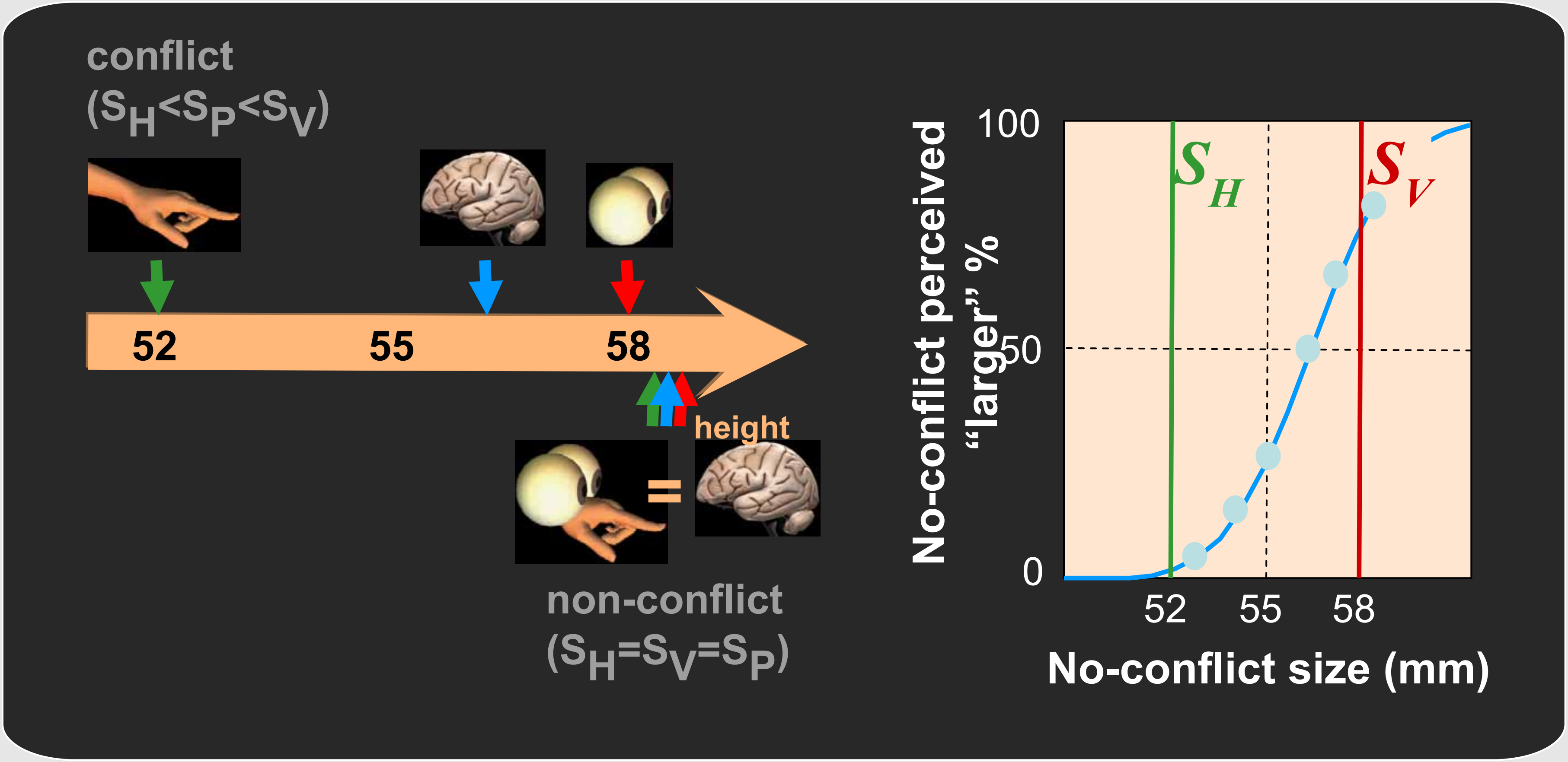




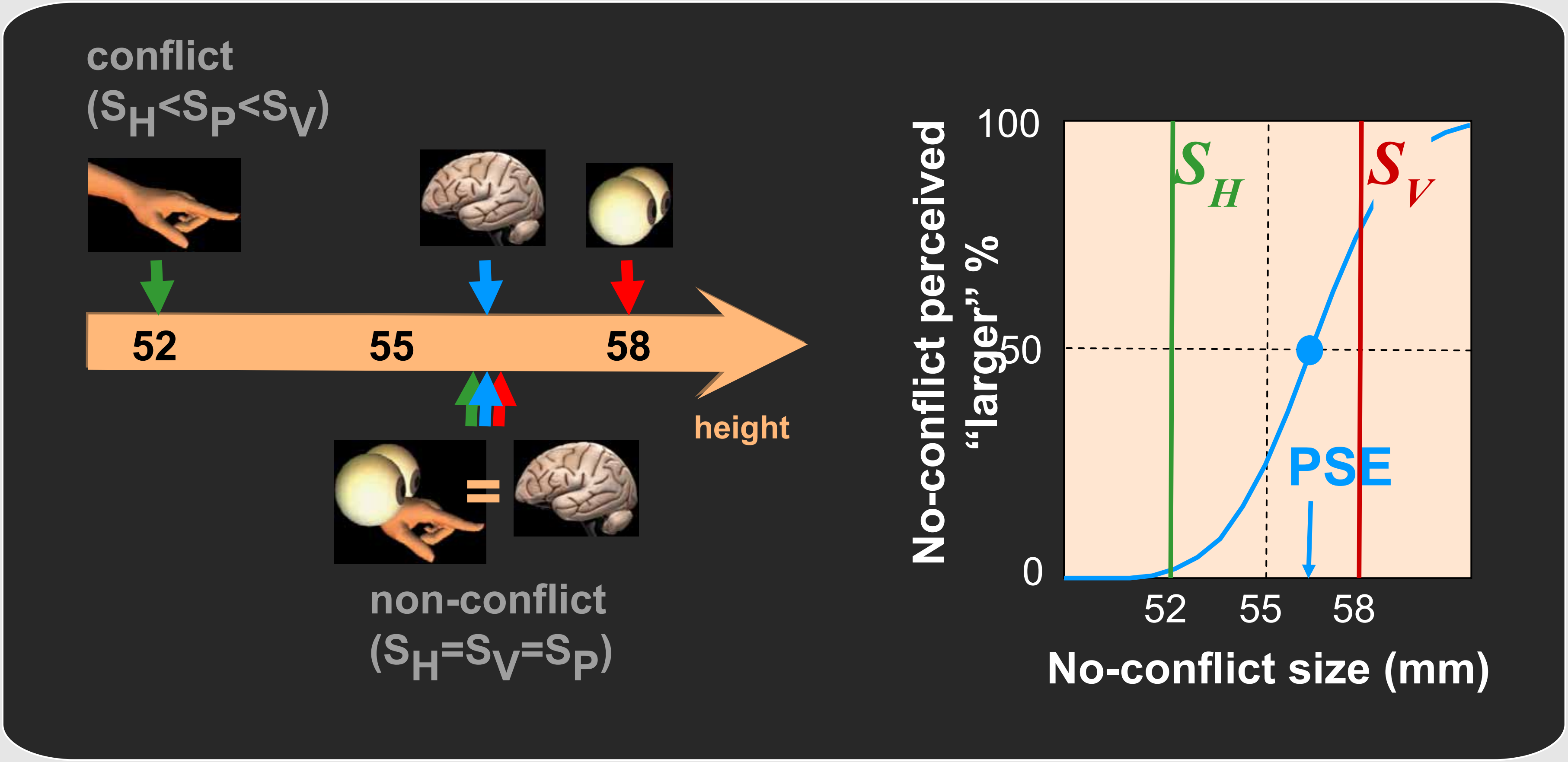
Combined Visual-Haptic Experiment



Combined Visual-Haptic Experiment



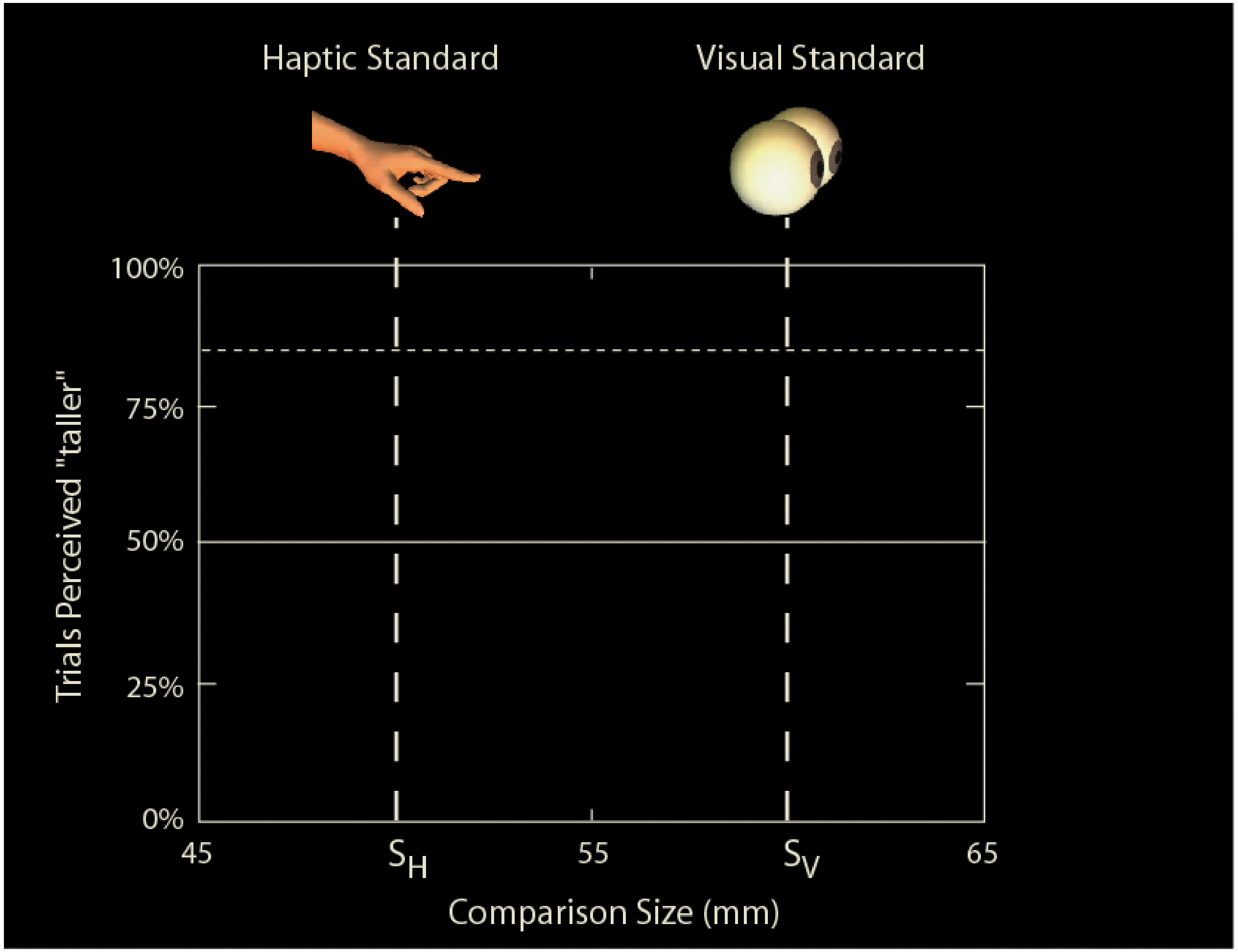
Combined Visual-Haptic Experiment



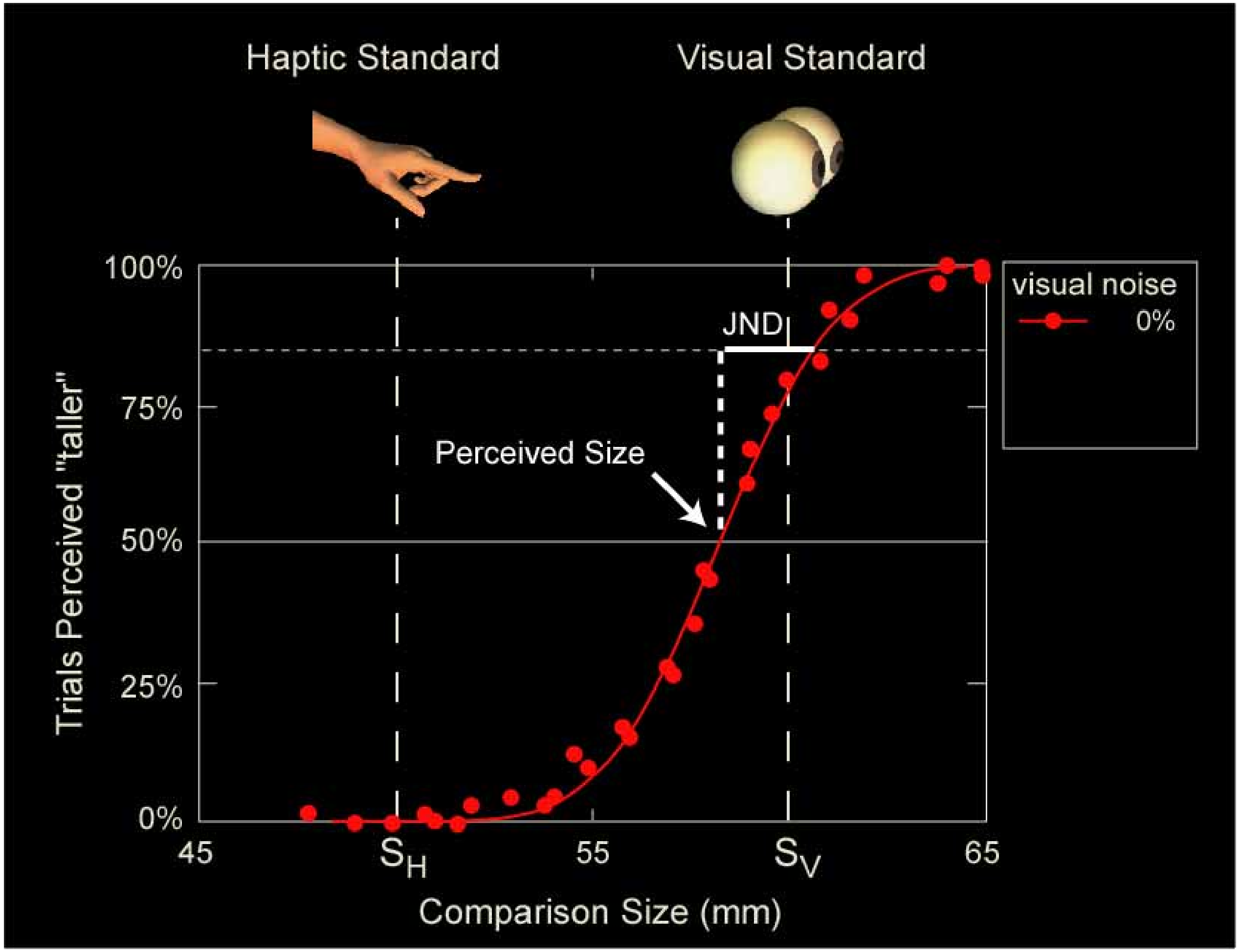
Point of subjective equality (PSE):

Value of no-conflict stimulus perceived as same size as conflict stimulus.

