Psychophysics

“The scientific study of the relation between stimulus and sensation”

- Fundamental to psychology
- Has become fundamental to understanding haptic devices and virtual environments


Subsequent slides based on A. M. Okamura’s Lecture 3 for 530.661 at JHU and Katherine Kuchenbecker’s slides from MEAM 625 at Univ. Pennsylvania, as well as examples from Dr. Provancher’s research.

Based on a slide by Steve Hsiao, a Neurophysiologist with the Zanvyl Krieger Mind/Brain Institute at Johns Hopkins University.
Sensory Dimensions

- Stimuli can differ on at least four basic dimensions.
  - **Intensity**: amplitude of signal
    - e.g., amplitude, frequency
  - **Quality**: kind or type, both between and within modalities
    - e.g., frequency change in vibratory stimulus
  - **Extension**: spatial aspects, such as size, location, and separation
    - e.g., size, location, separation between bumps
  - **Duration**: how long a sensation lasts
    - e.g., adaptation can occur for longer stimuli

Psychophysics Overview

- **Detection threshold (absolute limen)** *
  - Definition: the smallest energy of the stimulus that can be reliably detected
  - **Methods**: method of constant stimuli, method of limits, method of adjustment, signal detection theory, adaptive procedure
- **Discrimination (difference) threshold** *
  - Definition: the smallest difference between two stimuli that can be reliably detected
  - Also known as the just-noticeable-difference (JND)

Psychophysics Overview, Cont.

- **Weber’s law**: JND(%) = JND/Ref = constant
- detection = discrimination of noise and signal
- methods: same as detection experiment
- **Magnitude Scaling** *
  - Steven’s Power Law, $\psi = k I^a$

Detection

- The goal is to determine the minimum value of a stimulus necessary for detection.
- This value is known as the absolute threshold, stimulus threshold, Reiz Limen, RL, or absolute limen.
- **Threshold = Limen**

---

**Differential Sensitivity**

- The goal is to determine how much a stimulus needs to change in order to feel different from the nominal stimulus.
- This value is known as the difference threshold, Difference Limen, DL, just noticeable difference, and JND.

**Example: Difference thresholds for frequency**

![Graph showing difference thresholds for frequency](Goff 1967 in Gescheider 1985)

**Just Noticeable Difference: JND**

- The amount of change in a stimulus that creates a perceptible increment in sensation

**Example of differential JND**

- Stimulus intensity = 10 units
- Goes up to 12 units before observer notices a change
- Therefore, JND = 2 units at that stimulus level

---

**Weber’s Fraction, 1834**

*German Physiologist E.H. Weber*

- Linear relationship between differential threshold and stimulus intensity
  - For example: to feel different, 2 heavy weights must differ more than two light weights
- Not always perfect near the absolute threshold

**Weber’s law**

![Graph showing Weber’s law](Gescheider 1984)

**Revised Weber’s Law**

- When $\phi$ is very close to the absolute threshold, Weber’s fraction increases

$$\frac{\Delta \phi}{\phi + \alpha} = c$$
Classical Psychophysical Methods

- Determining sensory thresholds:
  - Method of Limits
  - Method of Adjustment
  - Method of Constant Stimuli

Method of Limits: Absolute

- Very common, fast method.
- Present ascending and descending stimulus series.
- Ask “Do you feel it?”
- Average the transition points to get absolute threshold.

Method of Limits: Differential

- Very common, fast method. Limited accuracy.
- Present ascending and descending stimulus series in comparison with a standard stimulus.
- Ask “Is it greater than, equal to, or less than the standard?”
- Average the transition points to get interval of uncertainty, then DL.

Method of Adjustment: Absolute

- Takes a moderate amount of time and delivers moderate accuracy.
- Give the subject control over the stimulus.
- Have subject increase stimulus from zero until detectable. Then decrease stimulus from very high until not detectable.
- Average reported values.

Need continuously variable stimuli.
**Method of Adjustment: Differential**

- Takes a moderate amount of time and delivers moderate accuracy.
- Give the subject control over the stimulus.
- Have subject increase stimulus from zero until detectable. Then decrease stimulus from very high until not detectable.
- DL is standard deviation of response distribution.

