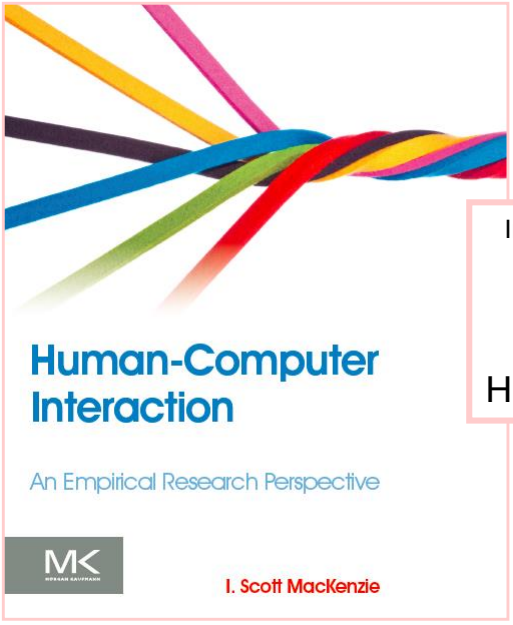


Interactive Systems Design
Prof. V. Fuccella

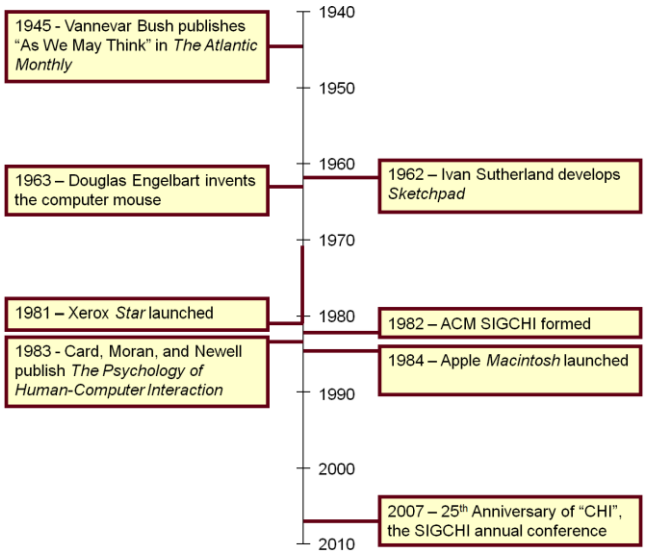
Chapter 1
Historical Perspective

Human-Computer Interaction

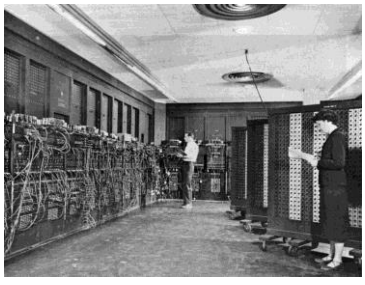
- Emerged in 1980s
- It owes a lot of older disciplines: **ergonomics**, cognitive and experimental psychology, sociology, antropology, computer science, linguistics
- Human factors (ergonomics):
  - Is both a science and a field of engineering;
  - It is concerned with human capabilities, limitations and performance, and with the design of systems which are efficient, safe, comfortable, and even enjoyable for the humans who use them;
  - It is also an art in the sense of respecting and promoting creative ways for practitioners to apply their skills in designing systems.
  - Human Factors in Computing Systems (CHI)

2

Significant Event Timeline

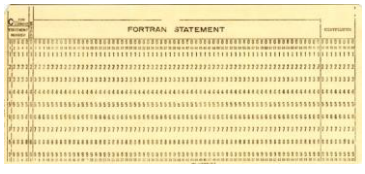


3

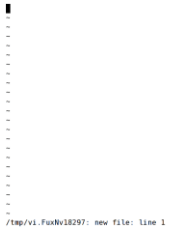


ENIAC (1940s)

UNIVAC (1950s)



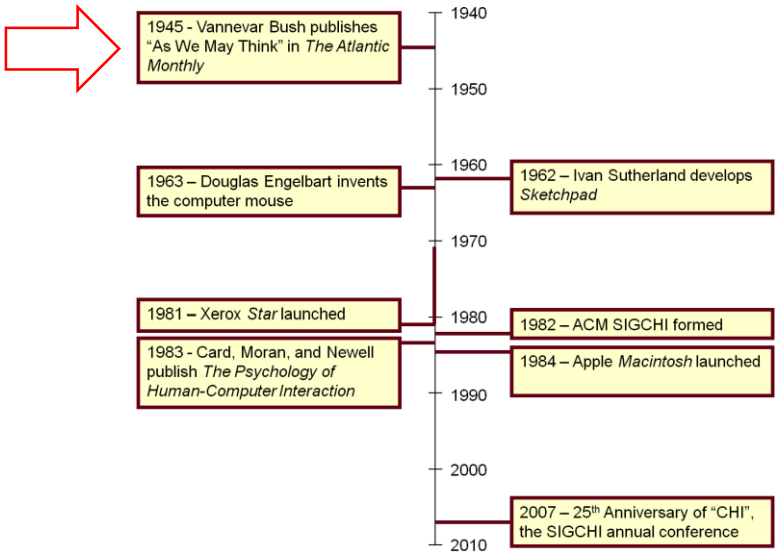
1920s - 1950s



VI (1976 - …)

4

Significant Event Timeline



5

"As We May Think"

Vannevar Bush (1945)

6

Reprinted in…

ACM DL DIGITAL LIBRARY

As we may think

Full Text: Pdf

Author: Vannevar Bush Director of the Office of Scientific Research and Development

Published in:
· Magazine
interactions Interactions Homepage archive
Volume 3 Issue 2, March 1996
Pages 35 - 46
ACM New York, NY, USA
table of contents doi>10.1145/227181.227186

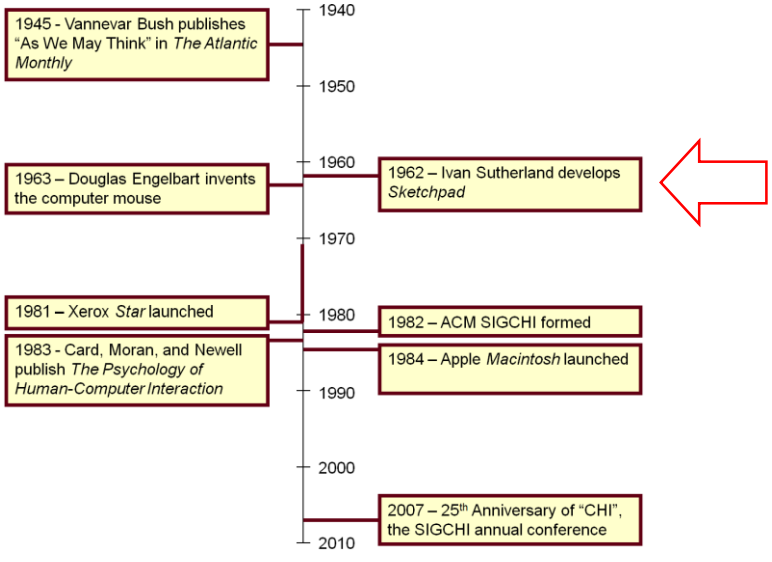
1996 Article

Bibliometrics
· Downloads (6 Weeks): 54
· Downloads (12 Months): 446
· Citation Count: 19

Click here

7

Significant Event Timeline



8

Sketchpad
Ivan Sutherland (1962)

9

Sketchpad

Heretofore, most interaction between man and computers has been slowed down by the need to reduce all communication to written statements that can be typed; in the past, we have been writing letters to rather than conferring with our computers. **(Sutherland, 1963)**

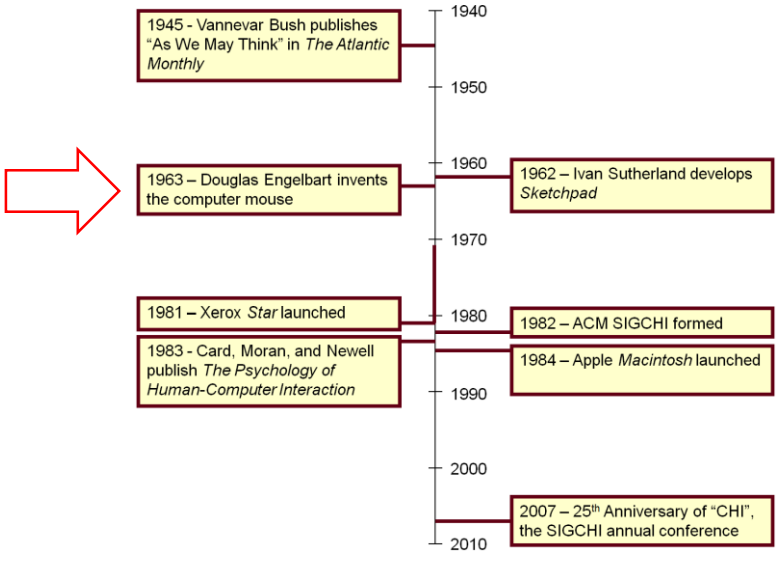
10

Sketchpad: "Direct Manipulation"

- Direct manipulation: correspond at least loosely to manipulation of physical objects
- Features:
  - Incremental action and rapid feedback
  - Reversibility
  - Exploration
  - Syntactic correctness of all actions
  - Replacing language with action
- Term coined by Ben Shneiderman[1]

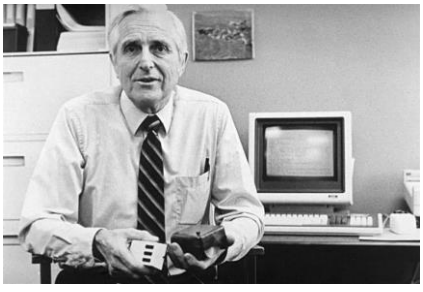
[1] Shneiderman, B., Direct manipulation: A step beyond programming languages, in *IEEE Computer*, 1983, August, 57-69.

Significant Event Timeline



12

Invention of the Mouse
Doug Engelbart (1963)



- **Turing award in 1997**
- **ACM SIGCHI Lifetime Achievement Award in 1998**

13

Read About Doug Engelbart at…

WIKIPEDIA
The Free Encyclopedia

Click here

DOUG ENGELBART INSTITUTE
"Boosting mankind's capability for coping with complex, urgent problems"
- Doug Engelbart

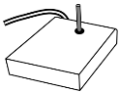
Click here

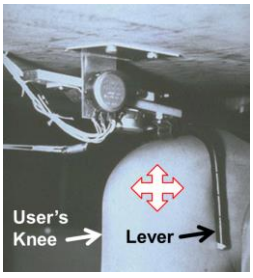
14

HCI's First User Study[1]

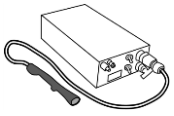
A comparative evaluation of…

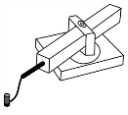

Mouse


Joystick



Knee-controlled lever


Lightpen


Grafacon

[1] English, W. K., Engelbart, D. C., & Berman, M. L. (1967). Display selection techniques for text manipulation. *IEEE Transactions on Human Factors in Electronics*, *HFE-8*(1), 5-15.

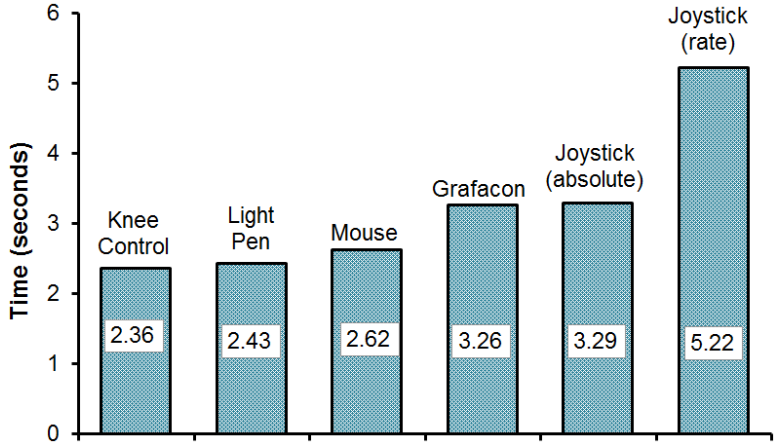
15

Experiment Design and Procedure

- Participants: 13
- Independent variable
  - "Input method" with six levels: mouse, light pen, Grafacon, joystick (position-control), joystick (rate-control), knee-controlled lever
- Dependent variables
  - Task completion time, error rate
  - (Note: task completion time = access time + motion time)
- Within-subjects, counterbalanced
- Task:
  - Press spacebar, acquire device, position cursor on target, select target

16

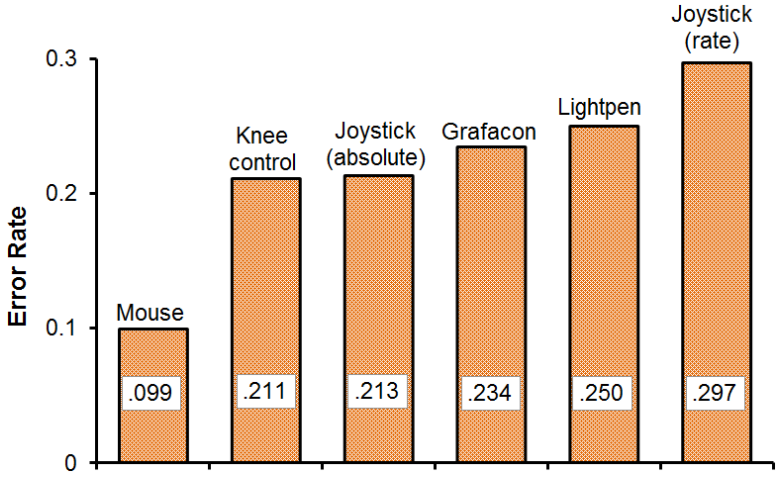
Results – Speed



Notes:
[1] Access time with the knee-controlled lever was zero (since the device is always "acquired").
[2] Light pen use is fatiguing, since the user's arm is held in the air in front of the display.

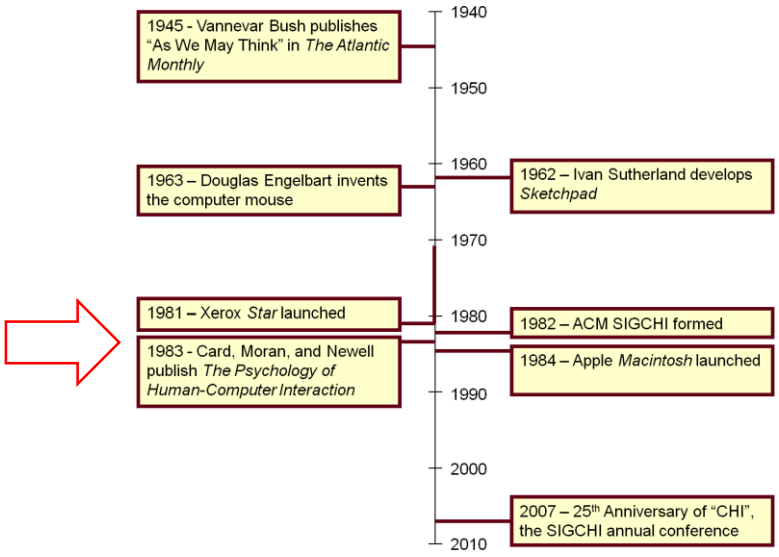
17

Results – Accuracy



18

Significant Event Timeline



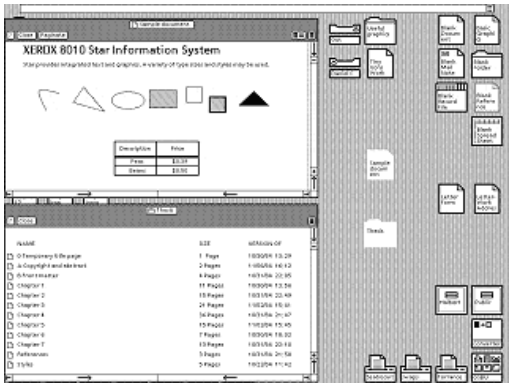
19

Xerox *Star*

- First commercially released computer system with a GUI (Graphical User Interface)
- It had windows, icons, menus and a pointing device (WIMP)
- It supported direct manipulation and what-you-see-is-what-you-get (WYSIWYG) interaction

20

Xerox *Star* (1981)



Price: 16,000 $

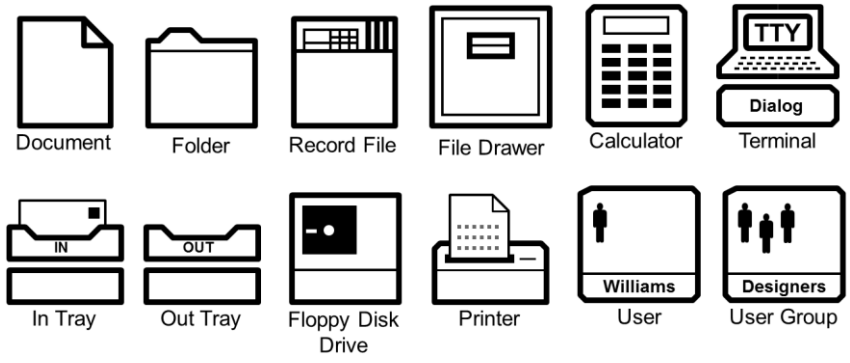
21

Desktop Metaphor

- Xerox Star used the Desktop Metaphor
  - Brings concepts from the office desktop to screen display: the user finds pictorial representations (icons) for things like documents, folders, trays and accessories
  - Metaphores are important in HCI: the user has existing knowledge from another domain
- Hidden details to increase usability: *open a document* instead of *invoke an editor*

22

Star GUI Icons



23

Birth of HCI - 1983

- Notable events:
  1. First ACM SIGCHI conference (1983)
  2. Publication of *The Psychology of Human-Computer Interaction* by Card, Moran, and Newell (1983)
  3. Apple *Macintosh* announced via brochures (December, 1983) and launched (January, 1984)

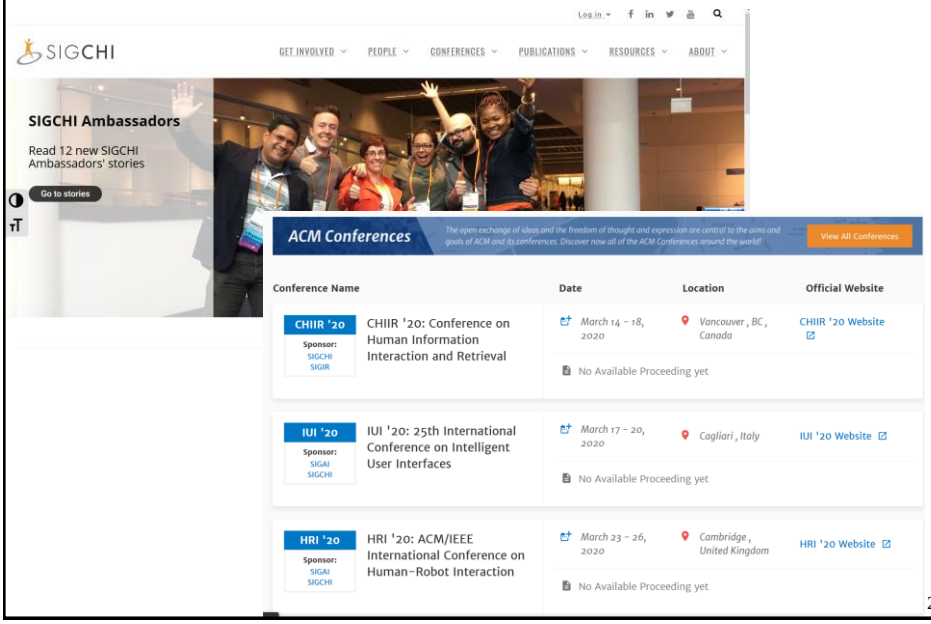
24

ACM SIGCHI Mission

The ACM Special Interest Group on Computer-Human Interaction is the world's largest association of professionals who work in the research and practice of computer-human interaction. This interdisciplinary group is composed of computer scientists, software engineers, psychologists, interaction designers, graphic designers, sociologists, and anthropologists, just to name some of the domains whose special expertise come to bear in this area. They are brought together by a shared understanding that designing useful and usable technology is an interdisciplinary process, and believe that when done properly it has the power to transform persons' lives.

25

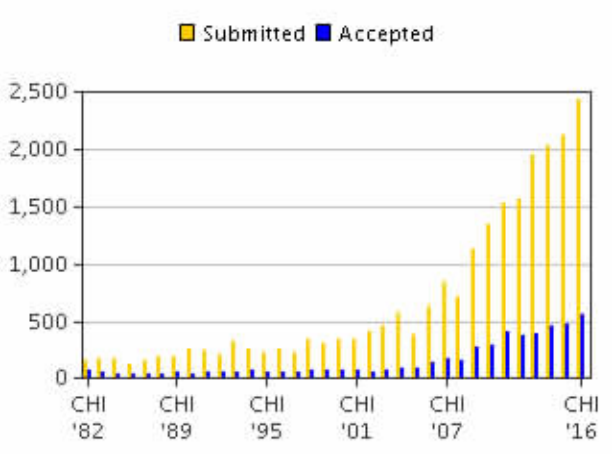
SIGCHI Web Site



26

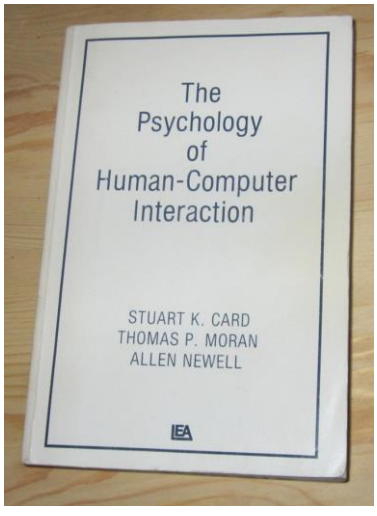
SIGCHI Conference Publications

The ACM CHI Conference on Human Factors in Computing Systems



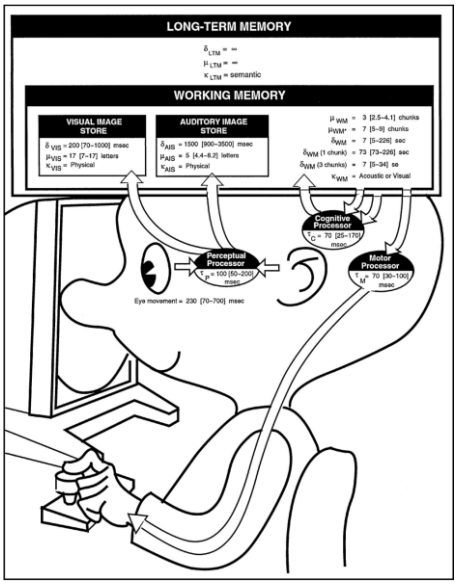
27

The Psychology of Human-Computer Interaction
Card, Moran, and Newell (1983)



28

The Model Human Processor

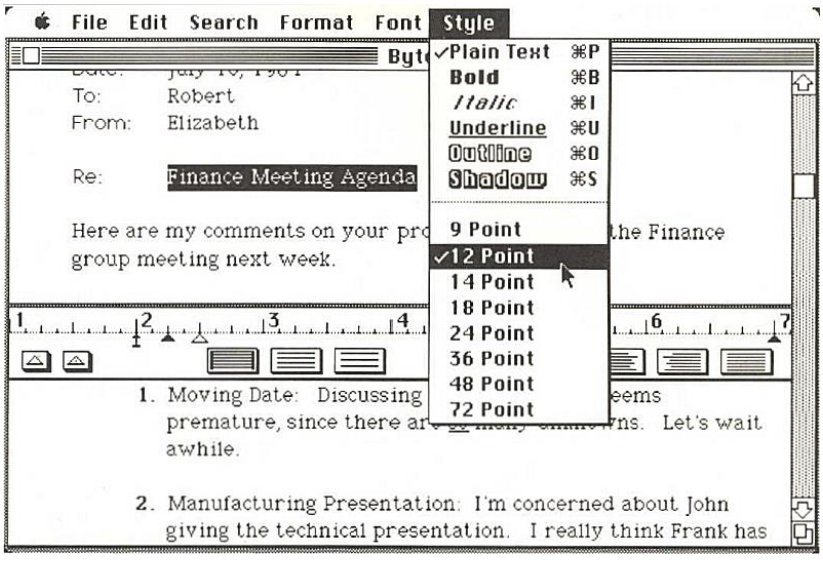


29

Apple *Macintosh* (1984)

30

MacWrite Software



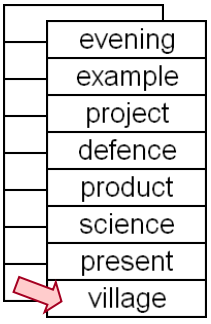
31

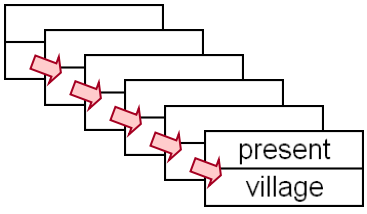
Apple *Macintosh* Commercial (1984)

32

Growth of HCI (1983-…)

- Example of an early research topic
  - Breadth vs. depth in menu design



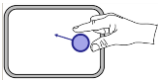
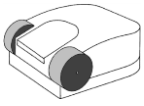
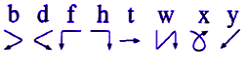


33

HCI Research

- Research precedes products
- Consider…
  - Two-finger gestures
    - Apple *iPhone*, 2007
  - Acceleration-sensing
    - Nintendo *Wiimote*, 2005
  - Wheel mouse
    - Microsoft *Intellimouse*, 1996
  - Single-stroke text input
    - Palm's *Graffiti*, 1995
- Were these ideas born out of engineering or design brilliance? Not really…

34

- Two-finger gestures: ~~2007?~~  **1978** [1]
- Acceleration-sensing: ~~2005?~~  **1998** [2]
- Wheel mouse: ~~1986?~~  **1993** [3]
- Single-stroke text input: ~~1995?~~  **1993** [4]

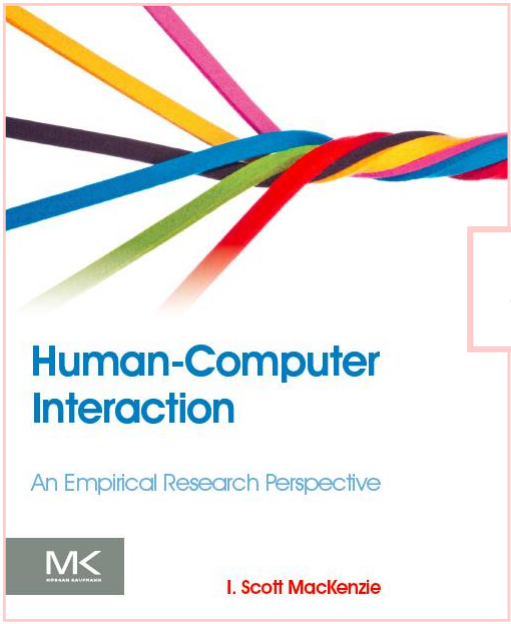
[1] Herot, C. F., & Weinzapfel, G. (1978). One-point touch input of vector information for computer displays. *Proc SIGGRAPH '78*, 210-216, New York: ACM.
[2] Harrison, B., Fishkin, K. P., Gujar, A., Mochon, C., & Want, R. (1998). Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. *Proc CHI '98*, 17-24, New York: ACM.
[3] Venolia, D. (1993). Facile 3D manipulation. *Proc CHI '93*, 31-36, New York: ACM.
[4] Goldberg, D., & Richardson, C. (1993). Touch-typing with a stylus. *Proc CHI '93*, 80-87, New York: ACM.

35

Connect to: http://join.quizizz.com

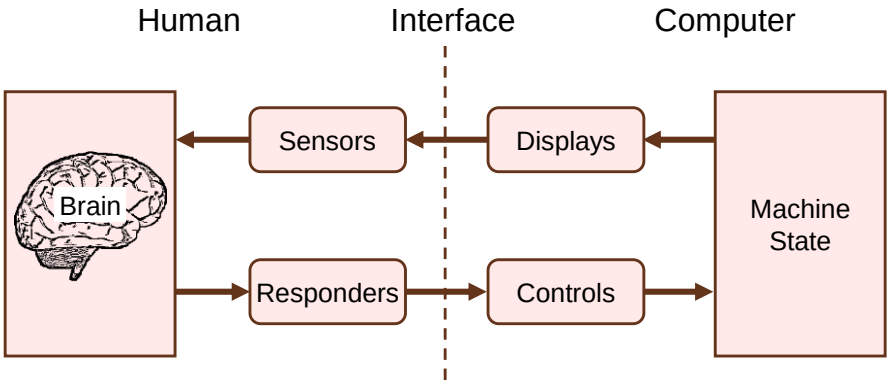
QUESTION TIME

36



Chapter 2
The Human Factor

Human Factors Model[1]

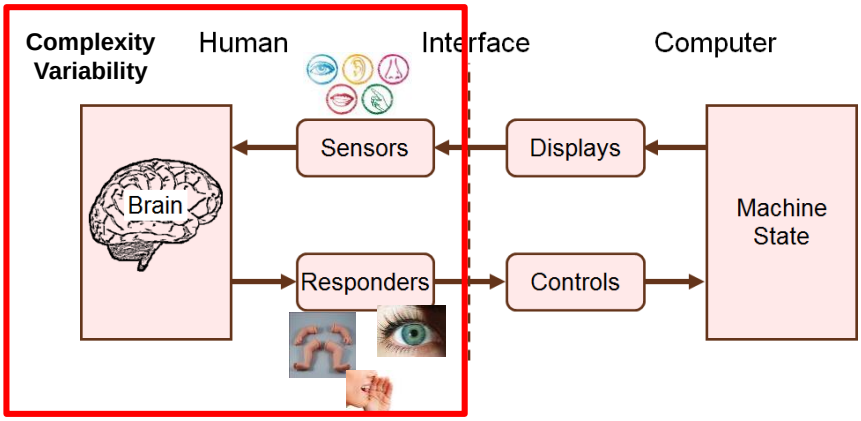


[1] Kantowitz, B. H., & Sorkin, R. D. (1983). *Human factors: Understanding people-system relationships*. New York. New York: Wiley.