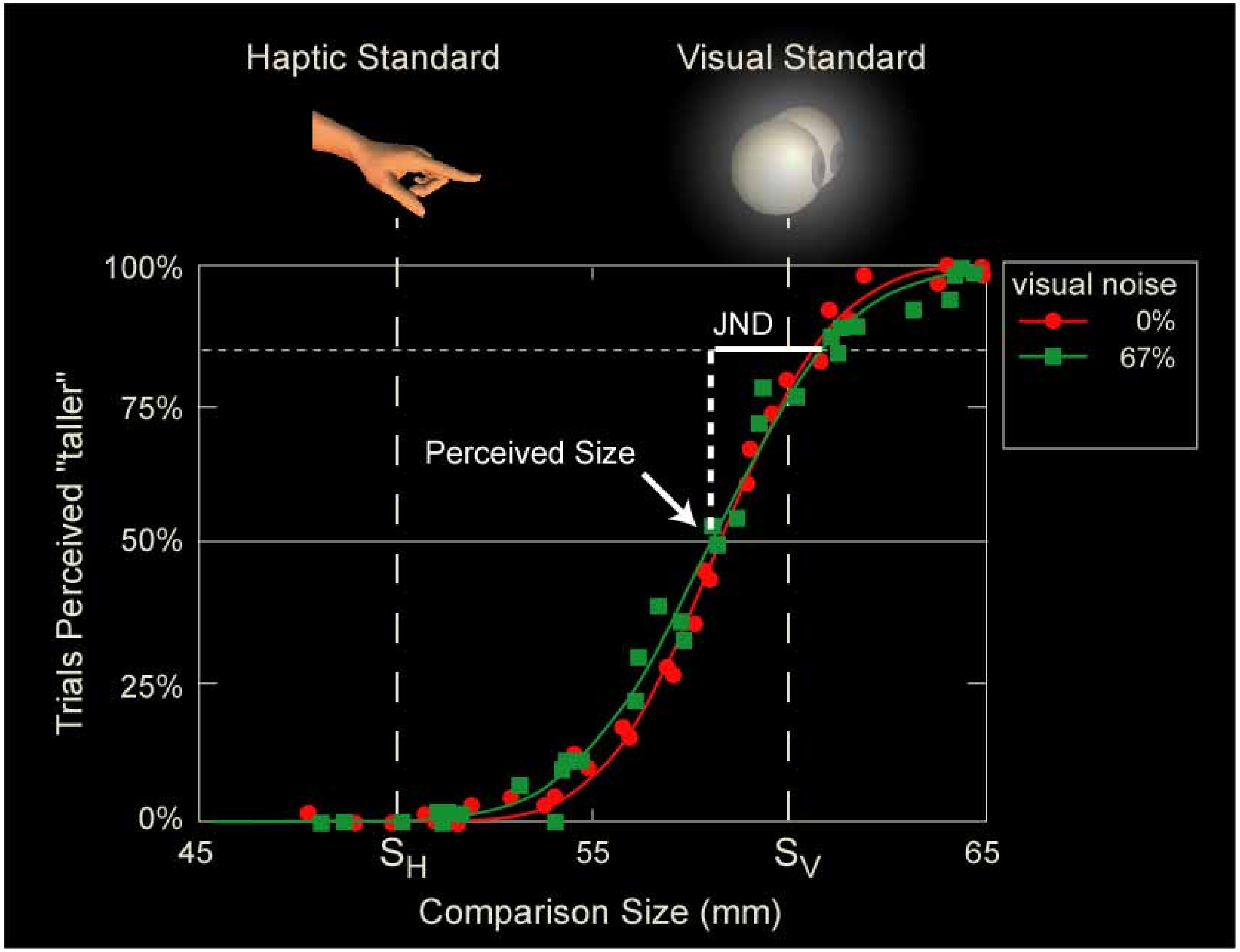
Visual-Haptic Integration



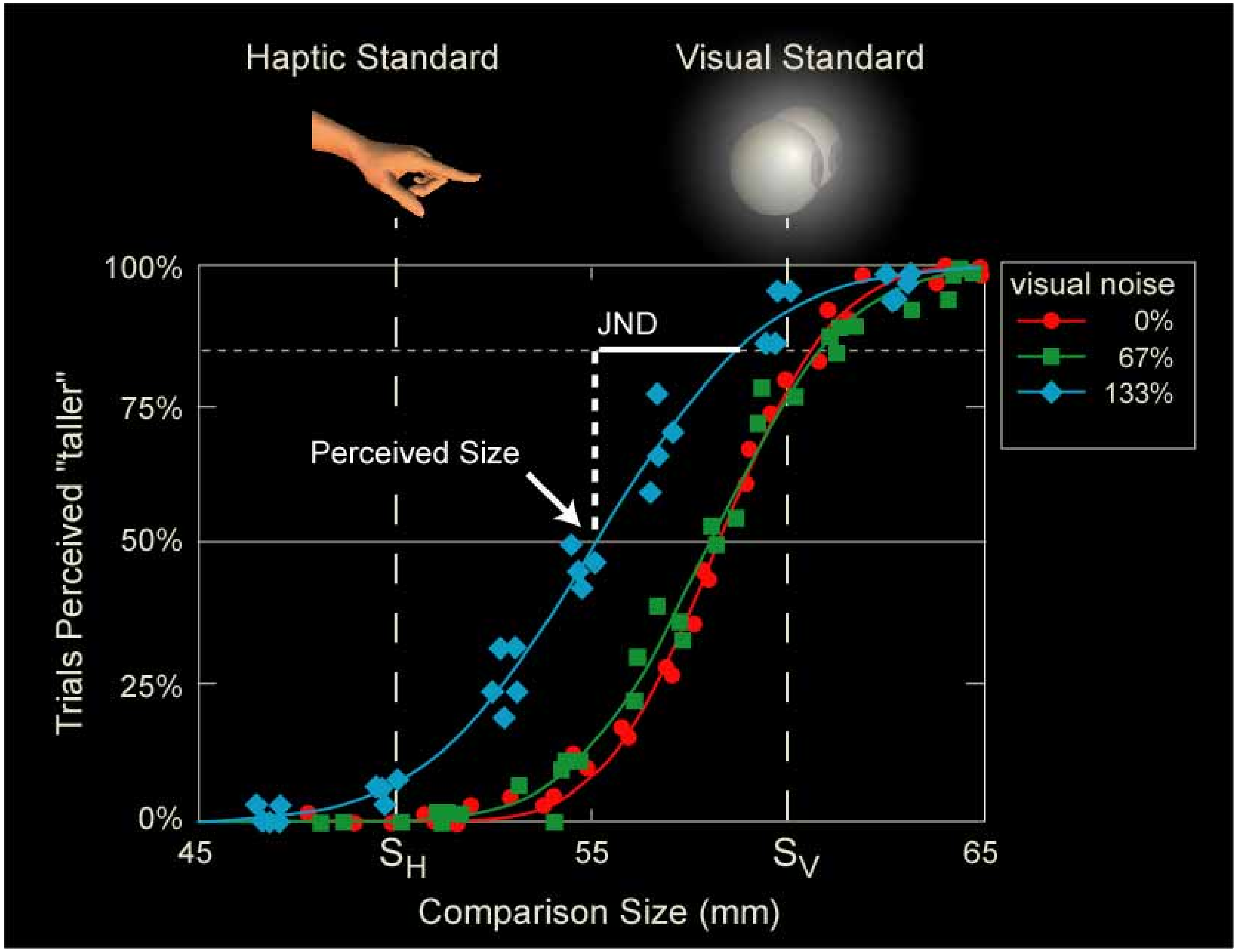
Visual-Haptic Integration



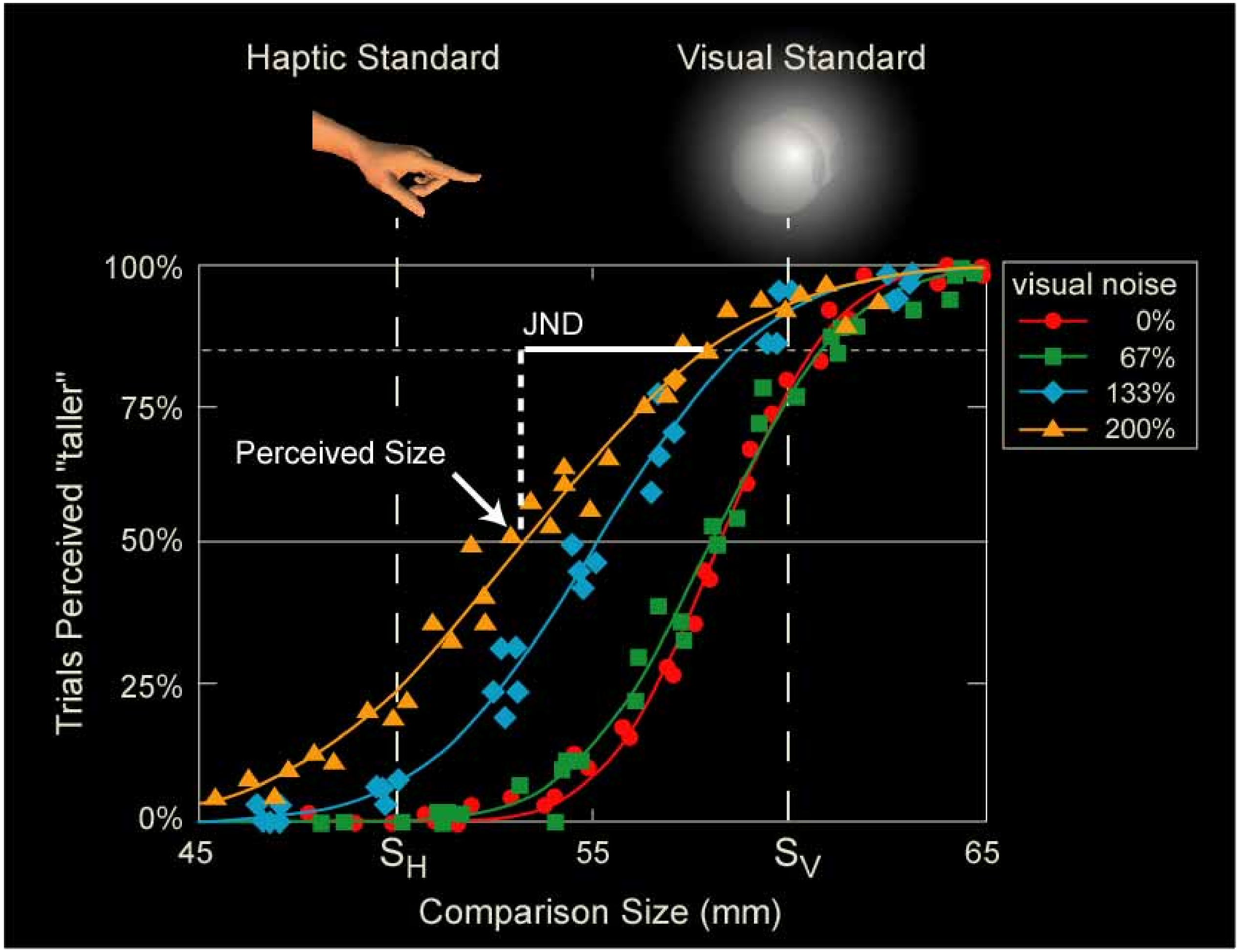
Visual-Haptic Integration



Visual-Haptic Integration



Visual-Haptic Integration

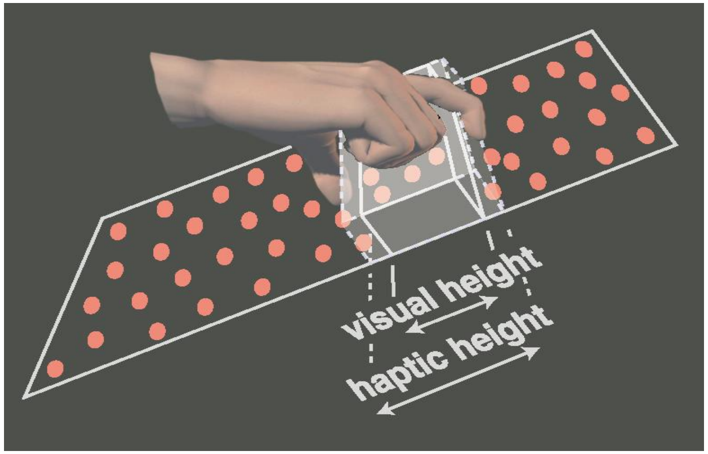


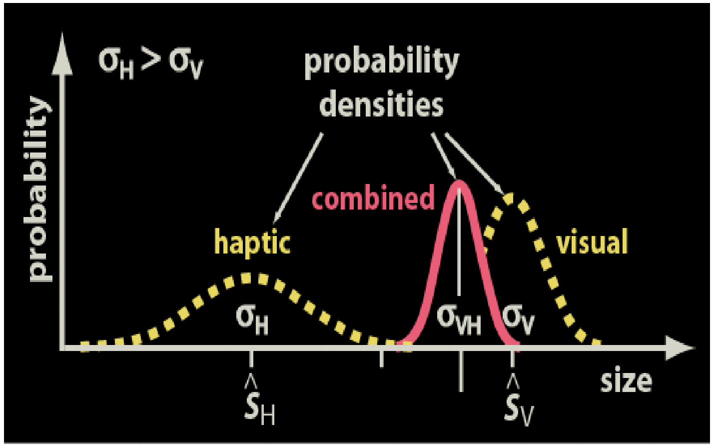
Summary
Visual-Haptic Integration

- the brain combines visual and haptic information in a statistically optimal way
  - Ernst, Banks & Bülthoff *Nature Neuroscience* 3 (1), 69-73 (2000)
  - Ernst & Banks *Nature* 415, 429-433 (2002)
  - Hillis, Ernst, Banks & Landy *Science* 298, 1627-1630 (2002)
  - Ernst & Bülthoff, *TICS* 8, 162-169 (2004)
- cues are weighted according to their reliability (variance)

$$\hat{S}_{VH} = w_V \hat{S}_V + w_H \hat{S}_H$$

- combination reduces variance
- explains "visual capture"
  - **the variance of visual size estimates is much smaller than haptic estimate**
  - **visual weight set to ~ 1.0**





Visual-Auditory Integration

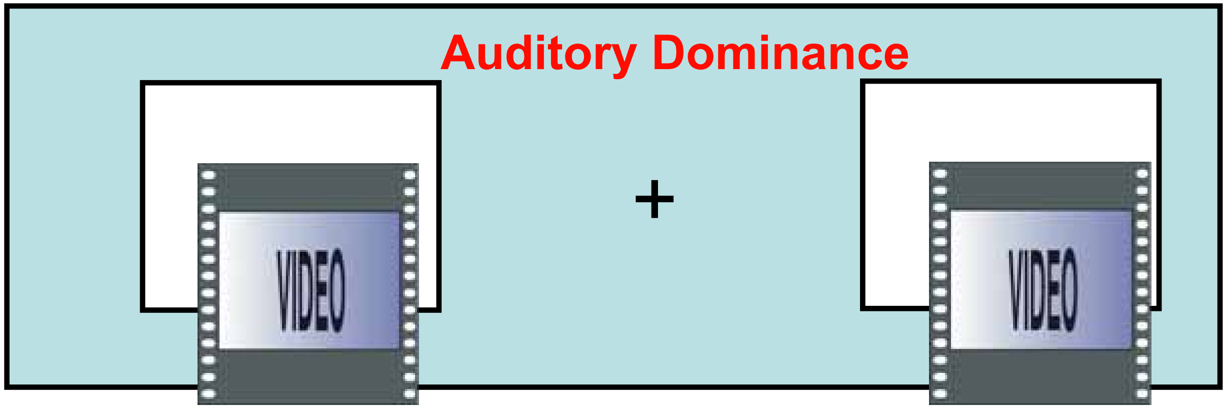
- Visual-Auditory Localization: Ventriloquist effect