Need continuously variable stimuli.
Need to present two stimuli simultaneously.

**Method of Constant Stimuli: Absolute**

- Slower but more accurate that method of limits or adjustment.
- Create a set of 3, 5, 7, or 9 stimuli neighboring your guess at the RL (abs. thresh.).
- Randomly present, about 100 trials.
- Ask “Do you feel it?”
- Use 50% point on s-shaped curve to get the RL.

**Method of Constant Stimuli: Differential**

- Slower but more accurate that method of limits or adjustment.
- Create a set of 3, 5, 7, or 9 stimuli neighboring your guess at the RL (abs. thresh.).
- Randomly present, about 100 trials.
- Randomly present pairs, one of which is always the standard.
- Ask “Is it greater than the ‘standard’?”
- Use average of 50% minus 25% and 75% minus 50% to get the DL (also called the JND).

**Magnitude Estimation**

- Subjects compare a broad set of stimuli to a standard stimulus.
- Report how much larger or smaller each stimulus is compared to the standard.
- Responses can be delivered as numbers, marks on a line, position in a spatial arrangement, etc.
- Can also do the inverse for some stimuli, i.e., magnitude production.
**Example of Magnitude Estimation**

Vibrotactile Feedback

- Sensory substitution – vibrotactile feedback to replace force feedback

![Graphs showing magnitude estimation](image)

[Murray, Klatzky, Khosla, Psychophysical Characterization and Testbed Validation of a Wearable Vibrotactile Glove for Telemanipulation Presence, 12(2) 2003]

© W. Provancher 2009

**Stevens’ Power Law**

\[ \psi = k I^a \]

- \( \psi \) Perceived magnitude, reported by subject
- \( I \) Stimulus intensity, applied by experimenter
- \( a \) Exponent, depends on type of stimulation e.g., \( a = 1 \) for visually apparent length, \( a = 3.5 \) for electric shock through fingers, \( a = 0.6 \) for 250 Hz vibration on finger
- \( k \) Proportionality constant, depends on type of stimulus and measurement units used

Based on slides by Wouter Bergmann Tiest, Utrecht University, and notes by J. Edward Colgate, Northwestern University.

© K. Kuchenbecker 2008

**Fitt’s Law**

- Used by interaction designers
- Fitt’s Law states that the time to acquire a target is a function of the distance to and size of the target
  \[ T = a + b \log_2 \left( \frac{D}{W} + 1 \right) \]
- Defines an index of difficulty
- For a haptic virtual environment or teleoperated system, you often want to show that you can minimize difficulty via haptic feedback

Adapted from A. M. Okamura’s lecture at the IEEE/IFRR School of Robotics Science on Haptic Interaction, September 29, 2006

© A. Okamura 2006

**Planning an Experiment**

- Develop (design, build, instrument, program) an experimental apparatus – a haptic interface with a purpose
- Determine the goal of the study
- Select the environment(s) with which your subjects will interact: virtual models, simple physical models, real patients, etc.
- Select your subject pool: university students, a broad sampling of people, novices or experts in a certain application domain

© A. Okamura 2006
### Experiment Procedure

- Develop a strict experimental procedure (called a protocol) (~1-4 weeks, evolved through pilot testing)
- Create a very clear set of instructions for your subjects (can be written, oral, video, etc.) (~1-3 weeks)
- Develop a questionnaire of relevant information for your subjects, including information such as age, gender, handedness (~1 week)
- Submit a Human Subjects Internal Research Board application; receive approval (generally takes 2-4 months)
- Perform experiment (~1-3 weeks), analyze data and document thoroughly (~1-3 weeks), write a paper (~2-6 weeks plus revisions from peer review)

Adapted from A. M. Okamura’s lecture at the IEEE/IFRR School of Robotics Science on Haptic Interaction, September 29, 2006 © K. Kuchenbecker 2008

### Example of Method of Constant Stimuli: Differential

- Curvature Discrimination with both physical and virtual stimuli
  - **Direct Discrimination**
  - **Virtual Discrimination**


### Experimental Design

- Standard psychophysical protocol
  - Measure perceptual thresholds (Just Noticeable Difference, JND)
  - Forced-choice, paired-comparison testing
  - Reduce order bias
    - Balance presentation order
    - Balance testing order
  - Isolate learning & fatigue effects

### Example

Which stimulus has a greater radius?