2

2. The Human Factor

Kantowitz, B. H., & Sorkin, R. D. (1983). *Human factors: Understanding people-system relationships.* New York. New York: Wiley.



3

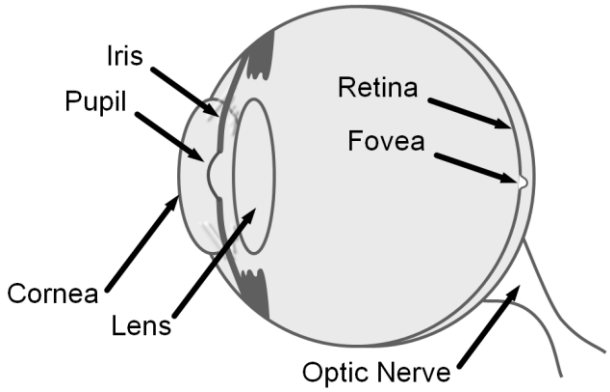
Human Senses

- Vision (sight)
- Hearing (audition)
- Touch (tactition)
- Smell
- Taste

4

Vision (The Eye)

- People obtain about 80% of their information through vision (the eye)



5

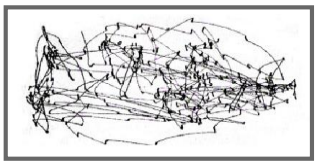
Fixations and Saccades

- Fixation
  - Eyes are stationary (dwell)
  - Take in visual detail from the environment
  - Long or short, but typically at least 200 ms
- Saccade
  - Rapid repositioning of the eye to fixate on a new location
  - Quick: ≈120 ms

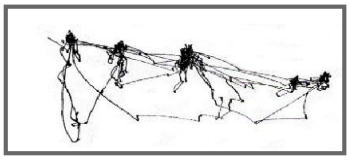
6

Yarbus' Eye Tracking Research (1965)[1]

The Unwanted Visitor
by Ilya Repin (1844-1930)



"Remember the position of people and objects in the room"

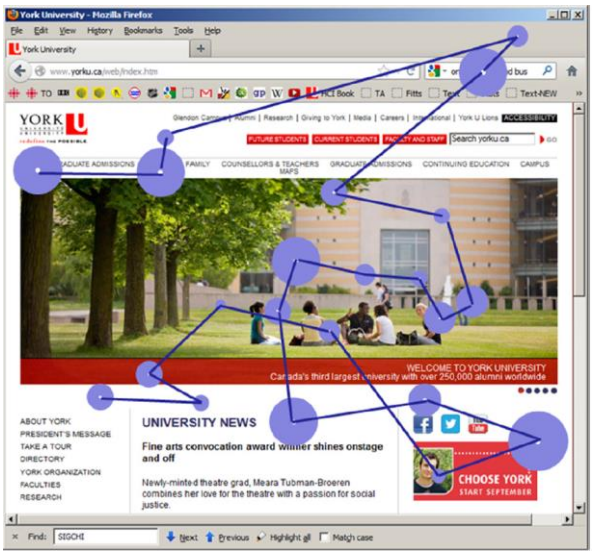


"Estimate the ages of the people"

[1] Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception, 1*, 7-27..

7

Eye Tracking Research



8

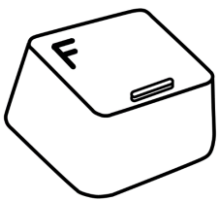
Hearing (Audition)

- Sound → cyclic fluctuations of pressure in a medium, such as air
- Created when physical objects are moved or vibrated
- Physical properties of sound…
  - Frequency
  - Intensity

9

Touch (Tactition)

- Part of somatosensory system, with…
- Receptors in skin, muscles, joints, bones
  - Sense of touch, pain, temperature, position, shape, texture, resistance, etc.
- Tactile feedback examples:



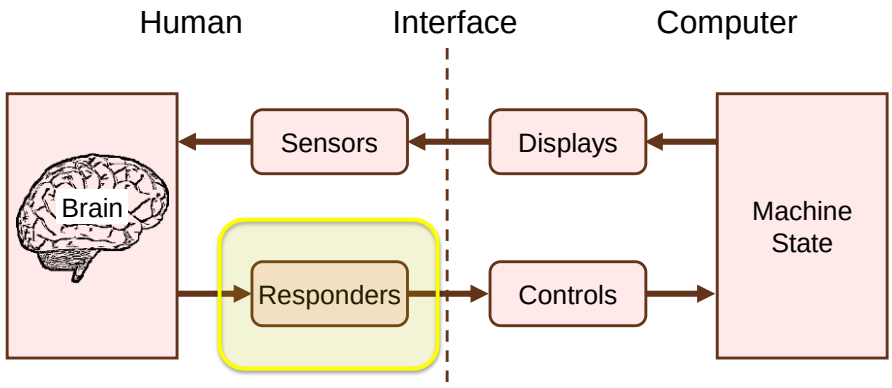
10

Smell and Taste

- Smell (olfaction)
  - Ability to perceive odours
  - Occurs through sensory cells in nasal cavity
- Taste (gustation)
  - Chemical reception of sweet, salty, bitter, and sour sensations
- Flavour
  - A perceptual process that combines smell and taste
- Only a few examples in HCI (e.g., Brewster et al., 2006; Bodnar et al., 2004)

11

Human Factors Model



12

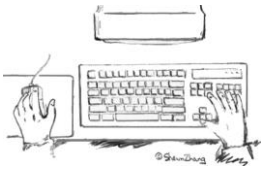
Responders

- Humans control their environment through responders, for example…
  - A finger to text or point
  - Feet to walk or run
  - Eyebrow to frown
  - Vocal chords to speak

13

Handedness

- Some users are left handed, others right handed. Important in interaction design.



- Handedness exists by degree
- Edinburgh Handedness Inventory used to measure handedness (next slide)

14

Edinburgh Inventory for Handedness[1]

	Left	Right
1. Writing	☐☐	☐☐
2. Drawing	☐☐	☐☐
3. Throwing	☐☐	☐☐
4. Scissors	☐☐	☐☐
5. Toothbrush	☐☐	☐☐
6. Knife (without fork)	☐☐	☐☐
7. Spoon	☐☐	☐☐
8. Broom (upper hand)	☐☐	☐☐
9. Striking a match	☐☐	☐☐
10. Opening box (lid)	☐☐	☐☐
Total (count checks)	☐	☐

Difference	Cumulative Total	**RESULT**
☐	☐	☐

Instructions

Mark boxes as follows:

x	preference
xx	strong preference
blank	no preference

Scoring

Add up the number of checks in the "Left" and "Right" columns and enter in the "Total" row for each column. Add the left total and the right total and enter in the "Cumulative Total" cell. Subtract the left total from the right total and enter in the "Difference" cell. Divide the "Difference" cell by the "Cumulative Total" cell (round to 2 digits if necessary) and multiply by 100. Enter the result in the "RESULT" cell.

Interpretation of RESULT

-100 to -40	left-handed
-40 to +40	ambidextrous
+40 to 100	right-handed

[1] Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsycholologia, 9*, 97-113.

15

Challenge (home)

- Use a spreadsheet to implement the Edimburgh Inventory (see right)
- Calculate your score

	Left	Right	
1. Writing		2	
2. Drawing		2	
3. Throwing		2	
4. Scissors		1	
5. Toothbrush		2	
6. Knife		1	
7. Spoon		2	
8. Broom		1	
9. Striking a match		1	
10. Opening box	1		
TOT	1	14	
DIFFERENCE	13		
CUMULATIVE TOT	15		
Result	**87**	**Right-handed**	

16

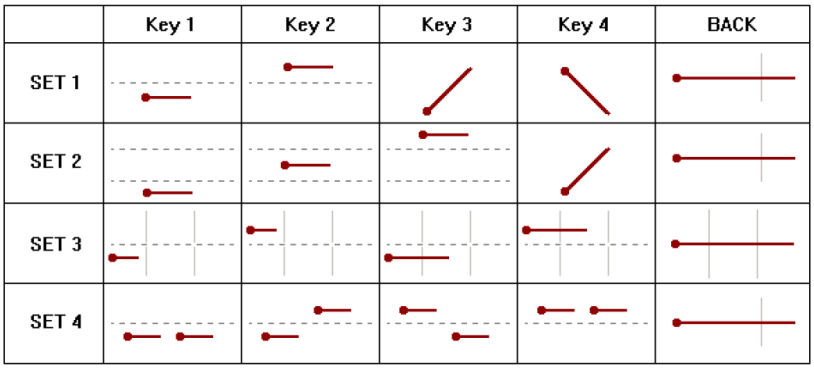
Human Voice

- Human vocal chords are responders
- Sounds created through combination of…
  - Movement in the larynx
  - Pulmonary pressure in the lungs
- Two kinds of vocalized sounds:
  1. Speech
  2. Non-speech
- Both with potential for computer control
  - Speech + speech recognition
  - Non-speech + signal detection (e.g., frequency, loudness, duration, change direction, etc.)

17

Non-speech Example[1]

- NVVI = non-verbal voice interaction



[1] Sporka, A., Felzer, T., Kruniawan, S., Polacek, O., Haiduk, P., & MacKenzie, I. S. (2011). CHANTI: Predictive text entry using non-verbal vocal input. *Proceedings of the ACM Conference on Human Factors in Computing Systems – CHI 2011*, 2463-2472.New York: ACM.

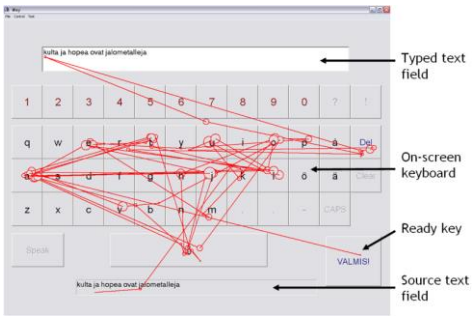
18

The Eye as a Responder

- As a controller, the eye is called upon to do "double duty"
  1. Sense and perceive the environment/computer
  2. Act as a controller via saccades and fixations

19

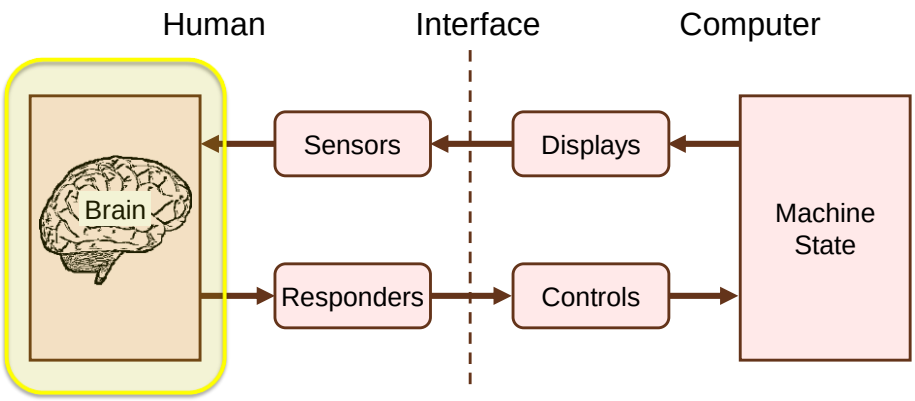
Example - Eye Typing[1]



[1] Majaranta, P., MacKenzie, I. S., Aula, A., & Räihä, K.-J. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Universal Access in the Information Society (UAIS), 5*, 199-208.

20

Human Factors Model



21

The Brain

- Most complex biological structure known
- Billions of neurons
- Sensors (human inputs) and responders (human outputs) are nicely mirrored, but it is the brain that connects them
- Brain functions:
  - Perception
  - Cognition
  - Memory
  - Language

22

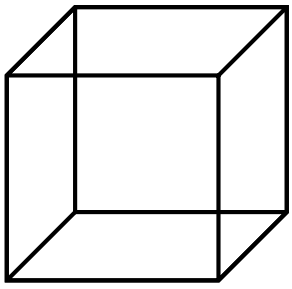
Perception

- 1st stage of processing for sensory input
- Psychophysics: branch of experimental psychology, studied since the 19th century
- Determine the *just noticeable difference* (JND): threshold below which the subject deems the two stimuli "the same"
- Experimental method:
  - Present subject with two stimuli, one after the other
  - Stimuli differ in a physical property (e.g., frequency)
  - Randomly vary the difference

23

Ambiguity

Necker cube



Which surface is at the front?

Rubin vase

Wine goblet or two faces?

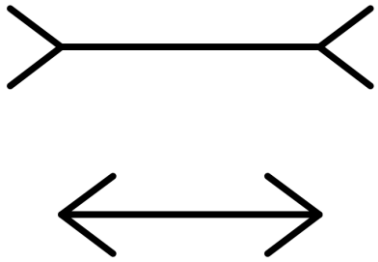
24

Illusion

Ponzo lines

Which black line is longer?

Műller-Lyer arrows



Which horizontal line is longer?

25

Cognition

- Cognition is the human process of conscious intellectual activity
  - E.g., thinking, reasoning, deciding
- Spans many fields
  - E.g., neurology, linguistics, anthropology
- Sensory phenomena → easy to study because they exist in the physical world
- Cognitive phenomena → hard to study because they exist within the human brain

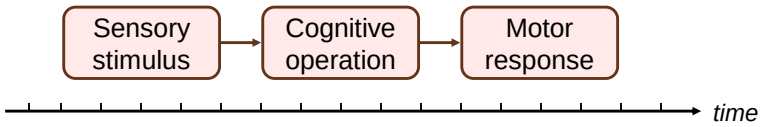
26

"Making a Decision"

- Not possible to directly measure the time for a human to "make a decision"
- When does the measurement begin and end?
- Where is it measured?
- On what input is the human deciding?
- Through what output is the decision conveyed?
- There is a sensory stimulus and motor response that bracket the decision (next slide)

27

Making a Decision – in Parts



Operation	Typical time (ms)
Sensory reception	1 – 38
Neural transmission to brain	2 – 100
Cognitive processing	70 – 300
Neural transmission to muscle	10 – 20
Muscle latency and activation	30 –70
Total:	113 - 528

Large variation!

28

Memory

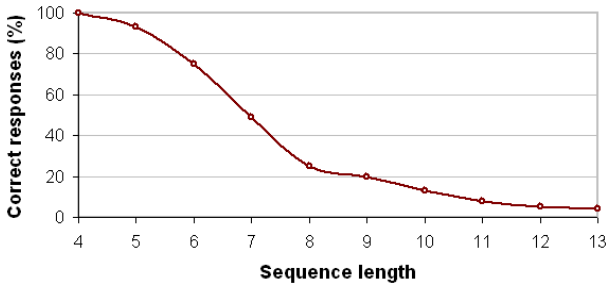
- Vast repository
- Long-term memory
  - Declarative/explicit area → information about events in time and objects in the external world
  - Implicit/procedural area → information about how to use objects and how to do things
- Short-term memory
  - Aka *working memory*
  - Information is active and readily available for access
  - Amount of working memory is small, about 7 (±2) units or chunks[1]

[1] Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

29

Short Term Memory Experiment

- Random sequences of digits recited to subjects
- Sequences vary from 4 to 13 digits
- After recitation, subjects copy sequence from memory to a sheet of paper
- Transcriptions on sheets scored (correct/incorrect)
- Results ($n \approx 60$):



30

Language

- The mental faculty that allows humans to communicate
  - As speech, available to (almost) all humans without effort
  - As writing, only available with considerable effort
- HCI interest: primarily in writing, creation of text

> Humankind is defined by language; but civilization is defined by writing.[1]

[1] Daniels, P. T., & Bright, W. (Eds.). (1996). *The world's writing systems*. New York: Oxford University Press.

31

Corpus

- One way to characterise written text is a corpus
- Large collection of representative text samples
- A corpus may be reduced to a word-frequency list:

Word Rank	English	French	German	Finnish	SMS English	SMS Pinyin
1	the	de	der	ja	u	wo (我)
2	of	la	die	on	i	ni (你)
3	and	et	und	ei	to	le (了)
4	a	le	in	että	me	de (的)
5	in	à	den	oli	at	bu (不)
…	…	…	…	…	…	…
1000	top	ceci	konkurrenz	muista	ps	jiu (舅)
1001	truth	mari	stieg	paikalla	quit	tie (贴)
1002	balance	solution	notwendig	varaa	rice	ji (即)
1003	heard	expliquer	sogenannte	vie	sailing	jiao (角)
1004	speech	pluie	fahren	seuran	sale	ku (裤)
…	…	…	…	…	…	…

32

Statistics and Language

- Native speakers intuitively understand the statistical nature of their language
- We…
  - Anticipate letters

    Questio_
  - Anticipate words:

    A picture is worth a thousand _____.
  - Anticipate entire phrases:

    To be or __ __ __.

33

SMS Shorthand

- The final frontier!
- A 13-year-old student's essay (excerpt)[1]

My smmr hols wr CWOT. B4, we used 2go2 NY 2C my bro, his GF & thr 3 :- kids FTF. ILNY, it's a gr8 plc.

- 102 characters

- Original (for the teacher to deduce)

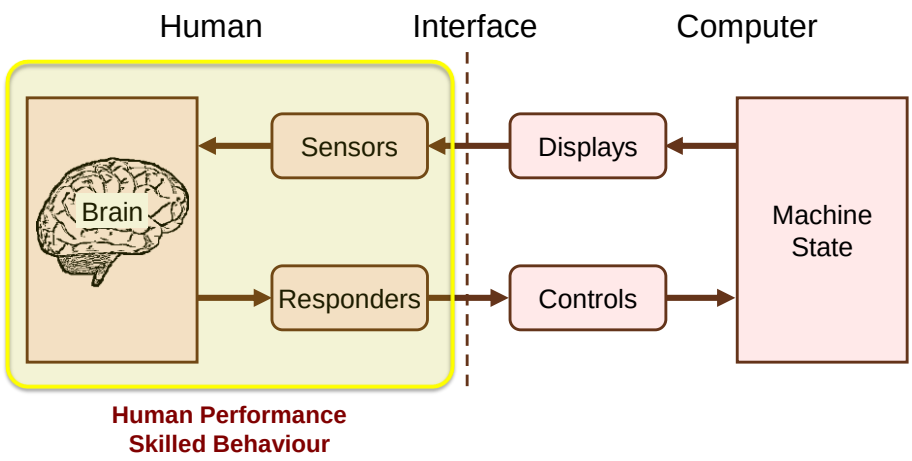
My summer holidays were a complete waste of time. Before, we used to go to New York to see my brother, his girlfriend and their three screaming kids face to face. I love New York. It's a great place.

- 199 characters

34

[1] http://news.bbc.co.uk/2/hi/uk_news/2814235.stm

Human Factors Model



35

Human Performance

- Humans use their sensors, brain, and responders to do things
- When the three work together to achieve a *goal*, human performance arises
- Examples:
  - Tying shoelaces
  - Folding clothes
  - Searching the web
  - Entering a text message on a mobile phone

36

Speed-accuracy Trade-off

- Fundamental property of human performance
- Go faster and errors increase
- Slow down and accuracy improves
- HCI research on a new interface or interaction technique must consider both the speed in doing tasks (achieving the goal!) and the accompanying accuracy

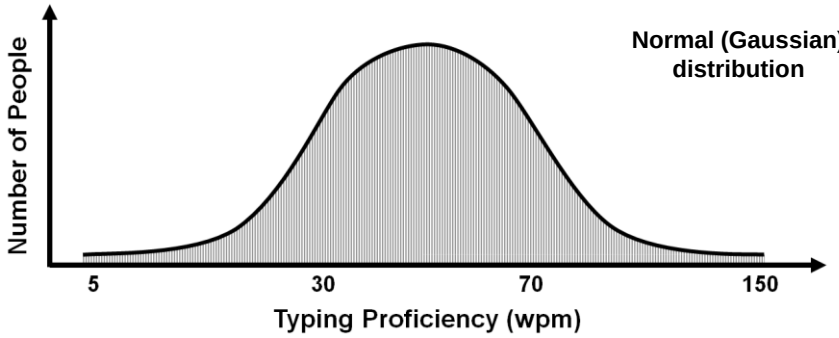
37

Human Diversity

- Human performance is highly complex:
  - Humans differ (age, gender, skill, motivation, etc.)
  - Environmental conditions affect performance
  - Secondary tasks often present
- Human diversity and human performance often shown in a distribution (next slide)

38

Human Diversity and Performance



Where are you on this chart?
Where is your mother?
Where is an 8-year old, just learning to use a computer?
Where is someone with a physical disability?
Where are you while using your mobile phone on a crowded bus (standing!)?

39

Reaction Time

- One of the most primitive manifestations of human performance is *simple reaction time*
- Definition: The delay between the occurrence of a single fixed stimulus and the initiation of a response assigned to it[1]
- Example: pressing a button in response to the onset of a stimulus light

[1] Fitts, P. M., & Posner, M. I. (1968). *Human performance*. Belmont, CA. Brooks/Cole Publishing Company.

40

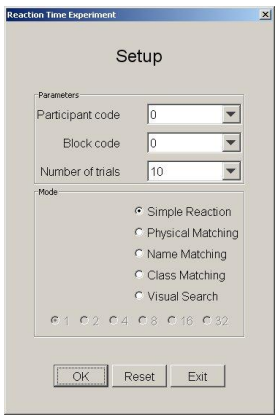
Sensory Stimuli and Reaction Time

- Delay time varies by type of sensory stimuli
- Approximate values[1]
  - Auditory → 150 ms
  - Visual → 200 ms
  - Smell → 300 ms
  - Pain → 700 ms

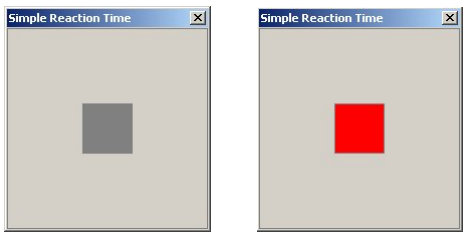
[1] Bailey, R. W. (1996). *Human performance engineering: Designing high quality, professional user interfaces for computer products, applications, and systems* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

41

Reaction Time Experiment



Simple Reaction Time



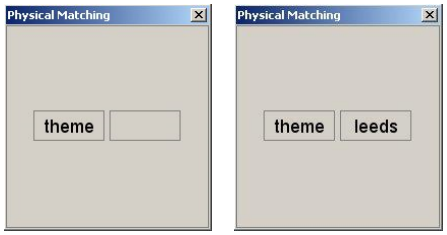
Use software from book's web site:

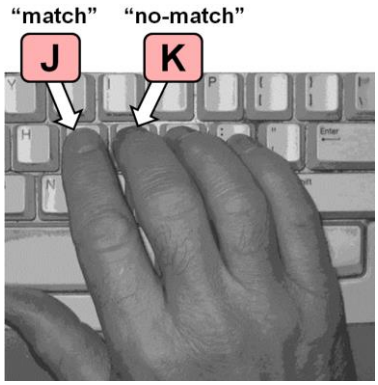
http://www.yorku.ca/mack/HCIbook/

42

Reaction Time Experiment (2)

Physical Matching

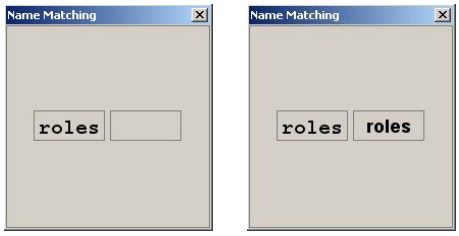




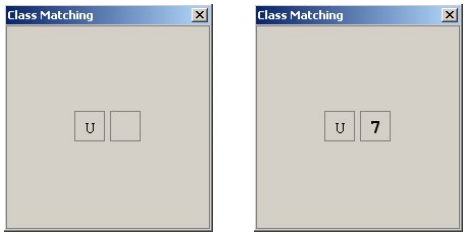
43

Reaction Time Experiment (3)

Name Matching

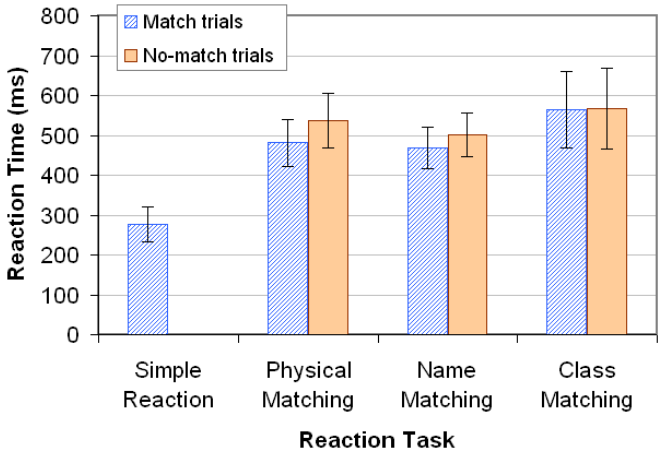


Class Matching



44

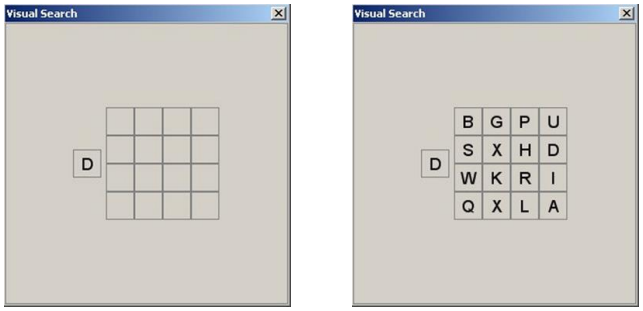
Experiment Results



45

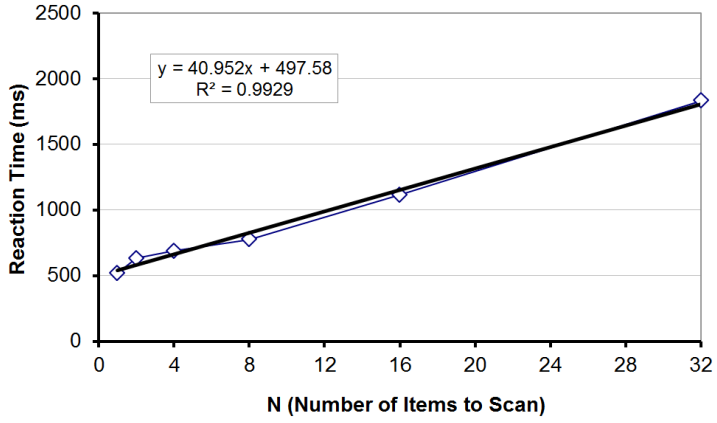
Visual Search

- A variation on physical matching
- User scans a collection of items looking for desired item
- Time increases with the number of items to scan
- Included in the demo software with N = 1, 2, 4, 8, 16, or 32 items



46

Experiment Results (1)



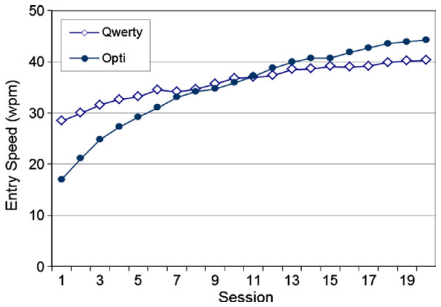
47

Skilled Behaviour

- For many tasks, human performance improves considerably and continuously with practice
- (Note: Very little improvement with practice in the simple reaction time tasks)
- In these tasks, there is interest in studying the progression of learning and the performance achieved according to the amount of practice
- Categories of skilled behavior:
  1. Sensory-motor skill (e.g., darts, gaming)
  2. Mental skill (e.g., chess, programming)
  - Some tasks required a lot of both (next slide)