Visual Dominance

The ventriloquist effect results from near-optimal bimodal integration
D. Alais & D. Burr, Curr. Biol. 2004

Edgar Bergen & Charlie McCarthy

- Visual-Auditory Temporal Judgments



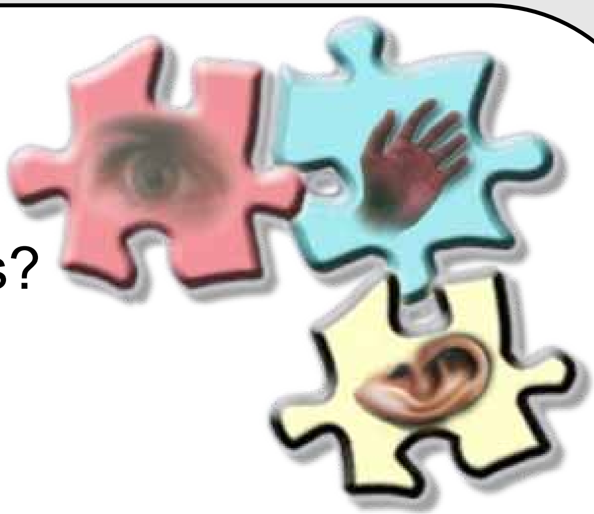
What you see is what you hear.
L. Shams, Y. Kamitani & S. Shimojo, Nature 2000

What you see is what you hear

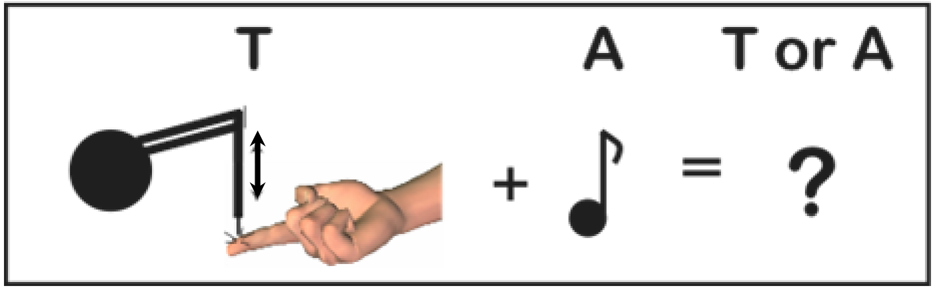
1

2

Integration of Temporal Events

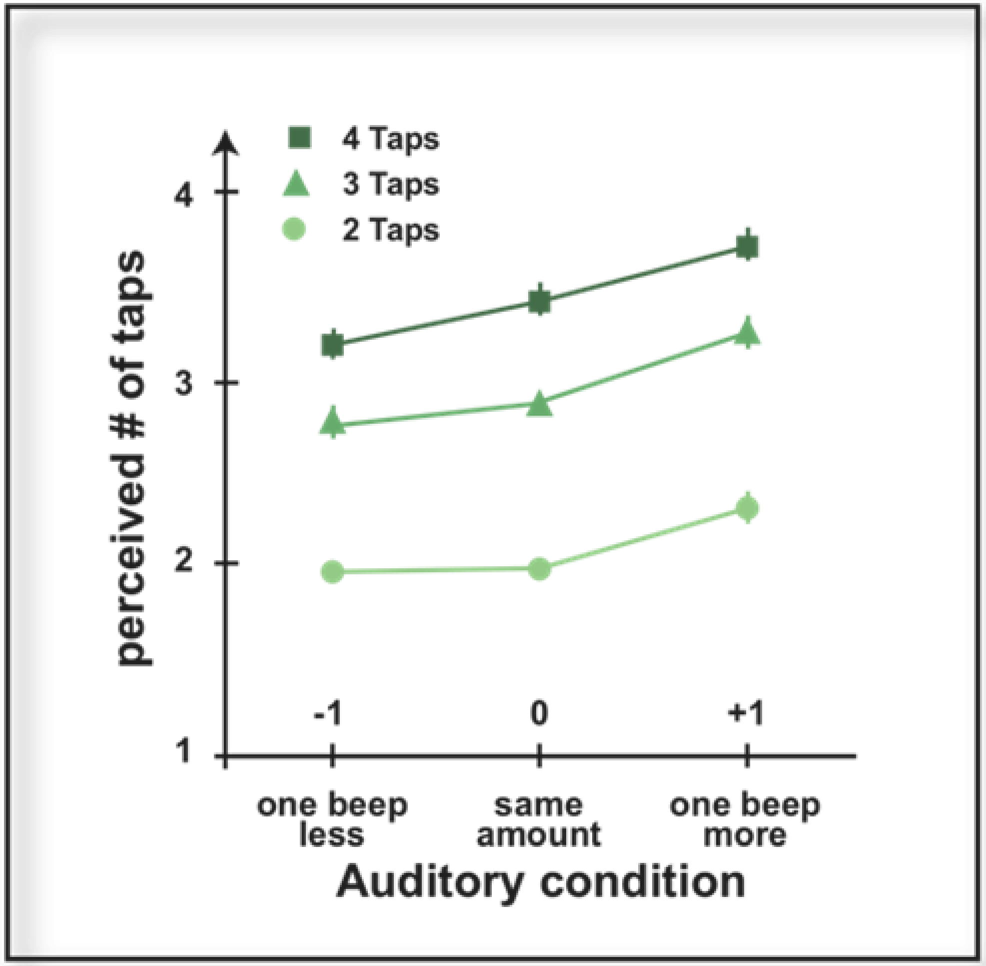


Is the **integration of temporal events** based on the **reliability of the signals**, similar to the integration of spatial signals?

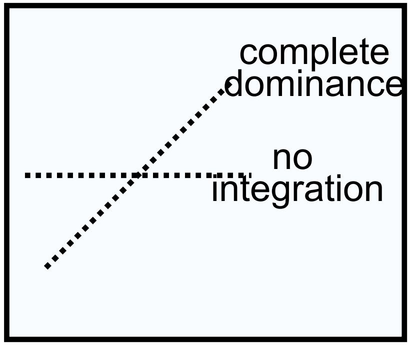


Instead of using vision and audition Ernst and Bresciani used **touch and audition**, because they are likely to be more comparable in terms of their reliabilities in processing temporal events.