12 comparisons per test
- Standard (reference) stimulus
- 6 comparison stimuli (presented 2x each)
- Paired-comparison, forced-choice
  - Data from similar tests were pooled

For Example

**Expected Form of Results**

\[ JND = \frac{JND_U + JND_L}{2} \]

Smaller JND is Better!

**Typical Experimental Results**

<table>
<thead>
<tr>
<th>Condition</th>
<th>JND 10 mm</th>
<th>JND 40 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Discrimination</td>
<td>0.84 mm</td>
<td>5.74 mm</td>
</tr>
<tr>
<td>Virtual Discrimination</td>
<td>1.35 mm</td>
<td>4.16 mm</td>
</tr>
</tbody>
</table>

**Secondary Experimental Effects**

- **Learning and fatigue**
  - Subjects performed better at the beginning of tests (up to 28% better)
  - Subjects fatigued possibly?

- **Presentation order effects**
  - More likely to choose incorrectly if stimuli small-to-large
    - 30% more likely for direct
    - 24% more likely for virtual

**User Testing Results**

Smaller JND is Better!
Statistics Reference

- This is a reference to allow you use and understand some basic statistics. It includes:
  - Notation
  - t-Tests
  - Effect Size
  - ANOVA
  - Preconditions and Assumptions
  - Problems with Null Hypothesis Testing
  - Other useful tools

Notation

- Statistics: measures that arise from your sample
  - y – some score in your sample
  - $s^2$ – variance of sample
  - s – standard deviation of sample

- Parameters: measures that arise from the population
  - $\sigma^2$ – variance of population
  - $\sigma$ – standard deviation of population
  - est. $\sigma$ – estimate of $\sigma$
  - est. $s^2$ – estimate of $\sigma^2$

- Other:
  - SS = Sums of Squares
    - $\Sigma (y_i - \text{mean}(y))^2$

- t-Tests
  - t-Test for two independent means
    - Test for difference between two group means, when the groups are totally independent.
    - $H_0$: $\mu_1 = \mu_2$
    - $H_A$: $\mu_1 \neq \mu_2$

  - t-Test for dependent groups (paired t-Test)
    - Test for difference between two group means, when the groups are not independent.
      - Repeated measures (within-subjects) design
      - Matched pairs design
    - $H_0$: $\mu_{\text{difference}} = 0$
    - $H_A$: $\mu_{\text{difference}} \neq 0$
t-Test: Interpreting Results

**t-Test: Two-Sample Assuming Equal Variances**

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.44139</td>
<td>1.32651</td>
</tr>
<tr>
<td>Variance</td>
<td>2.019462</td>
<td>0.73901</td>
</tr>
<tr>
<td>Observations</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Pooled Variance</td>
<td>1.379236</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>t Stat</td>
<td>-3.36607</td>
<td></td>
</tr>
<tr>
<td>P(\text{T &lt;= t}) one-tail</td>
<td>0.00172</td>
<td></td>
</tr>
<tr>
<td>t Critical one-tail</td>
<td>1.734064</td>
<td></td>
</tr>
<tr>
<td>P(\text{T &lt;= t}) two-tail</td>
<td>0.00344</td>
<td></td>
</tr>
<tr>
<td>t Critical two-tail</td>
<td>2.100922</td>
<td></td>
</tr>
</tbody>
</table>

If \( t < 0.05 \), reject the null hypothesis

If \( p < 0.001 \), just say "\( p < 0.001 \)"

don't give the exact value

"Reject the null hypothesis" or "Fail to reject the null hypothesis"

Never "Accept the null/alternative hypothesis"