48

Skilled Behaviour



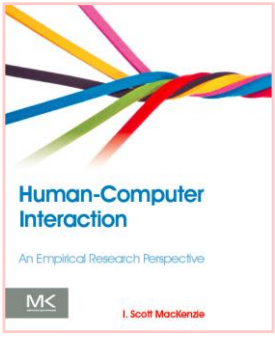
49

Connect to: http://join.quizizz.com

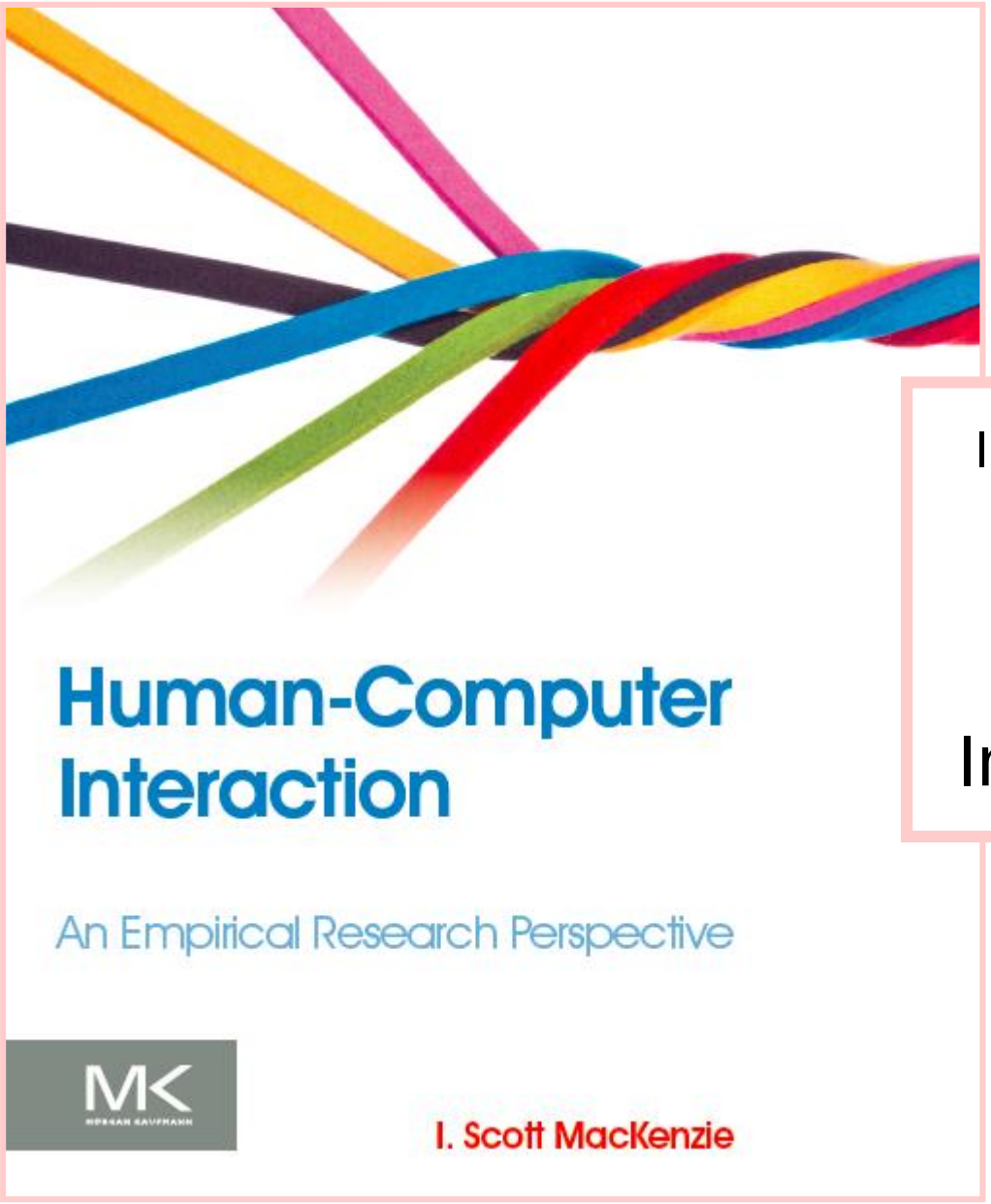
QUESTION TIME

50

Thank You


Human-Computer
Interaction
An Empirical Research Perspective
MK
I. Scott MacKenzie

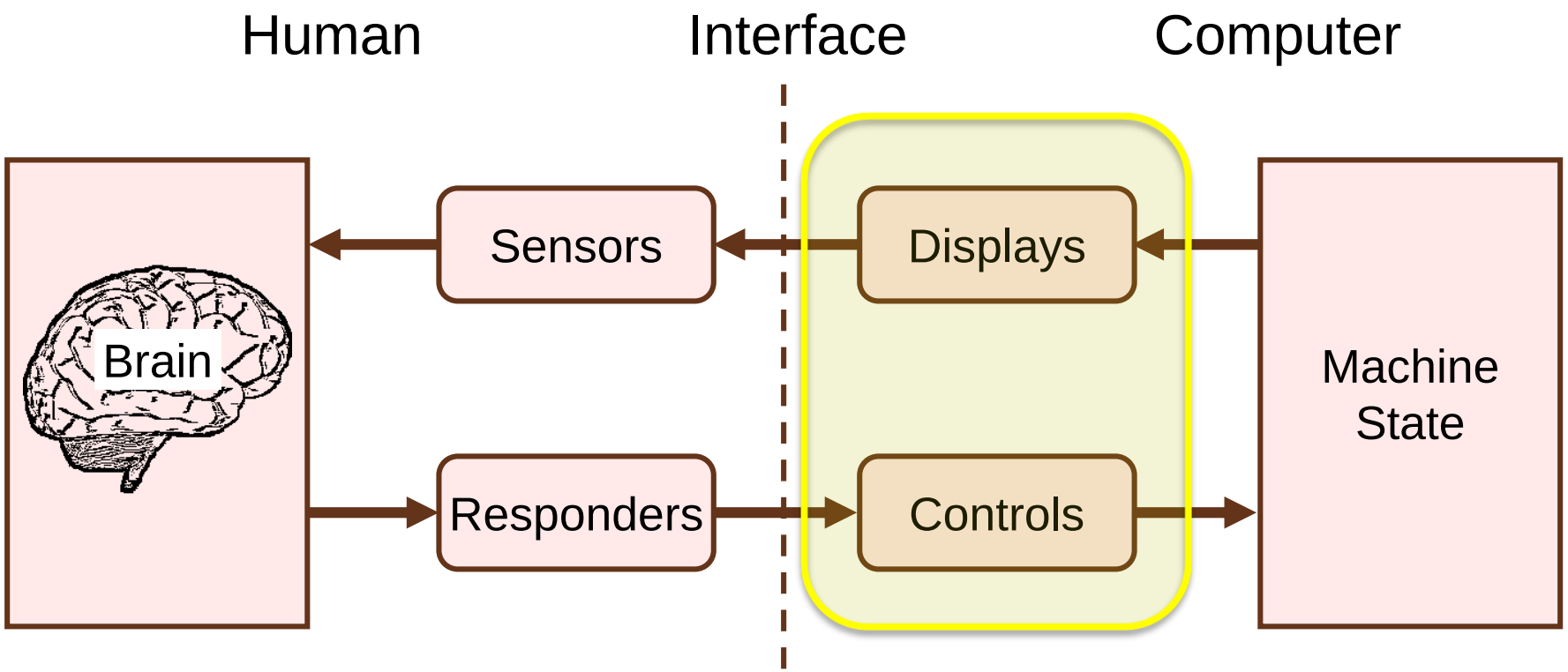
51



Interactive Systems Design
Prof. V. Fuccella

Chapter 3
Interaction Elements

Human Factors Model (revisited)



[1] Kantowitz, B. H., & Sorkin, R. D. (1983). *Human factors: Understanding people-system relationships*. New York. New York: Wiley.

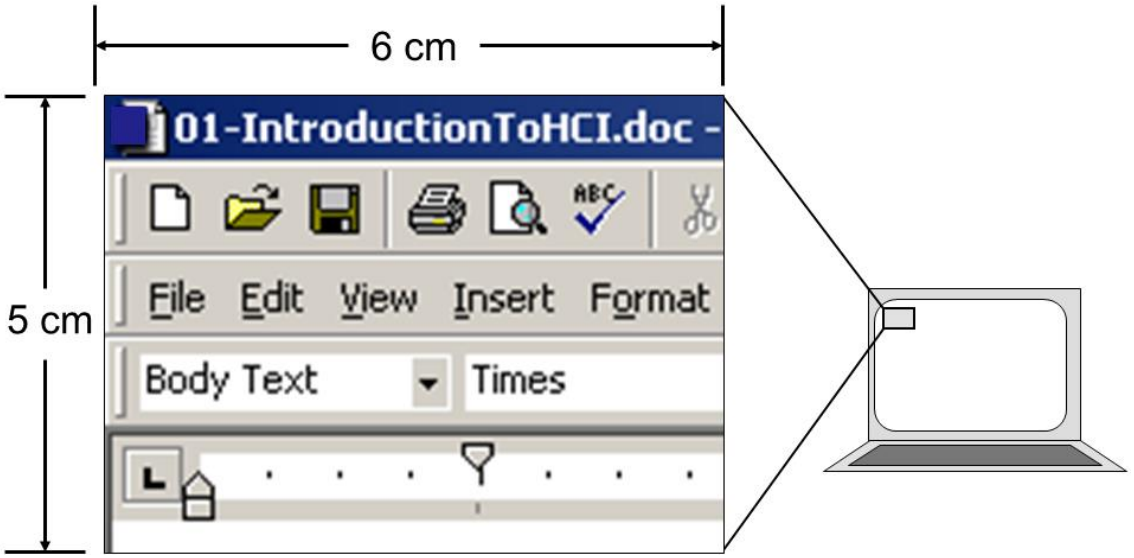
Hard Controls, Soft Controls

- In the past, controls were physical, single-purpose devices → *hard controls*
- Today's graphical displays are malleable
- Interfaces created in software → *soft controls*
- Soft controls rendered on a display
- Distinction blurred between soft controls and displays
- Consider controls to format ***this*** (see below)

Soft controls are also displays!

GUI Malleability

- Below is a 30 cm^2 view into a GUI
- >20 soft controls (or are they displays?)

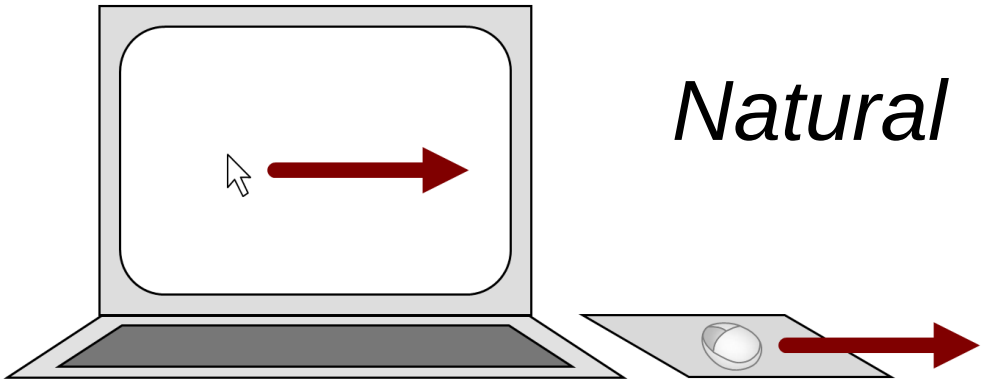


- Click a button and this space is morphed into a completely different set of soft controls/displays

Control-Display Relationships

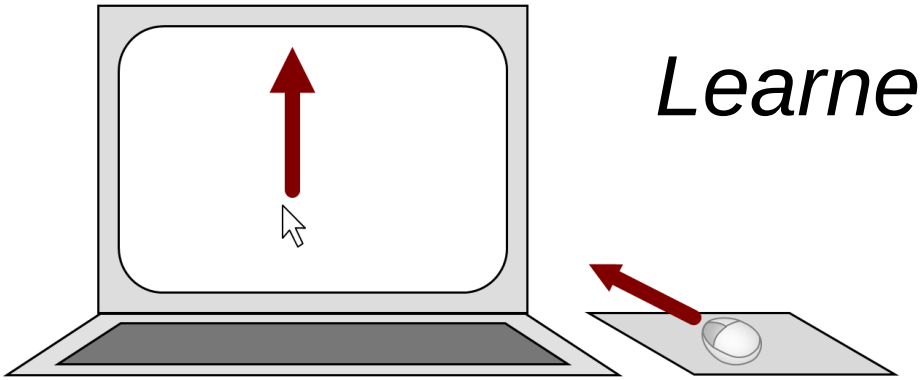
- Also called *mappings*
- Relationship between operation of a control and the effect created on a display
- At least three types:
  - Spatial relationships
  - Dynamic relationships
  - Physical relationships

Spatial Relationships



Natural

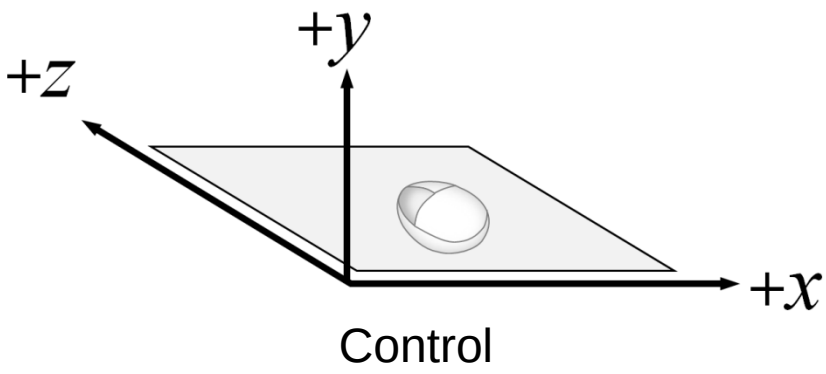
Spatial congruence
Control: right
Display: right



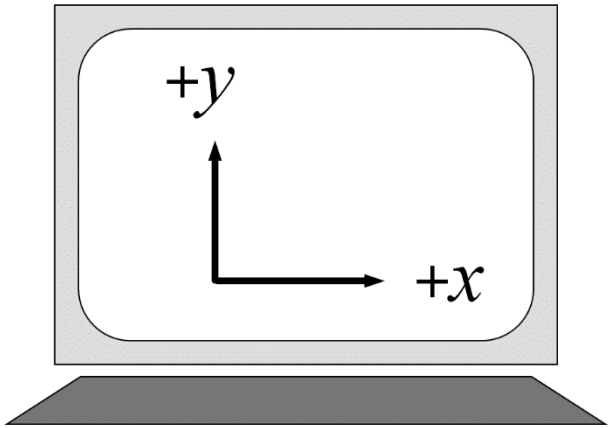
Learned

Spatial transformation
Control: forward
Display: up

Axis Labeling



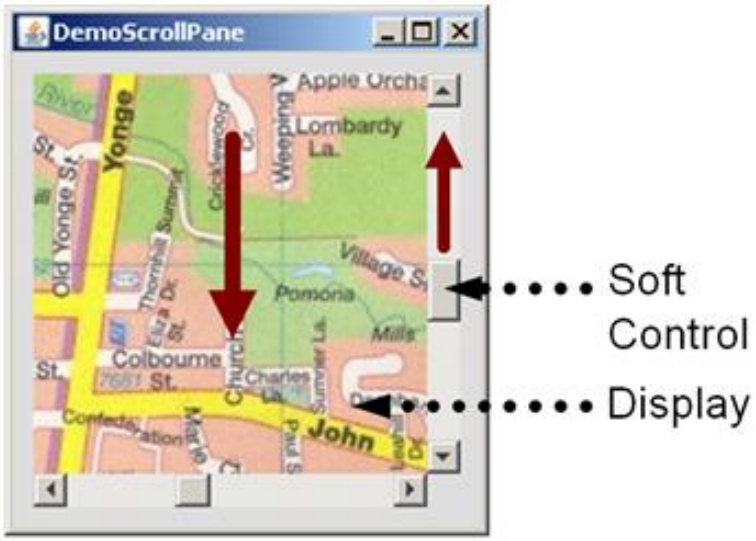
Control



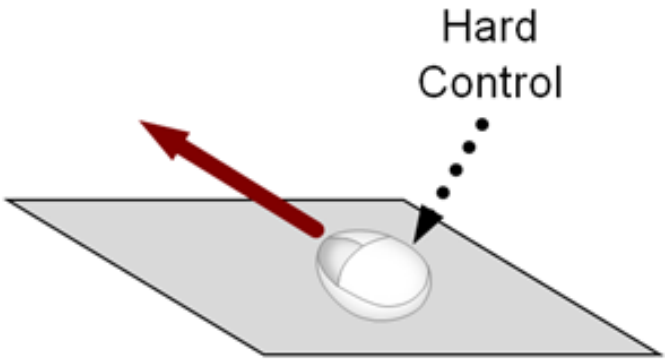
Display

Axis	Control (mouse)	Display (cursor)
x	+	+
y		+
z	+	

Third Tier



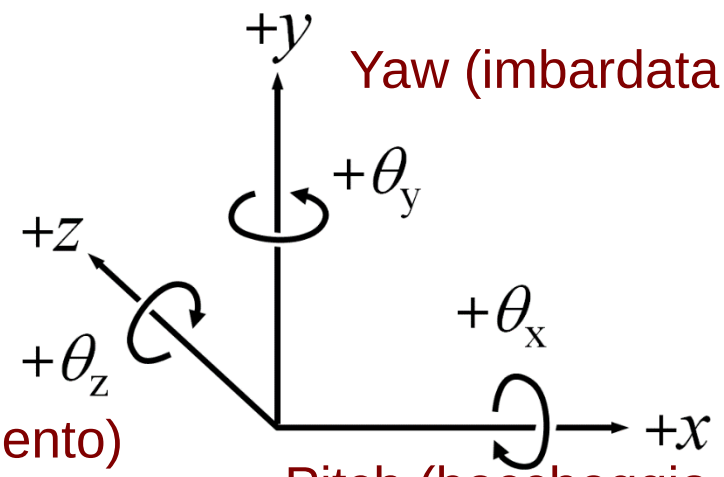
DOF	Hard Control	Soft Control	Display
x			
y		+	-
z	+		



3D

- In 3D there are 6 degrees of freedom (DOF). DOF are the number of variables needed to represent its position in the space
  - 3 DOF for position (x, y, z)
  - 3 DOF for orientation (θ_x, θ_y, θ_z)

In aeronautics…



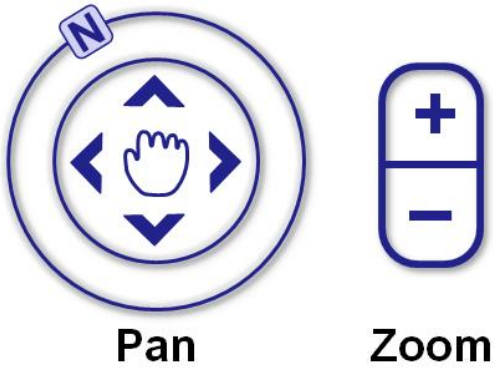
Yaw (imbardata o serpeggiamento)

Roll (rollio o coricamento)

Pitch (beccheggio o impennata)

3D in Interactive Systems

- Usually a subset of the 6 DOF are supported
- Spatial transformations are present and must be learned
- E.g., Google StreetView



Panning in Google StreetView

- (Switch to Google StreetView and demonstrate panning with the mouse)
- Spatial transformations:

DOF	Control	Display
x		
y		
z		
θ_x		
θ_y		
θ_z		

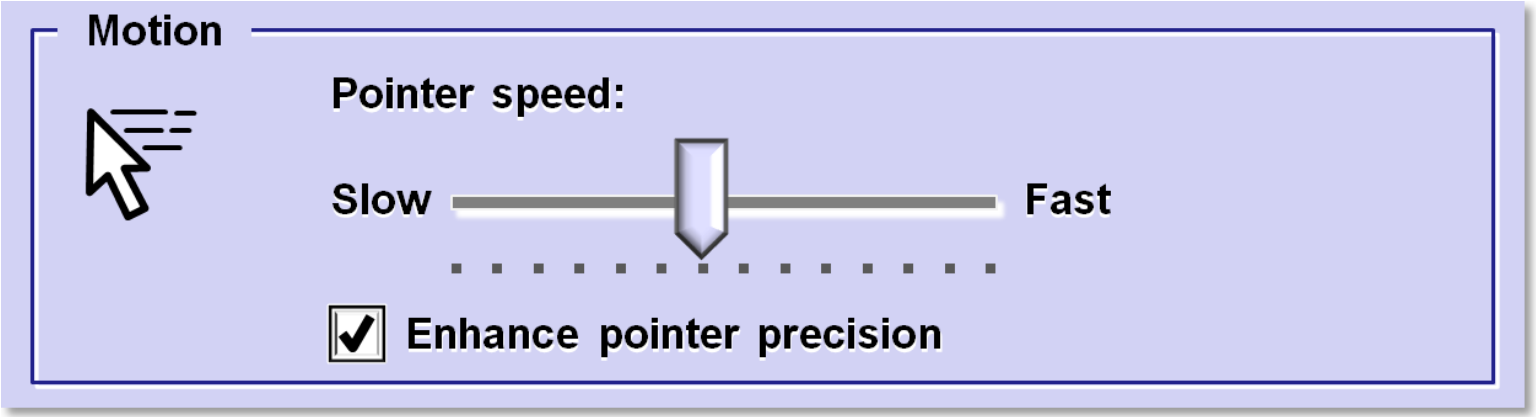
Panning in Google StreetView

- (Switch to Google StreetView and demonstrate panning with the mouse)
- Spatial transformations:

DOF	Control	Display
x	+	
y		
z	+	
θ_x		+
θ_y		-
θ_z		

CD Gain

- Quantifies the amount of display movement for a given amount of controller movement
- E.g., CD gain = 2 implies 2 cm of controller movement yields 4 cm of display movement
- For non-linear gains, the term *transfer function* is used
- Typical control panel to adjust CD gain:



CD Gain and User Performance

- Tricky to adjust CD gain to optimize user performance
- Issues:
  - Speed accuracy trade-off (what reduces positioning time tends to increase errors)
  - Opposing relationship between gross and fine positioning times:

Latency

- *Latency* (aka *lag*) is the delay between an input action and the corresponding response on a display
- Usually negligible on interactive systems (e.g., cursor positioning, editing)
- May be “noticeable” in some settings; e.g.,
  - Remote manipulation
  - Internet access (and other “system” response situations)
  - Virtual reality (VR)
- Human performance issues appropriate for empirical research

Property Sensed, Order of Control

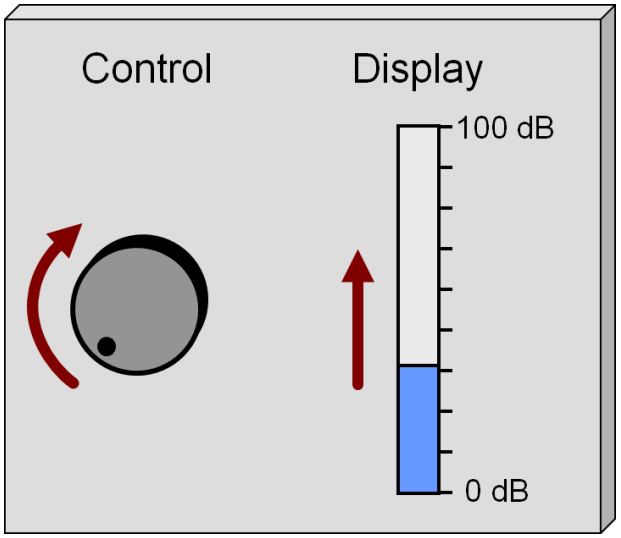
- Property sensed (property of control)
  - Position (graphics tablet, touchpad, touchscreen)
  - Displacement (mouse, joystick)
  - Force (joystick, pointing stick)
- Order of control (property of display)
  - Position (of cursor/object)
  - Velocity (of cursor/object)

Natural vs. Learned Relationships

- Natural relationships → spatially congruent
- Learned relationships → spatial transformation (relationship must be learned)

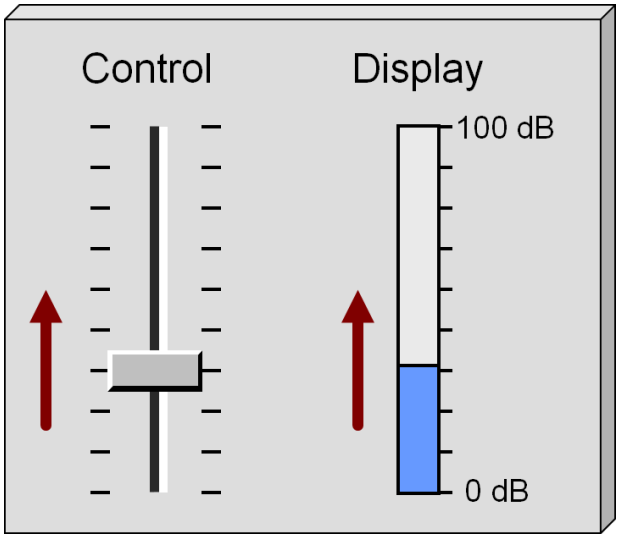
Examples

Learned relationship



DOF	Control	Display
x		
y		+
z		
θ_x		
θ_y		
θ_z	+	

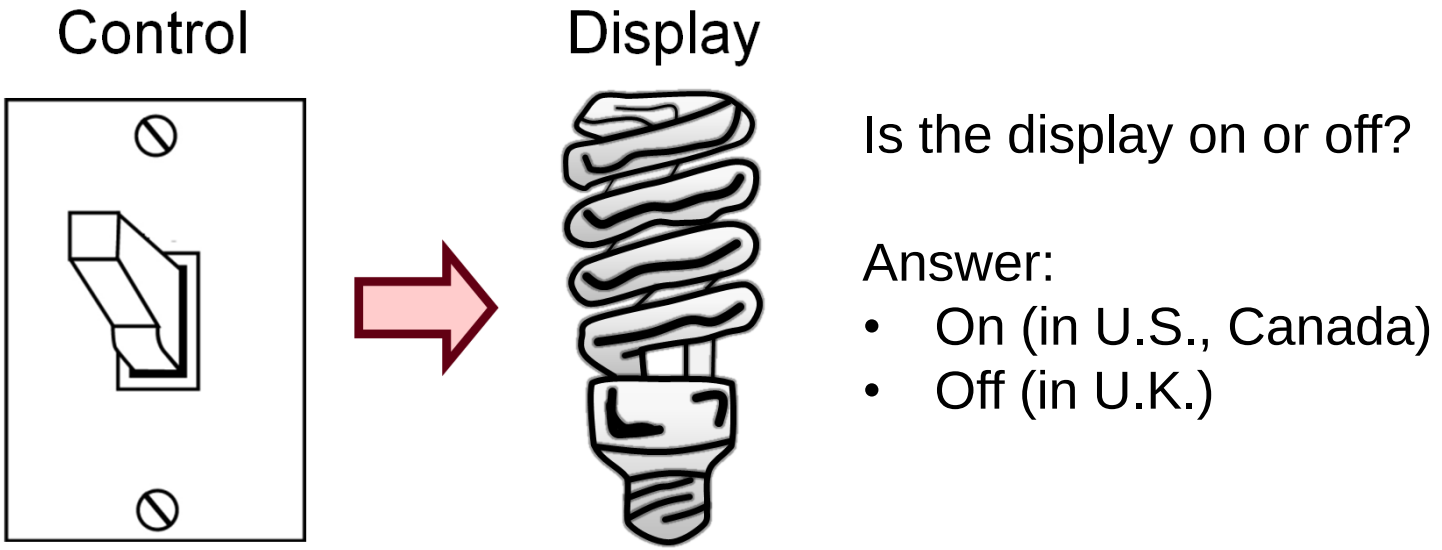
Natural relationship



DOF	Control	Display
x		
y	+	+
z		
θ_x		
θ_y		
θ_z		

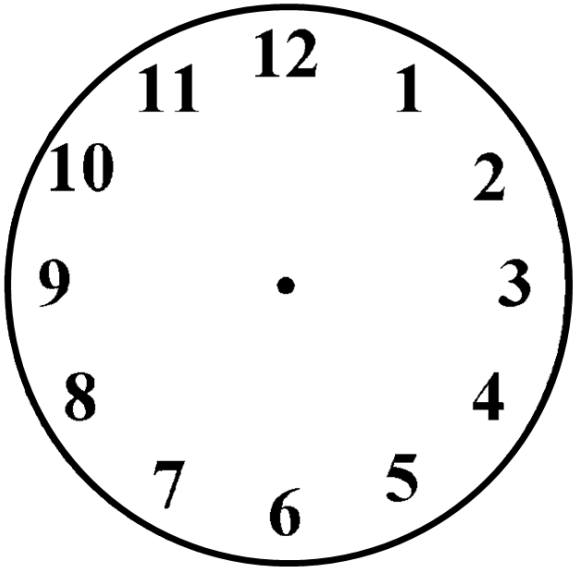
Learned Relationships

- Learned relationships seem natural if they lead to a *population stereotype* or *cultural standard*
- A control-display relationship needn't be a spatial relationship…



Clock Metaphor

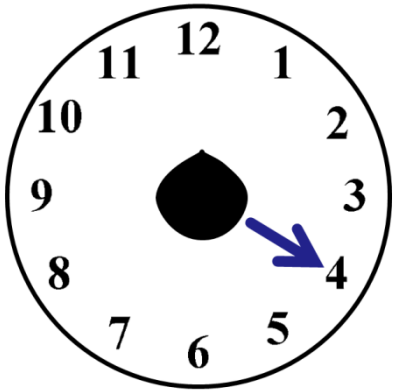
- Numeric entry on PDA[1]
- Users make straight-line strokes in direction of digit on clock face

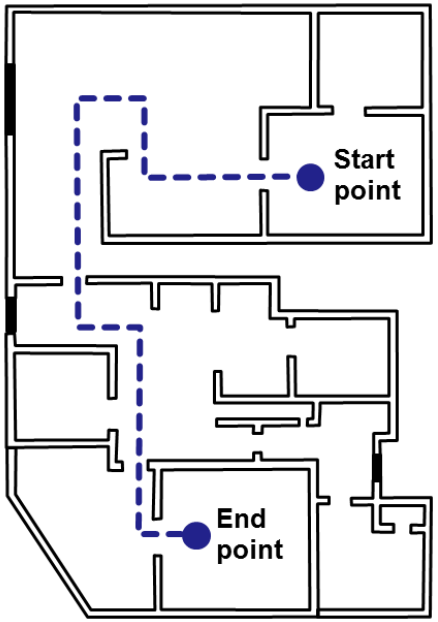


[1] McQueen, C., MacKenzie, I. S., & Zhang, S. X. (1995). An extended study of numeric entry on pen-based computers. *Proceedings of Graphics Interface '95*, 215-222, Toronto: Canadian Information Processing Society.

