Topic 8: Inspections, Analytics & Models (Part 3)

SECV2113 Human-Computer Interaction

Faculty of Computing

Universiti Teknologi Malaysia

01 INSPECTION: HEURISTIC EVALUATION

02 INSPECTION: WALK-THROUGHS

03 ANALYTICS AND A/B TESTING

04 PREDICTIVE MODELS

No Direct Involvement of Users

- The evaluation methods described in previous topics involved users
- The methods described in this topic do not directly involve users
- These methods are based on an understanding of users' behaviour but users are not directly involved

- **Three types of methods** are discussed:
  - Knowledge codified in heuristics
  - Data collected remotely
  - Models that predict people's performance

INSPECTION :

HEURISTIC EVALUATION

Inspections: Heuristic Evaluation & Walkthroughs

- Several kinds
- **Experts or researchers** use their knowledge of people and technology to review software usability
- **Critiques** can be formal or informal
- **Heuristic evaluation** is a review guided by a set of heuristics
- **Walkthroughs** involve stepping through a pre-planned scenario noting potential problems

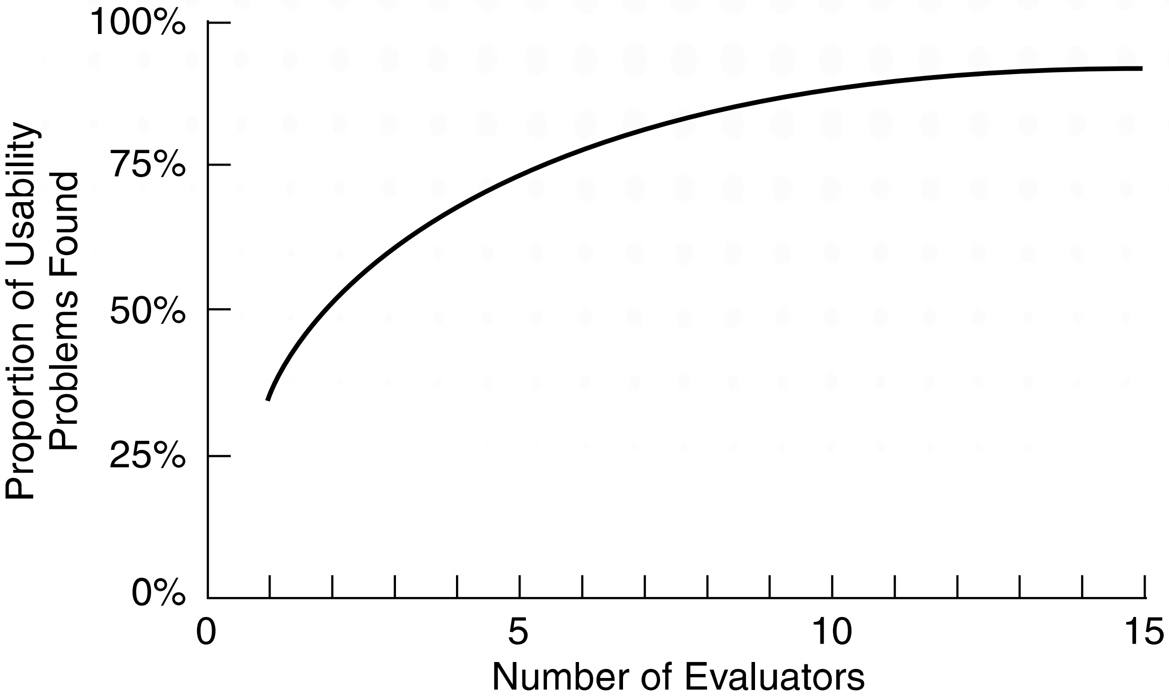
Heuristic Evaluation

- Developed by Jacob Nielsen in the early 1990s
- Based on heuristics distilled from an empirical analysis of 249 usability problems
- These heuristics have been revised for current technology by Nielsen and others for:
    - Mobile devices
    - Wearables
    - Virtual worlds
    - Social media
    - Other types of systems
- Design guidelines often form a basis for developing heuristics

Revised Version (2014) of Nielsen's Original Heuristics

- Visibility of system status
- Match between system and real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency
- Aesthestic and minimalist design
- Support error recovery
- Provide help and documentation

No of Evaluators and Problems



Curve showing the proportion of usability problems in an interface found by heuristic evaluation using different numbers of evaluators

Source: Nielsen and Mack, 1994. Courtesy of Wiley.