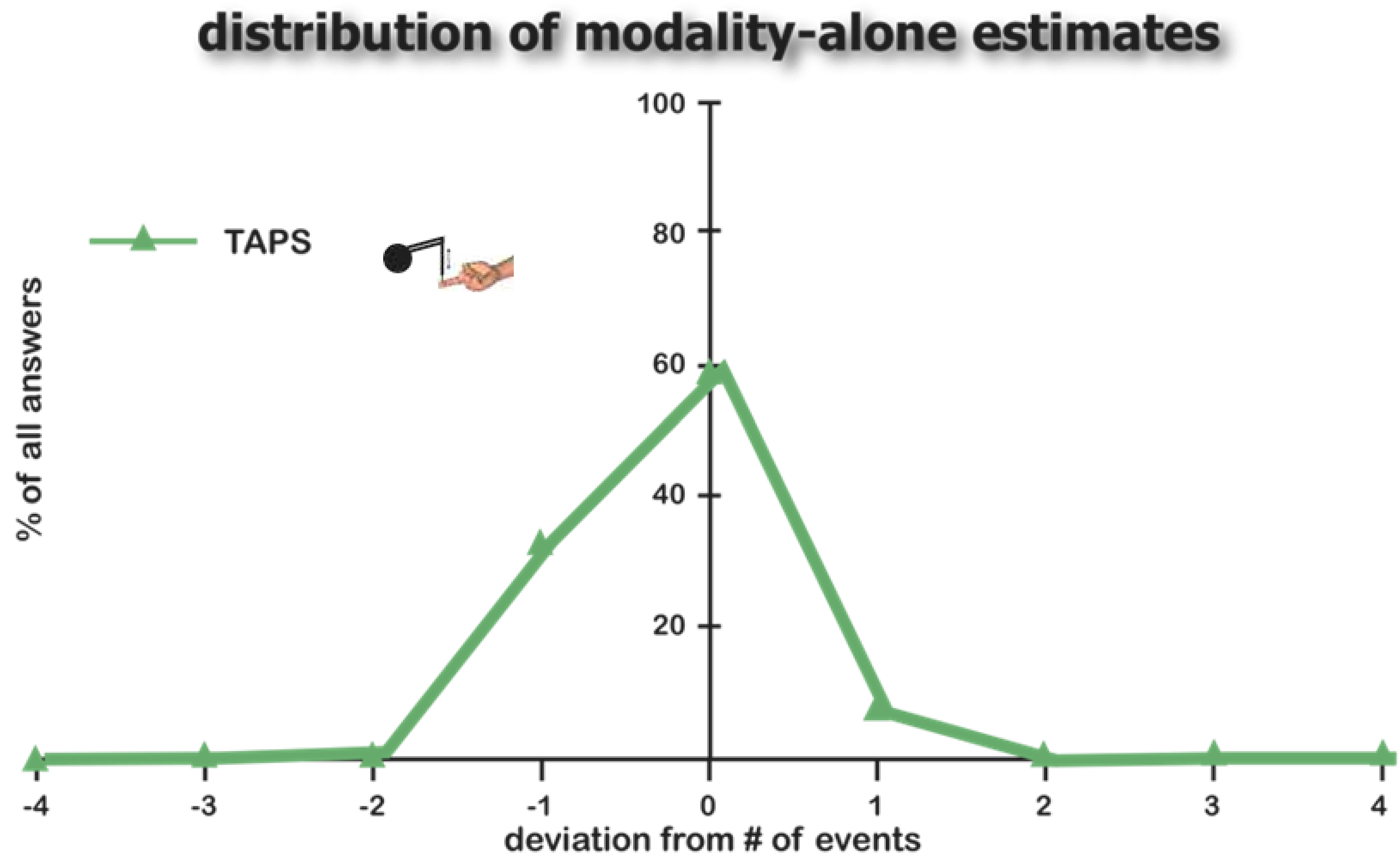
Auditory-Haptic Integration



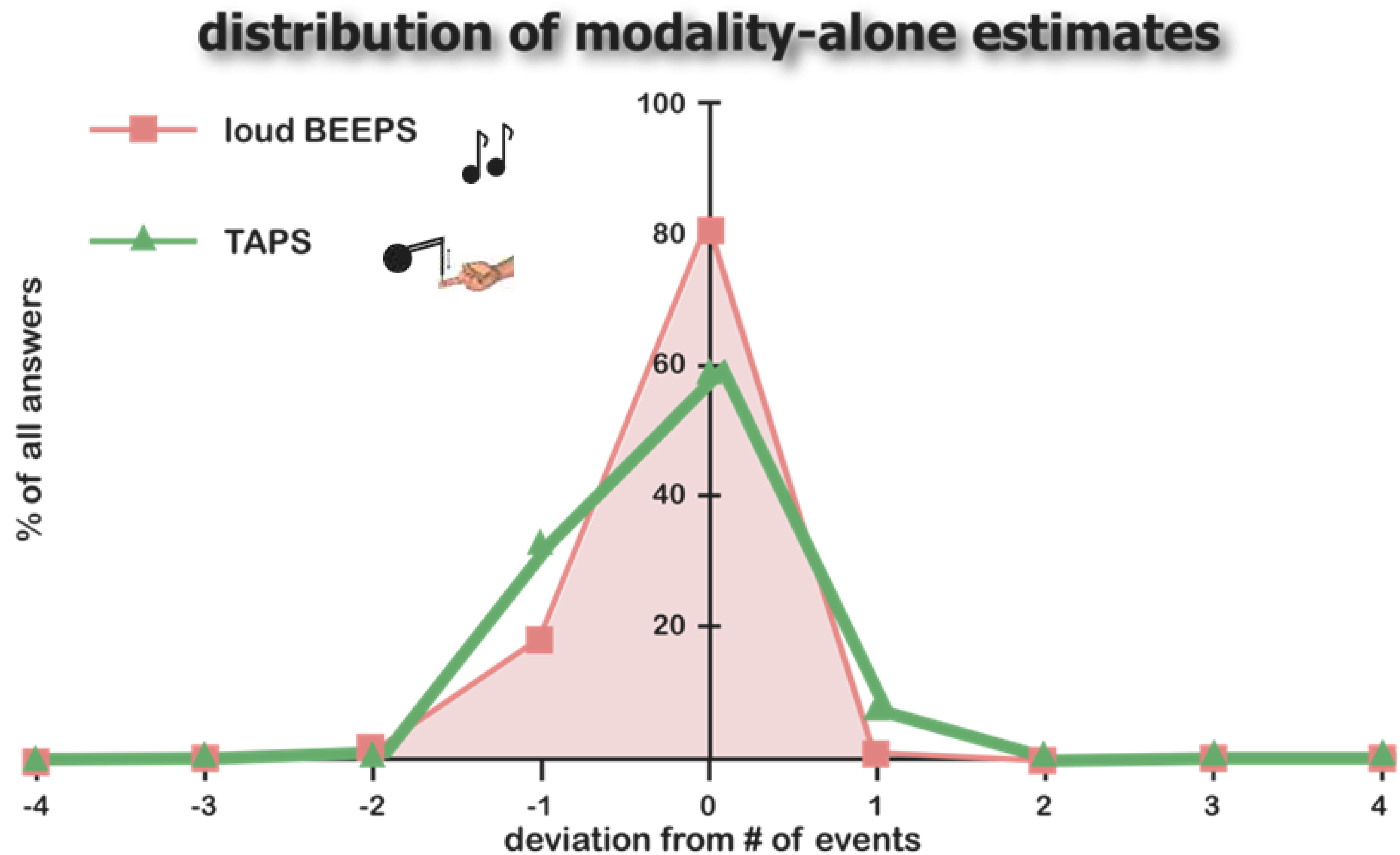
Predictions



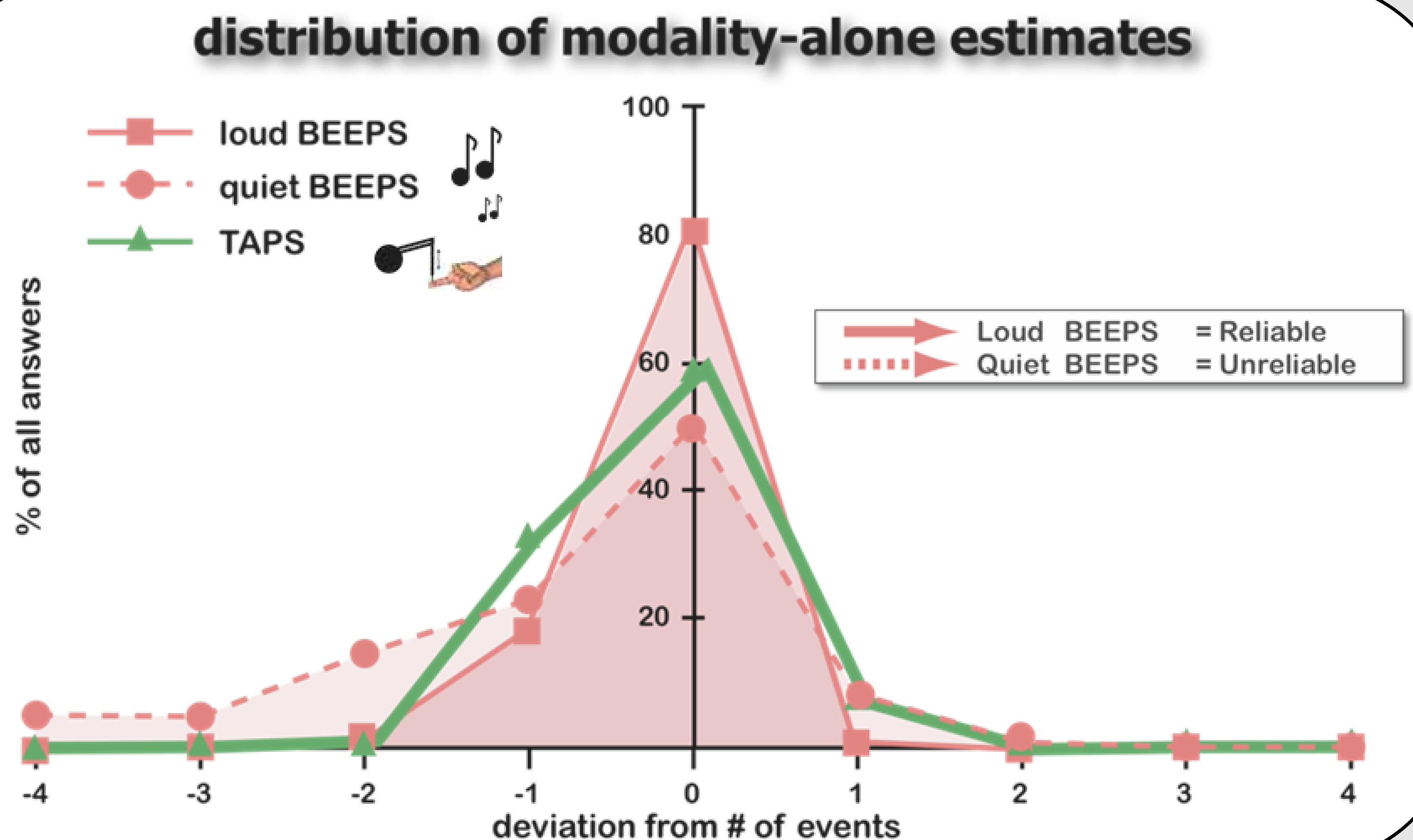
Signal Reliability



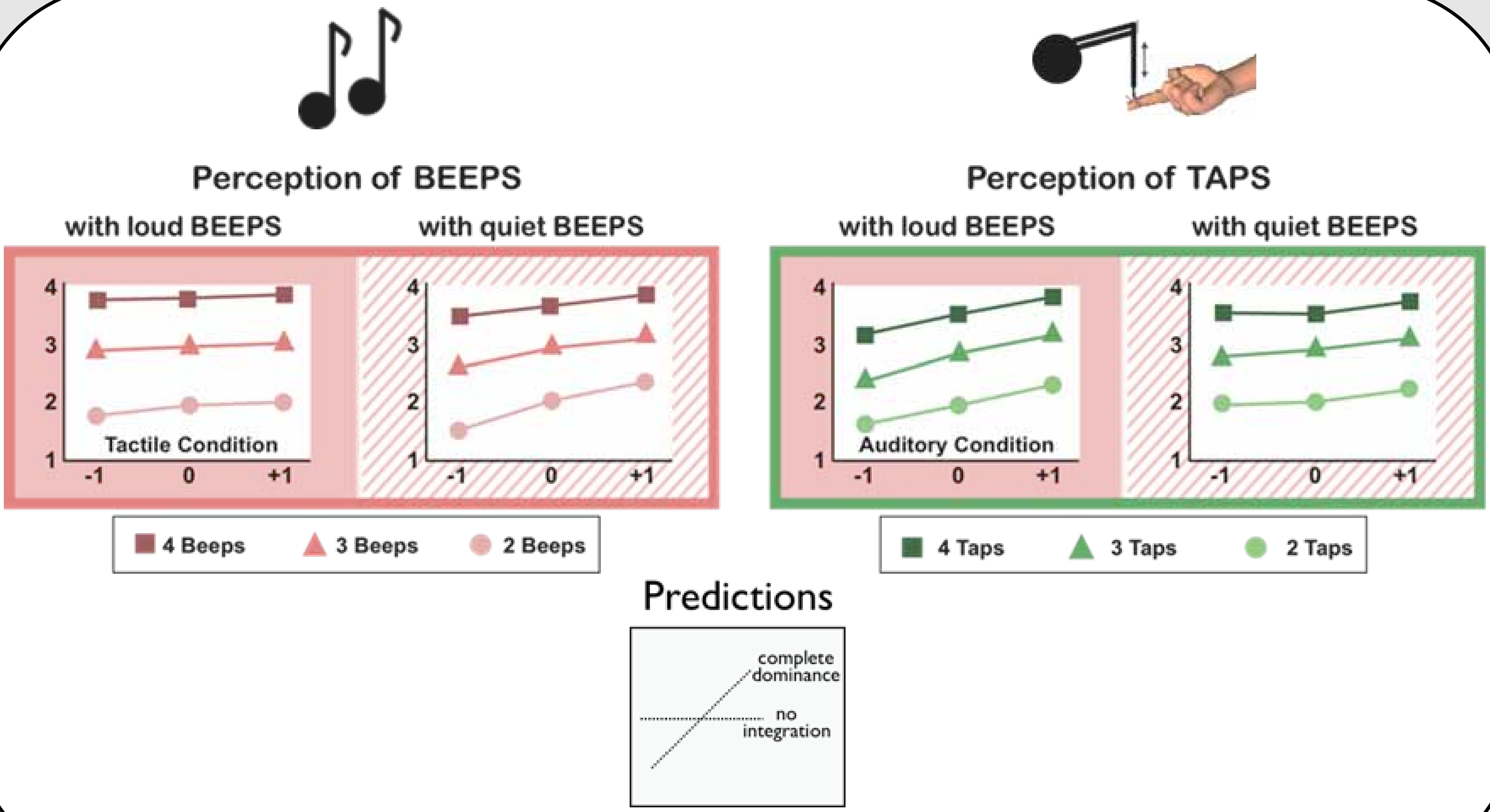
Signal Reliability



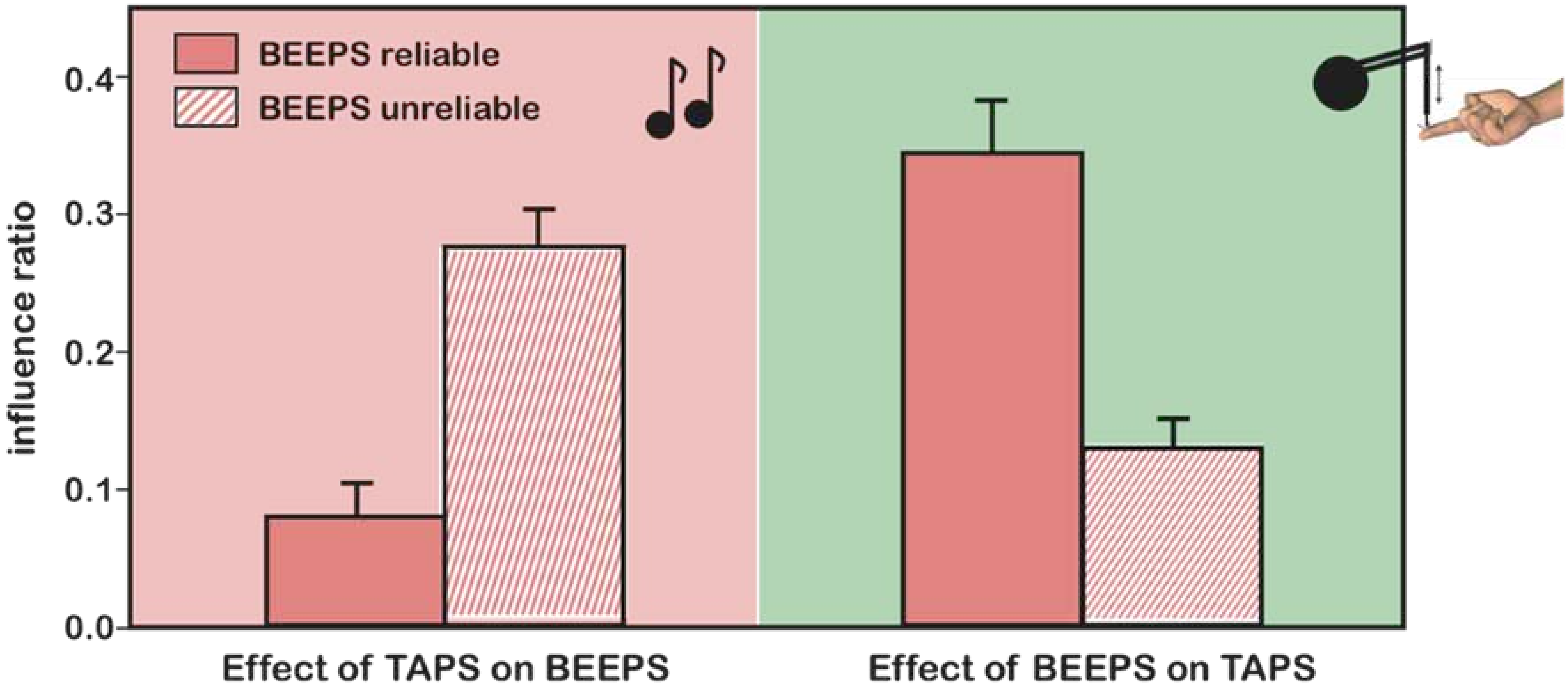
Signal Reliability



Effect of Reliability on Integration



Effect of Reliability on Integration

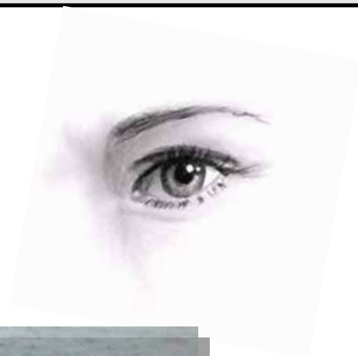


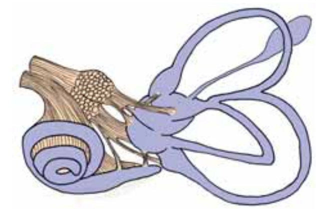
Conclusion: Integration of Temporal Events

- Tactile-auditory integration for sequences of temporal events
- The reliability of the signals determines the strength of bias

Marc Ernst Jean-Pierre Bresciani Abderrahmane Kheddarr
MPI Tübingen LSC Paris

Visual-Vestibular Integration

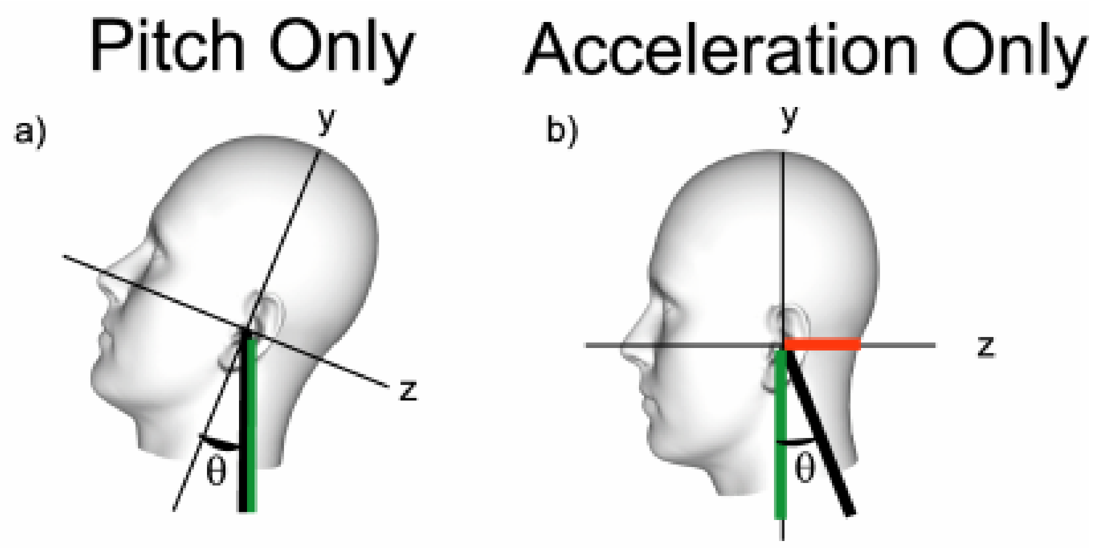




strong acceleration + low visibility

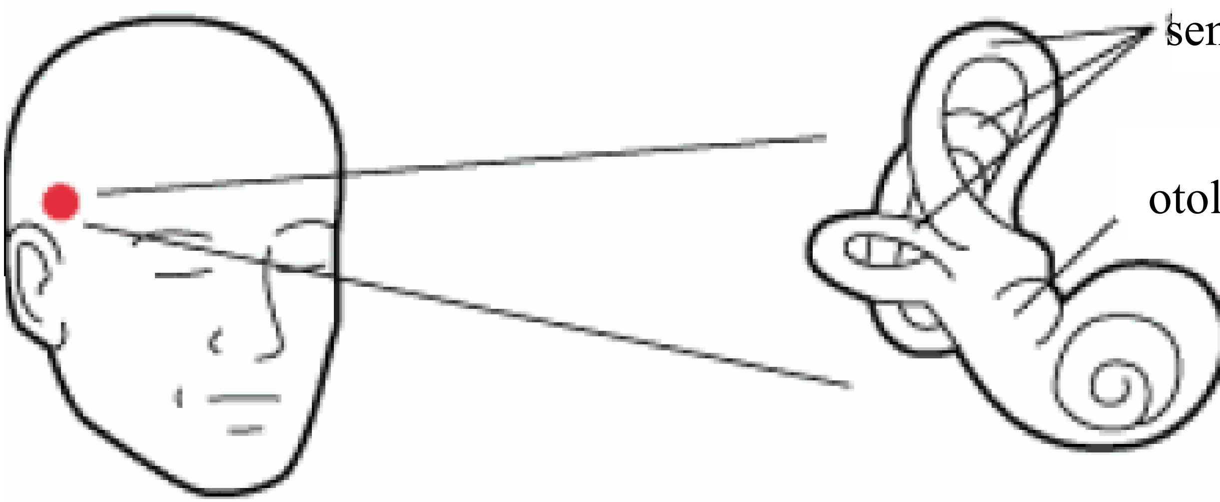
somatogravic illusion

Somatogravic Flight Illusion



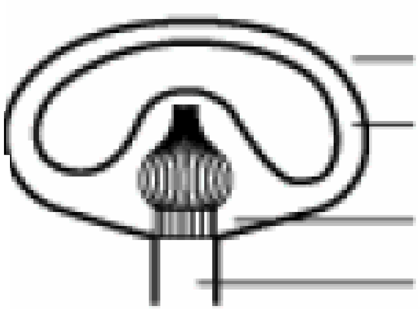
- Einstein:
  no sensor can distinguish gravity from translational acceleration.
- This holds of course for our otolith organs too
  - they sense gravity *minus* acceleration
  - they cannot distinguish certain tilts and motions