Clock Metaphor (2)

- Blind users carry a mobile locating device[1]
- Device provides spoken audio information about nearby objects (e.g. “door at 3 o’clock”)

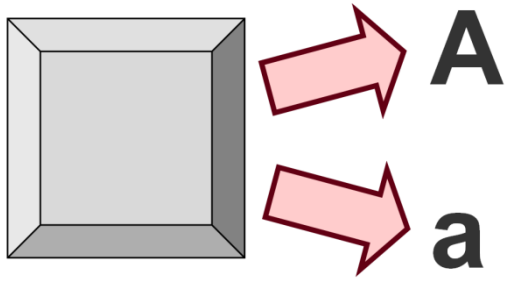




[1] Sáenz, M., & Sánchez, J. (2009). Indoor position and orientation for the blind. *Proceedings of HCI International 2007*, 236-245, Berlin: Springer.

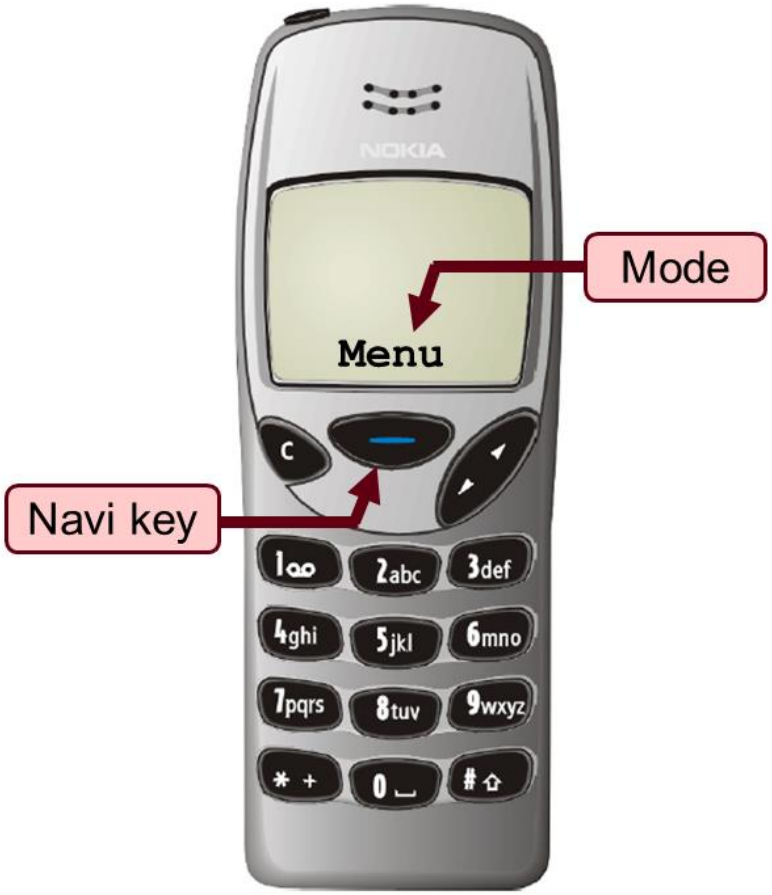
Modes

- A *mode* is a functioning arrangement or condition
- Modes are everywhere (and in most cases are unavoidable)
- Computer keyboards have modes
  - ≈100 keys + SHIFT, CTRL, ALT ➔ ≈800 key variations



Mobile Phone Example

- Navi key (first introduced on Nokia *3210*)
- Mode revealed by word above
- At least 15 interpretations: Menu, Select, Answer, Call, End, OK, Options, Assign, Send, Read, Use, View, List, Snooze, Yes

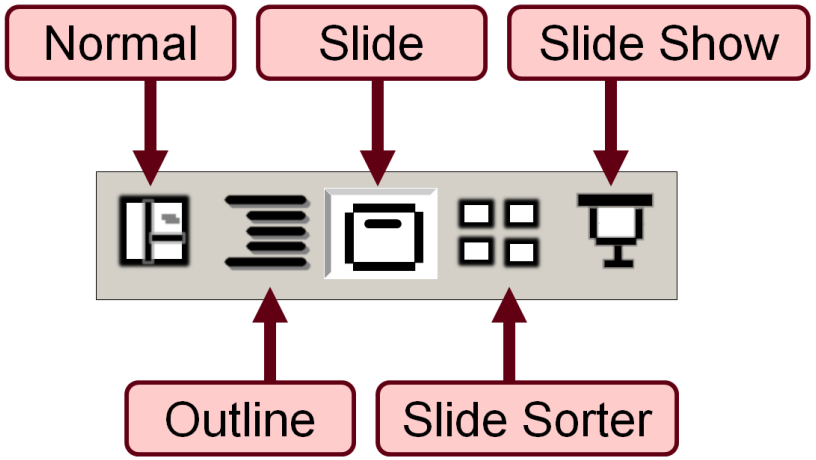


Contemporary LCD Monitor

- Similar to Navi key idea
- No labels for the four buttons above power button
- Function revealed on display when button pressed
- Possibilities explode

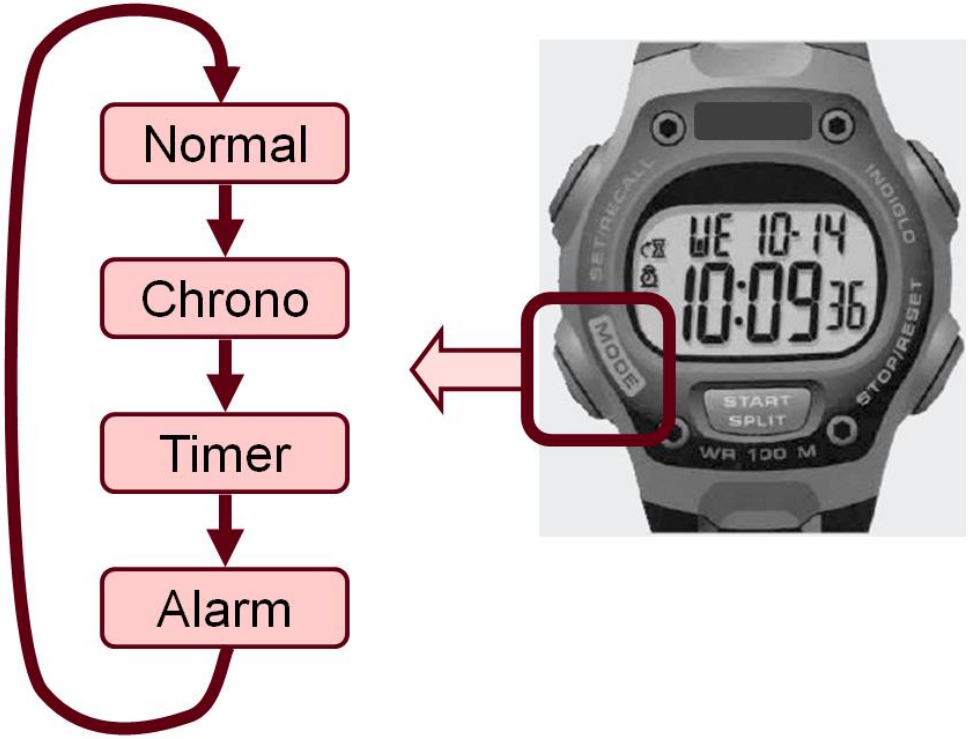
Mode Switching

- PowerPoint: Five view modes
- Switch modes by clicking soft button
- Current mode apparent by background shading
- Still problems lurk
- How to exit Slide Show mode?
  - PowerPoint → ESC
  - Firefox → ?



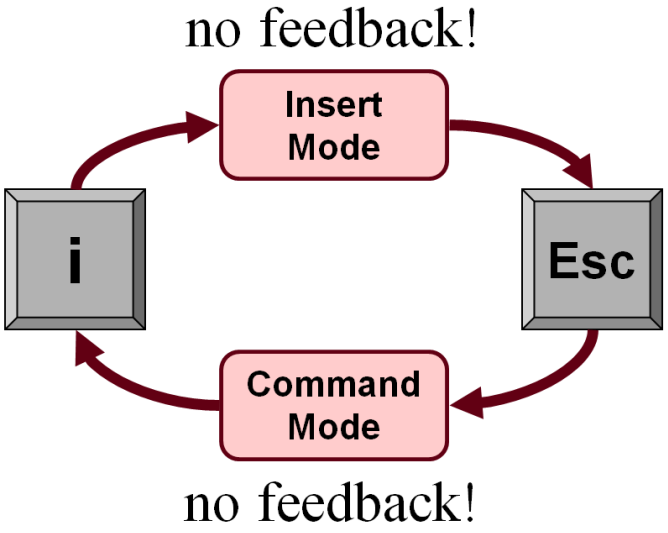
Mode Switching (2)

- Sports watch
- Single button cycles through modes



Mode Visibility

- Shneiderman: “offer information feedback”[1]
- Norman: “make things visible”[2]
- unix *vi* editor: Classic example of no mode visibility:

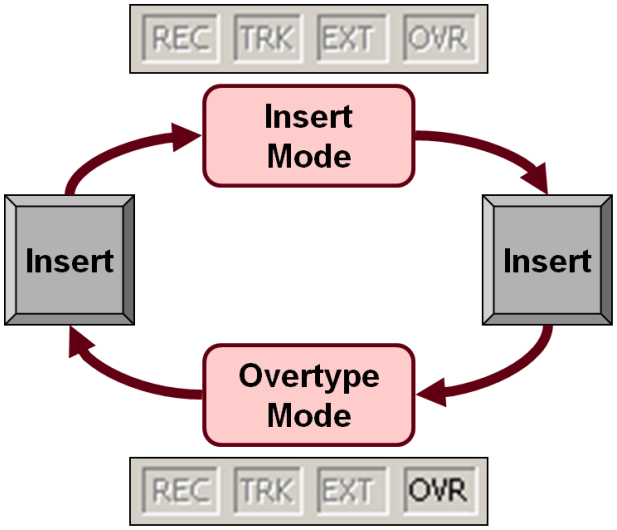


[1] Shneiderman, B., & Plaisant, C. (2005). *Designing the user interface: Strategies for effective human-computer interaction.* (4th ed.). New York: Pearson.

[2] Norman, D. A. (1988). *The design of everyday things*. New York: Basic Books.

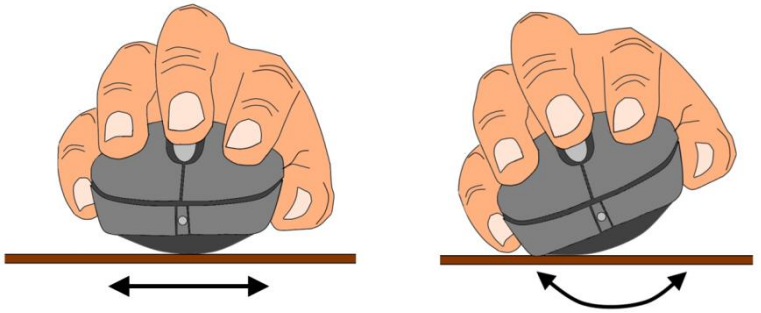
Mode Visibility (2)

- Insert vs. Overtype mode on MS/Word
- Some variation by version, but the user is in trouble most of the time

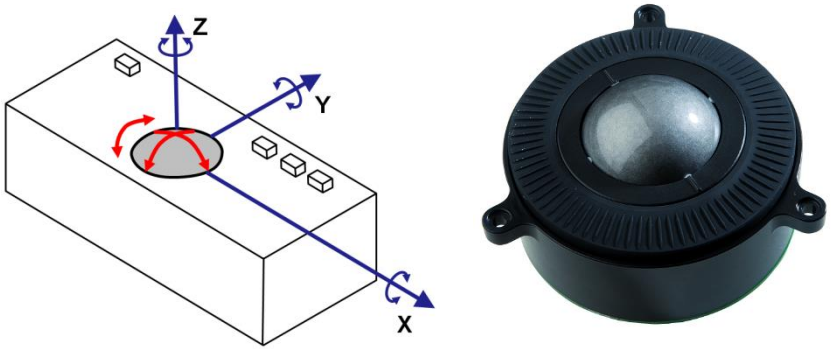


>2 Degrees of Freedom

- Examples in the HCI research literature
- 4 DOF *Rockin' Mouse*[1]



- Three-axis trackball[2]



[1] Balakrishnan, R., Baudel, T., Kurtenbach, G., & Fitzmaurice, G. (1997). The Rockin'Mouse: Integral 3D manipulation on a plane. *Proc CHI '97*, 311-318, New York: ACM.

[2] Evans, K. B., Tanner, P. P., & Wein, M. (1981). Tablet based valuators that provide one, two, or three degrees of freedom. *Computer Graphics, 15*(3), 91-97.

Separating the Degrees of Freedom

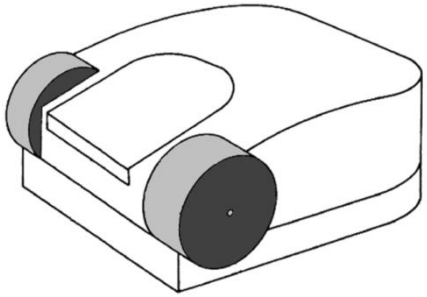
- More DOF is not necessarily better
- Must consider the context of use
- Etch-A-Sketch: separate 1 DOF *x* and *y* controllers:

Wheel Mouse

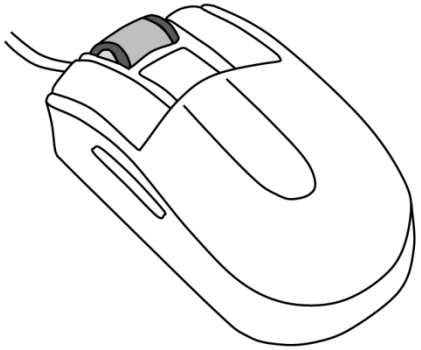
- Separate DOF via a wheel
- Successful introduction by Microsoft in 1996 with the *IntelliMouse* →

- Preceded by…

RollerMouse[1]



ProAgio[2]



[1] Venolia, D. (1993). Facile 3D manipulation. *Proc CHI '93*, 31-36, New York: ACM.
[2] Gillick, W. G., & Lam, C. C. (1996). U. S. Patent No. 5,530,455.

Mobile Context

- 1980s: born of mobile computing with PDAs
- 2007: launch of iPhone
- Touchscreens are the full embodiment of direct manipulation
  - No need for a cursor (cf. indirect input)
  - No control/display mappings (Spatial transformations, CD Gain, etc.)

Touch Input Challenges

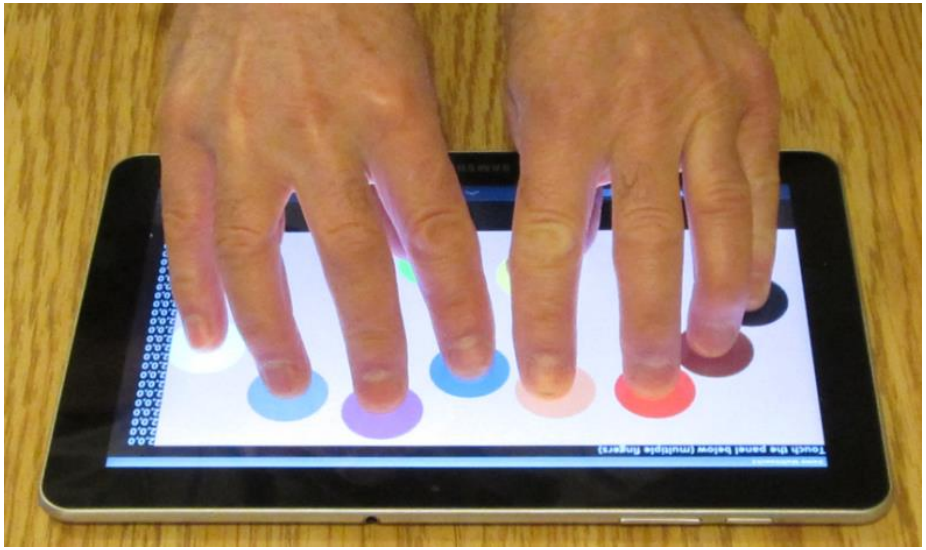
- Occlusion and accuracy ("fat finger problem")
- Early research → Offset cursor[1]
- Contemporary systems use variations; e.g., offset animation:

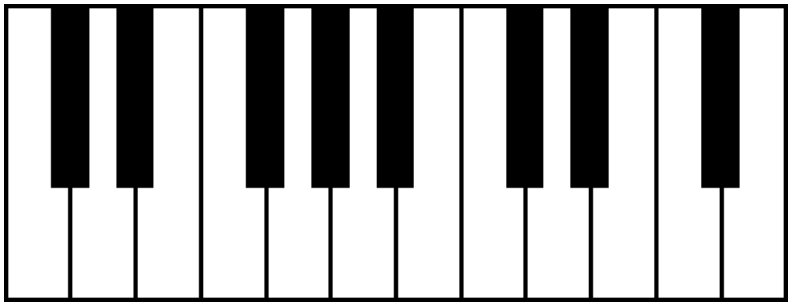
[1] Potter, R., Berman, M., & Shneiderman, B. (1988). An experimental evaluation of three touch screen strategies within a hypertext database. *Int J Human-Computer Interaction, 1* (1), 41-52.

Multitouch

Multitouch (>2)

- Piano keyboard: pressure data available





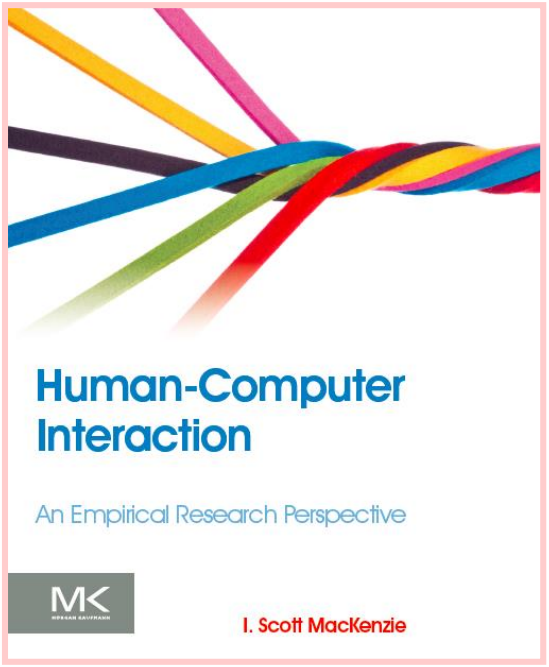
Accelerometers

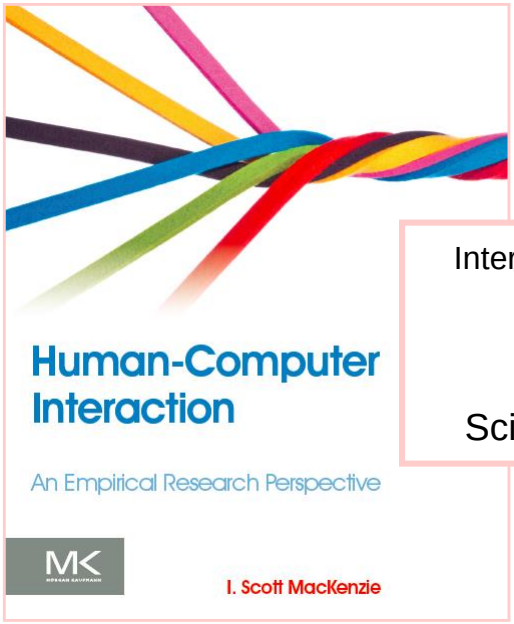
- Accelerometers enable tilt or motion as an input primitive
- Technology has matured; now common in mobile devices
- Many applications; e.g., spatially aware displays:

Connect to: http://join.quizizz.com

QUESTION TIME

Thank You


Human-Computer Interaction
An Empirical Research Perspective
I. Scott MacKenzie



Interactive Systems Design
Prof. V. Fuccella

Chapter 4
Scientific Foundations

Outline

- Research: theory and practice
- Research VS Engineering and Design
- Research Methods
- Scales of Measurement
- Research Questions and Validity of Experimental Procedures

Research – Definition

- Research is…

> Investigation or experimentation aimed at the discovery and interpretation of facts, the revision of accepted theories or laws in light of new facts.

- Example
  - Design and conduct a user study to test whether a new interaction technique improves on an existing interaction technique.

3

Experimentation

- A central activity in HCI research
- An experiment is sometimes called a *user study*
- Formal, standardized methodology preferred
  - Brings consistency to a body of work
  - Facilitates reviews and comparisons between different user studies

4

Research Must Be Published

- Publication is the final step
- Also an essential step
- *Publish or perish!*
  - Edict for researchers in all fields, and particularly in academia
- Until it is published, research cannot achieve its critical goal:
  - Extend, refine, or revise the existing body of knowledge in the field

5

Peer Review

- Research submitted for publication is reviewed by *peers* – other researchers doing similar research
- Only research meeting a high standard of scrutiny is accepted for publication
  - Are the results novel and useful?
  - Does the methodology meet the expected standards for the field?
- Accepted research is published (possibly online) and archived
- The final step is complete

6

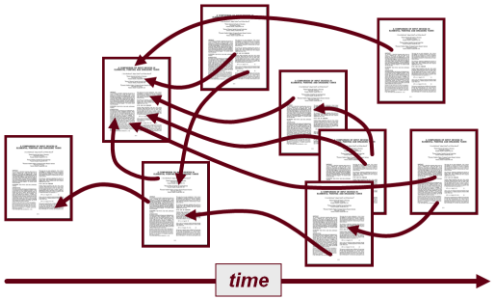
Patents

- Some research develops into bona fide inventions
- A researcher/company may wish to maintain ownership of (profit from) the invention
- Patenting is an option
- The patent application describes
  - Previous related work
  - How the invention addresses a need
  - The best mode of implementation
- If the application is granted, the patent is issued
- Note: A patent is a publication; thus patenting meets the must-publish criterion for research

7

Citations, References, Impact

- Citations, like hyperlinks, connect research to other research
- The number of citations to a research paper is an indication of the paper's impact
- Citations are used to evaluate publication venues (IF, SJR, SNIP)
- Can you spot the high-impact paper below? (arrows are citations)



8

Research Must Be Reproducible

- Research that cannot be replicated is useless
- A high standard of reproducibility is essential
- The research write-up must be sufficiently detailed to allow a skilled researcher to replicate the research if he/she desired
- The easiest way to ensure reproducibility is to follow a standardized methodology
- Many great advances in science pertain to methodology (e.g., Louis Pasteur's detailed disclosure of the methodology used in his research in microbiology)
- The most cited research paper is a "method paper"[1] (see Google Scholar for the latest citation count)

[1] Lowry, O. H., Rosenbrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *Journal of Biological Chemistry, 193*, 265-275.

9

Digital Libraries – H-index[1]

- Since the arrival of DL (Google Scholar, Scopus, etc.), citation counts are easy to gather
  - Can be gathered for papers, journals, etc.
  - Can also be gathered for researchers
- H-index is a measure of the impact of a researcher
- Calculation:
  - Rank a researcher's publications by the number of citations
  - the H-index is the point where the rank equals the number of citations;
- A researcher with H-index = *n* has *n* publications each with *n* or more citations

[1] Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, 102*, 16568-16572.

10

H-Index Calculation

Let's suppose a researcher has the publications reported in the table

Title	Citations
Title 1	4
Title 2	0
Title 3	12
Title 4	3
Title 5	5
Title 6	2
Title 7	0

1. We sort them by the number of citations

Title	Citations
Title 3	12
Title 5	5
Title 1	4
Title 4	3
Title 6	2
Title 2	0
Title 7	0

2. We count the rows until position <= # of citations

H-index = 3

11

H-index

- A respectable H-index (although debatable) is "number of years since PhD"
- Exercise: Open Scopus and search total number of papers, total number of citations and H-index of…
  - A professor you appreciate (or you don't)
  - A famous contemporary scientist

[1] Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, 102*, 16568-16572.

12

DL for Computer Science

- Google Scholar – all fields of science
  - https://scholar.google.it/
- ACM Digital Library – computer science
  - http://dl.acm.org/
- IEEE Explorer - engineering
  - http://ieeexplore.ieee.org/Xplore/home.jsp
- Science Direct – several fields of science
  - http://www.sciencedirect.com/

13

Quality of the Venue

- Not all scientific venues have the same prestige.
- It is possible to evaluate journals through:
  - **Impact Factor (IF):** measure reflecting the yearly average number of citations to recent articles published in that journal
  - **Scimago Journal Rank (SJR):** Similar to IF, but citations are weighted on the prestige (SJR) of the citing journal (similar to PageRank).
  - **Source Normalized Impact per Paper (SNIP):** citations received VS citations expected for the journal's subject field.

14

HCI Venues

- Conferences:
  - CHI: The ACM CHI Conference on Human Factors in Computing Systems
  - UIST: Annual ACM Symposium on User Interface Software and Technology
  - IUI: ACM International Conference on Intelligent User Interfaces
- Journals:
  - Human–Computer Interaction
  - TOCHI: ACM Transactions on Computer-Human Interaction
  - International Journal of Human-Computer Studies – Elsevier
  - Interacting with Computers - Oxford Journals
  - IEEE Transactions on Human-Machine Systems

15

Exercise

- Venues are also classified by scientific societies or committees of experts
  - In Italy, classification for computer science conferences were made by GRIN (GRuppo di INformatica)
  - The results can be used for public selections
- Exercise I: find the GRIN classification of the reported HCI conferences
- Exercise II: find the ranking of the reported HCI journals
  - SJR, SNIP
  - Classification of Journals (see e-learning platform)

16

Exercise

- Work in groups
- Use the main scientific search engines to find some (about 15) important papers describing a novel method for «text entry on smart watches» before 2018 (included).
- Evaluate your work:
  - Divide your documents into 2 sets:
    - Relevant: if present in the provided list
    - Not relevant: otherwise
  - Calculate *precision* (the fraction of retrieved documents that are relevant to the topic) and *recall* (fraction of the documents that are relevant to the topic that are successfully retrieved)

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

17

Organizing Literature

Study (1st author)	Number of Keys[a]	Direct/ Indirect	Scanning	Number of Participants	Speed[b] (wpm)	Notes
Bellman [2]	5	Indirect	No	11	11	4 cursors keys + SELECT key. Error rates not reported. No error correction method.
Dunlop [4]	4	Direct	No	12	8.90	4 letter keys + SPACE key. Error rates reported as "very low."
Dunlop [5]	4	Direct	No	20	12	4 letter keys + 1 key for SPACE/NEXT. Error rates not reported. No error correction method.
Tanaka-Ishii [25	3	Direct	No	8	12+	4 letters keys + 4 keys for editing, and selecting. 5 hours training. Error rates not reported. Errors corrected using CLEAR key.
Gong [7]	3	Direct	No	32	8.01	3 letter keys + two additional keys. Error rate = 2.1%. Errors corrected using DELETE key.
MacKenzie [16]	3	Indirect	No	10	9.61	2 cursor keys + SELECT key. Error rate = 2.2%. No error correction method.
Baljko [1]	2	Indirect	Yes	12	3.08	1 SELECT key + BACKSPACE key. 43 virtual keys. RC scanning. Same phrase entered 4 times. Error rate = 18.5%. Scanning interval = 750 ms.
Simpson [24]	1	Indirect	Yes	4	4.48	1 SELECT key. 26 virtual keys. RC scanning. Excluded trials with selection errors or missed selections. No error correction. Scanning interval = 525 ms at end of study.
Koester [10]	1	Indirect	Yes	3	7.2	1 SELECT key. 33 virtual keys. RC scanning with word prediction. Dictionary size not given. Virtual BACKSPACE key. 10 blocks of trials. Error rates not reported. Included trials with selection errors or missed selections. Fastest participant: 8.4 wpm.

[a] For "direct" entry, the value is the number of letter keys. For "indirect" entry, the value is the total number of keys.
[b] The entry speed cited is the highest of the values reported in each source, taken from the last block if multiple blocks.

[1] MacKenzie, I. S. (2009). The one-key challenge: Searching for an efficient one-key text entry method. *Proc ASSETS 2009*, 91-98, New York: ACM.

18

Research vs. Engineering vs. Design

- Researchers often work closely with engineers and designers, but the skills each brings are different
- Engineers and designers are in the business of building things, bringing together the best in *form* (design emphasis) and *function* (engineering emphasis)
- One can image that there is a certain tension, even trade-off, between form and function
- Sometimes, things don't go quite as planned →

20

Form Trumpeting Function

- The photo below shows part of a laptop computer
- The form is elegant – smooth, shiny, metallic
- The touchpad design (or is in engineering?) has a problem
- No tactile sense at the sides of the touchpad

The fix

21

Duct Tape To The Rescue

A true story!