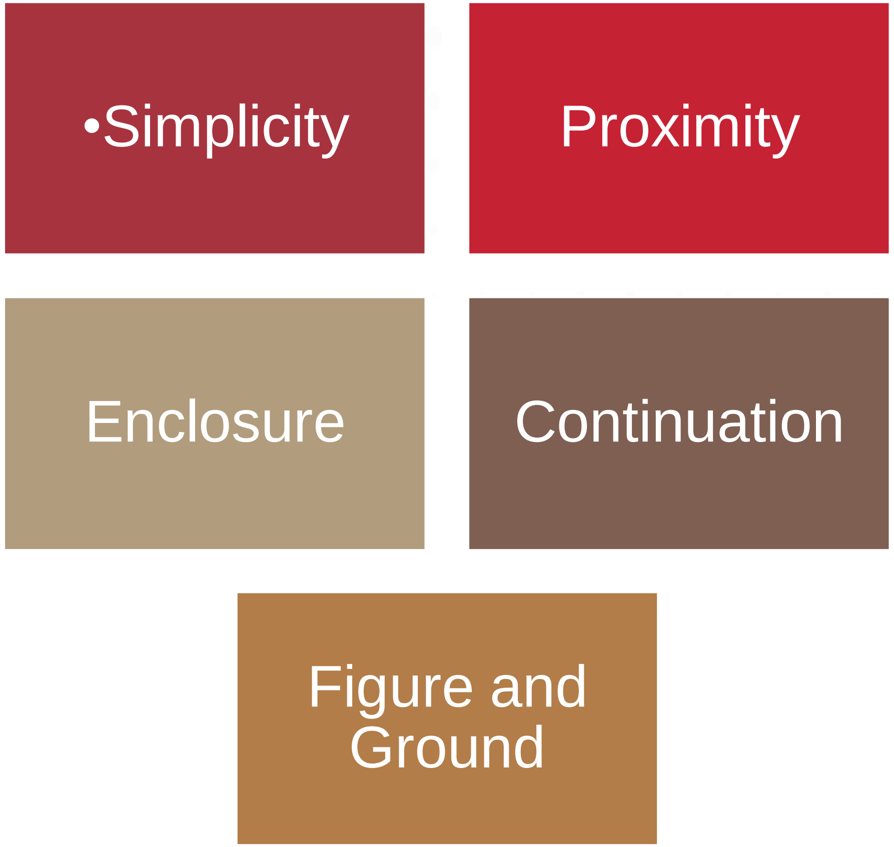
No of Evaluators

- Nielsen suggests that on average 5 evaluators identify 75-80 % of usability problems
- This often depends on the context or type of application (Cockton & Woolrych, 2001)
- Also depends on the skills and experience of the evaluators

Heuristics for Websites Focus on Key Criteria

- Ensure website is designed for people who will use it
- Clarity - avoid jargon, be concise & clear
- Minimise unnecessary complexity & reduce cognitive load
- Provide users with context – clear name & statement of purpose
- Promote positive & pleasurable experience

