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

- **convexity prior** (familiarity) dominates ambiguous interpretation

- **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

---


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,
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)

Visually and haptically specified shapes differ.
What shape is perceived?
Cube or elongated box?

View cube through distorting lens while exploring object haptically.

Irv Rock
### Rock & Victor (1964) Experimental Design

#### Stimulus Presentation

<table>
<thead>
<tr>
<th>Vision alone</th>
<th>Haptic alone</th>
<th>Conflict</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V$</td>
<td>$H$</td>
<td>$V$ $H$</td>
</tr>
</tbody>
</table>

#### Response Method

- **Drawing**

- **Vision alone**

- **Haptic alone**
Rock & Victor (1964)

Results

<table>
<thead>
<tr>
<th>Stimulus Presentation</th>
<th>Response Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision alone</td>
<td>Haptic alone</td>
</tr>
<tr>
<td>$V$</td>
<td>$H$</td>
</tr>
<tr>
<td>1.90</td>
<td>0.98</td>
</tr>
<tr>
<td>$V$</td>
<td>$H$</td>
</tr>
<tr>
<td>13.4</td>
<td>23.1</td>
</tr>
<tr>
<td>$V$</td>
<td>$H$</td>
</tr>
<tr>
<td>14.1</td>
<td>20.5</td>
</tr>
</tbody>
</table>
### Rock & Victor (1964) Results

<table>
<thead>
<tr>
<th>Stimulus Presentation</th>
<th>Response Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision alone</td>
<td>Haptic alone</td>
</tr>
<tr>
<td>$V$</td>
<td>$H$</td>
</tr>
<tr>
<td>1.90</td>
<td>0.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vision alone</th>
<th>Haptic alone</th>
<th>Conflict</th>
<th>Vision alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.4</td>
<td>23.1</td>
<td>14.1 mm</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vision alone</th>
<th>Haptic alone</th>
<th>Conflict</th>
<th>Haptic alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1</td>
<td>20.5</td>
<td>14.5 mm</td>
<td></td>
</tr>
</tbody>
</table>

*“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

Conflict
($S_H < S_P < S_V$)

Non-conflict
($S_H = S_V = S_P$)

No-conflict size (mm)

No-conflict perceived
"larger" %

$S_H$

$S_V$

50

100

0

52 55 58

Heinrich H. Bülthoff
MPI for Biological Cybernetics

Combined Visual-Haptic Experiment

- Conflict (\(S_H < S_P < S_V\))
  - Height
  - Non-conflict (\(S_H = S_P = S_V\))

No-conflict size (mm):
- 50 mm
- 100 mm

No-conflict perceived "larger" %:
- 55%
- 58%

\[\text{SVSH} \]
\[55 \quad 58 \quad 52\]

\[\text{SV}\]
\[\text{SH} \]

0 50 100

European Summer School 2006
Heinrich H. Bülthoff
MPI for Biological Cybernetics
Rauschholzhausen 9/11/2006
Combined Visual-Haptic Experiment

Conflict (SH<SP<SV)

Non-conflict (SH=SV=SP)

No-conflict perceived "larger" %

No-conflict size (mm)

52 55 58

SV

SH

0 50 100

height
Combined Visual-Haptic Experiment

\[ (S_H < S_P < S_V) \]

Conflict:
- \( S_H < S_P < S_V \)

Non-conflict:
- \( S_H = S_V = S_P \)

No-conflict size (mm):
- 50
- 100

No-conflict perceived "larger" %:
- 50
- 0

\( S_H \)
\( S_V \)

European Summer School 2006

Heinrich H. Bülthoff
MPI for Biological Cybernetics

Rauschholzhausen 9/11/2006
Combined Visual-Haptic Experiment

- No-conflict size (mm)
  - SH = 50
  - SV = 100
  - SP = 0

- No-conflict perceived "larger" %
  - SH = 55
  - SV = 52
  - SP = 58

- Conflict (SH < SP < SV)
- Non-conflict (SH = SV = SP)

- Height

- No-conflict size (mm)
  - SH
  - SV

Combined Visual-Haptic Experiment

Conflicts: $S_H < S_P < S_V$

Non-conflicts: $S_H = S_V = S_P$

No-conflict perceived "larger" %

No-conflict size (mm)
Combined Visual-Haptic Experiment

- **Conflict** ($S_H < S_P < S_V$)
- **Non-conflict** ($S_H = S_V = S_P$)

No-conflict perceived "larger" %

No-conflict size (mm)

- $S_H$
- $S_V$

Heinrich H. Bülthoff
MPI for Biological Cybernetics
Raischholzhausen 9/11/2006
Combined Visual-Haptic Experiment

Conflict ($S_H < S_P < S_V$)

Non-conflict ($S_H = S_V = S_P$)

No-conflict size (mm)

No-conflict perceived "larger" %

$S_H$

$S_V$

52 55 58
Combined Visual-Haptic Experiment

Point of subjective equality (PSE):
Value of no-conflict stimulus perceived as same size as conflict stimulus.