22

Research Milieu

- Engineering and design are about products
- Research is not about products
- Research is narrowly focused
- Research questions are small in scope
- Research is incremental, not monumental
  - Research ideas build on previous research ideas
  - Good ideas are refined, advanced (into new ideas)
  - Bad ideas are discarded, modified
- Products come later, much later

23

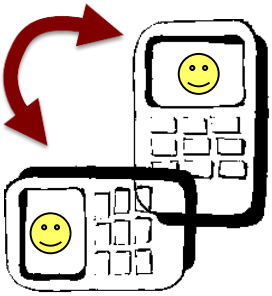
Example: Apple *iPhone* (2007)

iPhone Gestures:

- Tilt
- Flick gesture
- Multitouch

Tilt

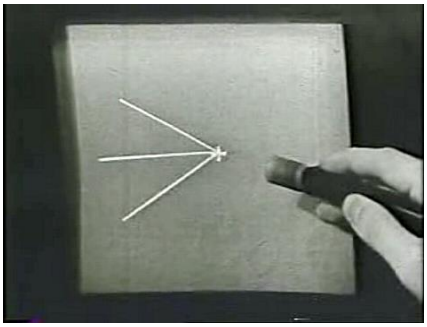
- Research on tilt as an interaction primitive dates at least to 1998[1]



[1] Harrison, B., Fishkin, K. P., Gujar, A., Mochon, C., & Want, R. (1998). Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. *Proc CHI '98*, 17-24, New York: ACM.

Flick

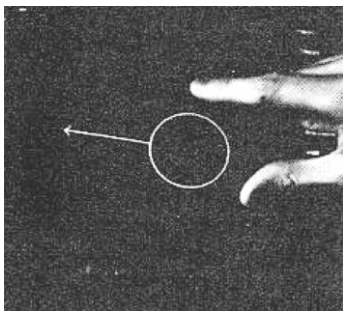
- Research on flick as an interaction primitive dates at least to 1963[1]



[1] Sutherland, I. E. (1963). Sketchpad: A man-machine graphical communication system. *Proceedings of the AFIPS Spring Joint Computer Conference*, 329-346, New York: ACM.

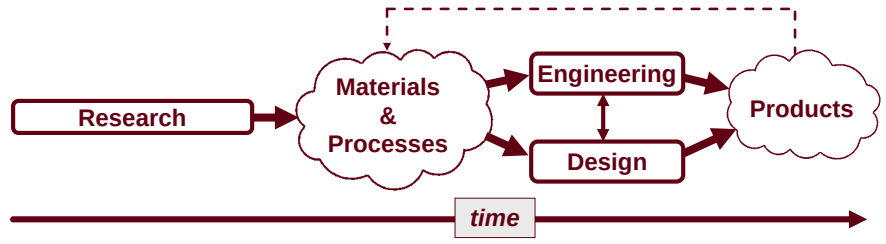
Multitouch

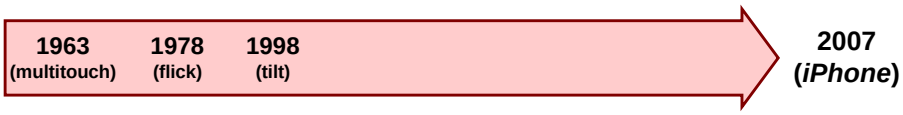
- Research on multitouch as an interaction primitive dates at least to 1978[1]



[1] Herot, C. F., & Weinzapfel, G. (1978). One-point touch input of vector information for computer displays. *Proceedings of SIGGRAPH 1978*, 210-216, New York: ACM.

So…





"Empirical" Research

- Empirical:
  - Originating in or based on observation or experience
  - Relying on experience or observation alone without due regard for system or theory (i.e., don't be blinded by pre-conceptions)
- Example: Nicolas Copernicus (1473-1543)
  - Prevailing system or theory: celestial bodies revolved around the earth
  - Copernicus made astronomical observations that cut against this view
  - Result: heliocentric cosmology (the earth and planets revolve around the sun)

29

Research Methods

- Qualitative VS Quantitative methods

	Qualitative	Quantitative
Objective	Exploratory, understanding things, develop ideas or hypotheses	Used to quantify attitudes, opinions, behaviors
Data collection	Unstructured, data are not numerical	Structured, data are numerical and can be statistically analized.
Sample size	Typically small	Typically large

- Main research methods
  - Observational method (qualitative)
  - Experimental method (quantitative)
  - Correlational method (quantitative)

30

Observational Method

- Focused on qualitative assessments
- Observe phenomena directly in the environment where they occur.
- Example methods:
  - Interviews, field investigations, contextual inquiries, case studies, field studies, focus groups, think aloud protocols, walkthroughs, cultural probes, etc.
- Relevance vs. precision
  - High in relevance (behaviours studied in a natural setting)
  - Low in precision (lacks control available in a laboratory)

31

Experimental Method

- Aka *scientific method*
- Controlled experiments conducted in lab setting
- Relevance vs. precision
  - Low in *relevance* (artificial environment)
  - High in *precision* (extraneous behaviours easy to control)
- At least two variables:
  - *Manipulated variable (*aka *independent variable)*
  - *Response variable* (aka *dependent variable*)
- Cause-and-effect conclusions possible (changes in the manipulated variable *caused* changes in the response variable)

32

Correlational Method

- Look for relationships between variables
- Observations made, data collected
  - Example: are user's privacy settings while social networking related to their age, gender, level of education, employment status, income, etc.
- Non-experimental
  - Interviews, on-line surveys, questionnaires, etc.
- Balance between relevance and precision (some quantification, observations not in lab)
- Cause-and-effect conclusions not possible

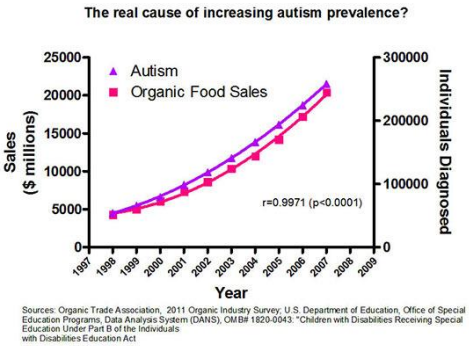
33

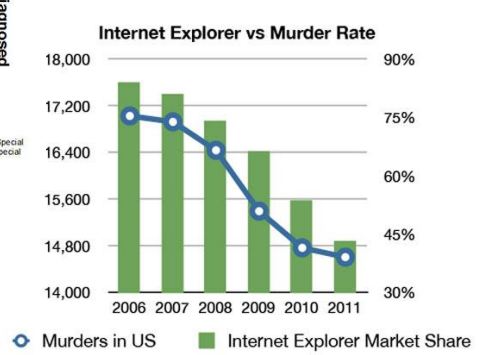
Relationships: Circumstantial & Causal

- As noted above…
  - Correlational methods → circumstantial relationships
  - Experimental methods → causal relationships
- Causal-and-effect conclusions not possible if the independent variable is a *naturally occurring attribute* of participants (e.g., gender, personality type, handedness, first language, political viewpoint)
- These attributes are legitimate independent variables
- But, they cannot be assigned to participants; hence causal relationships not valid
- HCI Example: T9 vs MultiTap (Favourite method)

34

Correlation VS Causation





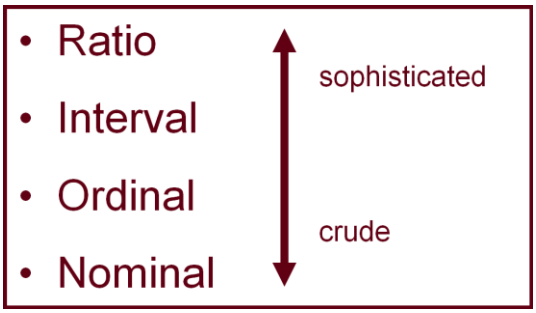
35

Observe and Measure

- Foundation of empirical research
- Observation is the starting point; observations are made…
  - By a human observer
  - By the apparatus
- Manual observation
  - Log sheet, notebooks
  - Screen capture, photographs, videos, etc.
- Measurement
  - With measurement, anecdotes (*April showers bring May flowers*) turn to empirical evidence
  - "*When you cannot measure, your knowledge is of a meager and unsatisfactory kind*" (Kelvin)

36

Scales of Measurement



37

Nominal Data

- Examples of nominal data (aka *categorical data*) : plate numbers, postal codes, job classification, etc.
- Can also be arbitrary codes assigned to attributes; e.g.,
  - 1 = male, 2 = female
  - 1 = mouse, 2 = touchpad, 3 = pointing stick
- The code needn't be a number; i.e.,
  - M = male, F = female
- Obviously, the statistical mean cannot be computed on nominal data
- Usually it is the count that is important
  - "Are females or males more likely to…"
  - "Do left handers or right handers have more difficulty with…"
  - Note: The count itself is a ratio-scale measurement

Ordinal Data

- Ordinal data associate an order or rank to an attribute
- More sophisticated than nominal data
  - Comparisons of "greater than" or "less than" possible
- The attribute is any characteristic or circumstance of interest; e.g.,
  - Users try three GPS systems for a period of time, then rank them: 1st, 2nd, 3rd choice
- (example on next slide)

Ordinal Data – HCI Example

How many email messages do you receive each day?

1. None (I don't use email)
2. 1-5 per day
3. 6-25 per day
4. 26-100 per day
5. More than 100 per day

Interval Data

- Equal distances between adjacent values
- Classic example: temperature (°F, °C)
- Statistical mean possible
  - E.g., the mean midday temperature during July
- Ratios not possible
  - Cannot say 10 °C is twice 5 °C
  - **Problem**: no absolute zero

Interval Data – HCI Example

- Responses on a Likert scale
- Likert scale characteristics:
  1. Statement soliciting level of agreement
  2. Responses are symmetric about a neutral middle value
  3. Gradations between responses are equal (more-or-less)
- Assuming "equal gradations", the statistical mean is valid (and related statistical tests are possible)
- Likert scale example → (next slide)

Interval Data – HCI Example (2)

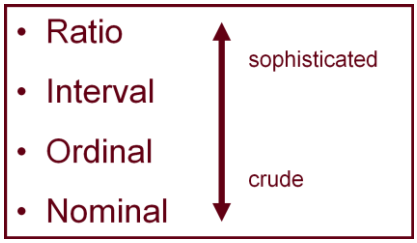
Please indicate your level of agreement with the following statements.

	Strongly disagree	Mildly disagree	Neutral	Mildly agree	Strongly agree
It is safe to talk on a mobile phone while driving.	1	2	3	4	5
It is safe to read a text message on a mobile phone while driving.	1	2	3	4	5
It is safe to compose a text message on a mobile phone while driving.	1	2	3	4	5

Ratio Data

- Most sophisticated of the four scales of measurement
- Absolute zero, therefore many calculations possible
- A "count" is a ratio-scale measurement
  - E.g., "time" (the number of seconds to complete a task)
- Enhance counts by adding further ratios where possible
  - Facilitates comparisons
  - Example – a 10-word phrase was entered in 30 seconds
    - Bad: t = 30 seconds
    - Good: Entry rate = 10 / 0.5 = 20 wpm

Scales of Measurement



OK to compute....	Nominal	Ordinal	Interval	Ratio
frequency distribution.	Yes	Yes	Yes	Yes
median and percentiles.	No	Yes	Yes	Yes
add or subtract.	No	No	Yes	Yes
mean, standard deviation, standard error of the mean.	No	No	Yes	Yes
ratio, or coefficient of variation.	No	No	No	Yes

45

Research Questions

- We conduct empirical research to answer (and raise!) questions about UI designs or interaction techniques
- Consider the following questions:
  - Is it viable?
  - Is it better than current practice?
  - Which design alternative is best?
  - What are the performance limits?
  - What are the weaknesses?
  - Does it work well for novices?
  - How much practice is required?

Testable Research Questions

- Preceding questions, while unquestionably relevant, are not testable
- Try to re-cast as testable questions (even though the new question may appear less important)
- Scenario…
  - You have invented a new text entry technique for touchscreen mobile phones, and you think it's pretty good. In fact, you think it is better than the Qwerty soft keyboard (QSK). You decide to undertake a program of empirical enquiry to evaluate your invention. What are your research questions?

Research Questions (2)

- Very weak

  Is the new technique any good?

- Weak

  Is the new technique better than QSK?

- Better

  Is the new technique faster than QSK?

- Better still

  Is the measured entry speed (in words per minute) higher for the new technique than for QSK after one hour of use?

Internal Validity

- Definition:
  - The extent to which the effects observed are due to the test conditions
- Statistically, this means…
  - Differences (in the means) are due to inherent properties of the test conditions
  - Variances are due to participant differences ("pre-dispositions")
  - Other potential sources of variance are controlled or exist equally or randomly across the test conditions

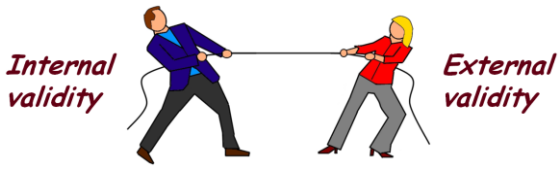
External Validity

- Definition:
  - The extent to which results are generalizable to other *people* and other *situations*
- People
  - The participants are *representative* of the broader intended population of users
- Situations
  - The *test environment* and *experimental procedures* are representative of real world situations where the interface or technique will be used

Experimental Procedure Example

- Scenario…
  - You wish to compare two text entry techniques for mobile devices
- External validity is improved if the experimental procedure mimics expected usage
- Test procedure should probably have participants…
  - Enter personalized paragraphs of text (e.g., a paragraph about a favorite movie)
  - Edit and correct mistakes as they normally would
- But… is internal validity compromised?

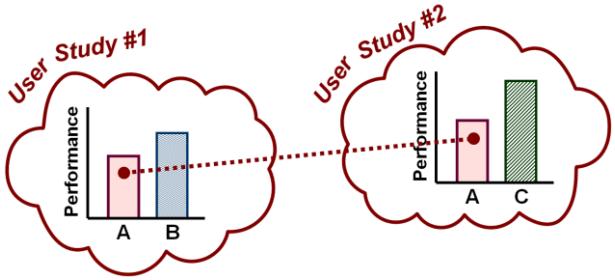
The Tradeoff



- There is tension between internal and external validity
- The more the test environment and experimental procedures are "relaxed" (to mimic real-world situations), the more the experiment is susceptible to uncontrolled sources of variation, such as pondering, distractions, fiddling, or secondary tasks

Comparative Evaluations

- Preferable to do a comparative evaluation rather than one-of
- More insightful results obtained
- Factorial experiments require comparison, because there must be at least one independent variable with at least two levels
- If one condition is a base line; comparisons possible between studies (assuming similar methodology)



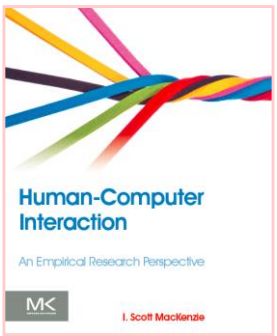
53

Connect to: http://join.quizizz.com

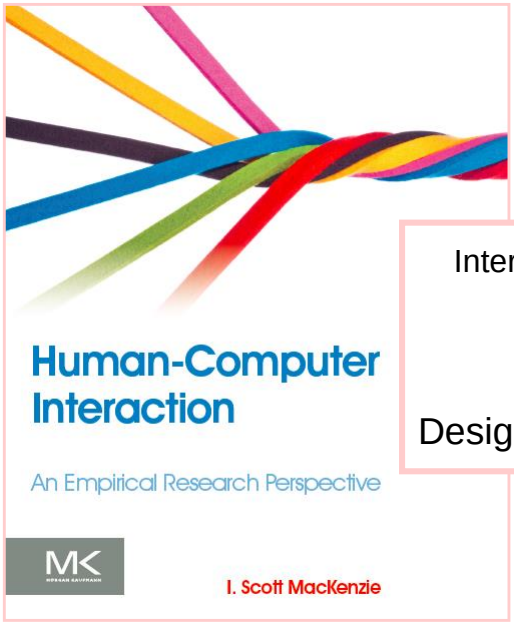
QUESTION TIME

54

Thank You



55



Interactive Systems Design
Prof. V. Fuccella

Chapter 5
Designing HCI Experiments

Outline

- Introduction
- Variables
- Design
  - Within-Subjects VS Between-Subjects
  - Counterbalancing
- Procedure
- Participants
- Questionnaires
- Longitudinal Studies

2

Methodology

- Learning to conduct and design an experiment is a skill required of all researchers in HCI
- Experiment design is the process of deciding participants, apparatus, tasks, order of tasks, procedures, variables, data collected, and so on
- Methodology is critical:

> Science is method. Everything else is commentary.[1]

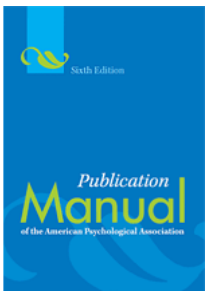
- What methodology?
  - Don't just make it up because it seems reasonable
  - Follow standards for experiments with human participants (next slide)

[1] This quote from Allen Newell was cited and elaborated on by Stuart Card in an invited talk at the ACM's SIGCHI conference in Austin, Texas (May 10, 2012).

3

APA

- American Psychological Association (APA) is the pre-dominant organization promoting research in psychology – the improvement of research methods and conditions and the application of research findings (http://www.apa.org/)
- *Publication Manual of the APA*[1], first published in 1929, teaches about the writing process and, implicitly, about the methodology for experiments with human participants
- Recommended by major journals in HCI



[1] APA. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: APA.

4

Ethics Approval

- *Ethics approval* is a crucial step that precedes every HCI experiment
- HCI experiments involve humans, thus…

> Researchers must respect the safety, welfare, and dignity of human participants in their research and treat them equally and fairly.[1]

- Proposal submitted to ethics review committee
- Criteria for approval:
  - research methodology
  - risks or benefits
  - the right not to participate, to terminate participation, etc.
  - the right to anonymity and confidentiality

[1] http://www.yorku.ca/research/students/index.html

5

Getting Started With Experiment Design

- Transitioning from the creative work in formulating and prototyping ideas to experimental research is a challenge

1. Formulate a research questions:

> Can a task be performed more quickly with my new interface than with an existing interface?

2. Begin with…

> What are the experimental variables?

Properly formed research questions inherently identify experimental variables (can you spot the independent variable and the dependent variable in the question above?)

6

Independent Variable

- An *independent variable* (IV) is a circumstance or characteristic that is manipulated in an experiment to elicit a change in a human response while interacting with a computer.
- "Independent" because it is independent of participant behavior (i.e., there is nothing a participant can do to influence an independent variable)
- The terms *independent variable* and *factor* are synonymous

7

Test Conditions

- An independent variable (IV) must have at least two levels
- The levels, values, or settings for an IV are the *test conditions* aka *experimental conditions*
- Name both the factor (IV) and its levels (test conditions):

Factor (IV)	Levels (test conditions)
Device	mouse, trackball, joystick
Feedback mode	audio, tactile, none
Task	pointing, dragging
Visualization	2D, 3D, animated
Search interface	Google, custom

8

Test Conditions

- An experiment with 2 independent variables, each with m and n levels respectively, is called an
  - m x n factorial design
- Example: TouchTap experiment
  - Design: within-subjects
  - IndependentVariables (factors):
    - Input Method in {TouchTap, widget-based technique};
    - Font Size in {1.75, 3.25, 4.75};
  - **2 x 3 within-subjects factorial design**
  - 6 test (experimental) conditions
- Experiments with >2 factors are also possible

9

Human Characteristics

- Human characteristics are *naturally occurring attributes*
- Examples:
  - Gender, age, height, weight, handedness, grip strength, finger width, visual acuity, personality trait, political viewpoint, first language, shoe size, etc.
- They are legitimate independent variables, but they cannot be "manipulated" in the usual sense
- Causal relationships are difficult to obtain due to unavoidable confounding variables

10

How Many IVs?

- An experiment must have at least one independent variable
- Possible to have 2, 3, or more IVs
- But the number of "effects" increases rapidly with the size of the experiment:

Independent Variables	Effects					Total
	Main	2-way	3-way	4-way	5-way	
1	1	-	-	-	-	1
2	2	1	-	-	-	3
3	3	3	1	-	-	7
4	4	6	3	1	-	14
5	5	10	6	3	1	25

- Advice: Keep it simple (1 or 2 IVs, 3 at the most)

11

Dependent Variable

- A *dependent variable* is a measured human behaviour
- "Dependent" because it depends on what the participant does
- Examples:
  - task completion time, speed, accuracy, error rate, throughput, target re-entries, task retries, presses of backspace, etc.
- Dependent variables must be clearly defined
  - Research must be reproducible!

12

Unique DVs

- Any observable, measurable behaviour is a legitimate dependent variable
  - So, feel free to "roll your own"
- Example: *negative facial expressions*[1]
  - Application: user difficulty with mobile games
  - Events logged included frowns, head shaking
  - Counts used in ANOVA, etc.
  - Clearly defined → reproducible

[1] Duh, H. B.-L., Chen, V. H. H., & Tan, C. B. (2008). Playing different games on different phones: An empirical study on mobile gaming. *Proceedings of MobileHCI 2008*, 391-394, New York: ACM.

13

Data Collection

- Obviously, the data for dependent variables must be collected in some manner
- Ideally, engage the experiment software to log timestamps, key presses, button clicks, etc.
- Planning and ***pilot testing*** important
- Ensure conditions are identified, either in the filenames or in the data columns
- Better to have two logs:
  - **High-level**: directly reports the values of dependent variables.
  - **Low-level**: reports all salient events of the trials

14

High-Level Log

0_0_A_stats.tema

```
[#]presented    transcribed presented_characters    transcribed_characters  input_time(sec) pause_time(sec) total_time(sec)  wpm mad numBksp numDelChars total_error cor_error   uncor_error
Mary had a little lamb  Mary had a little lamb  22  22  15.228  0.0 15.267  14.184397163120567  0   2   8   0.26666666666666666 0.26666666666666666 0.0
suburbs are sprawling up everywhere suburbs are sprawling up everywhere 35  35  11.287  0.0 11.308  29.768760520953307  0   0   0   0.0 0.0 0.0
pay off a mortgage for a house  Pay off a mortgage for a house  30  30  9.106   0.0 9.123   35.58093564682627   1   0   0   0.03333333333333333 0.0 0.03333333333333333
taking the train is usually faster  taking the train is usually faster  34  34  9.148   0.0 9.166   36.729339746392654  0   0   0   0.0 0.0 0.0
exceed the maximum speed limit  exceed the maximum speed limit  30  30  12.513  0.0 12.54   23.01606329417406   0   1   5   0.14285714285714285 0.14285714285714285 0.0
the laser printer is jammed the laser printer is jammed 27  27  7.998   0.0 9.268   36.00900225056264   0   0   0   0.0 0.0 0.0
a big scratch on the tabletop   a big scratch on the tabletop   29  29  8.49    0.0 8.507   39.57597173144876   0   0   0   0.0 0.0 0.0
Mary had a little lamb  Mary had a little lamb  22  18  13.056  0.0 13.075  13.786764705882353  4   2   7   0.3793103448275862  0.2413793103448276  0.13793103448275862
the music is better than it sounds  the music is better than it sounds  34  34  9.57    0.0 9.599   38.87147335423197   0   0   0   0.0 0.0 0.0
microscopes make small things look big  microscopes make small things look big  38  38  10.461  0.0 10.49   30.972182391740752  0   0   0   0.0 0.0 0.0
```


Low-Level Log

0_0_A_events.tema

```
#Log opened: Fri Apr 12 09:16:35 CEST 2013
#just like it says on the can good
#1365751074947
0,<Entr>,pos@0
220,<Entr>,pos@1
#REJ,
#just like it says on the can good
#1365751188459
0,j,pos@0
452,jus,pos@0
1219,just,pos@0
1970,just,pos@0
1981,<Sp>,pos@4
2561,l,pos@5
2828,li,pos@5
3567,lilo,pos@5
7975,lilo,pos@5
7988,<Entr>,pos@9
8451,<Entr>,pos@10
#REJ,just lilo
#just like it says on the can good
```

Control Variable

- A *control variable* is a circumstance (not under investigation) that is kept constant while testing the effect of an independent variable
- More control means the experiment is less generalizable (i.e., less applicable to other people and other situations)
- Research question: Is there an effect of font color or background color on reading comprehension?
  - Independent variables: font color, background color
  - Dependent variable: comprehension test scores
  - Control variables
    - Font size (e.g., 12 point)
    - Font family (e.g., Times)
    - Ambient lighting (e.g., fluorescent, fixed intensity)
    - Etc.

17

Random Variable

- A *random variable* is a circumstance that is allowed to vary randomly
- More variability is introduced in the measures (that's bad!), but the results are more generalizable (that's good!)
- Research question: Does user stance affect performance while playing *Guitar Hero?*
  - Independent variable: stance (standing, sitting)
  - Dependent variable: score on songs
  - Random variables
    - Prior experience playing a real musical instrument
    - Prior experience playing *Guitar Hero*
    - Amount of coffee consumed prior to testing
    - Etc.

18

Control vs. Random Variables

- There is a trade-off which can be examined in terms of internal validity and external validity (see below)

Variable	Advantage	Disadvantage
Random	Improves external validity by using a variety of situations and people.	Compromises internal validity by introducing additional variability in the measured behaviours.
Control	Improves internal validity since variability due to a controlled circumstance is eliminated	Compromises external validity by limiting responses to specific situations and people.

- Remember:
  - Internal validity: The extent to which the effects observed are due to the test conditions
  - External validity: The extent to which results are generalizable to other *people* and other *situations*

19

Confounding Variable

- A *confounding variable* is a circumstance that varies systematically with an independent variable
- Research question: In an eye tracking application, is there an effect of "camera distance" on task completion time?
  - Independent variable: Camera distance (near, far)
    - Near camera (A): inexpensive camera mounted on eye glasses
    - Far camera (B): expensive camera mounted above system display
  - Dependent variable: task completion time
  - But, "camera" is a confounding variable: camera A for the near setup, camera B for the far setup
  - Are the effects due to camera distance or to some aspect of the different setups?

20

Design

Within-subjects VS Between-subjects

- Two ways to assign conditions to participants:
  - *Within-subjects* → each participant is tested on each condition
  - *Between-subjects* → each participant is tested on one condition only
  - Example: An IV with three test conditions (A, B, C):

Within-subjects

Participant	Test Condition		
1	A	B	C
2	A	B	C

Between-subjects

Participant	Test Condition
1	A
2	A
3	B
4	B
5	C
6	C

21

Within-subjects, Between-subjects

Pros and Cons

	Within-Subjects	Between-Subjects
Participants	Fewer 	More
Variation due to participants	Less 	More
Balance groups	No need 	Needed
Order effects	There may be	Not present 
Amount of trials per participant	High	Low 

22

Within-subjects, Between-subjects

- Sometimes…
  - A factor must be assigned within-subjects
    - Examples: input method, …
  - A factor must be assigned between-subjects
    - Examples: gender, handedness
- With two factors, there are three possibilities:
    - both factors within-subjects
    - both factors between-subjects
    - one factor within-subjects + one factor between-subjects (this is a *mixed design*)

23

Order Effects, Counterbalancing

- Only relevant for within-subjects factors
- The issue: *order effects* (aka *learning effects*, *practice effects*, *fatigue effects*, *sequence effects*)
- Order effects avoided by *counterbalancing*:
  - Participants divided into groups
  - Test conditions are administered in a different order to each group
  - Possible counterbalancing schema: Latin square
  - Distinguishing property of a Latin square → each test condition occurs precisely once in each row and column (next slide)

24

Latin Squares

2 x 2

A	B
B	A

3 x 3

A	B	C
B	C	A
C	A	B

4 x 4

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

5 x 5

A	B	C	D	E
B	C	D	E	A
C	D	E	A	B
D	E	A	B	C
E	A	B	C	D

25

Balanced Latin Square

- With a balanced Latin square, each condition precedes and follows any other condition an equal number of times
- Only possible for even-orders
- Top row pattern: A, B, *n*, C, *n* – 1, D, *n* – 2, …

4 x 4

A	B	D	C
B	C	A	D
C	D	B	A
D	A	C	B

6 x 6

A	B	F	C	E	D
B	C	A	D	F	E
C	D	B	E	A	F
D	E	C	F	B	A
E	F	D	A	C	B
F	A	E	B	D	C

26

Test Conditions with multiple factors

- The different test conditions can be different levels of the same factor.
- They can also be a tuple of levels for more factors.
- A Factorial Design with multiple factors
  - A: A1, A2, …, An
  - B: B1, B2, …, Bm
- leads to a number of possible test conditions given by:
  - A x B (cartesian product: size = n x m)
- A subset of all the possible test conditions can be chosen

27

Example

- An experimenter seeks to determine if three editing methods (A, B, C) differ in the amount of time to do a common editing task:

  Replace one 5-letter word with another, starting one line away.

- Conditions are assigned within-subjects
  - Method A: arrow keys, backspace, type
  - Method B: search and replace dialog
  - Method C: point and double click with the mouse, type
- Twelve participants are recruited and divided into three groups (4 participants/group)
- Methods administered using a 3×3 Latin Square

28

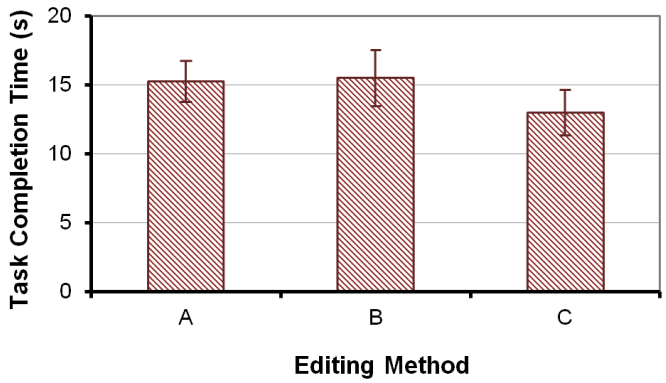
Results - Data

Participant	Test Condition			Group	*Mean*	*SD*
	A	B	C			
1	12.98	16.91	12.19	1 **ABC**	14.7	1.84
2	14.84	16.03	14.01			
3	16.74	15.15	15.19			
4	16.59	14.43	11.12			
5	18.37	13.16	10.72	2 **BCA**	14.6	2.46
6	15.17	13.09	12.83			
7	14.68	17.66	15.26			
8	16.01	17.04	11.14			
9	14.83	12.89	14.37	3 **CAB**	14.4	1.88
10	14.37	13.98	12.91			
11	14.40	19.12	11.59			
12	13.70	16.17	14.31			
Mean	15.2	15.5	13.0			
SD	1.48	2.01	1.63			

Group effect is small
∴ Counterbalancing worked!