Heuristics for UI Design based on Gestalt Principles



•Simplicity

Proximity

Enclosure

Continuation

Figure and Ground

Heuristics for Ambient Displays

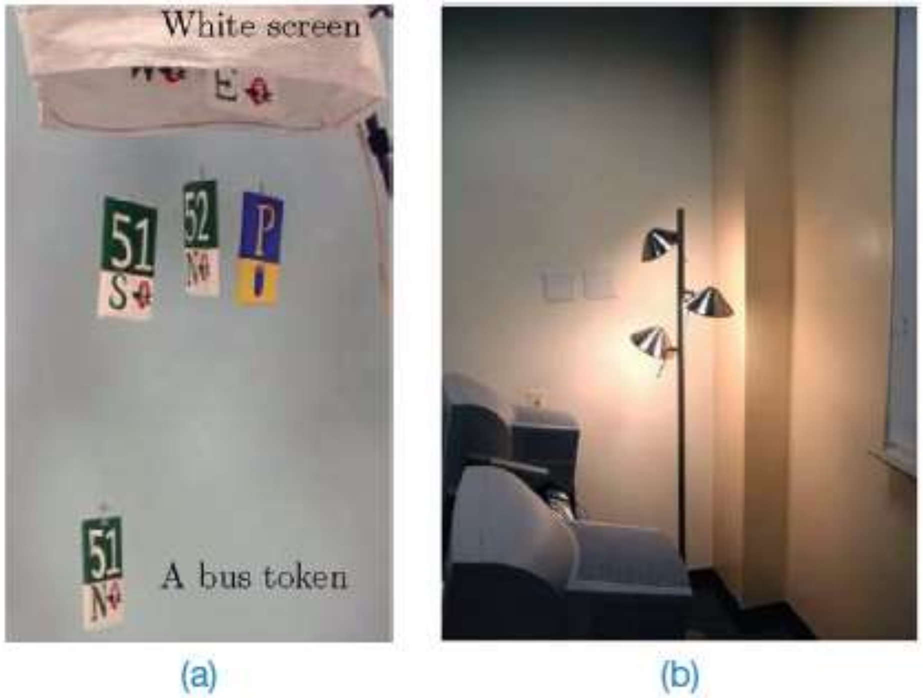


Figure 15.2 Two ambient devices: (a) bus indicator, (b) lightness and darkness indicator

Source: J. Mankoff, A. K. Dey, G. Hsich, J. Kientz, Lederer and A. Morgan (2003) Heuristic evaluation of ambient devices. In *Proceedings of CHI 2003*, ACM Fig.1, p. 170. ©2003 Association for Computing Machinery, Inc. Reprinted by permission.

Doing Heuristic Evaluation

1. Briefing session to tell evaluators what to do

2. Evaluation period of 1-2 hours in which:
   - Each evaluator works separately
   - Take one pass to get a feel for the product
   - Take a second pass to focus on specific features

3. Debriefing session in which evaluators work together to prioritize problems

Severity Ratings in Heuristic Evaluation

- **Severity ratings** can be used to allocate the most resources to fix the most serious problems and can also provide a rough estimate of the need for additional usability efforts.
- If the severity ratings indicate that several disastrous usability problems remain in an interface, it will probably be inadvisable to release it.
- But one might decide to go ahead with the release of a system with several usability problems if they are all judged as being cosmetic in nature.

Source: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

Contributing factors to a problem

The severity of a usability problem is a combination of three factors:

1. The **frequency** with which the problem occurs: Is it common or rare?

2. The **impact** of the problem if it occurs: Will it be easy or difficult for the users to overcome?

3. The **persistence** of the problem: Is it a one-time problem that users can overcome once they know about it or will users repeatedly be bothered by the problem?

If a problem has most of these factors, it should be rated more severe

Source: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

Severity Rating Scale

Nielsen (1994) proposed the following 0 to 4 rating scale to rate the severity of usability problems:

0 = I don't agree that this is a usability problem at all

1 = Cosmetic problem only: need not be fixed unless extra time is available on project

2 = Minor usability problem: fixing this should be given low priority

3 = Major usability problem: important to fix, so should be given high priority

4 = Usability catastrophe: imperative to fix this before product can be released

Severity Rating Scale (cont.)

Barnum, C. (2011) proposed a 3-point severity rating scale:

1 = Catastrophe—user cannot complete task; or user can complete the process but expresses extreme irritation at the process; or user needs assistance (business goal is to make the process independent of user assistance).

2 = Serious problem—user is frustrated but gets through it; suggests that others may be less inclined to put up with the inconvenience or that frustration level will be high.

3 = Cosmetic problem—user may hesitate or pick the wrong option, but user corrects it without incident; or user expresses minor irritation or annoyance, but it doesn’t affect ability to complete tasks.

Problem Prioritisation

- Besides the rating scales shown in previous slides, there are other examples of severity rating scales that has been proposed over the years.
- To effectively understand the impact of each usability problems found, some researchers estimate its severity using dual ratings, for example in terms of usability principles and the ease to solve the problem.
- This can further help in prioritising the problem areas

Source: https://irp-cdn.multiscreensite.com/b18f9b96/files/uploaded/HE_NWApp_v01.pdf

Prioritisation Steps(Dual rating example)