Vestibular System

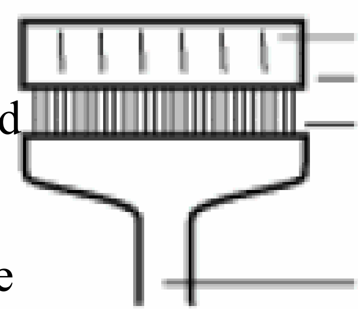


from Greek:
otos - ear
lithos - stone

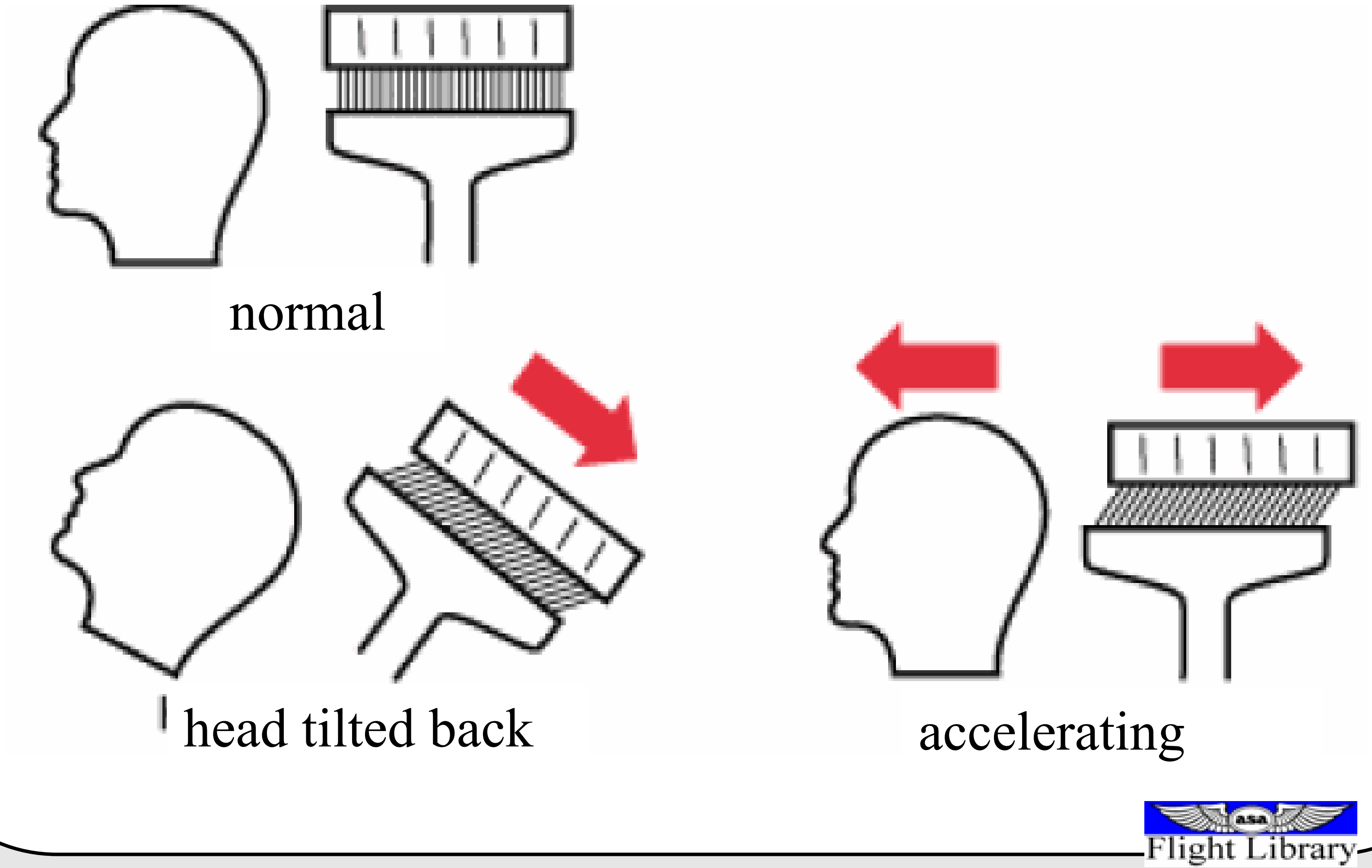
semicircular canal



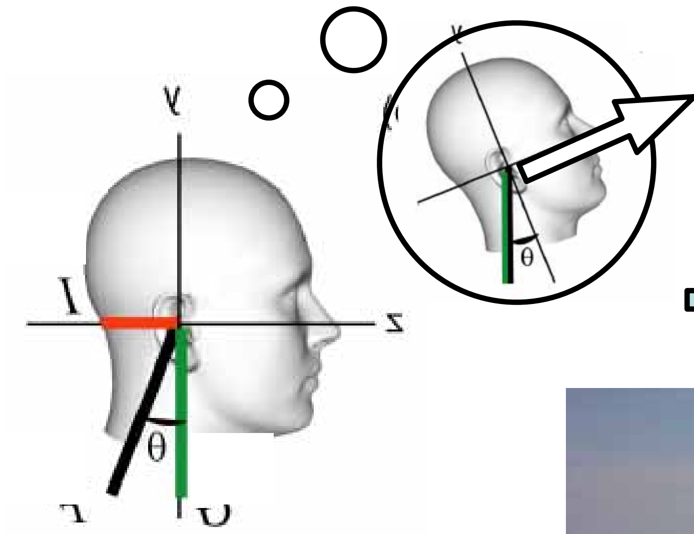
otolith organ



Otolithic System



Consequence
of somatogravic illusion



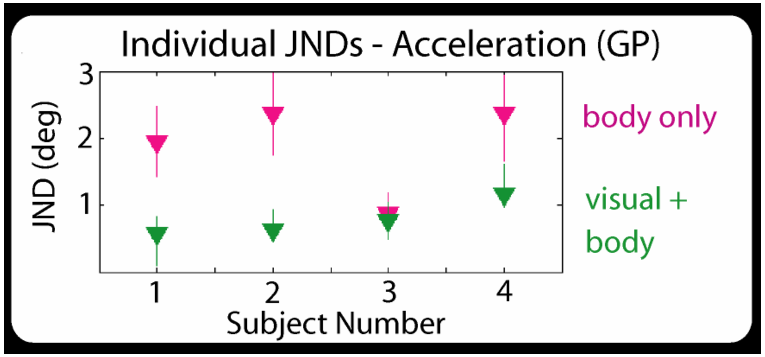
Consequence

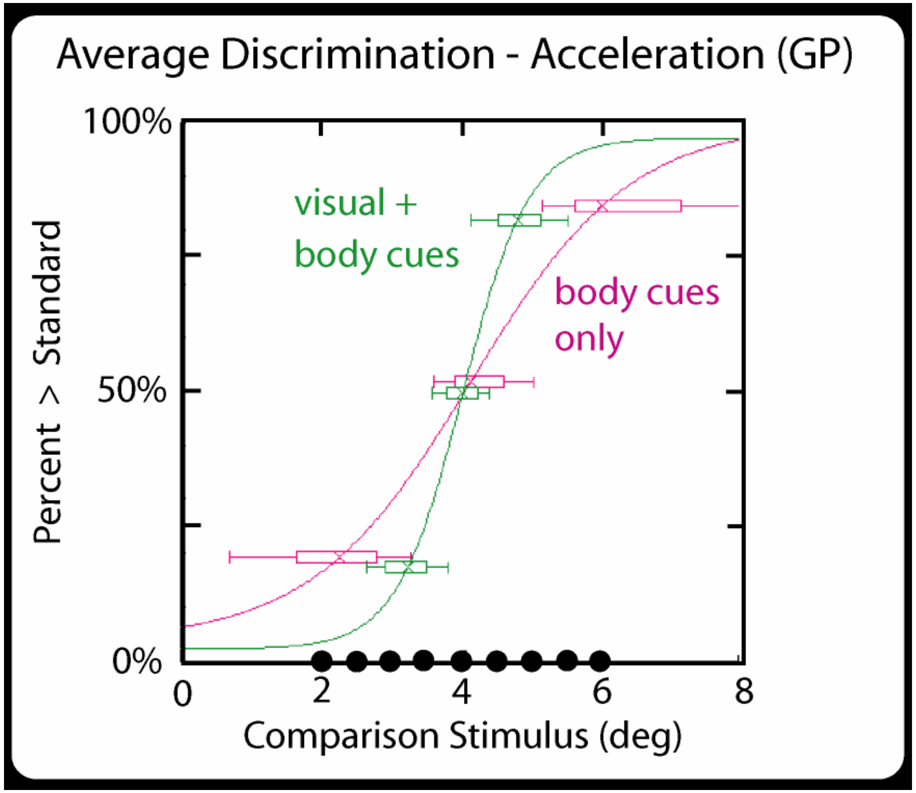
Pogen MacNeilage,
Daniel Berger,
Marty Banks,
Heinrich Bülthoff

Multimodal Integration: visual – body cues

MacNeilage, Berger, Banks & Bülthoff, VSS 2004

- Which cues are best to override the **strong but wrong cues** from the vestibular system?
- Examine influence of **visual** forward-movement on **perceived tilt** during real tilt





Implications for Cockpit Design

- artificial horizons are not always sufficient to overcome the somatogravic illusion
- we need an attitude indicator which overrides the somatogravic illusion
- **visual and auditory capture**
  might be the answer how to improve
  flight safety in low visibility conditions

High-Level Integration

Attention and Awareness

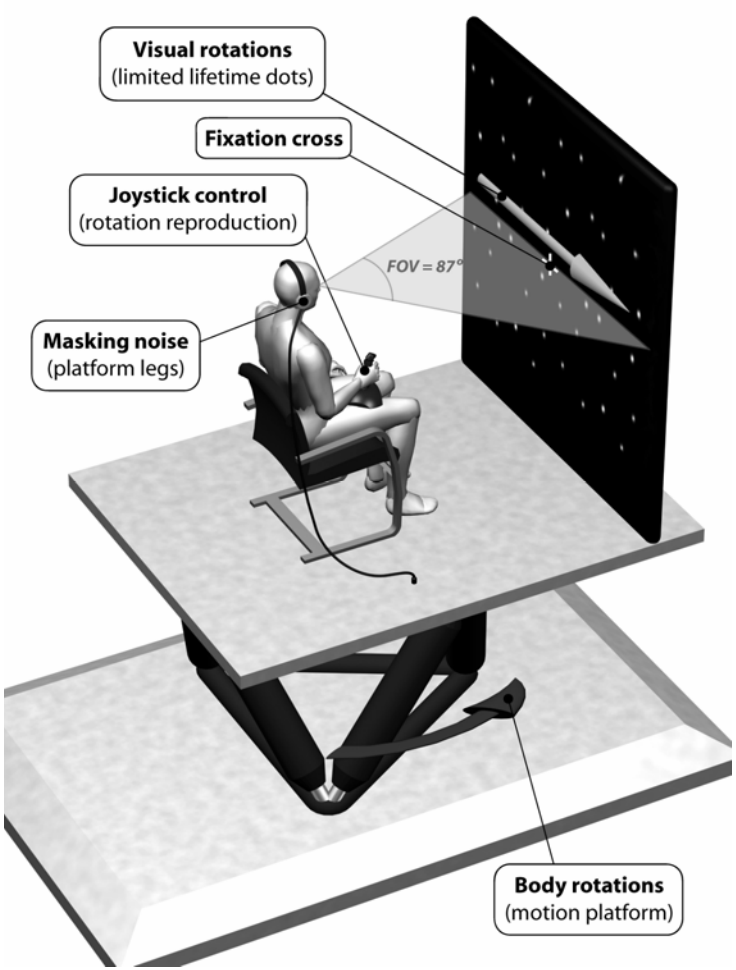
- Perception of self-motion is **multimodal**:
  - Visual modality
  - Vestibular modality
  - Other body senses

    Pilots: seat-of-the-pants feeling

- How are the different cues combined for the perception of self-motion?
- Does integration change if we **attend** to one cue or become **aware** of conflicts between cues?