<table>
<thead>
<tr>
<th>No-conflict size (mm)</th>
<th>No-conflict perceived &quot;larger&quot; %</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td></td>
</tr>
</tbody>
</table>

No-conflict perceived size
SH = height
SV
SP

conflict
(S_H < S_P < S_V)

non-conflict
(S_H = S_V = S_P)
Visual-Haptic Integration
Visual-Haptic Integration

![Graph showing visual-haptic integration](image_url)
Visual-Haptic Integration

Haptic Standard

Visual Standard

Trials Perceived "taller"

Comparison Size (mm)

Perceived Size

JND

visual noise
0%
67%

S_H
55
S_V
65

0%
25%
50%
75%
100%
Visual-Haptic Integration

The figure illustrates the relationship between haptic and visual perception of size. The curves represent the percentage of trials in which participants perceived the haptic or visual stimuli as taller than a standard stimulus. The JND (Just Noticeable Difference) is indicated, showing the point at which the visual noise affects the perceived size. The visual noise conditions are 0%, 67%, and 133%, each represented by different symbols and lines.
Visual-Haptic Integration

Haptic Standard  

Visual Standard

Perceived Size

Comparison Size (mm)

Trials Perceived "taller"

visual noise
- 0%
- 67%
- 133%
- 200%

JND

S_H  45

S_V  55

S_V  65
Summary

Visual-Haptic Integration

• the brain combines visual and haptic information in a statistically optimal way

• 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 Localization: Ventriloquist effect

Visual Dominance

The ventriloquist effect results from near-optimal bimodal integration

• Visual-Auditory Temporal Judgments

Auditory Dominance

What you see is what you hear.
What you see is what you hear
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

- complete dominance
- no integration

Graph showing perceived number of taps across different auditory conditions and tap counts.

- 4 Taps
- 3 Taps
- 2 Taps

Conditions:
- one beep less
- same amount
- one beep more

Auditory condition vs. perceived # of taps chart.
Signal Reliability

distribution of modality-alone estimates

% of all answers

deviation from # of events

TAPS
Effect of Reliability on Integration

Perception of BEEPS
- with loud BEEPS
- with quiet BEEPS

Perception of TAPS
- with loud BEEPS
- with quiet BEEPS

Predictions
- complete dominance
- no integration

4 Beeps ▲ 3 Beeps □ 2 Beeps

4 Taps ▲ 3 Taps □ 2 Taps
Effect of Reliability on Integration

- Effect of TAPS on BEEPS
  - BEEPS reliable
  - BEEPS unreliable

- Effect of BEEPS on TAPS

Influence ratio vs. time
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
MPI Tübingen

Jean-Pierre Bresciani

Abderrahmane Kheddar
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 tubes

Otolith organ

Semicircular canal

Otolith organ

Canal wall
Endolymph fluid
Sensory hairs
Vestibular nerve

Otolith crystals
Gelatinous membrane
Sensory hairs
Vestibular nerve
Otolithic System

- normal
- head tilted back
- accelerating
Consequence of somatogravic illusion

Pogen MacNeilage, Daniel Berger, Marty Banks, Heinrich Bülthoff
• 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

![Graph showing the average discrimination of acceleration (GP)](image)

<table>
<thead>
<tr>
<th>JND (deg)</th>
<th>Subject Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

- Visual + body cues
- Body cues only

![Graph showing percent > standard vs comparison stimulus (deg)](image)
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

- Presentation phase (3s)
- Delay (1s)
- Active imitation phase (return)
- Button press

Rotation angle vs. time graph:
- Platform turn
- Visual turn = platform turn * gain factor
- Gain factor

Equation:

visual turn = platform turn * gain factor
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 weights

Response offsets

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

- Platform off
- Rotations
- Translations
- Rot & Trans

Mean right/left distance
Mean right/left velocity
Mean roll angle
Mean front/back distance
Mean front/back velocity
Mean pitch angle

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
Motionbase™
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 2 m/sec
commercial platform
with safety approval
but needs visual display
- HMD
- dome projection
- wall projection
To extend the motion range even further, we will put the RoboCoaster on a linear sledge.
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 \textit{minus} acceleration – so they can’t distinguish certain tilts and motions.
• Yet the brain can usually resolve the ambiguity.
• How?

\[ g = g - a \]

Tilt without translation

\[ g - a \]

Translation without tilt
Tilt-Translation Disambiguation

- Temporal properties: tilt may be sustained, acceleration seldom is (Paige & Tomko 1991).
- 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.

Chronos 3D Eye Tracker

- 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
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.
• 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)
• 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-Günter 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)
• **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.