1. Identify the severity of an issue based on its ranking

Severity Ranking	
Rating	**Definition**
0	This is not a usability problem at all
1	Cosmetic problem only: need not be fixed unless extra time is available on project
2	Minor usability problem: fixing this should be given low priority
3	Major usability problem: important to fix, so should be given high priority
4	Usability catastrophe: imperative to fix this before product can be released

2. For each issue identified, rank its ease of fixing

Ease of Fixing Ranking	
Rating	**Definition**
0	Problem would be extremely easy to fix.
1	Problem would be easy to fix. Involves specific interface elements and solution is clear.
2	Problem would require some effort to fix. Involves multiple aspects of the interface or would require team of developers to implement changes before next release or solution is not clear.
3	Problem would be difficult to fix. Requires much development effort to finish before next release, involves multiple aspects of interface. Solution may not be immediately obvious or may be disputed.

Understanding the Impact

(1) Overview: Tabulate the identified issue with how frequent the issue occurs, its severity ranking and its ease of fixing

Usability / UI Issue	Heuristic #	Frequency	Severity ranking (0 – 4)	Ease of fixing ranking (0 – 3)

(2) Identify which heuristic is the most problematic

Heuristic	Frequency
h1	
h2	
h3	
h4	
h5	
h6	
h7	
h8	
h9	
h10	
	(total)

(3) Identify issues by their severity & ease of fixing

Severity	Freq	Ease of Fixing	Freq
0	2	0	6
1	0	1	9
2	5	2	5
3	4	3	1
4	10		
	total = 21		total = 21

Advantages and Problems

- Few ethical and practical issues to consider because users are not involved
- Can be difficult and expensive to find expert evaluators with knowledge of application domain and the people using it
- Biggest problems:
  - Important problems may get missed
  - Many trivial problems are often identified, such as false alarms
  - Evaluators have biases

Turning Design Guidelines and Golden Rules into Heuristics

- Ask questions like the following:

“Does the application include a visible title page, section or site? Does the user always know where they are in the application? Does the user always know what the system or application is doing? Are the links clearly defined? Can all actions be visualized directly (i.e., no other actions are required)?”

Granollers, 2018, p. 62

Evaluating Accessibility using WCAG Guidelines

- Web Content Accessibility Guidelines (WCAG) (Lazar et al., 2015)
- Guidelines can be used as heuristics for evaluating websites
- Governments and large corporations have to make their websites accessible by law
- Four key concepts **(POUR):**
  - **P**erceivable – eg, provide text alternatives for non-text content
  - **O**perable – eg, make functionality available via a keyboard
  - **U**nderstandable – eg, make content appear and be readable in predictable ways
  - **R**obust – eg ensure compatibility with other tools

Source: WCAG 2.1 at a Glance.

INSPECTION

:

COGNITIVE WALKTHROUGH

Cognitive Walkthroughs

- Focus on ease of learning
- Designer presents an aspect of the design and usage scenarios
- Evaluator is told the assumptions about people using the system, context of use, task details
- One or more evaluators walk through the design prototype with the scenario
- Evaluators are guided by three questions

The **THREE (3)** Questions

1. Will the correct action be sufficiently evident to the people using the system?
2. Will the they notice that the correct action is available?
3. Will the they associate and interpret the response from the action correctly?

As the evaluators work through the scenario, they note problems

Pluralistic Walkthroughs

- Variation on the cognitive walkthrough theme
- Performed by a carefully managed team
- A panel of evaluators begins by working separately
- This is followed by a managed discussion that leads to agreed decisions
- The approach lends itself well to participatory design
- There are also other adaptations of basic cognitive walkthroughs