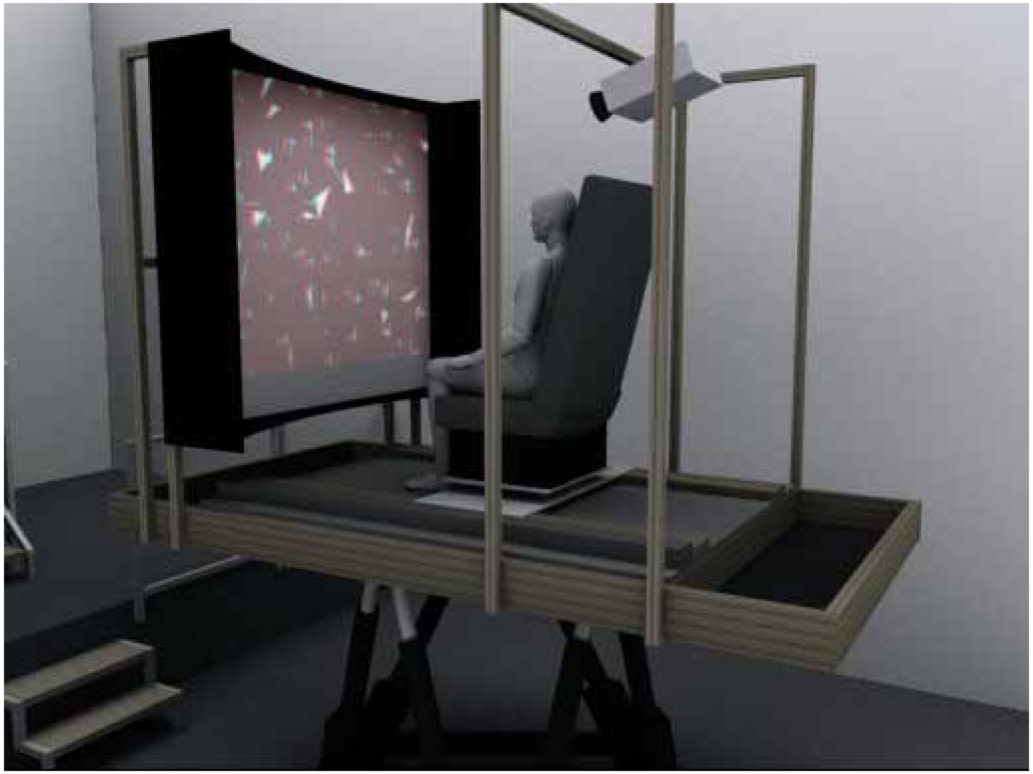
Daniel Berger, PhD thesis 2005

Multi-sensory Integration in the motion lab



Passive Rotation

scene and body



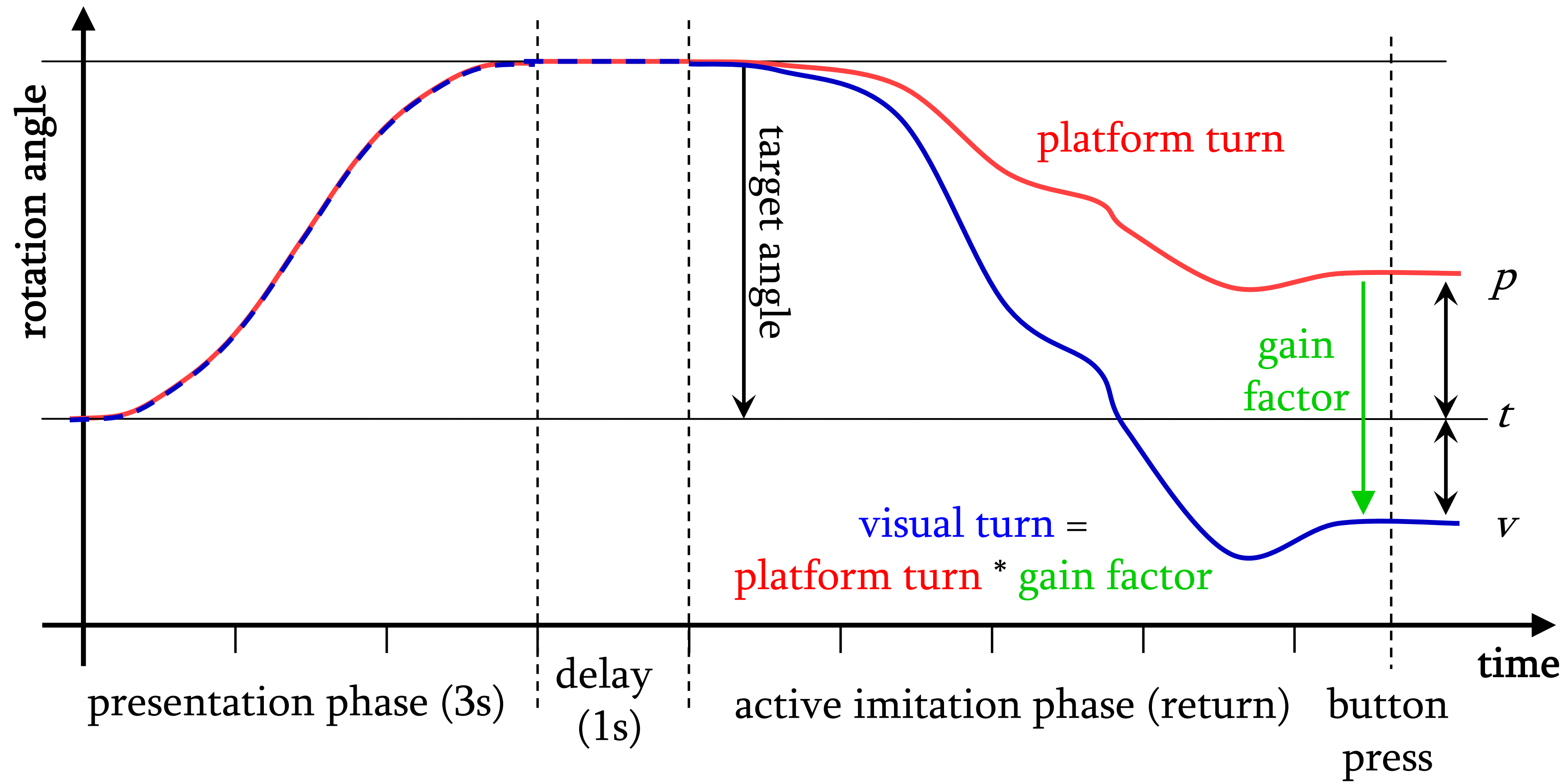
Active return to start position

gain between scene and body rotation

Cue conflicts between visual and body cues

- **Task**: Return an upright (yaw) rotation which has been presented with **both visual and body motion**
- **Trick**: During active return, a gain factor is introduced between the modalities
- **Report**: After each return, the participant has to indicate whether a **cue conflict was noticed** or not
  - Participants were told explicitly about possible cue conflicts

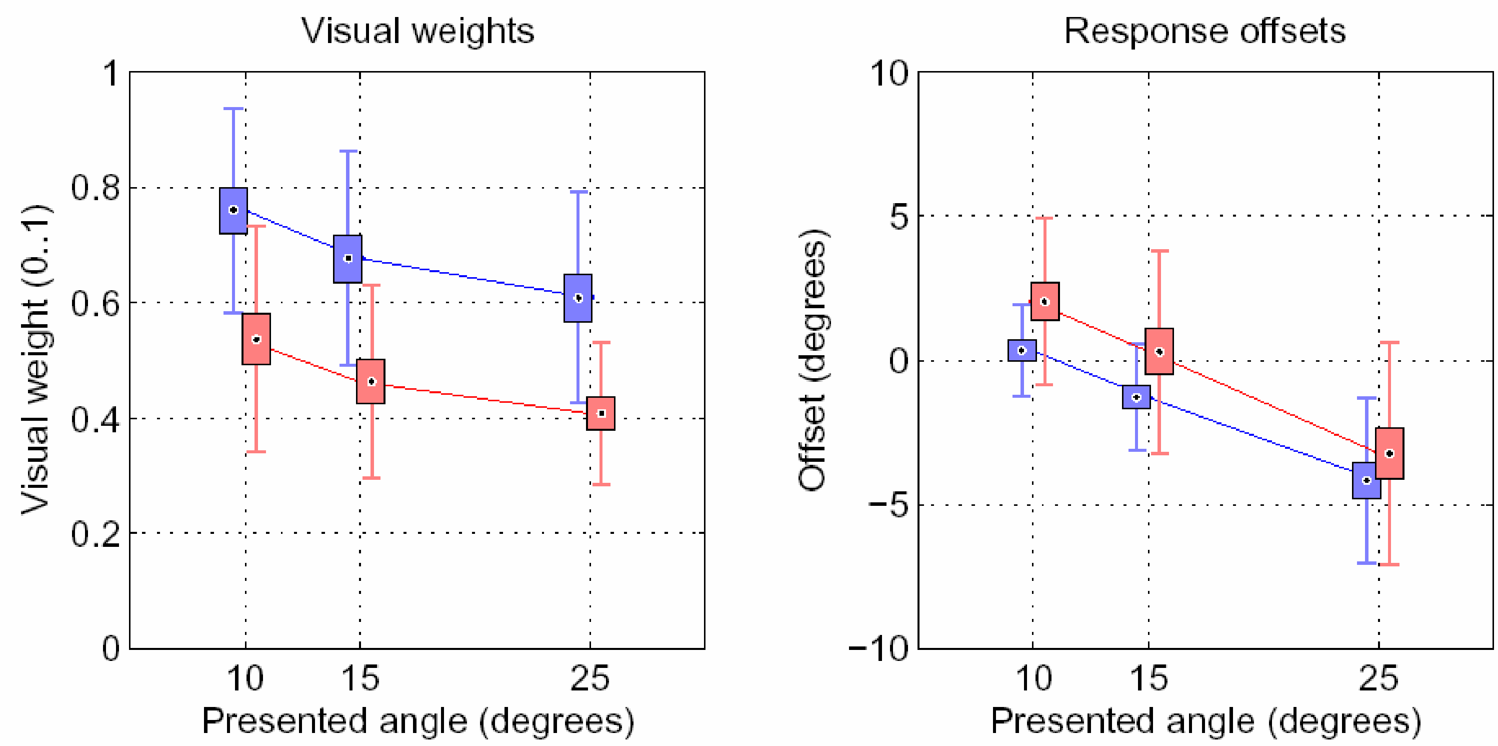
Movement compensation



Experimental Design

- **Method**: **Passively presented turn angles**: either 10°, 15° or 25°
- **Task**: return to start position (joystick control of motion)
  - **Gain factors** during active return: either 0.35, 0.71, 1.0, 1.41 or 2.85
- **Design: Two attention conditions**:
  - "turn back the platform" (**RBody**)
  - "turn back within the visual scene" (**RVis**)
- **Response**:
  - visual scene turned faster
  - platform turned faster
  - no difference noticed (**conflict awareness**)
- 20 participants (10 female, 10 male)
- 120 trials per block (3 angles, 5 factors, 8 repetitions)

Cue Weights and Attention



- Visual attention (RVis): "Return within visual scene"
- Body attention (RBody): "Return platform"

Conclusions

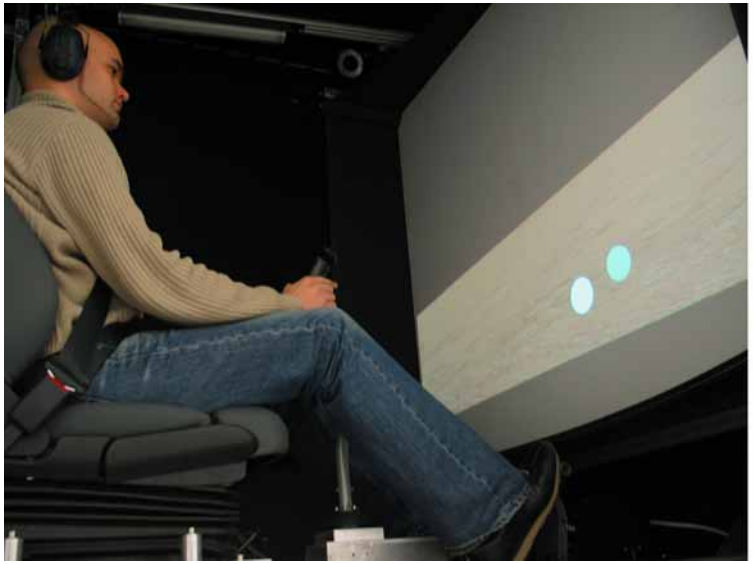
- Higher visual weights for small turns
- Attention modulates the weighting
- Attended cues have higher weights
- The influence of attention on cue weights is stronger in trials in which a cue conflict is detected by the participant

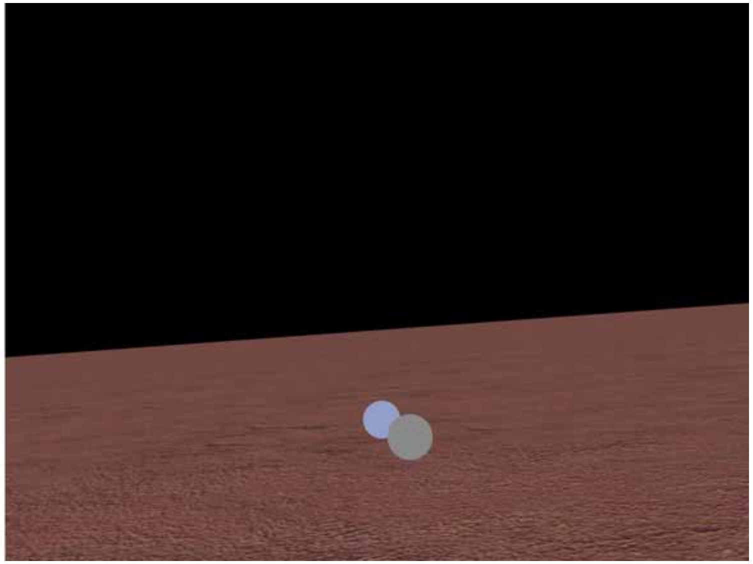
Multimodal Integration for Control tasks

- control task pose a whole new set of problems for integration
- new research direction of our lab
- how are cues integrated during active control of orientation in space
  - 3D maze navigation *(Manuel Vidal, 2004)*
  - body sway (*Cunningham et al, 2006*)
  - flight control *(Beykirch et al, 2006)*
  - helicopter stabilization *(Terzibas and Berger, 2005)*

Integration for Control Tasks
Helicopter stabilization

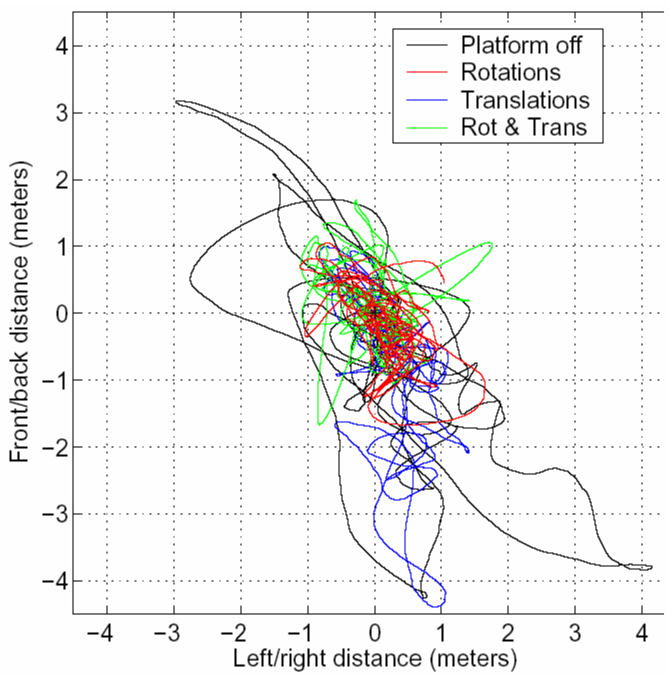
- Trained participants stabilize a simulated helicopter at a target spot
  - target ball on ground and ball representing the helicopter position
- Four different body motion cueing conditions:
  - **R:** platform (pitch and roll) rotations
  - **T:** translations (front-back and left-right)
  - **T & R:** both rotations & translations, platform off

