Visual Tasks

- Testing hypotheses
- Estimation
- Selection of actions
Example

Size?
Distance?
Color?
Some examples of cues to depth or shape
Shading and contour
Texture

1. Density
2. Foreshortening
3. Size
Linear perspective
Uncrossed disparity

Zero retinal disparity

Crossed disparity

Binocular Disparity

- Fixation point
- Interpret as far
- Images of fixation point are fused
- Interpret as close
Fixation point

zero disparity

uncrossed disparity

crossed disparity
Wheatstone stereoscope (c. 1838)

Sir Charles Wheatstone
Dual mirror stereoscope
Cues in conflict

When we have multiple cues to depth or shape the cues may conflict
What cues?
Ames room
Ames room
Ames room
Cue Integration

When we have multiple cues to the same thing:

What to do?
Can we benefit?
Rock & Victor (1964)

View object through distorting lens while exploring object haptically.

Visually and haptically specified shapes differed. What shape is perceived?
Rock & Victor (1964)
Experimental Design

<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>Vision alone</td>
<td>Haptic alone</td>
</tr>
<tr>
<td>$V$</td>
<td>$H$</td>
</tr>
<tr>
<td>Vision alone</td>
<td>Haptic alone</td>
</tr>
<tr>
<td>$V$</td>
<td>$H$</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>
<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>

Legend:
- $V$: Vision
- $H$: Haptic
- 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>13.4</td>
<td>23.1</td>
</tr>
<tr>
<td>14.1</td>
<td>20.5</td>
</tr>
</tbody>
</table>

Drawing

Vision alone

Haptic alone

Visual Capture
(How) should we combine cues?

\[ S_H \] haptic size estimate

\[ S_V \] visual size estimate

random variables
Modeling Cue Combination
\[ S_H : \text{Gaussian}(s, \sigma_H) \]
\[ S_V : \text{Gaussian}(s, \sigma_V) \]
Thought Example

\[ S_H : \text{Gaussian}\ (10\text{ cm}, 1\text{ cm}) \]

\[ S_V : \text{Gaussian}\ (10\text{ cm}, 2\text{ cm}) \]

$haptic\ cue\ is\ less\ variable$

$\$10\ if\ you\ are\ within\ 1\ cm\ of\ s$

Which cue?

Chances of winning?
matlab
Possible Rules

1. Fixed hierarchy rule
   Vision > Auditory > Haptic > …. hypothetical order
   Use first available cue in order above.

2. Best single cue
   Use cue with lowest variance.
   Discard others.

3. Combine the cues somehow …
   does this make any sense?
Can we do better by combining cues?

\[ S = wS_H + (1 - w)S_V \]

\[ 0 \leq w \leq 1 \]

“weighted linear combination”
All Three Rules are weighted linear combinations

\[ S = wS_H + (1-w)S_V \]

1. Fixed hierarchy rule \( w = 0 \) (Vision)
2. Best single cue \( w = 1 \) (Haptic)
3. Combine the cues somehow \( w = ? \)
Can we do better by combining cues?

\[ S = wS_H + (1-w)S_V \]

\[ 0 \leq w \leq 1 \]

What is the “best” value of \( w \)?
We have to decide what our goal is.

We combine two cue SH and SV to get a new cue S.

Goal: We want S to be unbiased and to have minimum variance.

ASSUMPTION: UMVUE
\[ E\left[ S \right] = wE\left[ S_H \right] + (1-w)E\left[ S_V \right] \]

\[ = ws + (1-w)s = s \]

unbiased
Review on variance

\[ \sigma^2 \]

\[ \text{Var}[sX] = s^2 \text{Var}[X] \]
\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \]

X, Y \textit{independent} variables

ASSUMPTION: \( S_H \) and \( S_V \) independent
\[ \text{Var}[S] = w^2 \text{Var}[S_H] + (1-w)^2 \text{Var}[S_V] \]

\[ = w^2 \sigma_v^2 + (1-w)^2 \sigma_h^2 \]

A parabola in \( w \) up-facing or down?
\[ \text{Var} \left[ S \right] = w^2 \sigma_H^2 + (1-w)^2 \sigma_V^2 \]
\[ \text{Var}[S] = w^2 \sigma_H^2 + (1 - w)^2 \sigma_V^2 \]
\[
\frac{\partial \text{Var}[S]}{\partial w} = 2w\sigma_H^2 - 2(1-w)\sigma_V^2 = 0
\]

\[
W = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_H^2}
\]

Minimum or maximum?
\[ W = \frac{\sigma^2}{\sigma^2_V + \sigma^2_H} \]

some examples ....
\[
\text{Var}\left[ S \right] = w^2 \sigma_H^2 + (1 - w)^2 \sigma_V^2
\]
Rock & Victor (1964)

View object through distorting lens while exploring object haptically

Visually and haptically specified shapes differed. What shape is perceived?