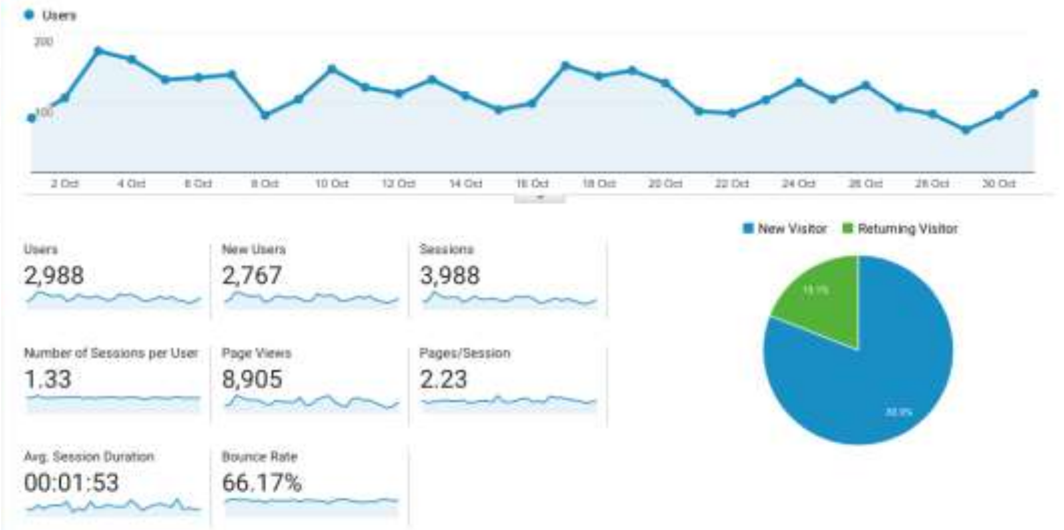
ANALYTICS

& A/B TESTING

Web Analytics

- A form of interaction logging that analyses users’ activities on websites
- Web designers use the analysis to improve their designs
- When designs don’t meet peoples’ needs, they will not return to the site
- They become one-time users
- Web analytics enable designers to track the activities of people on their site
- They can see how many people come to the site, how long they stay, and where they go next
- Web analytics offer designers a “big picture” about how their site performs based on users’ activity
- Google provides is one of the most well-known analytics

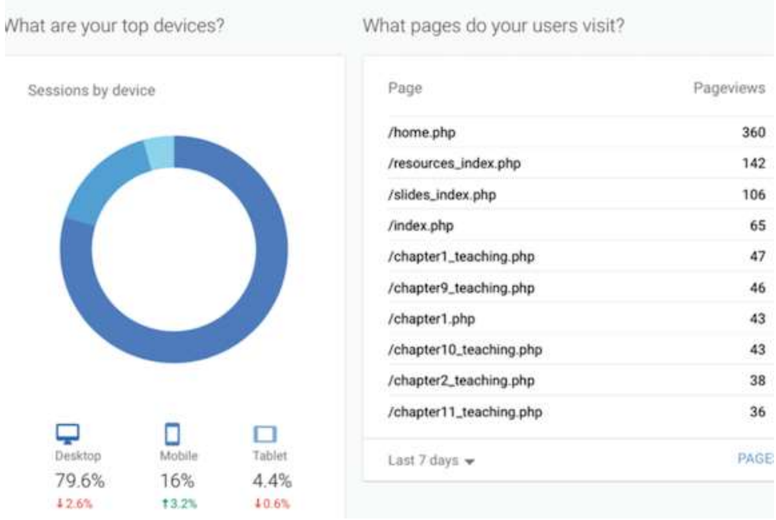
Segment of Google Analytics for Interaction Design 5e website, December 2022



Users	New Users	Sessions
2,988	2,767	3,988

Number of Sessions per User	Page Views	Pages/Session
1.33	8,905	2.23

Avg. Session Duration	Bounce Rate
00:01:53	66.17%


What are your top devices?