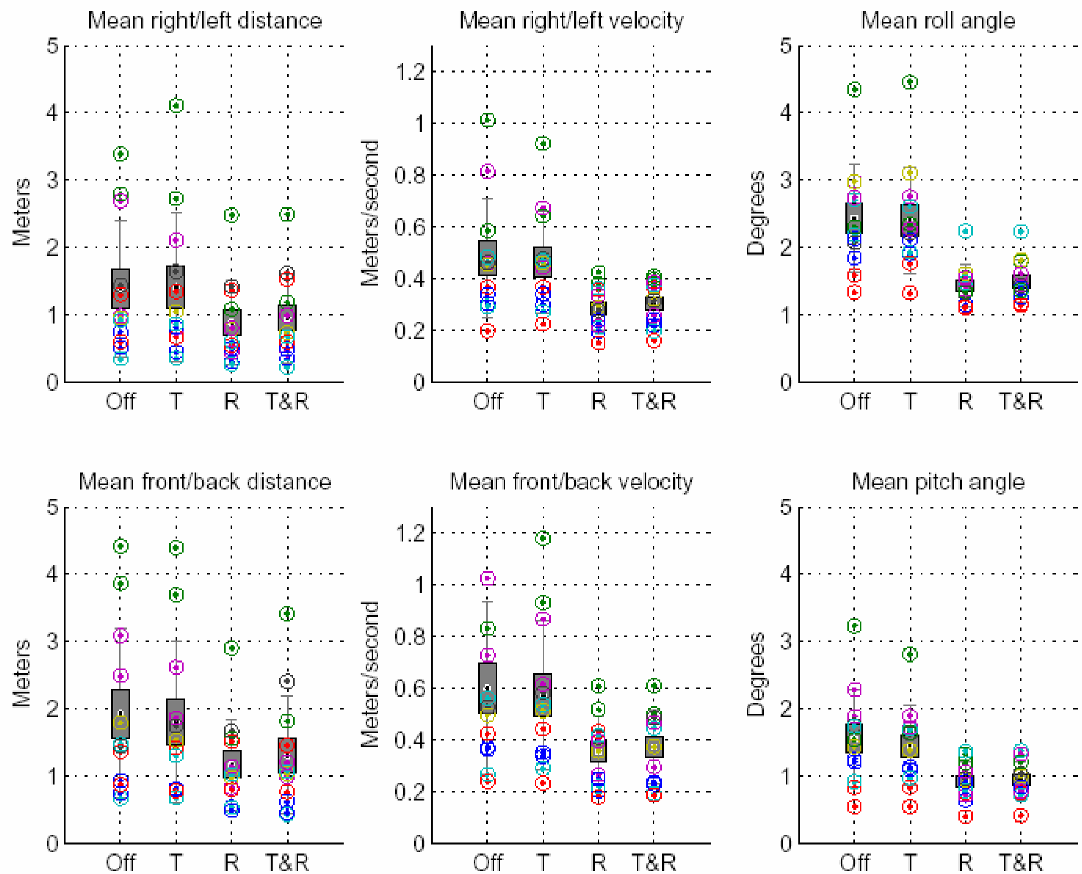




Results

Example Trajectories





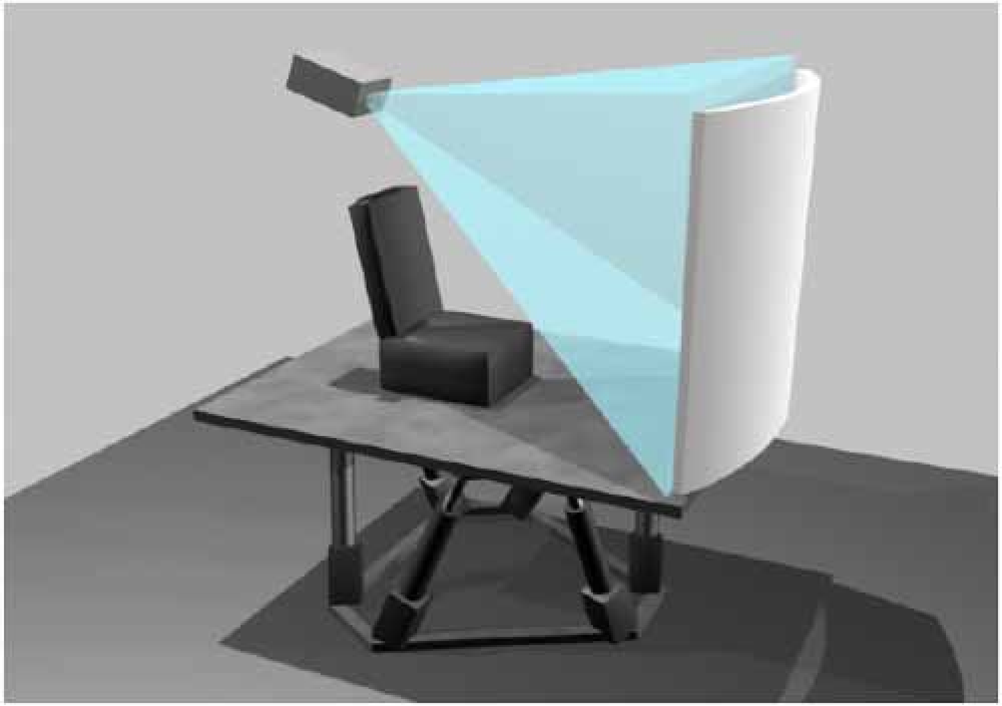
Off: platform off, T: translations, R: rotations

Results

- Stabilization performance significantly better in all six measures **with body rotations**
- **Body translations** had no significant effect on stabilization performance
  - Possibly translation simulation was not realistic enough (due to **range limitations** of the Stewart motion platform, wash-out filters had to be used)
- Better motion range with new motion simulator

Limitations of traditional motion simulators

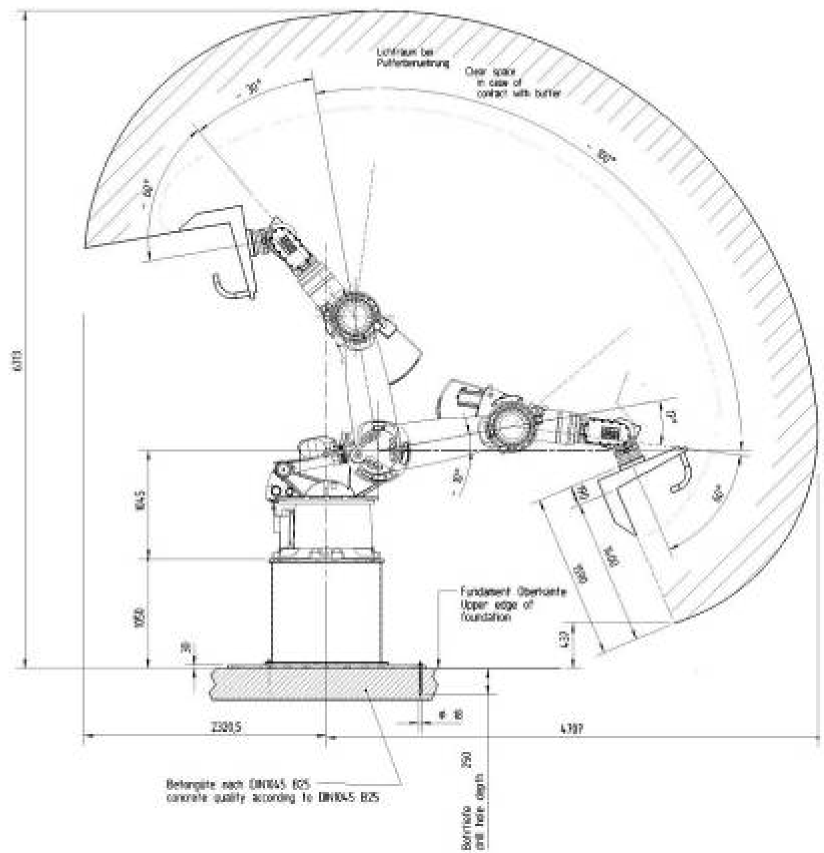
- 6-DOF but limited motion range
  - x: ± 0.46m
  - y: ± 0.43m
  - z: ± 0.25m
  - yaw: ± 44 deg
  - pitch: ± 33 deg
  - roll: ± 28 deg
- Recent attempts to overcome these limitations
- Kuka Robocoaster



Stewart Platform at MPI Tübingen

MotionbaseTM

The MPI Motion Simulator
based on the KUKA Robocoaster©



Ultimate Motion Simulator
KUKA Robocoaster©

New Perception-Action Lab CYBERNEUM

- ***RoboLab***
  - Visual-vestibular integration
  - Egomotion simulation
  - Flight and driving simulator
- **TrackingLab** (15x12m)
  - Free Space Locomotion
  - Motion Capture (Vicon)
  - Motion Tracking (Vicon)
- **Cyberwalk** (EU IST-FP 6)
  - Omnidirectional Treadmill

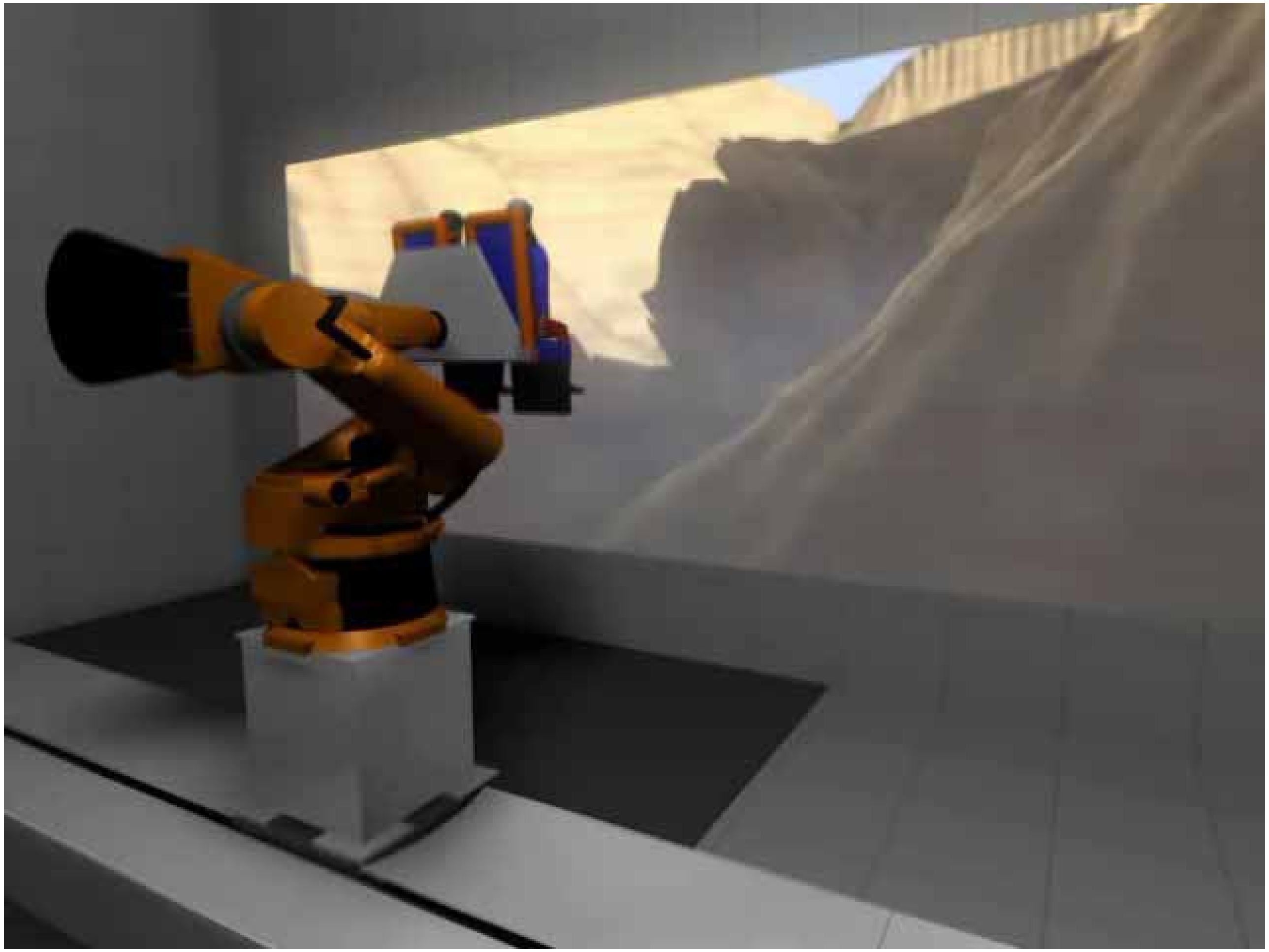
MPI Motion Simulator

- robot arm with low inertia
  - width 11,20 m
  - height 6,75 m
  - weight 3600 kg
- brushless AC drives
- max speed 2m/sec
- commercial platform with safety approval
- but needs visual display
  - HMD
  - dome projection
  - wall projection

Next Generation Flight Simulator

To extend the motion range even further we will put the RoboCoaster on a linear sledge

Ultimate Flight Simulator



Conservative estimate for motion range based on mechanical stops:
x: 1.6m (+ 8m)
y: 2.7m (+ 8m)
z: 1.2m
yaw: (unlimited)
pitch: ± 45 deg
roll: unlimited

x or y can be extended by **8m** with the linear sled

Future Plans in collaboration with DLR

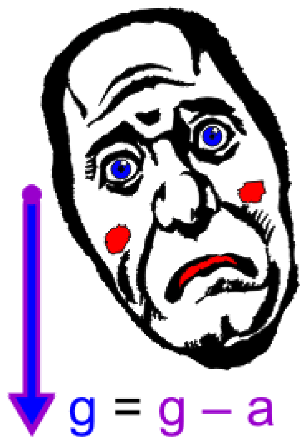
- Similar to the DLR Mars Mission using an Elumens™ projection system we plan to build a spherical projector on a modified single seat Robocoaster.
- Extend the motion range by replacing mechanical end stops with a failsafe end stop and closed-loop control system.

Towards a Better Understanding of Motion Simulation: A human perspective

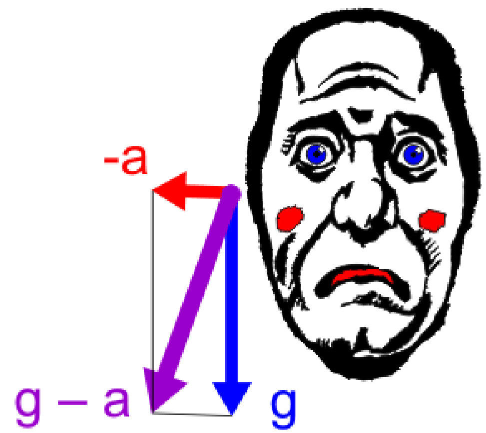
With this new setup, we hope to further elucidate the way the human brain processes the various cues to egomotion, **thus adding a human perspective to simulator design.**

1. Project
Tilt-Translation Ambiguity

- Einstein said, no sensor can distinguish gravity from translational acceleration.
- This holds also for our otolith organs – they sense gravity *minus* acceleration – so they can't distinguish certain tilts and motions.
- Yet the brain can usually resolve the ambiguity.
- How?