29

Results - Chart



30

Example: TouchTap

- Design: within-subjects
- IndependentVariables (factors):
  - Input Method in {TouchTap, widget-based technique};
  - Font Size in {1.75, 3.25, 4.75};
- **2 x 3 within-subjects factorial design**
- 6 test conditions

Method\Font	small	medium	large
TouchTap	T-s (A)	T-m (B)	T-l (C)
Widget-based	W-s (D)	W-m (E)	W-l (F)

- We can counterbalance with a balanced latin square
- If necessary, to avoid confusing participants, we only change the input method once (see the paper)

31

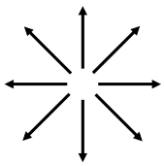
Other Techniques

- Instead of using a Latin square, all orders (n!) can be used; 3 case →
- Conditions can be randomized
- Randomizing best if the tasks are brief and repeated often; examples (Fitts' law, see below)

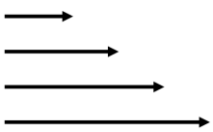
A	B	C
A	C	B
B	C	A
B	A	C
C	A	B
C	B	A

Target size

Movement direction



Movement distance



32

Challenge (homework)

- Work in project groups
- Modify the table summarizing your literature
- Look at the main factorial experiment described in each of the 3 papers and add the following columns to the table:
  - Factorial design: Within-subjects / Between subjects
  - Participants number
  - Independent variables: report levels and factors
  - Counterbalancing schema
  - Dependent variables: for each of them also report unit of measure, gathering method (computer log, manual annotation, …)

33

Experiment Task

- Recall the definition of an independent variable:
  - a circumstance or characteristic that is manipulated in an experiment to *elicit a change in a human response* while interacting with a computer
- The experiment task must "elicit a change"
- Qualities of a good task: *represent, discriminate*
  - Represent activities people do with the interface
  - Discriminate among the test conditions

34

Task Examples

- Usually the task is self-evident (follows directly from the research idea)
- Research idea → a new graphical method for entering equations in a spreadsheet
  - Experiment task → insert an equation using (a) the graphical method and (b) the conventional method
- Research idea → an auditory feedback technique for programming a GPS device
  - Experiment task → program a destination location into a GPS device using (a) the auditory feedback method and (b) the conventional method

35

Knowledge-based Tasks

- Most experiment tasks are *performance-based* or *skill-based* (e.g., inserting an equation, programming a destination location; see previous slide)
- Sometimes the task is *knowledge-based* (e.g., "Use an Internet search interface to find the birth date of Albert Einstein.")
- In this case, participants become contaminated (in a sense) after the first run of task, since they have acquired the knowledge
- Experimentally, this poses problems (beware!)
- A creative approach is needed (e.g., for the other test condition, slightly change the task; "…of William Shakespeare")

36

Procedure

- The *procedure* encompasses everything that occurs with participants
- The procedure includes the experiment task (obviously), but everything else as well…
  - Arriving, welcoming
  - Signing a consent form
  - Instructions given to participants about the experiment task (next slide)
  - Demonstration trials, practice trials
  - Rest breaks
  - Administering of a questionnaire or an interview

37

Instructions

- Very important (best to prepare in advance; write out)
- Often the goal in the experiment task is "to proceed as quickly and accurately as possible but at a pace that is comfortable"
- Other instructions are fine, as per the goal of the experiment or the nature of the tasks, but…
- Give the same instructions to all participants
- If a participant asks for clarification, do not change the instructions in a way that may cause the participant to behave differently from the other participants

38

Participants

- Researchers want experimental results to apply to people not actually tested – a population
- Population examples:
  - Computer-literate adults, teenagers, children, people with certain disabilities, left-handed people, engineers, musicians, etc.
- For results to apply generally to a population, the participants used in the experiment must be…
  - Members of the desired population
  - Selected at random from the population
- True random sampling is rarely done (consider the number and location of people in the population examples above)
- Some form of *convenience sampling* is typical

39

How Many Participants?

- Too few → experimental effects fail to achieve statistical significance
- Too many → statistical significance for effects of no practical value
- The correct number… (drum roll please)
  - Use the same number of participants as used in similar research[1]
- Also consider the counterbalancing schema

[1] Martin, D. W. (2004). *Doing psychology experiments* (6th ed.). Pacific Grove, CA. Belmont, CA: Wadsworth.

40

Questionnaires

- Questionnaires are used in most HCI experiments
- Two purposes:
  - Collect information about the participants
    - Demographics (gender, age, first language, handedness, visual acuity, etc.)
    - Prior experience with interfaces or interaction techniques related to the research
  - Solicit feedback, comments, impressions, suggestions, etc., about participants' use of the experimental apparatus
- Questionnaires, as an adjunct to experimental research, are usually brief

41

Information Questions

- Questions constructed according to how the information will be used

Which browser do you use? _____ **Open-ended**

Which browser do you use? **Closed**
☐ Mozilla *Firefox* ☐ Google *Chrome*
☐ Microsoft *IE* ☐ Other (_____)

Please indicate your age:_____ **Ratio-scale data**

Please indicate your age? **Ordinal-scale data**
☐ < 20 ☐ 20-29 ☐ 30-39
☐ 40-49 ☐ 50-59 ☐ 60+

42

Participant Feedback

- Using NASA Task Load Index (TLX):

Frustration: I felt a high level of insecurity, discouragement, irritation, stress, or annoyance.

1	2	3	4	5	6	7
Strongly disagree			Neutral			Strongly agree

- ISO 9241-9:

Eye fatigue:

1	2	3	4	5	6	7
Very high						Very low

43

System Usability Scale

- 10 Questions in 5 (or 7) point Likert scale
- Odd items are positive, even are in negative form;
- Scores are in range 0-100:
  1. Sum the score contributions from each item (0 to 4).
     - For items 1,3,5,7,and 9 the score contribution is the scale position minus 1.
     - For items 2,4,6,8 and 10, the contribution is 5 minus the scale position.
  2. Multiply the sum of the scores by 2.5 to obtain the overall value of SU.

Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.

44

Challenge

- Design SUS questionnaire for Tapping VS Gesture Writing experiment:
  - Search the questions in the literature
  - Prepare a spreadsheet with score calculation
  - Populate with participant's data (your colleague)
  - Report the score in the experiment spreadsheet
- Homework: Design the initial questionnaire with demographics and previous experience for your project work.

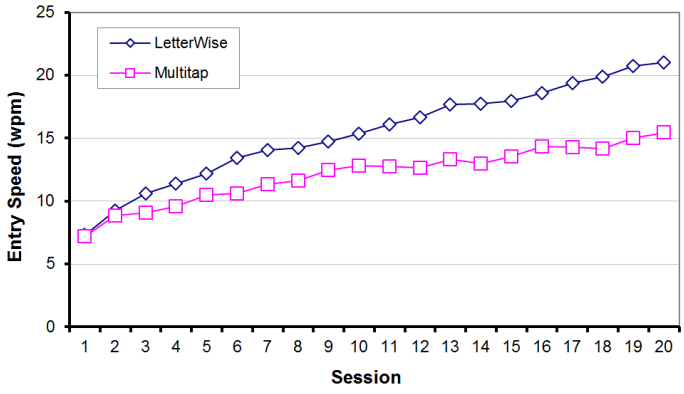
45

Longitudinal Studies

- Sometimes instead of "balancing out" learning effects, the research seeks to promote and investigate learning
- If so, a *longitudinal study* is conducted
- "Practice" is the IV
- Participants are practiced over a prolonged period of time
- Practice units: blocks, sessions, hours, days, etc.
- Example on next slide

46

Longitudinal Study – Results[1]

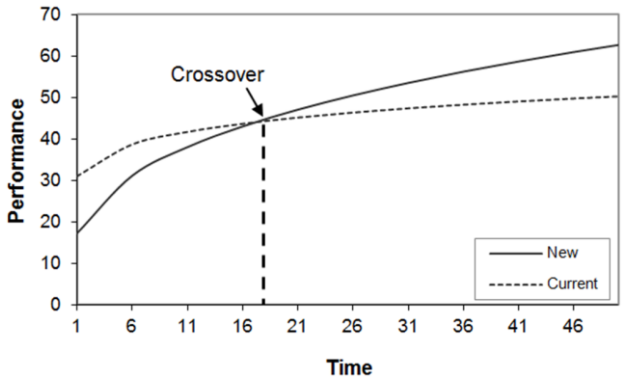


[1] MacKenzie, I. S., Kober, H., Smith, D., Jones, T., & Skepner, E. (2001). LetterWise: Prefix-based disambiguation for mobile text entry. *Proceedings of the ACM Symposium on User Interface Software and Technology - UIST 2001*, 111-120, New York: ACM.

47

The New vs. The Old

- Sometimes a new technique will initially perform poorly in comparison to an established technique
- A longitudinal study will determine if a crossover point occurs and, if so, after how much practice (see below)



48

Running the Experiment

- **Pilot testing** with 1 or 2 participants to: smooth the protocol, check the amount of time for each participant, etc.
- The experiment begins…
  - Participants sign the consent form fill questionnaire
  - Instructions are given
- The experimenter is the *public face* of the experiment:
  - Must portray him/her self as neutral
  - Should avoid to be overly attentive or conveying indifference

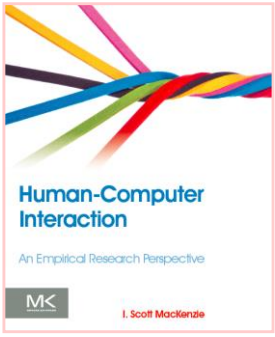
49

Connect to: http://join.quizizz.com

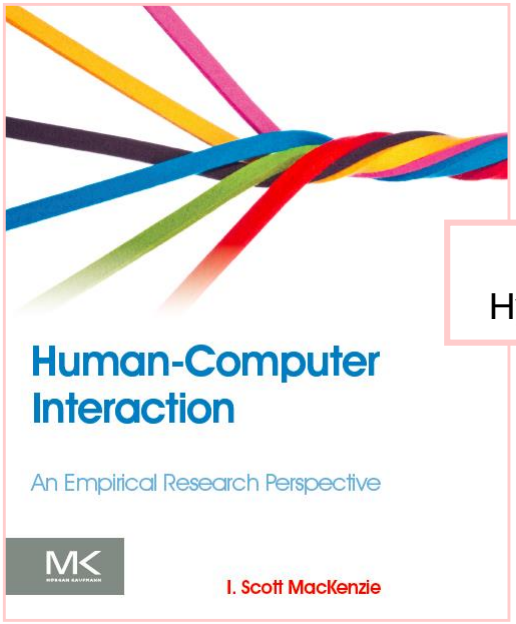
QUESTION TIME

50

Thank You


Human-Computer
Interaction
An Empirical Research Perspective
MK
I. Scott MacKenzie

51



Chapter 6
Hypothesis Testing

Objectives

- Main objective: know enough of statistics to analyze your project's experiment data
  - Use simple tools to run statistical tests to determine whether your results were significant
- You will not learn theoretical foundations of statistics
- We will treat statistic tools as “black boxes”. You will learn how to…
  - …give input and interpret output
  - …write results in your report

2

What is Hypothesis Testing?

- … the use of statistical procedures to answer research questions
- Typical research question (generic):

> Is the time to complete a task less using Method A than using Method B?

- For hypothesis testing, research questions are statements:

> There is no difference in the mean time to complete a task using Method A vs. Method B.

- This is the *null hypothesis* (assumption of "no difference")
- Statistical procedures seek to reject or accept the null hypothesis (details to follow)

3

Statistical Procedures

- Two types:
  - Parametric
    - Data are assumed to come from a distribution, such as the normal distribution, *t*-distribution, etc.
  - Non-parametric
    - Data are not assumed to come from a distribution
- Lots of debate on assumptions testing and what to do if assumptions are not met (avoided here, for the most part)
- A reasonable basis for deciding on the most appropriate test is to match the type of test with the measurement scale of the data (next slide)

4

Measurement Scales vs. Statistical Tests

Measurement Scale	Defining Relations	Examples of Appropriate Statistics	Appropriate Statistical Tests
Nominal	• Equivalence	• Mode • Frequency	• Non-parametric tests
Ordinal	• Equivalence • Order	• Median • Percentile	
Interval	• Equivalence • Order • Ratio of intervals	• Mean • Standard deviation	• Parametric tests • Non-parametric tests
Ratio	• Equivalence • Order • Ratio of intervals • Ratio of values	• Geometric mean • Coefficient of variation	

- Parametric tests most appropriate for…
  - Ratio data, interval data
- Non-parametric tests most appropriate for…
  - Ordinal data, nominal data (although limited use for ratio and interval data)

5

Tests Presented Here

- Parametric
  - Analysis of variance (ANOVA)
    - Used for ratio data and interval data
    - Most common statistical procedure in HCI research
  - Post hoc tests
    - Useful in case we have more than two test conditions
    - To know between which pair of test conditions thare was a significant difference
- We will see…
  - One-Way ANOVA (single factor)
    - 2 Test conditions
    - >2 Test conditions (post-hoc test needed)
    - 2 Test conditions with Between-subject design
  - Two-way ANOVA (two factors)

6

Analysis of Variance

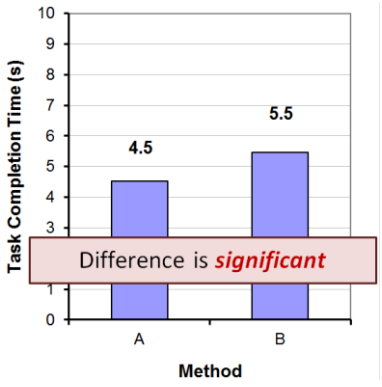
- The *analysis of variance* (ANOVA) is the most widely used statistical test for hypothesis testing in factorial experiments
- Goal → determine if an independent variable has a significant effect on a dependent variable
- Consider the following research question

> Is the time to complete a task less using Method A than using Method B?

- Let's explain through two simple examples (next slide)

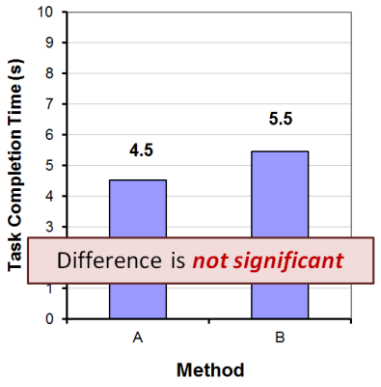
7

Example #1



"Significant" implies that in all likelihood the difference observed is due to the test conditions (Method A vs. Method B).

Example #2

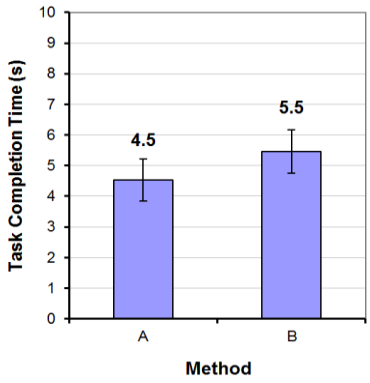


"Not significant" implies that the difference observed is likely due to chance.

File: `06-AnovaDemo.xlsx`

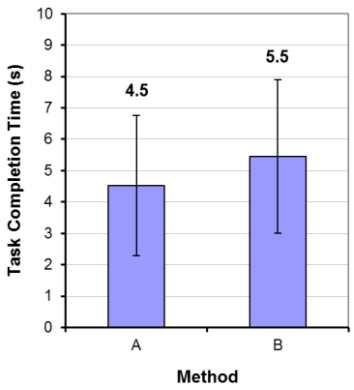
8

Example #1



"Significant" implies that in all likelihood the difference observed is due to the test conditions (Method A vs. Method B).

Example #2

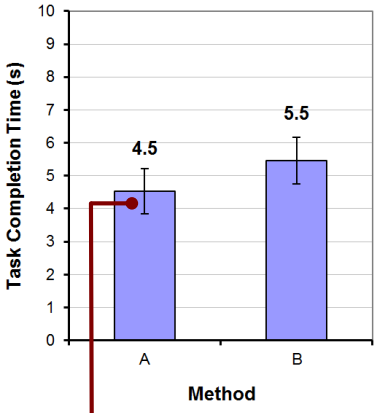


"Not significant" implies that the difference observed is likely due to chance.

File: `06-AnovaDemo.xlsx`

9

Example #1 - Details



Error bars show ±1 standard deviation

Note: Within-subjects design

Participant	Method	
	A	B
1	5.3	5.7
2	3.6	4.8
3	5.2	5.1
4	3.6	4.5
5	4.6	6.0
6	4.1	6.8
7	4.0	6.0
8	4.8	4.6
9	5.2	5.5
10	5.1	5.6
Mean	4.5	5.5
SD	0.68	0.72

Note: *SD* is the square root of the variance

10

Example #1 – ANOVA[1]

ANOVA Table for Task Completion Time (s)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	5.080	.564				
Method	1	4.232	4.232	9.796	.0121	9.796	.804
Method * Subject	9	3.888	.432				

Probability of obtaining the same (or a more extreme) result if the null hypothesis is true

Reported as…

$F_{1,9} = 9.80, p < .05$

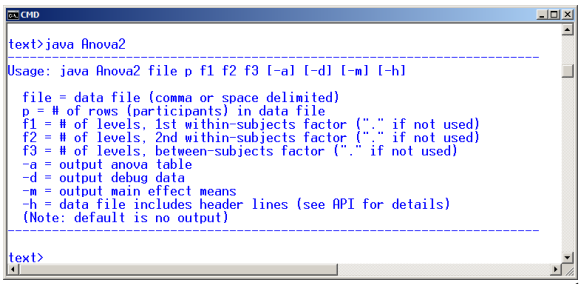
Thresholds for "p"
- .05
- .01
- .005
- .001
- .0005
- .0001

[1] ANOVA table created by *StatView* (now marketed as *JMP*, a product of SAS; www.sas.com)

Anova2 Software

- **HCI:ERP** web site includes analysis of variance Java software: Anova2
- Operates from command line on data in a text file
- Extensive API with demos, data files, discussions, etc.
- Download and demonstrate

Demo



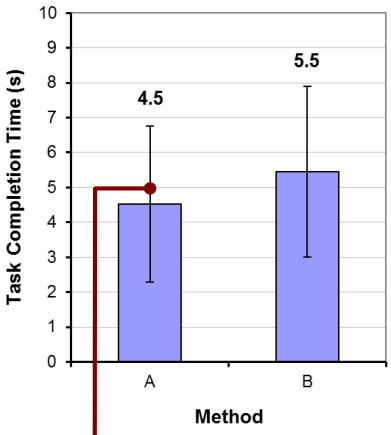
12

How to Report an *F*-statistic

The mean task completion time for Method A was 4.5 s. This was 20.1% less than the mean of 5.5 s observed for Method B. The difference was statistically significant ($F_{1,9} = 9.80$, $p < .05$).

- Notice in the parentheses
  - Uppercase for *F*
  - Lowercase for *p*
  - Italics for *F* and *p*
  - Space both sides of equal sign
  - Space after comma
  - Space on both sides of less-than sign
  - Degrees of freedom are subscript, plain, smaller font
  - Three significant figures for *F* statistic
  - No zero before the decimal point in the *p* statistic (except in Europe)

Example #2 - Details



Error bars show ±1 standard deviation

Participant	Method	
	A	B
1	2.4	6.9
2	2.7	7.2
3	3.4	2.6
4	6.1	1.8
5	6.4	7.8
6	5.4	9.2
7	7.9	4.4
8	1.2	6.6
9	3.0	4.8
10	6.6	3.1
Mean	4.5	5.5
SD	2.23	2.45

Example #2 – ANOVA

ANOVA Table for Task Completion Time (s)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	9	37.372	4.152				
Method	1	4.324	4.324	.626	.4491	.626	.107
Method * Subject	9	62.140	6.904				

Probability of obtaining the same (or a more extreme) result if the null hypothesis is true

Reported as…

$F_{1,9} = 0.626$, ns

Note: For non-significant effects, use "ns" if $F < 1.0$, or "$p > .05$" if $F > 1.0$.

Example #2 - Reporting

The mean task completion times were 4.5 s for Method A and 5.5 s for Method B. As there was substantial variation in the observations across participants, the difference was not statistically significant as revealed in an analysis of variance ($F_{1,9} = 0.626$, ns).

16

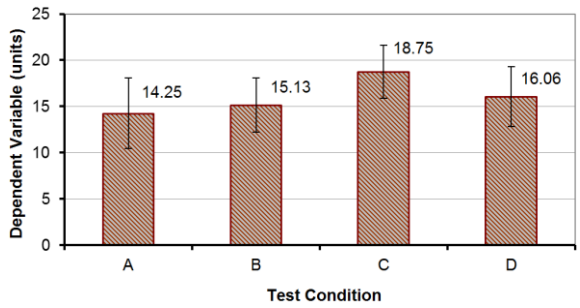
Challenge

- Run ANOVA on the data of the experiment Tapping VS Gesturing
- For each Dependent Variable
  - Which is the null hypothesis?
  - Appropriately format input file
  - Run ANOVA and write the phrase to report the result

17

More Than Two Test Conditions

Participant	Test Condition			
	A	B	C	D
1	11	11	21	16
2	18	11	22	15
3	17	10	18	13
4	19	15	21	20
5	13	17	23	10
6	10	15	15	20
7	14	14	15	13
8	13	14	19	18
9	19	18	16	12
10	10	17	21	18
11	10	19	22	13
12	16	14	18	20
13	10	20	17	19
14	10	13	21	18
15	20	17	14	18
16	18	17	17	14
Mean	14.25	15.13	18.75	16.06
SD	3.84	2.94	2.89	3.23



18

ANOVA

ANOVA Table for Dependent Variable (units)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	15	81.109	5.407				
Test Condition	3	182.172	60.724	4.954	.0047	14.862	.896
Test Condition * Subject	45	551.578	12.257				

- There was a significant effect of Test Condition on the dependent variable ($F_{3,45} = 4.95$, $p < .005$)
- Degrees of freedom
  - If n is the number of test conditions and m is the number of participants, the degrees of freedom are…
  - Effect → $(n - 1)$
  - Residual → $(n - 1)(m - 1)$
  - Note: single-factor, within-subjects design

19

Post Hoc Comparisons Tests

- A significant F-test means that at least one of the test conditions differed significantly from one other test condition
- Does not indicate which test conditions differed significantly from one another
- To determine which pairs differ significantly, a post hoc comparisons tests is used
- Examples:
  - Fisher PLSD, Bonferroni/Dunn, Dunnett, Tukey/Kramer, Games/Howell, Student-Newman-Keuls, orthogonal contrasts, Scheffé
- Scheffé test on next slide

20

Scheffé Post Hoc Comparisons

Scheffe for Dependent Variable (units)
Effect: Test Condition
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
A, B	-.875	3.302	.9003	
A, C	-4.500	3.302	.0032	S
A, D	-1.813	3.302	.4822	
B, C	-3.625	3.302	.0256	S
B, D	-.938	3.302	.8806	
C, D	2.688	3.302	.1520	

- Test conditions A:C and B:C differ significantly (see chart three slides back)

21

Between-subjects Designs

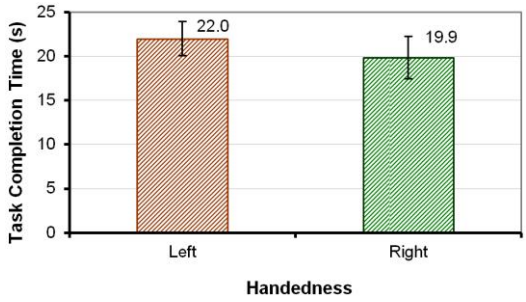
- Research question:
  - *Do left-handed users and right-handed users differ in the time to complete an interaction task?*
- The independent variable (handedness) must be assigned between-subjects
- Example data set →

Participant	Task Completion Time (s)	Handedness
1	23	L
2	19	L
3	22	L
4	21	L
5	23	L
6	20	L
7	25	L
8	23	L
9	17	R
10	19	R
11	16	R
12	21	R
13	23	R
14	20	R
15	22	R
16	21	R
Mean	20.9	
SD	2.38	

22

Summary Data and Chart

Handedness	Task Completion Time (s)	
	Mean	*SD*
Left	22.0	1.93
Right	19.9	2.42



23

ANOVA

ANOVA Table for Task Completion Time (s)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Handedness	1	18.063	18.063	3.781	.0722	3.781	.429
Residual	14	66.875	4.777				

- The difference was not statistically significant ($F_{1,14}$ = 3.78, $p > .05$)
- Degrees of freedom:
  - Effect → $(n - 1)$
  - Residual → $(m - n)$
  - Note: single-factor, between-subjects design

24

Two-way ANOVA

- An experiment with two independent variables is a *two-way design*
- ANOVA tests for
  - Two main effects + one interaction effect
- Example
  - Independent variables
    - Device → D1, D2, D3 (e.g., mouse, stylus, touchpad)
    - Task → T1, T2 (e.g., point-select, drag-select)
  - Dependent variable
    - Task completion time (or something, this isn't important here)
  - Both IVs assigned within-subjects
  - Participants: 12
  - Data set (next slide)

25

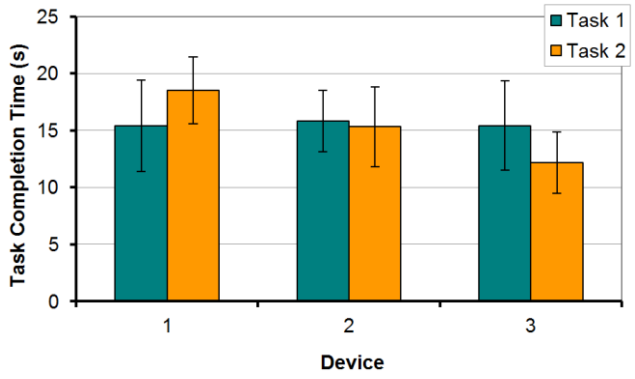
Data Set

Participant	Device 1		Device 2		Device 3	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
1	11	18	15	13	20	14
2	10	14	17	15	11	13
3	10	23	13	20	20	16
4	18	18	11	12	11	10
5	20	21	19	14	19	8
6	14	21	20	11	17	13
7	14	16	15	20	16	12
8	20	21	18	20	14	12
9	14	15	13	17	16	14
10	20	15	18	10	11	16
11	14	20	15	16	10	9
12	20	20	16	16	20	9
Mean	15.4	18.5	15.8	15.3	15.4	12.2
SD	4.01	2.94	2.69	3.50	3.92	2.69

26

Summary Data and Chart

	Task 1	Task 2	*Mean*
Device 1	15.4	18.5	17.0
Device 2	15.8	15.3	15.6
Device 3	15.4	12.2	13.8
Mean	15.6	15.3	15.4



27

ANOVA

ANOVA Table for Task Completion Time (s)

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Subject	11	134.778	12.253				
Device	2	121.028	60.514	5.865	.0091	11.731	.831
Device * Subject	22	226.972	10.317				
Task	1	.889	.889	.076	.7875	.076	.057
Task * Subject	11	128.111	11.646				
Device * Task	2	121.028	60.514	5.435	.0121	10.869	.798
Device * Task * Subject	22	244.972	11.135				

Can you pull the relevant statistics from this chart and craft statements indicating the outcome of the ANOVA?

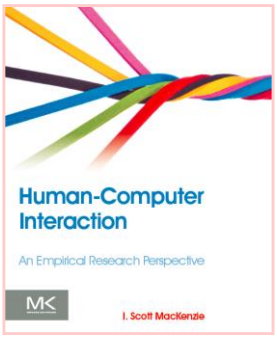
28

ANOVA - Reporting

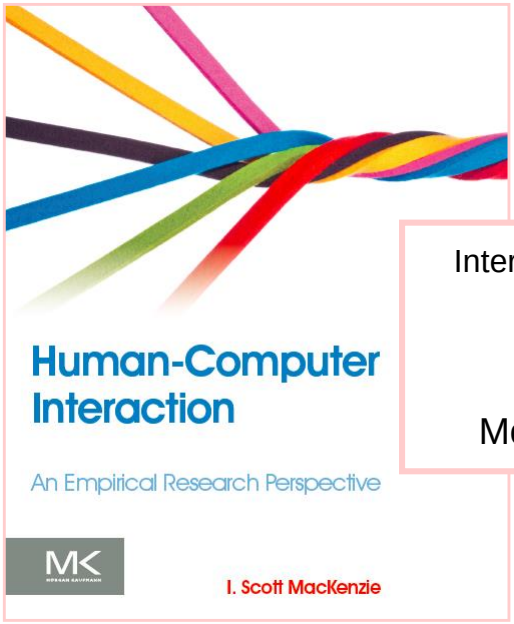
The grand mean for task completion time was 15.4 seconds. Device 3 was the fastest at 13.8 seconds, while device 1 was the slowest at 17.0 seconds. The main effect of device on task completion time was statistically significant ($F_{2,22} = 5.865$, $p < .01$). The task effect was modest, however. Task completion time was 15.6 seconds for task 1. Task 2 was slightly faster at 15.3 seconds; however, the difference was not statistically significant ($F_{1,11} = 0.076$, ns). The results by device and task are shown in Figure x. There was a significant Device × Task interaction effect ($F_{2,22} = 5.435$, $p < .05$), which was due solely to the difference between device 1 task 2 and device 3 task 2, as determined by a Scheffé post hoc analysis.