Sessions by device

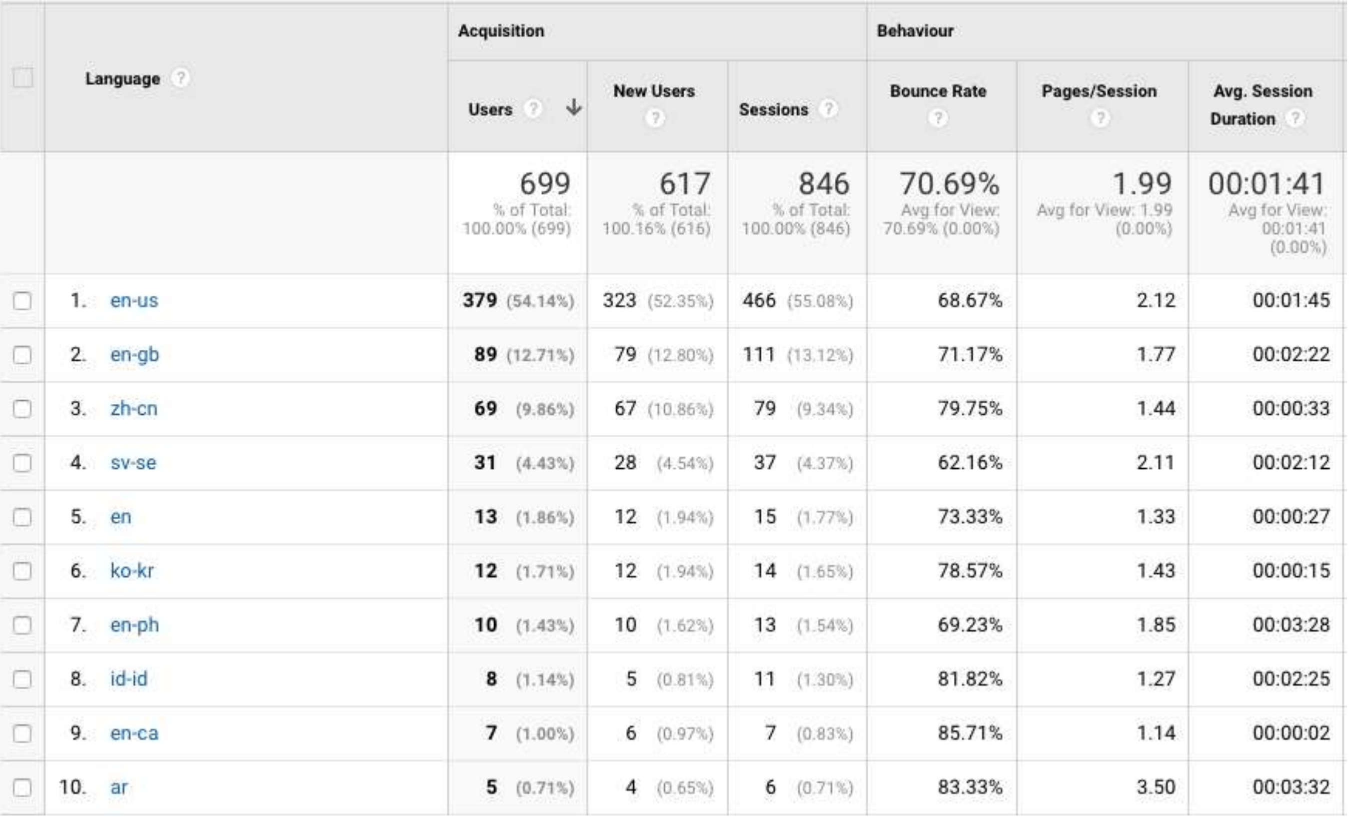
Desktop	Mobile	Tablet
79.6%	16%	4.4%
↓2.6%	↑3.2%	↓0.6%

What pages do your users visit?

Page	Pageviews
/home.php	360
/resources_index.php	142
/slides_index.php	106
/index.php	65
/chapter1_teaching.php	47
/chapter9_teaching.php	46
/chapter1.php	43
/chapter10_teaching.php	43
/chapter2_teaching.php	38
/chapter11_teaching.php	36

Last 7 days ▾

PAGE



□	Language ?	Acquisition			Behaviour		
		Users ? ↓	New Users ?	Sessions ?	Bounce Rate ?	Pages/Session ?	Avg. Session Duration ?
		699 % of Total: 100.00% (699)	617 % of Total: 100.16% (616)	846 % of Total: 100.00% (846)	70.69% Avg for View: 70.69% (0.00%)	1.99 Avg for View: 1.99 (0.00%)	00:01:41 Avg for View: 00:01:41 (0.00%)
□	1. en-us	379 (54.14%)	323 (52.35%)	466 (55.08%)	68.67%	2.12	00:01:45
□	2. en-gb	89 (12.71%)	79 (12.80%)	111 (13.12%)	71.17%	1.77	00:02:22
□	3. zh-cn	69 (9.86%)	67 (10.86%)	79 (9.34%)	79.75%	1.44	00:00:33
□	4. sv-se	31 (4.43%)	28 (4.54%)	37 (4.37%)	62.16%	2.11	00:02:12
□	5. en	13 (1.86%)	12 (1.94%)	15 (1.77%)	73.33%	1.33	00:00:27
□	6. ko-kr	12 (1.71%)	12 (1.94%)	14 (1.65%)	78.57%	1.43	00:00:15
□	7. en-ph	10 (1.43%)	10 (1.62%)	13 (1.54%)	69.23%	1.85	00:03:28
□	8. id-id	8 (1.14%)	5 (0.81%)	11 (1.30%)	81.82%	1.27	00:02:25
□	9. en-ca	7 (1.00%)	6 (0.97%)	7 (0.83%)	85.71%	1.14	00:00:02
□	10. ar	5 (0.71%)	4 (0.65%)	6 (0.71%)	83.33%	3.50	00:03:32