**Effect Size**

- Raw effect size is generally best.
- Normalized effect size:
  - Cohen's \( d \) – effect size in sample. Most common.
  - Hedges's \( g \) – estimate of effect size in population
  - Glass's \( \Delta \) – compare to control group

If \( p < 0.05 \), reject the null hypothesis

\[
d = \frac{\bar{Y}_1 - \bar{Y}_2}{S_{\text{pooled}}}\]

\[
\Delta = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{est.} \sigma_{\text{control}}}
\]

where \( S_{\text{pooled}} = \sqrt{\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2}} \)

\[
ge = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{est.} \sigma_{\text{pooled}}}
\]

where \( \text{est.} \sigma_{\text{pooled}} = \sqrt{\frac{(N_1 - 1)\text{est.} \sigma_1^2 + (N_2 - 1)\text{est.} \sigma_2^2}{N_1 + N_2 - 2}} \)

**ANOVA (F-test)**

- Use to analyze several groups means for differences.
- When only 2 groups, same as t-Test
- Does not tell you which group mean is different.
  - \( H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \ldots \)
  - \( H_A: \) At least one of the group means is different from all the rest

**ANOVA: Three Main Types**

- **Single Factor / 1-Way**
  - One independent variable
- **Two Factor / 2-Way with Replication**
  - Multiple measure of each condition
  - Two independent variables
  - Looks at interaction between independent variables
- **Two Factor / 2-Way without Replication**
  - Only one measure of each condition
  - Otherwise, like Two Factor with Replication
- **Multi-Factor also possible**
ANOVA: Interpreting Results

**Summary**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Count</th>
<th>Sum</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>10</td>
<td>28.9054</td>
<td>2.89054</td>
<td>0.858315</td>
</tr>
<tr>
<td>Group 2</td>
<td>10</td>
<td>22.0589</td>
<td>2.20589</td>
<td>0.382291</td>
</tr>
<tr>
<td>Group 3</td>
<td>10</td>
<td>-4.4139</td>
<td>-0.44139</td>
<td>2.019462</td>
</tr>
<tr>
<td>Group 4</td>
<td>10</td>
<td>13.2651</td>
<td>1.32651</td>
<td>0.73901</td>
</tr>
</tbody>
</table>

If \( p < 0.05 \), reject the null hypothesis.

**ANOVA**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>( F )</th>
<th>( P )</th>
<th>( F ) crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>62.30891</td>
<td>3</td>
<td>20.77194</td>
<td>20.77442</td>
<td>0.526981</td>
<td>2.866266</td>
</tr>
<tr>
<td>Within Groups</td>
<td>35.9917025</td>
<td>36</td>
<td>0.99977</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>98.3006124</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**How to report these results:**

\[
F(df_1, df_2) = F_{obtained}, \ p = \ldots
\]

\[
e.g., \ F(3, 36) = 20.77, \ p < 0.001
\]

"Reject the null hypothesis" or "Fail to reject the null hypothesis"

Never "Accept the null/alternative hypothesis"

‌

Precondition and Assumptions

- All samples are independent
- Essential
- Data must be normally distributed
- How to tell
  - Look at histogram of data
  - Shapiro-Wilk test for Non-Normality
  - If not normally distributed
    - Transform data to make normal
- Variance the same for all groups
- How to tell
  - Test: F-ratio, Levene’s F
  - If variances not equal
    - t-Test: use t-test for unequal variance (Welch’s t)
  - ANOVA: nothing you can do

Note: t-test and F-test are robust to violations of normality and variance, especially if the sample size is large (>30) and approximately the same for all groups.

Other Useful Tools

- **Confidence Interval** – Better than Null Hypothesis Testing
  - "95% probability that the interval contains the true value of the population mean."
  - Not: "95% probability that the true value of the population mean is within the interval."
  - Do not use Microsoft Excel confidence interval.
- **Linear Correlation**
  - Statistic of merit: \( R^2 \)
  - Test significance of \( R^2 \) with t-test for single group mean with Degrees of Freedom = N-2.