29

Thank You



30



Interactive Systems Design
Prof. V. Fuccella

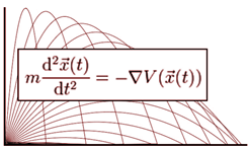
Chapter 7
Modeling Interaction

Introduction

- A *model* is a simplification of reality
- Consider…

Architect's scale model of a building

Physicist's model for the trajectory of a tossed ball



- Both are simplifications of complex phenomena
- The architect's model is a *description* → provides insight into space usage, movement of people, light, shade, etc.
- The physicist's model is a *prediction* → gives the ball's position as a function of time

2

Why Use Predictive Models

- Card et al. presented perhaps the first predictive model in HCI.[1] In many respects, their work was straight-forward experimental research; but they went further:
  - "While these empirical results are of direct use in selecting a pointing device, it would obviously be of greater benefit if a theoretical account of the results could be made. For one thing, the need for some experiments might be obviated; for another, ways of improving pointing performance might be suggested."
- This is a call for the use of predictive models in HCI
- They went on to present predictive models using Fitts' law (which we meet shortly)

[1] Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics, 21*, 601-613.

5

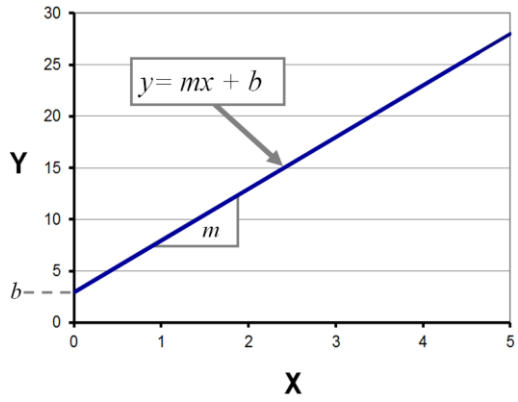
Predictive Model Examples

- Linear prediction equation
- Fitts' law
- Choice reaction time
- Keystroke-level model (KLM)

6

Linear Prediction Equation

- The basic prediction equation expresses a linear relationship between a predictor variable (*x*) and a criterion variable (*y*):



7

Linear Regression

- A linear prediction equation is built using a statistical procedure know as *linear regression*
- Goal:
  - Given a set of *x*-*y* sample points, find the coefficients *m* and *b* (previous slide) for the line that minimizes the squared distances (*least squares*) of the points from the line
- The result is a prediction equation that gives the best estimate of *y* in terms of *x*
- The assumption, of course, is that the relationship is linear
- Want the details? Just enter "linear regression" or "least squares" into Google or Wikipedia

8

Example

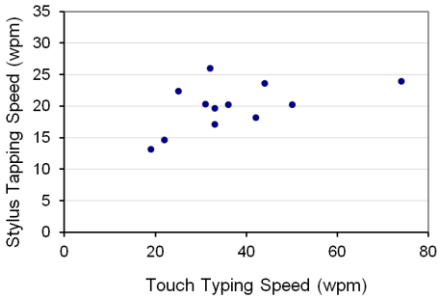
- A research project investigated text entry on soft keyboards[1]
- The research also asked…
  - *Can stylus tapping entry speed be predicted from touch typing entry speed?*
- Touch typing speed is the predictor variable (*x* - measured in a pre-test)
- Stylus typing speed is the criterion variable (*y* - measured experimentally)
- Data and scatter plot

[1] MacKenzie, I. S., & Zhang, S. X. (2001). An empirical investigation of the novice experience with soft keyboards. *Behaviour & Information Technology, 20*, 411-418.

9

Data and Scatter Plot

Participant	Stylus Tapping Speed (wpm)	Touch Typing Speed (wpm)
P1	18.2	42
P2	23.6	44
P3	26.0	32
P4	20.3	50
P5	20.3	36
P6	17.1	33
P7	24.0	74
P8	14.7	22
P9	20.3	31
P10	19.7	33
P11	22.4	25
P12	13.1	19

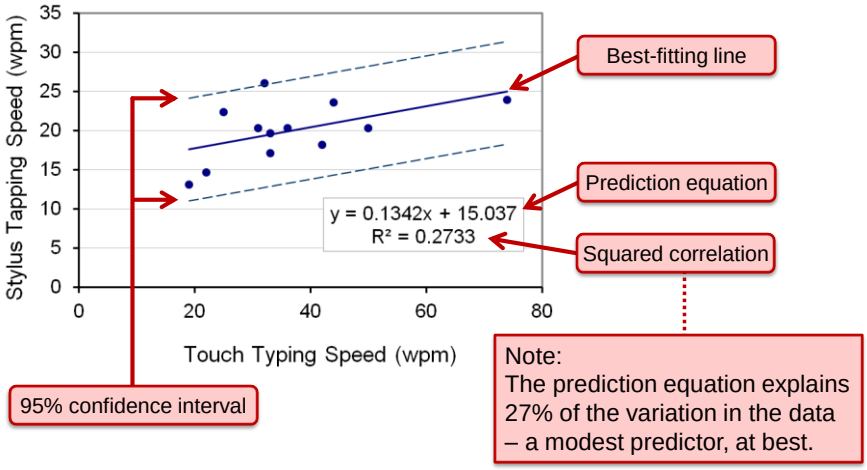


There seems to be a relationship: Faster touch typists seem to be faster at stylus tapping.

Questions:
What is the prediction equation?
How strong is the relationship?

10

Prediction Equation



11

Predictive Model Examples

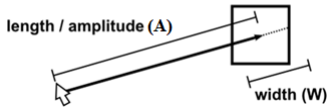
- Linear prediction equation
- Fitts' law
- Choice reaction time
- Keystroke-level model (KLM)

12

Fitts' Law

- One of the most widely used models in HCI
- Model for rapid aimed movements (e.g., moving a cursor toward a target and selecting the target)
- Main application: Use a Fitts' law prediction equation to analyse and compare design alternatives
- Origins: Two highly-cited papers in experimental psychology, one from 1954[1], the other for 1964[2]

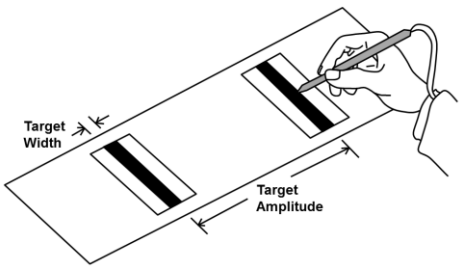
$$MT = a + b\log_2\left(\frac{A}{W} + 1\right)$$



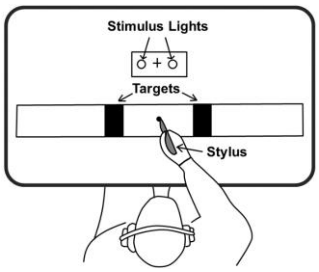
[1] Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381-391.
[2] Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology, 67*, 103-112. **. It is frequently used to predict the movement of**

13

Fitts' Law – Task Paradigms



Serial task



Discrete task

These sketches were adapted from Fitts' 1954 and 1964 papers. It is easy to imagine comparable tasks implemented on computing technology.

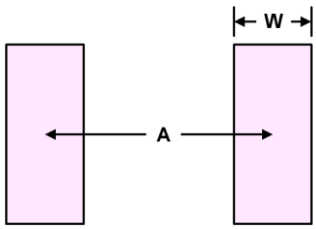
14

Fitts' Index of Difficulty (*ID*)

- Fitts' index of difficulty (*ID*) is a measure of the difficulty of a target selection task:

$$ID = \log_2\left(\frac{A}{W} + 1\right)$$

Units: bits



- Fitts hypothesized that the relationship between movement time (*MT*) and *ID* is linear:

$$MT = a + b\ ID$$

15

Fitts' Law Models for Pointing Devices

- A research project compared four pointing devices, including two for remote pointing[1]
- Twelve participants performed a series of serial target selection tasks using the four devices
- For our purpose, we'll look at the data and models for two of the devices:

Interlink *RemotePoint*

Microsoft *Mouse 2.0*

[1] MacKenzie, I. S., & Jusoh, S. (2001). An evaluation of two input devices for remote pointing. *Proceedings - EHCI 2000*, 235-249, Heidelberg, Germany: Springer-Verlag.

16

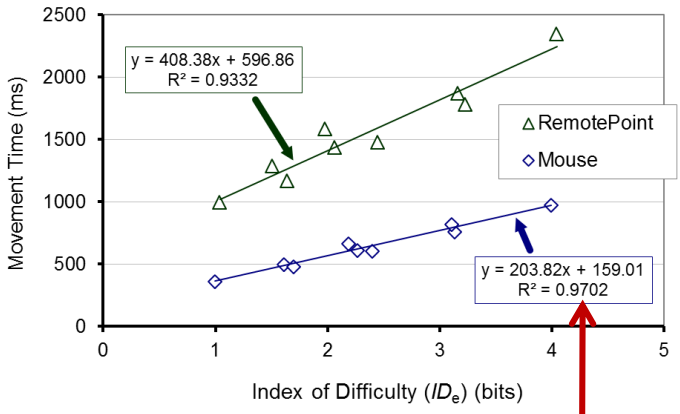
Experiment Conditions and Observations

Conditions | Mouse Observations | RemotePoint Observations

A (pixels)	W (pixels)	ID (bits)	Mouse W_e (pixels)	Mouse ID_e (bits)	Mouse MT (ms)	Mouse TP (bits/s)	RemotePoint W_e (pixels)	RemotePoint ID_e (bits)	RemotePoint MT (ms)	RemotePoint TP (bits/s)
40	10	2.32	11.23	2.19	665	3.29	13.59	1.98	1587	1.25
40	20	1.58	19.46	1.61	501	3.21	21.66	1.51	1293	1.17
40	40	1.00	40.20	1.00	361	2.76	37.92	1.04	1001	1.04
80	10	3.17	10.28	3.13	762	4.11	10.08	3.16	1874	1.69
80	20	2.32	18.72	2.40	604	3.97	25.21	2.06	1442	1.43
80	40	1.58	35.67	1.70	481	3.53	37.75	1.64	1175	1.40
160	10	4.09	10.71	3.99	979	4.08	10.33	4.04	2353	1.72
160	20	3.17	21.04	3.11	823	3.77	19.09	3.23	1788	1.81
160	40	2.32	41.96	2.27	615	3.69	35.97	2.45	1480	1.65
		Mean	23.25	2.38	644	3.60	23.51	2.35	1555	1.46

For model building… *x* sample points, *y* sample points (Mouse); *x* sample points, *y* sample points (RemotePoint)

Fitts' Law Prediction Equations



Squared correlations are very high.
Yes, the *MT-ID* relationship is linear!

Predictive Model Examples

- Linear prediction equation
- Fitts' law
- Choice reaction time
- Keystroke-level model (KLM)

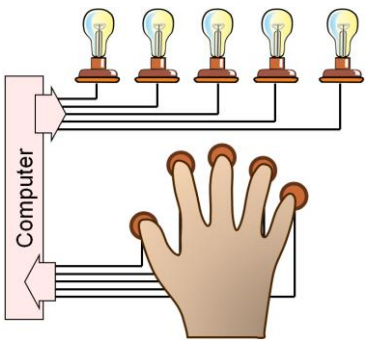
19

Choice Reaction Time

- Given *n* stimuli, associated one-for-one with *n* responses, the time to react to the onset of a stimulus is the *choice reaction time*
- Modeled by the Hick-Hyman law:[1] [2]

$$RT = a + b \log_2(n + 1)$$

- Coefficients:
  - $a \approx 200$ ms
  - $b \approx 150$ ms/bit
- An Information processing model (like Fitts' law)



[1] Hick, W. E. (1952). On the rate of gain of information. *Quarterly J Exp Psychol, 4*, 11-36.
[2] Hyman, R. (1953).Stimulus information as a determinant of reaction time. *J Exp Psychol, 45*, 188-196.

20

HCI Applications

- Not many, but…
  - A telephone operator selects among ten buttons when a light behind a button turns on[1]
  - Time to select items in a hierarchical menu (visual search eliminated by practicing participants to expert levels)[2]
  - Activation time for mode switching with non-dominant hand in a tablet interface[3]
- Difficult to apply because additional behaviours are often present, such as visual search or movement

[1] Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum.
[2] Landauer, T. K., & Nachbar, D. W. (1985). Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width. *Proc CHI '85*, 73-77, ACM.
[3] Ruiz, J., Bunt, A., & Lank, E. (2008). A model of non-preferred hand mode switching. *Proceedings of Graphics Interface 2008*, 49-56, Toronto: Canadian Information Processing Society.

21

Predictive Model Examples

- Linear prediction equation
- Fitts' law
- Choice reaction time
- Keystroke-level model (KLM)

22

Keystroke-Level Model (KLM)[1] [2]

- One of the earliest and most comprehensive models in HCI
- Developed specifically for predicting human performance with interactive computing systems
- Predicts expert error-free task completion times

[1] Card, S. K., Moran, T. P., & Newell, A. (1980, July). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, *23*, 396-410.
[2] Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Erlbaum.

23

Why Use the KLM?

- Consider a task such as:

  Replace one 5-letter word with another, starting one line away.

- There are at least three ways to do the task:
  1. arrow keys, backspace, type
  2. search and replace dialog
  3. point and double click with the mouse, type
- The KLM can predict the time for each method
- If used at the design stage, design alternatives may be considered and compared → design choices follow

24

A KLM Prediction

- A task is broken into a series of subtasks
- Total predicted time is the sum of the subtask times:

$$t_{EXECUTE} = t_K + t_P + t_H + t_D + t_M + t_R$$

- Operators:
  - K → keystroking P → pointing H → homing
  - D → drawing M → mental prep R → system response
- Some operators are omitted or repeated, depending on the task (e.g., if *n* keystroking operations are required, t_K becomes $n \times t_K$)
- Operator values (next slide)

25

KLM Operators and Values

Operator	Description	Time (s)
K	PRESS A KEY OR BUTTON Pressing a modifier key (e.g., shift) counts as a separate operation, Time varies with typing skill: Best typist (135 wpm) Good typist (90 wpm) Average skilled typist (55 wpm) Average non-secretary typist (40 wpm) Typing random letters Typing complex codes Worst typist (unfamiliar with keyboard)	 0.08 0.12 0.20 0.28 0.50 0.75 1.20
P	POINT WITH A MOUSE Empirical value based on Fitts' law. Range from 0.8 to 1.5 seconds. Operator does *not* include the button click at the end of a pointing operation	1.10
H	HOME HAND(S) ON KEYBOARD OR OTHER DEVICE	0.40
D(n_D,l_D)	DRAW n_D STRAIGHT-LINE SEGMENTS OF TOTAL LENGTH l_D. Drawing with the mouse constrained to a grid.	.9 n_D + .16 l_D
M	MENTALLY PREPARE	1.35
R(*t*)	RESPONSE BY SYSTEM Different commands require different response times. Counted only if the user must wait.	*t*

26

Original KLM Experiment

- The KLM was validated in an experiment with fourteen tasks performed using various methods and systems
- Example: Task 1 → *Replace a 5-letter word with another word (one line from previous task)*
- Using one system, POET, the task was broken down as follows:

Jump to next line	**M K**[LINEFEED]
Issue Substitute command	**M K**[S]
Type new word	**K**[word]
Terminate new word	**M K**[RETURN]
Type old word	**K**[word]
Terminate old word	**M K**[RETURN]
Terminate command	**K**[RETURN]

- 4 mental operations + 15 keystroking operations, hence →

27

KLM Prediction (Example)

$$t_{EXECUTE} = 4 \times t_M + 15 \times t_K$$

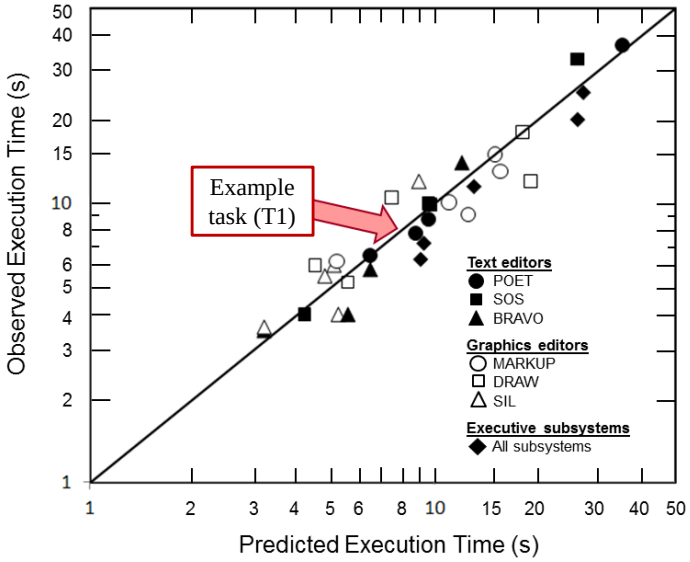
- M set to 1.35 seconds (two slides back)
- K set to 0.23 seconds, based on a 5-minute pre-test
- So…

$$t_{EXECUTE} = 4 \times 1.35 + 15 \times 0.23$$
$$= 8.85 \text{ seconds}$$

- This is the prediction
- What about the observation? (next slide)

28

Results



29

Modern Applications

- Mouse interaction was just emerging when the KLM was introduced
- An obvious KLM update is to replace the pointing constant (t_P) with a Fitts' law prediction equation, as appropriate for the device (e.g., mouse vs. touchpad) and task (e.g., point-select vs. drag-select)
- For example, using the Fitts' law equation given earlier for the mouse…

$$t_P = 0.159 + 0.204 \times \log_2\left(\frac{A}{W} + 1\right)$$

30

Pointing Operator – Update

- For example, a mouse point-select operation over 3.2 cm to click a 1.2 cm wide toolbar button should take about…

$$t_P = 0.159 + 0.204 \times \log_2\left(\frac{3.2}{1.2} + 1\right) = 0.45 \text{ seconds}$$

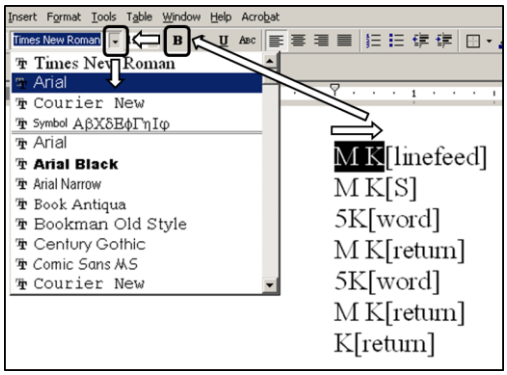
- If the same task involves moving the pointer 44.6 cm, the prediction becomes…

$$t_P = 0.159 + 0.204 \times \log_2\left(\frac{44.6}{1.2} + 1\right) = 1.22 \text{ seconds}$$

31

Pointing Operator – Example

- Develop KLM mouse and keyboard predictions for the GUI screen below
- Task: Change the font and style for "M K" to bold, Arial



32

Mouse Analysis

- Operations:

Mouse Subtasks	KLM Operators	t_P (s)
Drag across text to select "M K"	**M P**[2.5, 0.5]	0.686
Move pointer to Bold button and click	**M P**[13, 1]	0.936
Move pointer to Font drop-down button and click	**M P**[3.3, 1]	0.588
Move pointer down list to Arial and click	**M P**[2.2, 1]	0.501
	$\sum t_P$ =	2.71

- Prediction:

$$t_{\text{EXECUTE}} = 4 \times t_M + \sum t_P = 4 \times 1.35 + 2.71 = 8.11 \text{ seconds}$$

33

Keyboard Analysis

- Operations:

Keyboard Subtasks	KLM Operators
Select text	**M K**[shift] **3K**[→]
Convert to boldface	**M K**[ctrl] **K**[b]
Activate Format menu and enter Font sub-menu	**M K**[alt] **K**[o] **K**[f]
Type *a* ("Arial" appears at top of list)	**M K**[a]
Select "Arial"	**K**[↓] **K**[enter]

- Prediction:

$$t_{\text{EXECUTE}} = 4 \times t_M + 12 \times t_K = 4 \times 1.35 + 12 \times 0.75 = 14.40 \text{ seconds}$$

Use "typing complex codes" (t_K = 0.75 s)

34

Contemporary Uses of the KLM

- The KLM continues to be widely used in HCI
- Some contemporary uses include…
  - Attention shifts with mobile phones
  - Stylus-based circling gestures
  - Managing folders and messages in e-mail applications
  - Predictive text entry on mobile phones
  - Task switching in multi-monitor systems
  - Mode switching on tablet PCs
  - Distractions in in vehicle information systems (IVIS)

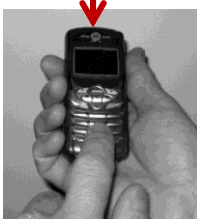
See **HCI:ERP** for citations

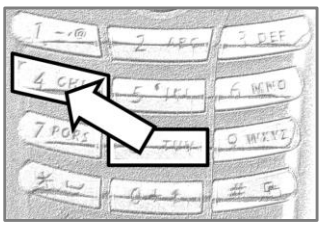
35

KLM and Mobile Phone Keypad

- Since input uses a single finger or thumb, the keystroking operator can be replaced with a pointing operator[1]

$$t_{P(\text{index finger})} = 0.165 + 0.052 \log_2\left(\frac{A}{W} + 1\right)$$





$$t_{P(\text{thumb})} = 0.176 + 0.064 \log_2\left(\frac{A}{W} + 1\right)$$

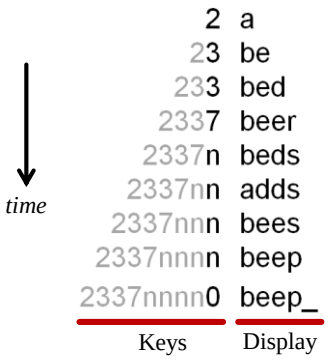
See **HCI:ERP** for further discussion

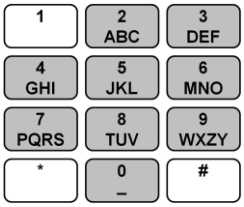
[1] Silfverberg, M., MacKenzie, I. S., & Korhonen, P. (2000). Predicting text entry speed on mobile phones. *Proc CHI 2000*, 9-16, New York: ACM.

36

KLM and Predictive Text Entry

- Interesting because of the combination of keystroking/pointing operations and mental operations
- Consider entering "beep" on a mobile phone keypad using predictive text entry (T9):



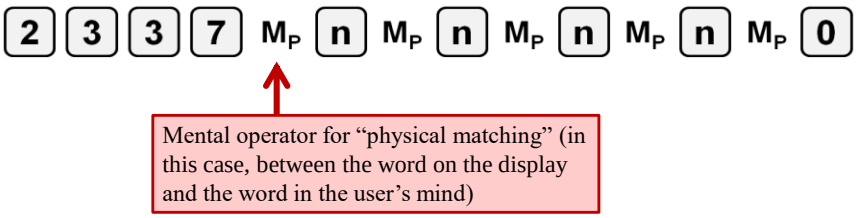


Mobile phone keypad (for reference)

KLM model?

KLM Operators for "beep"

- Perhaps the following…



- Two questions:
  1. *Where should mental operators be placed?*
  2. *What value should the mental operator assume?*

Expert Behaviour

- Experts know the T9 key sequences for common words (i.e., no need for M_P at end of word):
  - *the* → 8430
  - *of* → 630
  - *and* → 2630
- But, how far down a word frequency list[1] does such behaviour extend?
- What about ambiguous words?
  - *if* → 43n0
  - *no* → 66n0
  - *beep* → 2337nnnn0

"beep" is at rank 20,767. Even experts will likely require the M_P operator at the end of the word (unless, for some reason, "beep" is a word they commonly enter).

[1] The word-frequency list used here is the 64,000 word list described by Silfverberg et al. (2000).

39

Ambiguous Words

- Below is a list of the top ten words requiring a press of next

Rank	Word	Keystrokes	Higher Ranking Colliding Word (rank)
47	if	43n0	he (15)
51	no	66n0	on (13)
63	then	8436n0	them (57)
72	me	63n0	of (2)
78	these	84373n0	there (35)
105	go	46n0	in (6)
118	us	87n0	up (56)
159	home	4663n0	good (115)
227	night	64448n0	might (141)
298	war	927n0	was (10)

M_P operators?

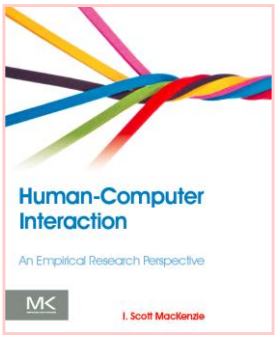
40

Heuristics for M_P Operator

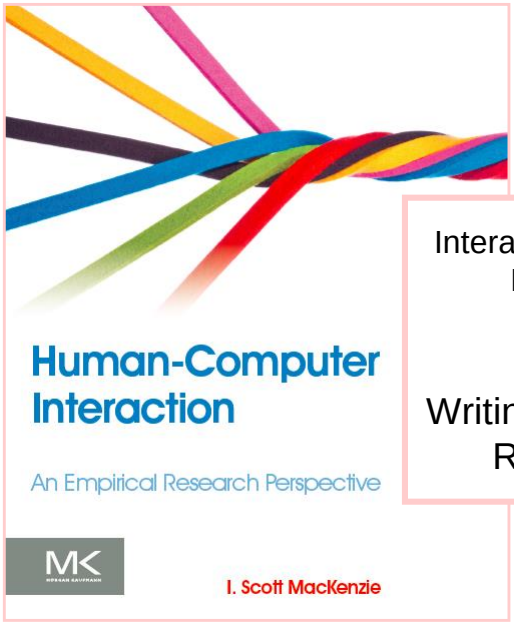
- It is not possible to precisely know when a user will hesitate to perform an M_P operation – a physical match between the word on the display and the word in the user's mind
- Two approaches for KLM modeling with M_P:
  - All-in → include M_P at every reasonable juncture
  - All-out → exclude all M_P operations
- The two approaches will produce upper bound (all-in) and lower bound (all-out) predictions

41

Thank You



42



Interactive Systems Design
Prof. V. Fuccella

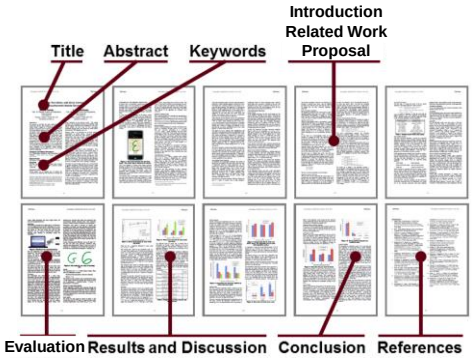
Chapter 8
Writing and Publishing a Research Paper

Context

- Publication is the final (and essential!) step in a research project
- And so, we finish with a chapter on writing and publishing a research paper

Parts of a Research Paper[1]

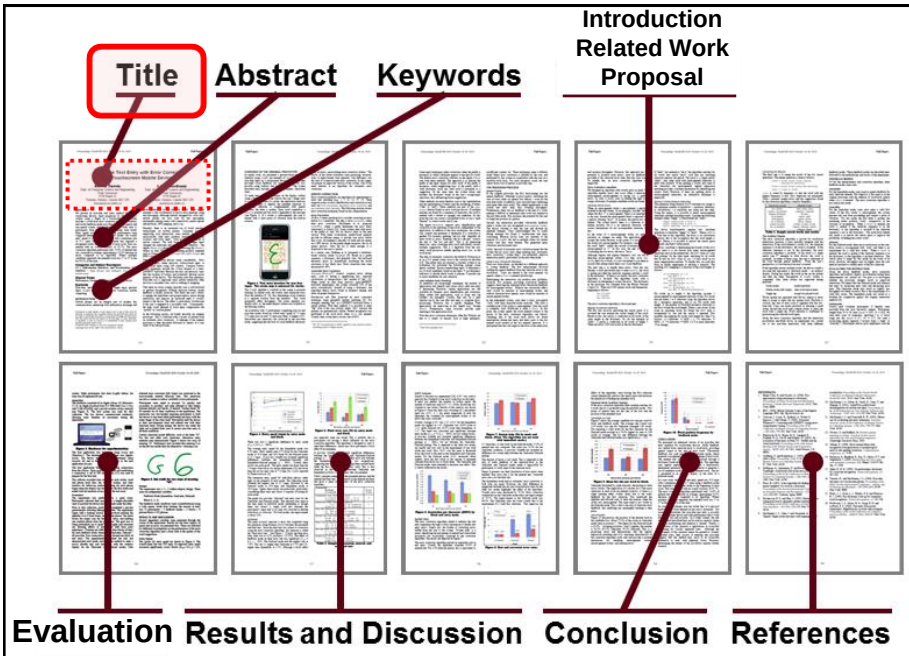
Backdrop paper



[click here] to view the backdrop paper
(nordichi2010.pdf)

[1] Tinwala, H., & MacKenzie, I. S. (2010). Eyes-free text entry with error correction on touchscreen mobile devices. *Proc NordiCHI 2010*, 511-520, New York: ACM.