Why visual capture?
Humans integrate visual and haptic information in a statistically optimal fashion

Marc O. Ernst & Martin S. Banks

Vision Science Program/School of Optometry, University of California, Berkeley 94720-2020, USA
Visual/Haptic Setup

- CRT displaying 3D image
- head & chin rest
- stereoglasses
- line of sight
- opaque surface mirror
- virtual visual & haptic scene
- force-feedback devices (PHANToMs)
Visual Capture?

Why should vision be the “gold standard” all other modalities are compared to?

\[ S_{VH} = w_V S_V + w_H S_H \]

Weights

\[ w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2} \]

Variance

\[ \frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2} \]
Visual Capture?

Why should vision be the “gold standard” all other modalities are compared to?

\[ S_{VH} = w_V S_V + w_H S_H \]

Weights

\[ w_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2} \]

Variance

\[ \frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2} \]
Visual Capture?

Why should vision be the “gold standard” all other modalities are compared to?

\[ S_{VH} = w_V S_V + w_H S_H \]

Weights
\[ W_V = \frac{\sigma_H^2}{\sigma_V^2 + \sigma_H^2} \]

Variance
\[ \frac{1}{\sigma_{VH}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2} \]
Visual Capture?

Why should vision be the “gold standard” all other modalities are compared to?

\[
S_{VH} = w_V S_V + w_H S_H
\]

Weights

\[
W_V = \frac{\sigma^2_H}{\sigma^2_V + \sigma^2_H}
\]

Variance

\[
\frac{1}{\sigma^2_{VH}} = \frac{1}{\sigma^2_V} + \frac{1}{\sigma^2_H}
\]
Experimental Outline

1) determine (& manipulate) within-modality variances
   • discrimination thresholds (2-IFC, constant stimuli)

2) make predictions for combined performance
   • using MLE model to predict weights & combined variance.

3) measure combined performance & compare to prediction
   • similar to within-modality 2-IFC discrimination task (get PSE and thresholds)
2-IFC Task

Timeline

Three Conditions

Visual-alone  Haptic-alone  Visual-Haptic

Standard  Comparison

1. Interval  1 sec
2. Interval  1 sec

no feedback!
STOP: How do we estimate the variance (or SD) of a cue?

\[ X \sim \text{Gaussian}(s, \sigma) \]
Determining Within-Modality Variance

Within-Modal Discrimination

Proportion of Trials Perceived "taller"

Comparison Height (mm)
Determining Within-Modality Variance

Within-Modal Discrimination

Proportion of Trials Perceived "taller"

Comparison Height (mm)

Threshold

standard
Determining Within-Modality Variance

![Graph showing within-modal discrimination](image-url)
Determining Within-Modality Variance
Determining Within-Modality Variance
Determining Within-Modality Variance

Within-Modal Discrimination

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

Proportion of Trials Perceived "taller"

Comparison Height (mm)

Threshold

standard
From Variance to Threshold

Predicted weights for combined performance from within-modal data

\[ w_V = \frac{\sigma^2_H}{\sigma^2_V + \sigma^2_H} \]

\[ JND_i = \sqrt{2} \cdot \sigma_i \]

\[ w_V = \frac{JND^2_H}{JND^2_V + JND^2_H} \]

Predicted combined threshold from within-modal data

\[ \frac{1}{\sigma^2_{VH}} = \frac{1}{\sigma^2_V} + \frac{1}{\sigma^2_H} \]

\[ JND_i = \sqrt{2} \cdot \sigma_i \]

\[ \frac{1}{JND^2_{VH}} = \frac{1}{JND^2_V} + \frac{1}{JND^2_H} \]
Visual-Haptic Discrimination
Visual-Haptic Discrimination
Visual-Haptic Discrimination
Visual-Haptic Discrimination

Haptic Standard

Visual Standard

Perceived Size

Trials Perceived "taller"

Comparison Size (mm)

JND

visual noise
- 0%
- 67%
- 133%
Visual-Haptic Discrimination

[Graph showing the comparison between haptic and visual standards, with curves indicating perceived size and JND (Just Noticeable Difference) for different visual noise levels.]
Empirical Thresholds and Weights
Individual Differences

Predicted Visual Weight vs Empirical Visual Weight for different noise levels:
- Blue circles: 0% noise
- Red circles: 133% noise

Individuals marked: LAS, JWW, RSB, MOE, KML, HTE
Conclusions

**Humans integrate visual and haptic information in a statistically optimal fashion**

Marc O. Ernst* & Martin S. Banks

Vision Science Program/School of Optometry, University of California, Berkeley 94720-2020, USA

- Combination reduces variance.
- Linear weighting scheme for visual-haptic perception.
- Explains behavior like apparent “visual capture” or visual dominance.
ASSUMPTIONS

• Cues are Gaussian and Independent

• Goal: UMVUE

What if they are not?


What if we have other goals?

next lecture