Sensory Cue Integration

Laurence T. Maloney

Perceptual Tasks

Testing hypotheses
signal detection theory
psychometric function

Estimation this lecture
Selection of actions next lecture
Perceptual Estimation Tasks

Sources of information

Visual Perceptual Information

Retinal images – modeled as cues
Cast Shadows

Texture
1. Density
2. Foreshortening
3. Size

Linear perspective
Wheatstone stereoscope (1838)

Dual-Mirror Stereoscope

left retinal image  right retinal image

fixation points
The images of a single vertical line on the two retinas.

The images of a single vertical line on the two retinas.
There are many depth cues and more than one may be relevant for a given depth estimate.
Multiple Depth Cues

Cues in conflict?
Do we use only one cue?
(presumably the “best” one)
Can we combine cue information to do “better” than any single cue?

What cues?

Ames room
Ames room

Multiple Depth Cues

Cues in conflict?
Do we use only one cue?  
(presumably the "best" one)
Visually and haptically specified shapes differed. What shape is perceived?

**Experimental Design**

<table>
<thead>
<tr>
<th>Stimulus Presentation</th>
<th>Response Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision alone</td>
<td>Vision alone</td>
</tr>
<tr>
<td>Haptic alone</td>
<td>Haptic alone</td>
</tr>
<tr>
<td>Conflict</td>
<td>Drawing</td>
</tr>
<tr>
<td></td>
<td>Vision alone</td>
</tr>
</tbody>
</table>

**Results**

<table>
<thead>
<tr>
<th>Stimulus Presentation</th>
<th>Response Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision alone</td>
<td>Vision alone</td>
</tr>
<tr>
<td>Haptic alone</td>
<td>Haptic alone</td>
</tr>
<tr>
<td>Conflict</td>
<td>Drawing</td>
</tr>
<tr>
<td></td>
<td>Vision alone</td>
</tr>
</tbody>
</table>

- 1.90 0.90 1.05
- 13.4 23.1 14.1 mm
- 14.1 20.5 14.6 mm
The results of Rock & Victor (1964) suggest we use only one of two available cues to size. True?

How is that possible?
Modeling Cue Integration

How should we integrate cues?

\[ S_H \text{ haptic size estimate} \]
\[ S_V \text{ visual size estimate} \]

*Gaussian random variables*

*Independent*

\[ S_H : \text{Gaussian}(s, \sigma_H) \]
\[ S_V : \text{Gaussian}(s, \sigma_V) \]

true location

\[ s \]
Which cue is “better”? We need a criterion: what counts as “better”?

Let’s Play a Game

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

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

$10 if you are within 1\,\text{cm} of s$

Which cue?

Chances of winning?
Let's Play a Game

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

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

\[ S_H \text{ 68%  $6.80} \]

\[ S_V \text{ 38%  $3.80} \]

Can we do better by combining cues?

\[ S = wS_V + (1-w)S_H \]

*weighted linear combination*

\[ E[S] = wE[S_V] + (1-w)E[S_H] \]

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

*unbiased*
\[
E[S] = wE[S_v] + (1-w)E[S_h] \\
= ws + (1-w)s = s
\]

Gaussian \( S \) : \text{Gaussian}(s, \sigma(w))

---

Review on variance \( \sigma^2 \)

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

\( X, Y \) independent variables

---

\[
\text{Var}[S] = w^2\text{Var}[S_v] + (1-w)^2\text{Var}[S_h] \\
= w^2\sigma_v^2 + (1-w)^2\sigma_h^2 \\
= \sigma_h^2 - 2\sigma_h^2w + (\sigma_h^2 + \sigma_v^2)w^2
\]

a parabola in \( w \)
$Var[S] = w^2 \sigma_v^2 + (1-w)^2 \sigma_H^2$

$\sigma_H^2 = + \sigma_v^2$

$\sigma_v^2$

$0 \quad w \quad 1$

$\frac{\partial Var[S]}{\partial w} = 2w \sigma_v^2 + 2(1-w) \sigma_H^2 = \neq 0$

$w = \frac{\sigma_H^2}{\sigma_v^2 + \sigma_H^2}$
\[ W = \frac{\sigma_H^2}{\sigma_v^2 + \sigma_H^2} \]

Optimal Cue Integration

\[ Var[S] = w^2 \sigma_v^2 + (1 - w)^2 \sigma_H^2 \]

\[ = \frac{\sigma_v^2 \sigma_H^2}{\sigma_v^2 + \sigma_H^2} \]

\[ Var[S] = w^2 \sigma_v^2 + (1 - w)^2 \sigma_H^2 \]

Some examples ....