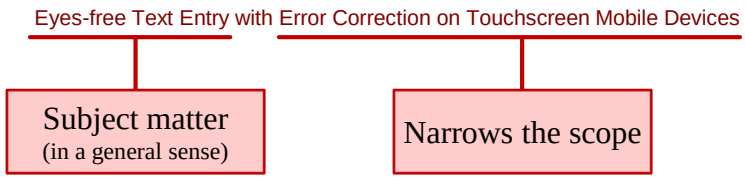
3



4

Title

- Every word tells!
- The title must…
  - Identify the subject matter of the paper
  - Narrow the scope of the work
  - (A title should be neither too broad nor too narrow.)
- Backdrop paper title:



5

Authors and Affiliations

- … follow the title
- Format as per the template file

Download the SIGCHI template file (for conference papers)

`(proceedings.tex)`

Title

Authors and affiliations

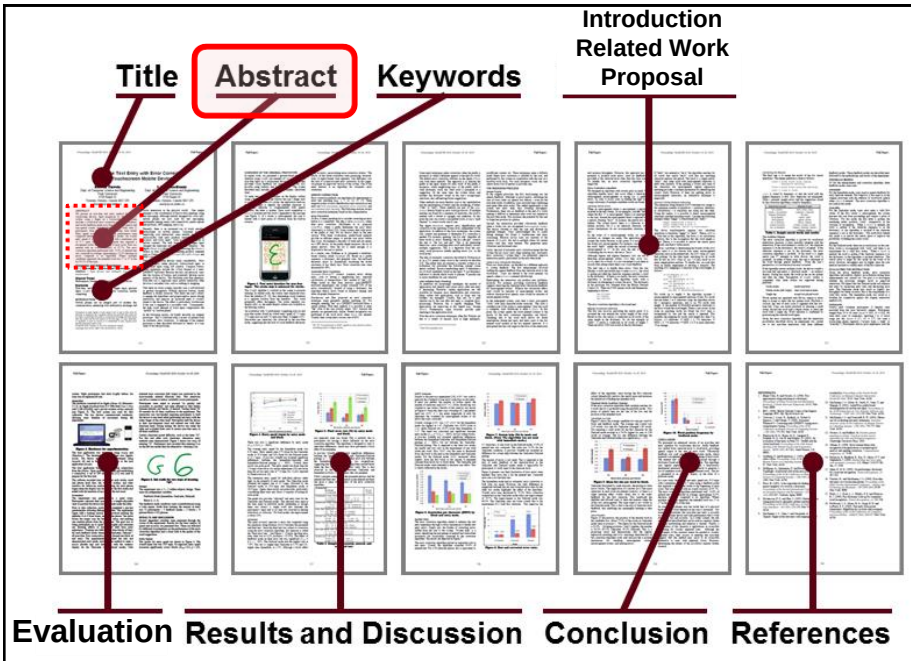
From the SIGCHI template file…

SIGCHI Conference Proceedings Format

1st Author Name	2nd Author Name	3rd Author Name
Affiliation	Affiliation	Affiliation
Address	Address	Address
e-mail address	e-mail address	e-mail address
Optional phone number	Optional phone number	Optional phone number

Details matter! Ensure the font family, font size, font style, and positioning are correct.

6



7

Abstract

- Written last
- Typically a word limit (e.g., 150 words)
- A single paragraph, no citations
- The abstract's mission is to tell the reader…
  1. What you did
  2. What you found
- Give the most salient finding(s)
- Common fault:
  - Treating the abstract as an introduction to the subject matter (don't!)

8

Abstract Example[1]

What was done

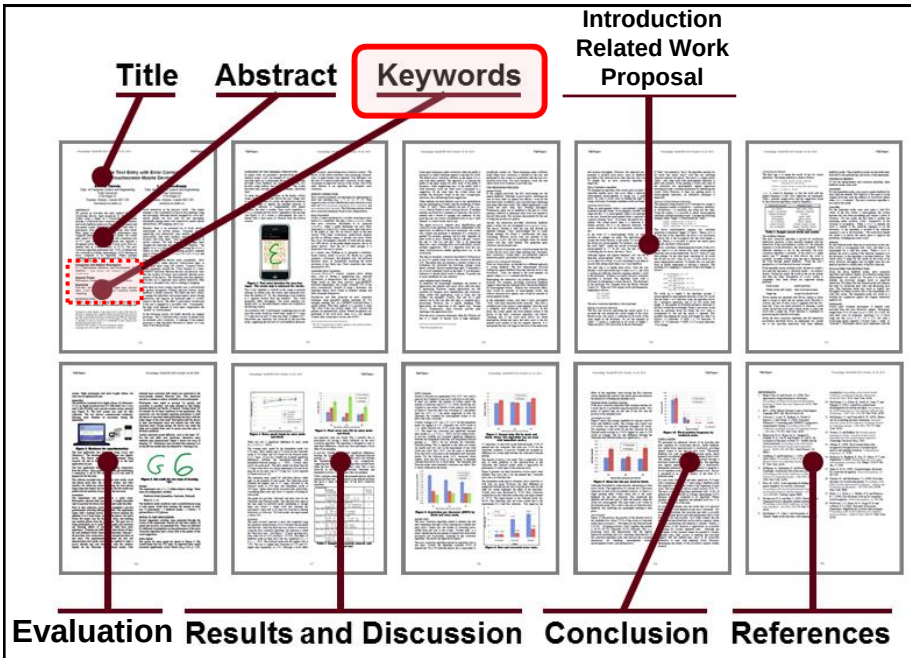
This study addresses to what extent spatial mnemonics can be used to assist users to memorize or infer a set of text input chords. Users mentally visualize the appearance of each character as a 3x3 pixel grid. This grid is input as a sequence of three chords using one, two, or three fingers to construct each chord. Experiments show that users are able to use the strategy after a few minutes of instruction, and that some subjects enter text without help after three hours of practice. Further, the experiments show that text can be input at a mean rate of 5.9 words per minute (9.9 words per minute for the fastest subject) after 3 hours of practice. On the downside, the approach suffers from a relatively high error rate of about 10% as subjects often resort to trial and error when recalling character patterns.

What was found

(144 words)

[1] Sandnes, F. E. (2006). Can spatial mnemonics accelerate the learning of text input chords? *Proceedings of the Working Conference on Advanced Visual Interfaces - AVI 2006*, 245-249, New York: ACM.

9



10

Keywords

- Used for database indexing and searching
- Chosen by the author(s)
- Backdrop paper:

Keywords
Eyes-free, text entry, touchscreen, finger input, gestural input, *Graffiti,* auditory display, error correction, mobile computing.

11

Computing Classification System

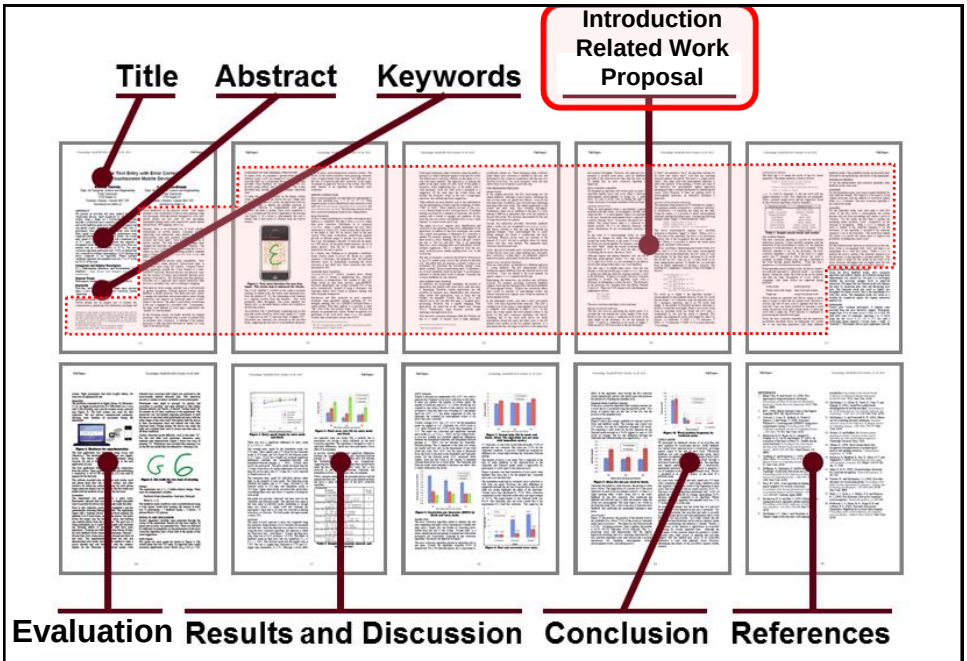
- Since 1998, ACM conference and journal papers are required to also include categories, subject descriptors, and general terms (the latter are optional for conference papers)
- Provided by the ACM (not the author)
- Backdrop paper:

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *input devices and strategies (e.g., mouse, touchscreen)*

General Terms
Performance, Design, Experimentation, Human Factors

Click here to view the ACM's how-to guide (if Internet connection available)
(http://www.acm.org/about/class/how-to-use)

12



13

Introduction

- Opening section of the research paper
- Headings vary (e.g., Introduction, Background, …)
- Expected Content:
  - Introduction to the topic of research
  - UI problem or challenge statement
  - Citation of most notable (if existing) solutions
  - Contribution of the work
    - What is novel and interesting about the research?
    - Anticipation of the impending solution (which is developed and evaluated in the rest of the paper)
  - Overview of the paper

14

Overview of Paper

- Usually an overview of the entire paper is given a the end of the introduction
- Backdrop paper:

> In the following section, we briefly describe our original prototype. This is followed with a review of related work on automatic error correction. A redesign of the original prototype is then described followed by details of a user study to test the prototype.

(5th paragraph)

15

Related Work

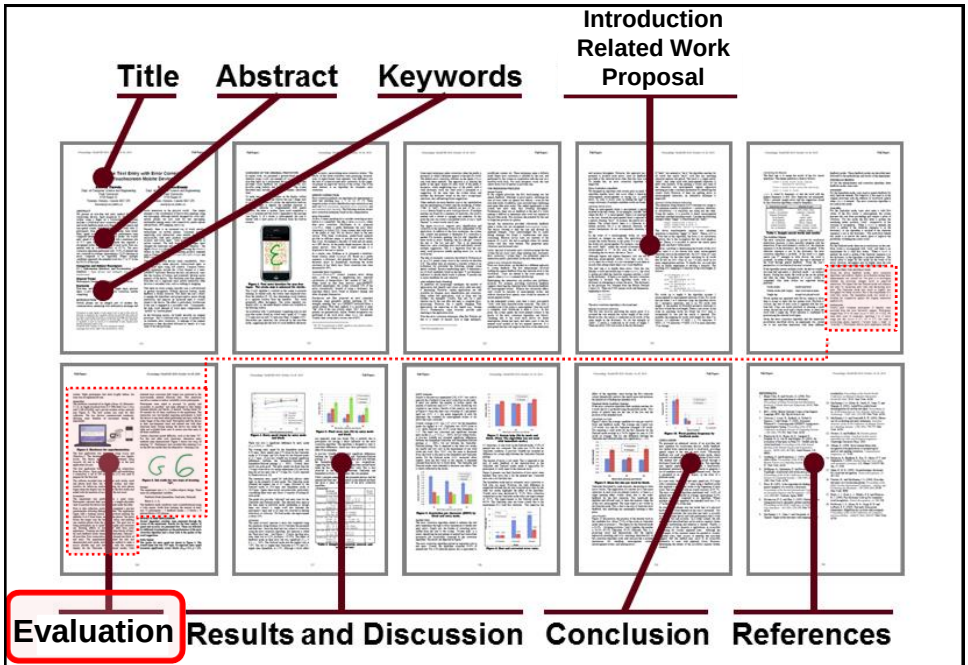
- Literature review: discussion of related work
  - Shortly describe each work
  - Highlight differences and similarities with your solution
- Include citations (with full bibliographic information in reference section at end)
- Possibly include a table summarizing the main features of related work
- Organization: possibly classify related approaches according to a common important feature
- Usually, in short papers the section is not present and a small number of related work is summarized in the introduction

16

Proposal

- Describe in details the proposed solution
- An interactive technique can be difficult to describe
  - Use images
  - Reference a movie (e.g. on Youtube)
- Sections and sub-sections
  - No rules (organize in any manner that seems reasonable)
  - It's your story to tell!

17



18

Evaluation

- Tells the reader how the experiment was designed and carried out
- Headings vary (Evaluation, Method, Methodology, Experiment, User Study, …)
- In style, the evaluation section must be straight-forward: simple, clear, predictable (like a recipe)
- Research must be replicable (as already noted)
  - The section must provide sufficient information that a skilled researcher could replicate the experiment if he/she chose

19

Predictability

- The organization of evaluation section must be predictable
- Allows a reader to scour papers quickly to find key points in the design of the experiment
- Convention dictates that the method section contains the following sub-sections (and in the following order):
  - Participants
  - Apparatus
  - Procedure
  - Design

20

Participants

- The Participants sub-section tells the reader the number of participants and how they were selected
- Were they volunteers or were they paid?
- Demographic information is also given (e.g., age, gender, related experience, …)
- Other details, as appropriate (e.g., income, highest level of education, visual acuity, …)
- This section is usually short, however…

21

Apparatus

- The Apparatus sub-section describes the system (hardware and software)
- Headings vary (e.g., Materials, Interface, …)
- Reproducibility extremely important
  - Give all the details necessary
- Use screen snaps or photos of the interface
- If technical details were disclosed in the Introduction, just refer the reader back to an earlier section (e.g., “the software included the algorithm described in the preceding section”)

22

Procedure

- The Procedure sub-section tells the reader exactly what happened with each participant
- Things to note:
  - Instructions
  - Task description
  - Demonstration or practice
  - Questionnaire administering
  - Trial repetitions, rest breaks, total time
  - etc.

23

Experiment Task

- Procedure section describes the task:
  - What was the task?
  - What was the goal of the task?
  - When did timing begin and end?
  - Were errors recorded?
  - Were participants instructed to, or allowed to, correct errors?
  - How were errors corrected?
  - Did participants correct errors at their discretion?
  - Were rest breaks allowed, encouraged, or enforced?
  - Etc. (give all the details!)

24

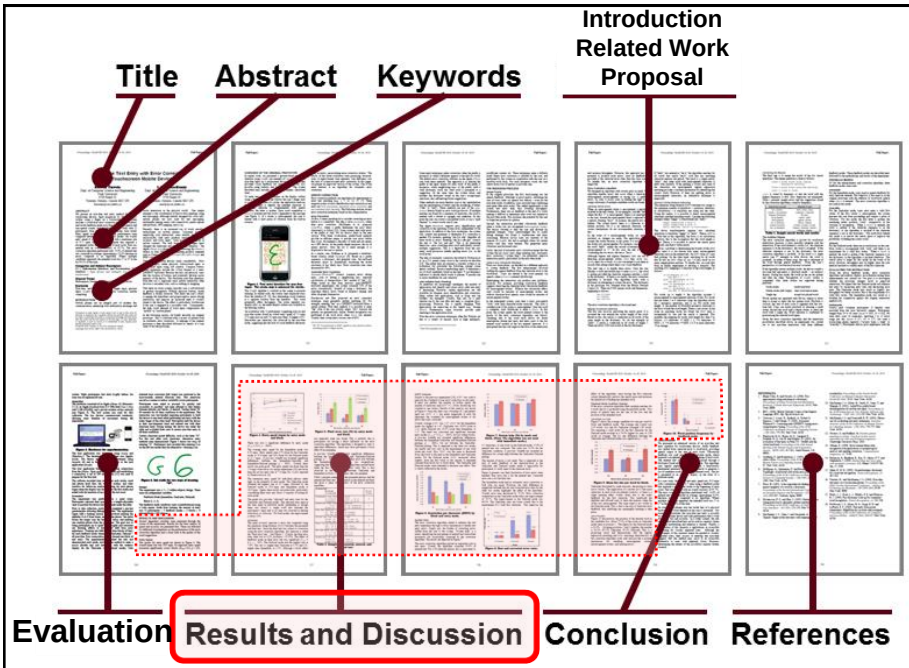
Design

- The Design sub-section summarizes the experiment in terms of the variables, assignment of conditions, etc.
- For short papers, these details are sometimes given in the Procedure section
- Common beginning…
  - "The experiment was a 3 × 2 within-subjects design…"
- Conclude with a big-picture summary:

> Aside from training, the amount of entry was 12 participants × 3 feedback modes × 3 blocks × 4 phrases/block = 432 phrases.

(Backdrop paper)

25



26

Results and Discussion

- Results and discussion are usually combined
- Same level heading as Evaluation (results are not part of the evaluation)
- If there were outliers or any data filtering or transformations, state this up front
- Statistical approach and tests sometimes conveyed in an opening paragraph
- No strict rules, but a common approach is to organize this section by dependent variables, beginning with the most important (e.g., speed, task completion time)

27

Results and Discussion (2)

- For each dependent variable, begin with a broad observation, then progress to finer details
- Give the effect size in absolute and/or relative terms:

> The mean task completion time for method A was 2.7 seconds. Method B was 9.1% slower with a mean task completion time of 3.0 seconds.

28

Results and Discussion (3)

- Report results of statistical analysis

> As expected, entry speed increased significantly across blocks ($F_{2,18} = 6.2, p < .05$). There was also a significant difference by entry mode ($F_{2,18} = 32.3, p < .0001$).

29

Results and Discussion (4)

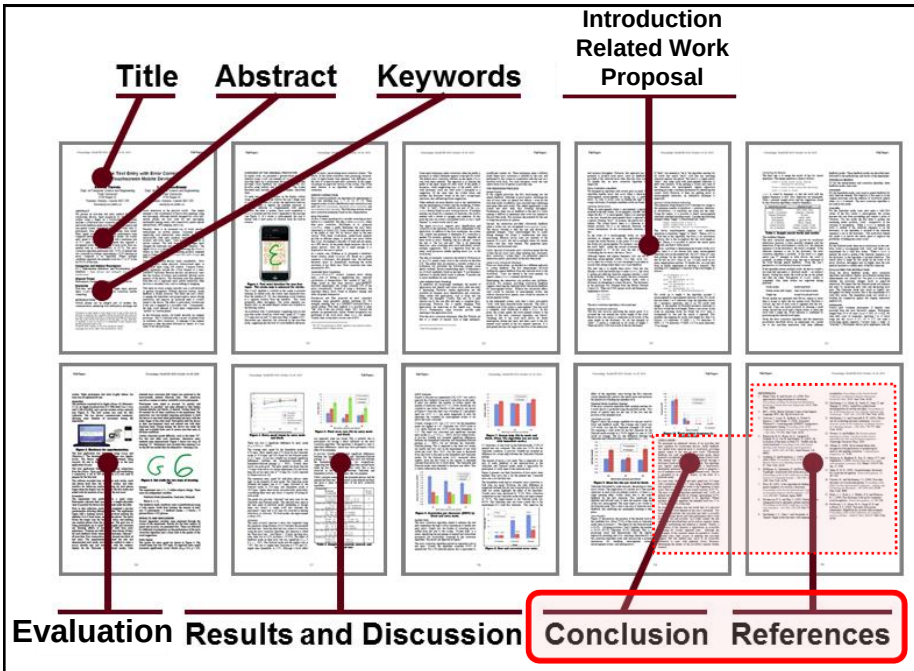
- Discuss and explain the results:
  - What caused the differences in the measurements across experimental condition?
  - What detail in the interaction cause one method to be faster/slower than the other?
  - Did one condition require more input actions?
  - **Important:** report *limitations* of the proposed approach and of the experimental methodology
    - Were participants confused?
    - Was the method hard to learn?
    - Did participants experience fatigue or discomfort?
    - etc.

30

Results and Discussion (5)

- Do not give too many results
  - It is your job to distinguish what is important and relevant from what is unimportant
- Compare
  - Draw comparisons with related work (cited, of course)
- Visuals
  - Use as appropriate, to illustrate and create interest
  - Line charts, bar charts, etc.
- Participant feedback
  - Interviews, questionnaires, etc.
  - Analyse, discuss

31



32

Conclusion, References

- Conclusion
  - Summarize what you did
  - Restate contribution and/or significant findings
  - Identify topics for further work (but avoid developing new ideas in the Conclusion section)
- Acknowledgment
  - Optional (thank people who helped, funding agencies)
- References
  - Full bibliographic information for papers cited
  - Format as required (details matter!)

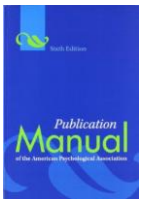
33

Preparing the Manuscript

34

Formatting Rules

- Consult template files or other requirements for conference or journal submissions
- A good source: APA Publication Manual[1] →
- APA's on-line FAQ:
  - *When do you use a comma?*
  - *When do you use double quotation marks?*
  - *Do you use brackets in the same way you use parentheses?*
  - *When are numbers expressed in words?*
  - *etc.*



Click here to view FAQs about APA style (if Internet connection available)
(http://www.apastyle.org/learn/faqs/)

[1] APA. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: APA.

36

Dictionary

- The final source for spelling
  - British or American spelling fine; be consistent
- Also, use a dictionary to determine…
  - When to capitalize (*Internet*)
  - When to use a hyphen (*e-mail*)
  - When not to use a hyphen (*online*)
  - When to set as two words (*screen snap*)
  - When to set a single word (*database*)

Click here to view Merriam-Webster dictionary (if Internet connection available)
(http://www.merriam-webster.com/)

37

Citations and Reference Lists

- Format citations and references as required for the type of submission
- Next slide gives examples for typical conference proceedings

38

REFERENCES

1. Aula, A., Khan, R. M., and Guan, Z., How does search behavior change as search becomes more difficult? *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems - CHI 2010*, (New York: ACM, 2010), 35-44.
2. Brajnik, G., Yesilada, Y., and Harper, S., The expertise effect of web accessibility evaluation methods, *Human-Computer Interaction*, *26*, 2011, 246-283.
3. Brown, T., *Change by design: How design thinking transforms organizations and inspires innovation*. New York: HarperCollins, 2009.
4. Buxton, W., There's more to interaction than meets the eye: Some issues in manual input, in *User centered system design: New perspectives on human-computer interaction*, (D. A. Norman and S. W. Draper, Eds.). Hillsdale, NJ: Erlbaum, 1986, 319-337.
5. ESA, Electronic Software Association, *Industry facts*, http://www.theesa.com/facts/, (accessed February 4, 2012).

Conference paper

Journal paper

Book

Book chapter

Internet document

39

Checklist (see previous slide)

- References are numbered.
- References are ordered alphabetically by 1st author's surname.
- For each author, the surname comes first, followed a comma, then the initials for the given names. Include a space between the initials if there is more than one (e.g., "Smith, B. A." not "Smith, B.A.")
- For the title of the publication, only capitalize the first word, the first word in a secondary title (e.g., after a colon), and proper nouns.
- Always include the year. Substitute "in press" for accepted but not-yet-published papers.
- Always include pages (except for complete books or web pages).
- For the name of the publication, set in italics and capitalize all keywords (e.g., *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems – CHI 2011*).
- For journal publications, include the volume number in italics.
- If space permits, use the full name for conferences and journals. If space is tight, use abbreviated names for conferences and journals (e.g., *Proc CHI '99*). Do not mix full and abbreviated names; use one style or the other. If using abbreviated names, be consistent.
- Give the location and name of the publisher for conference papers and books (e.g., "New York: ACM"). Use the most economical yet understandable expression of the location (e.g., "New York," not "New York: NY"; but use "Cambridge: MA") and publisher (e.g., "Springer" not "Springer Publishing Company").
- Use *align left* (ragged right) for the reference list. (Note: The rest of the manuscript is justified.)
- Only include works that are cited in paper.
- Study and imitate!
- Be consistent.

40

Citation Examples

Basic citation:

A previous experiment [5] confirmed that…

Group multiple citations together:

Our results are consistent with previous findings [e.g., 5, 7, 12].

Do not treat citations as nouns:

It was proposed in [5] that… *** Incorrect ***
It was proposed by Smith and Jones [5] that… *** Correct ***

Exception (within parentheses):

There are many user studies on this topic (see [6] for a review).

41

Citation Examples (2)

Quotations require a page number:

> Smith and Jones argue, "the primary purpose of research is publication" [14, p. 125].

Include page numbers when citing a point from a book:

> Norman defines six categories of slips [15, pp. 105-110].

Use "et al." if there are three or more authors:

> Douglas et al. [5] describe an empirical evaluation using an isometric joystick.

42

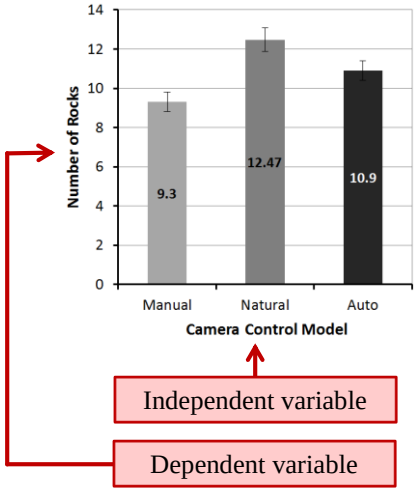
Visual Aids

- Visual aids include charts, photos, drawings, sketches, etc.
- A powerful way to convey ideas and results
- Use generously
- Examples…

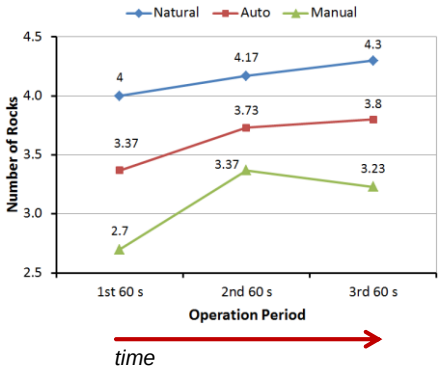
44

Results

Bar chart



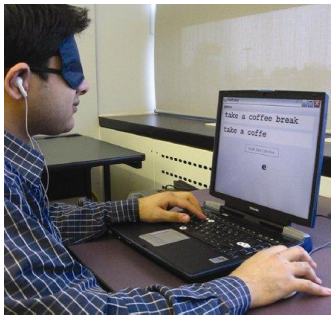
Line chart

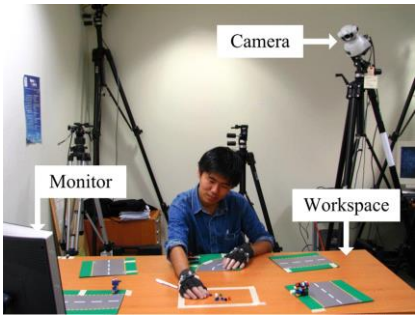


45

Experiment Procedure

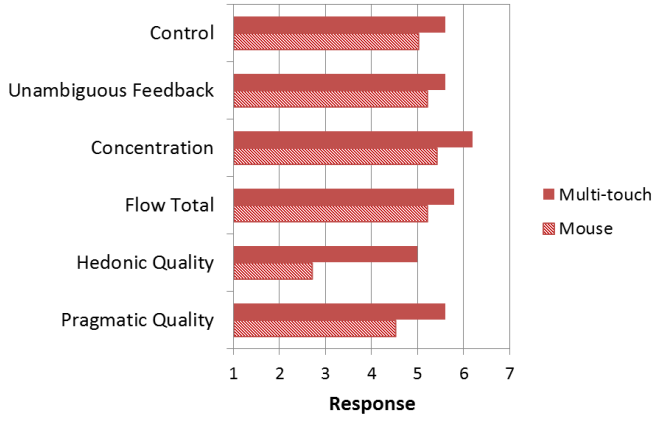
- A photo provides clarity about the experimental procedure





46

Questionnaire Responses



47

Writing for Clarity

- The goal in writing a research paper is communication
- Effective communication demands clarity:
  - *A clear mind attacking a clearly stated problem and producing clearly stated conclusions*[1]
- From the SIGCHI template file under Language, Style, and Content…
  - *Write in a straightforward style*
  - *Avoid long or complex sentence structures*
- Easer said than done

[1] Day, R. A., & Gastel, B. (2006). *How to write and publish a scientific paper* (6th ed.). Westport, CT: Greenwood Publishing.

48

Examples

Original	Revised
In order to do this	To do this
Should be able to understand	Should understand
The software used was our	The software was our
Stacking objects one on top of the other	Stacking objects
Prior gaming experience	Gaming experience
[illegible] in mind	[illegible] mind
Two differen[illegible]	[illegible]
We ran an exploratory pilot study	We ran a pilot study
At their own discretion	At their discretion
Studies conducted in the past have found	Studies have found

Click here to see complete list

(OmitNeedlessWords-Rule_17.doc)

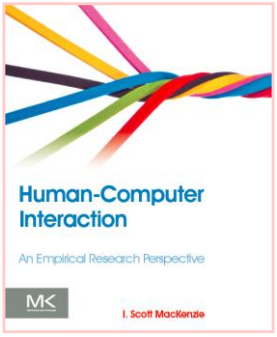
49

Resources

- On the craft and art of scholarly writing, the following are recommended: (1st three also good for research methodology)
  1. APA. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: APA.
  2. Day, R. A., & Gastel, B. (2006). *How to write and publish a scientific paper* (6th ed.). Westport, CT: Greenwood Publishing.
  3. Martin, D. W. (2004). *Doing psychology experiments* (6th ed.). Pacific Grove, CA. Belmont, CA: Wadsworth.
  4. Strunk, W., Jr., & White, E. B. (2000). *The elements of style* (4th ed.). Needham Heights, MA: Pearson.

50

Thank You


Human-Computer Interaction
An Empirical Research Perspective
MK
I. Scott MacKenzie

51