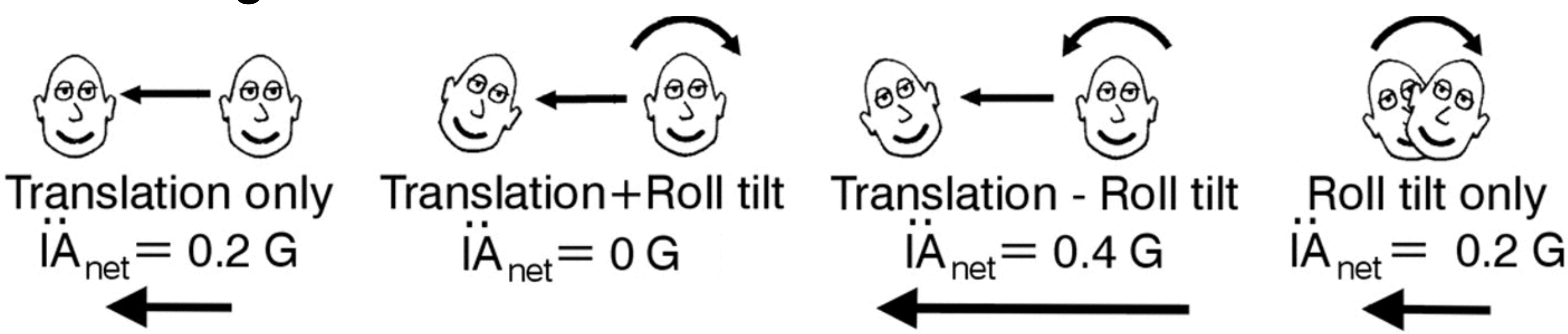
Tilt without translation



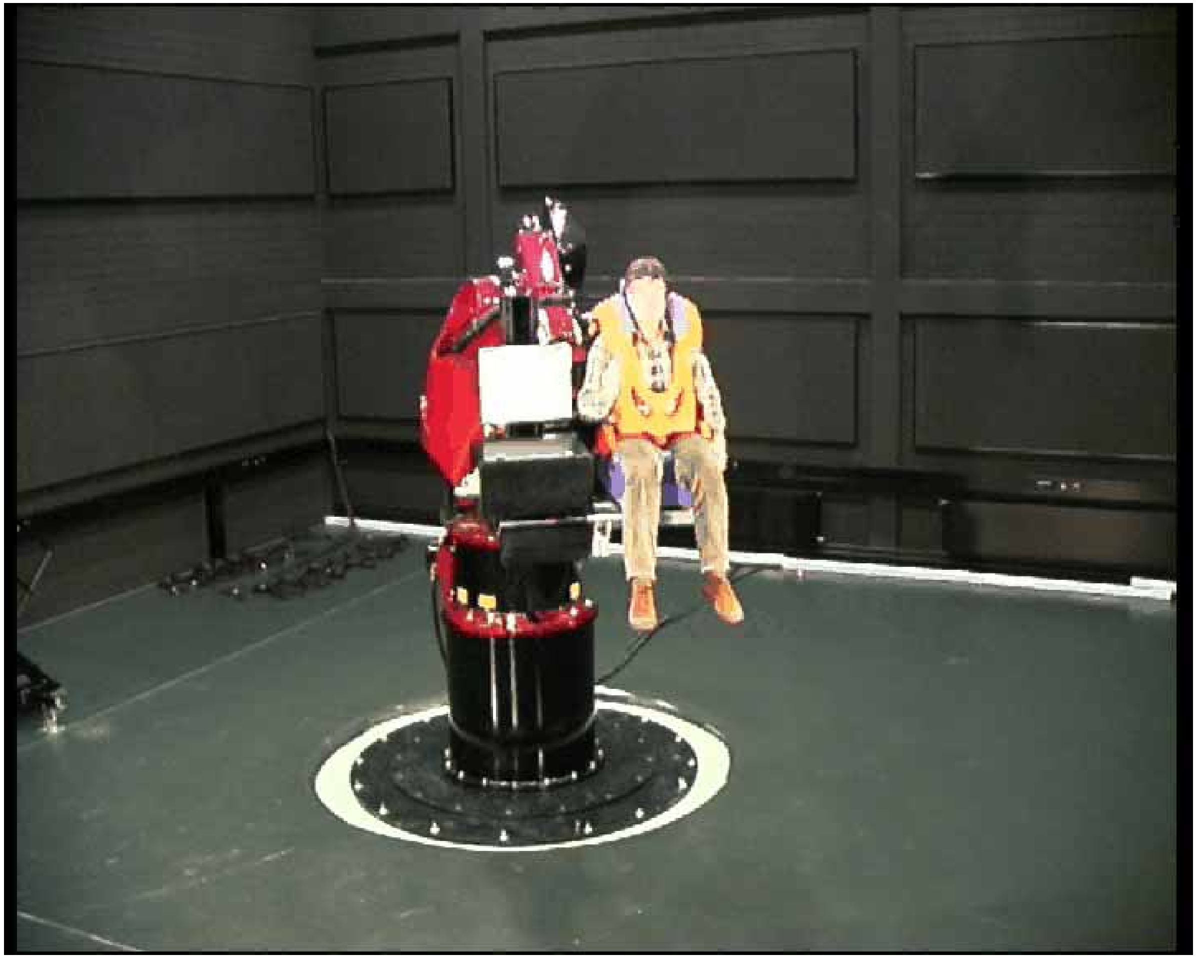
Translation without tilt

Tilt-Translation Disambiguation

- Temporal properties: tilt may be sustained, acceleration seldom is (Paige & Tomko 1991).
- Multisensory integration: semicircular canals detect rotation (Angelaki & Hess 2000, Zupan et al. 2002).
- We'll test these theories in humans and more precisely than in previous studies, e.g., we'll control the magnitude of the gravitoinertial vector.

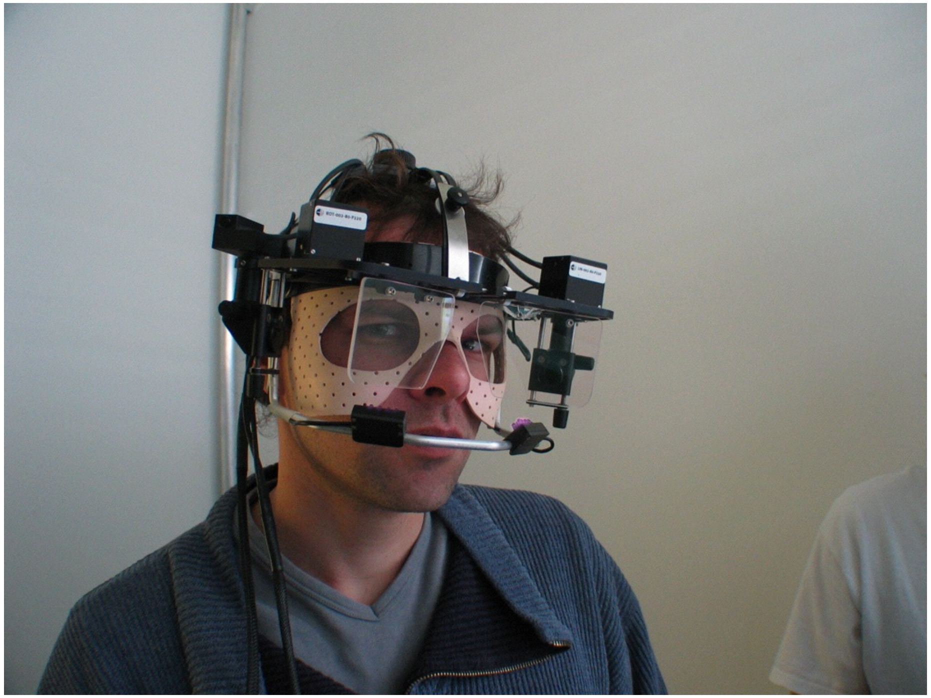


Tilt-Translation Experiment



Response Measure

Use eye-moment as a physiological response measure for acceleration response of otoliths.
x,y + cyclotorsion of eye movement movements



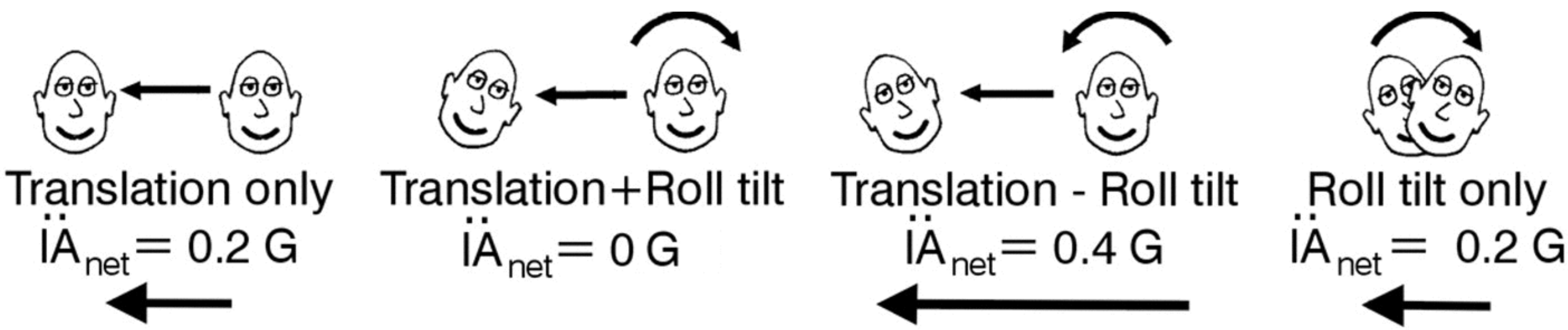
- CMOS image sensors
- 400 Hz
- Face mask
- Bite bar
- torsional eye recording
- Resolution: < 0.1°
- Linearity: < 2.5% +/- 20° H/V
  <4.0% +/- 20° T
- Accelerometers: 1mg

Chronos 3D Eye Tracker

Tilt-Translation Disambiguation

Further questions:

- Can disambiguation adapt?
- Do oculo-motor adaptations affect perception?
- Can adaptation training modify spatial disorientation in pilots?

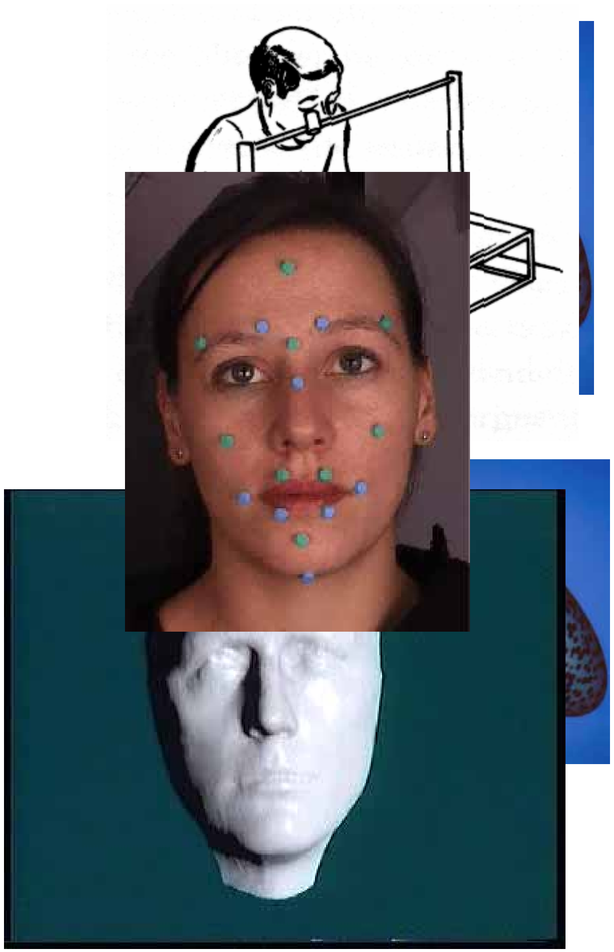


Long Term Goals

- Pursue a practical understanding of the neural processing of multi-sensory self-movement information.
- Explore relationship between physiologic (eye movements) and perceptual data of self-motion in humans.
- Evaluate existing models of this type of response.

Review

- Sensor fusion takes place at all levels
  - early vision
    - stereo, shading, texture, motion, …
  - between modules
    - vision, vestibular, auditory, …
  - higher cognitive levels
    - recognition, emotion
- Sensor fusion use all modes of interaction
  - accumulation or joint regularization
    - shape-from-x
  - cooperation or strong fusion
    - shading and texture
  - disambiguation
    - stereo disambiguates shading (rotating mask)



Conclusion

- sensor fusion works in many cases in a statistical optimal way
- the brain seems to know Bayesian statistics
- but there are also many other ways how the brain can make sense of the world

Members of the new perception-action lab


Daniel Berger


Karl Beykirch


Astros Chatziastros

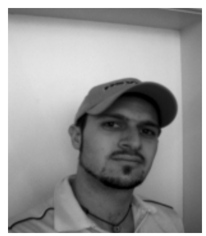

Reinhard Feiler


Gerald Franz


Paolo Pretto


Franck Caniard


Hans-Günther Nusseck


Andreas Wacker


Bernhard Riecke


Cengiz Terzibas


Jörg Schulte-Pelkum


Michael Weyel


Marc Ernst

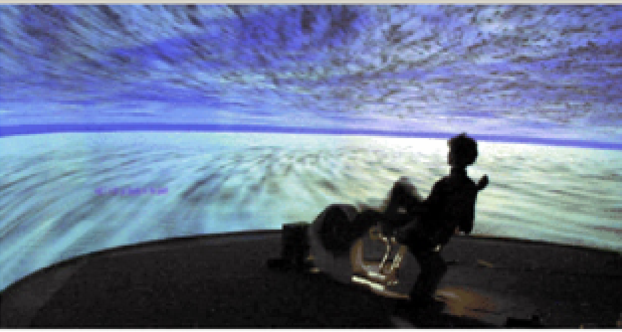

Harald Teufel


Heinrich Bülthoff


Jan Wiener


Manuel Vidal

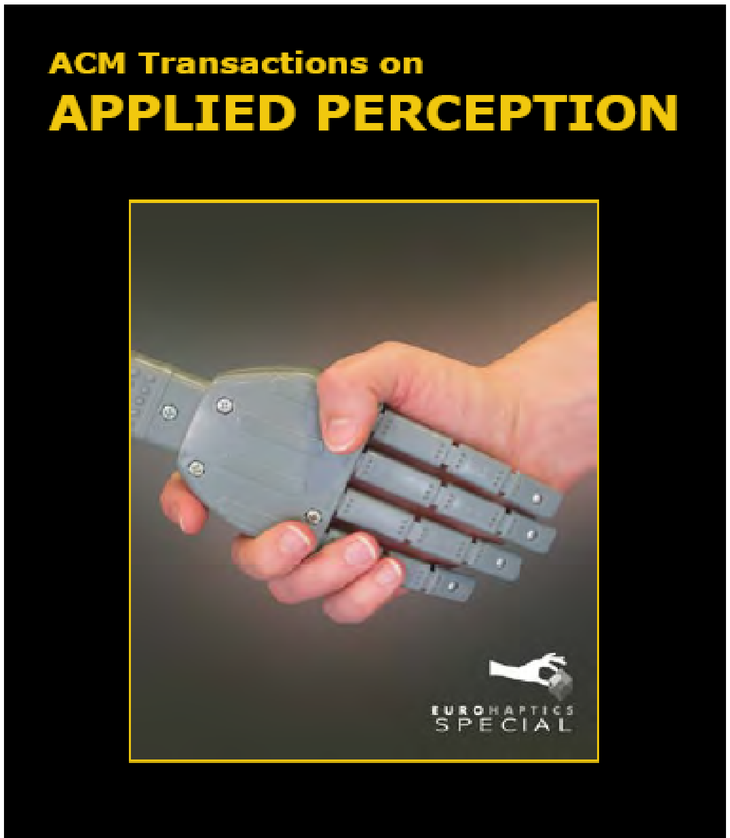
perception of ego-motion



research topics:

- visual cues for steering
- interaction between gaze, attention, and driving direction
- spatial updating
- visual-vestibular sensor fusion
- reference frames and field of view
- cognitive influence on reflexive behavior
- perceptually oriented ego-motion simulation (POEMS)

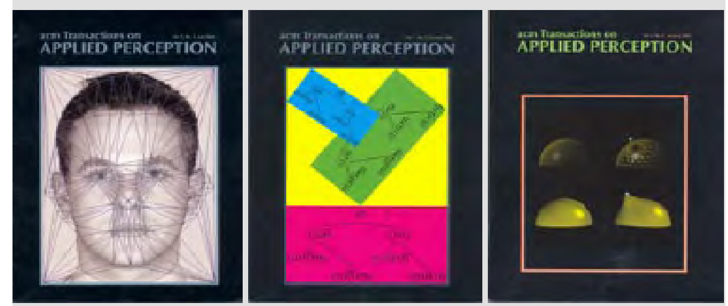
ACM Transactions on Applied Perception



An interdisciplinary journal to foster synergy between computer science and perception

Audio
Vision
Haptics
Graphics
Visualization

In full color
Since July 2004



www.acm.org/tap

APGV

Applied Perception in Graphics and Visualization
Boston, before SIGGRAPH

- Interdisciplinary meeting to foster synergy
  - between perception and computer science
  - every year
- alternating between SIGGRAPH and ECVP
  This symposium seeks to provide a forum
  for the wider exchange of ideas and information
  - between members of the graphics and visualization communities who are using insights from visual/auditory/haptic perception to advance the design and guide the evaluation of methods for more effective visual/auditory/haptic representation,
  - and members of the vision sciences community who are using computer graphics to facilitate the investigation of fundamental processes of perception.