Oruc et al (2003) Vis Res
Let's Play a Game

\[ S_H : \text{Gaussian}(10 \text{ cm}, 1\text{ cm}) \]
\[ S_V : \text{Gaussian}(10 \text{ cm}, 2\text{ cm}) \]

\[ S_H : 1 \quad 68\% \quad $6.80 \]
\[ S_V : 4 \quad 38\% \quad $3.80 \]
\[ S : 0.8 \quad 74\% \quad $7.40 \]

Minimizing the variance of our unbiased estimate leads to the highest expected value in our “game”.

The criterion used in the cue integration literature is **minimum variance unbiased**.

**Beware**: If distributions are not Gaussian then the weighted sum rule of combination may not be the minimum variance unbiased estimator.

---

Rock & Victor (1964)

View object through distorting lens while exploring object haptically

**Visual capture?**

Visually and haptically specified shapes differed. What shape is perceived?
Humans integrate visual and haptic information in a statistically optimal fashion

Marco G. Ernst & Martin S. Banks

Vision Science Program/Department of Ophthalmology, University of California, Berkeley, CA 94720-3800, USA

---

Visual/Haptic Setup

- CBT displaying SC image
- head & chin rest
- stereoglasses
- line of sight
- upper surface mirror
- virtual visual & haptic screen
- force-feedback platform (PUMA 300)

---

Visual Capture?

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

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

Weights

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

Variance

\[ \frac{1}{\sigma^2_{\text{vis}}} = \frac{1}{\sigma^2_V} + \frac{1}{\sigma^2_H} \]
Visual Capture?

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

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

Weights

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

Variance

\[ \frac{1}{\sigma_{uv}^2} = \frac{1}{\sigma_V^2} + \frac{1}{\sigma_H^2} \]
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

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

![Diagram showing proportion of trials perceived as taller vs. comparison height (mm)]

Determining Within-Modality Variance

![Diagram showing proportion of trials perceived as taller vs. comparison height (mm)]

Determining Within-Modality Variance

![Diagram showing proportion of trials perceived as taller vs. comparison height (mm)]
From Variance to Threshold

Predicted weights for combined performance from within-modal data

Visual-haptic variance estimators weights

\[ w_V = \frac{\sigma_H}{\sigma_V} \]

Predicted combined threshold from within-modal data

Visual-haptic variance

\[ \frac{1}{\sigma_{\text{V-H}}} = \frac{1}{\sigma_V} + \frac{1}{\sigma_H} \]

Visual-Haptic Discrimination

Visual-Haptic Discrimination
Empirical Thresholds and Weights

**Discrimination Thresholds**

- Haptic (empirical)
- Visual (empirical)
- Combined (predicted)
- Combined (empirical)

**Weights & PSEs**

- Visual capture
- Haptic capture

Individual Differences

- Empirical Visual Weight vs. Predicted Visual Weight
- 0% noise, 133% noise
- LAS, RSB, MOE, HTE, JWW, KML

Conclusions

- Combination reduces variance.
- Linear weighting scheme for visual-haptic perception.
- Explains behavior like "visual capture" or visual dominance.
  - i.e., vision is given a weight of ~ 1.0 if the variance of the visual estimate is less than the variances of the other modalities.
Reliability / Precision

\[ r_V = \frac{1}{\sigma_V^2} \]
\[ r_H = \frac{1}{\sigma_H^2} \]
\[ r_{opt} = \frac{1}{\sigma_{opt}^2} \]

Just a change in notation.

For the optimal integrated cue

Reliability / Precision

\[ r_V = \frac{1}{\sigma_V^2} \]
\[ r_H = \frac{1}{\sigma_H^2} \]
\[ r_{opt} = \frac{1}{\sigma_{opt}^2} \]

\[ w_{opt} = \frac{\frac{r_V}{r_V + r_H}}{r_V + r_H} \]

The weight given to a cue is proportional to its reliability

\[ r_{opt} = r_V + r_H \]

Cue reliabilities add under optimal integration