A/B Testing

- A type of large-scale experiment
- Offers another way to evaluate a website, application or app running on a mobile device
- Often used for evaluating changes in design of social media applications
- Compares how two groups of users perform on two versions of a design
- Can involve thousands of people
- May create ethical dilemmas if users don’t know they are part of the test
- Care is needed to ensure that other issues are not affecting users’ behavior

PREDICTIVE MODELS

Predictive Models

- Provide a way of evaluating products or designs without directly involving users
- Less expensive than testing with users and some kinds of inspections
- Usefulness limited to systems with predictable tasks, for example, voicemail systems, smartphones, and dedicated mobile devices
- Based on knowledge of expert error-free behavior

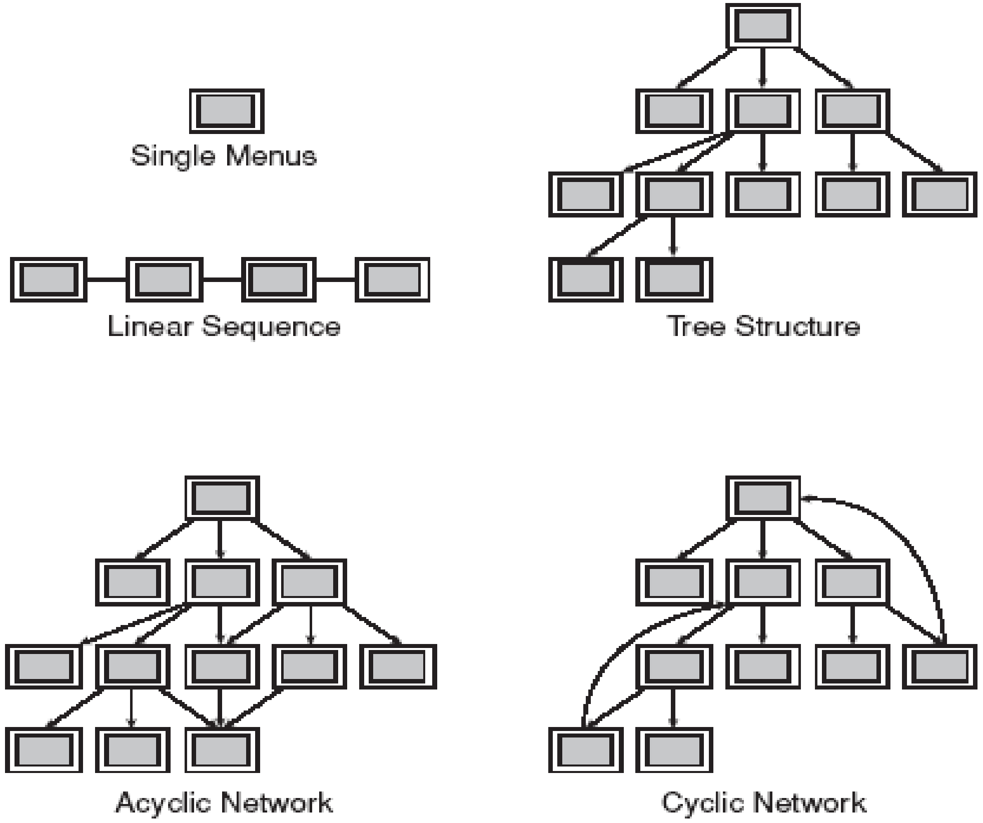
Fitts’ Law (1954)

- Fitts’ Law predicts that the time to point at an object using a device is a function of the distance from the target object and the object’s size
- The further away and the smaller the object, the longer the time to locate it and point to it
- It is particularly useful for determining where on a screen to position an object
- Fitts’ Law is useful for evaluating systems for which the time to locate an object is important, for example, smartphones, handhelds, and mobile devices

Fitts' Law:
A Model of Human Motor Performance

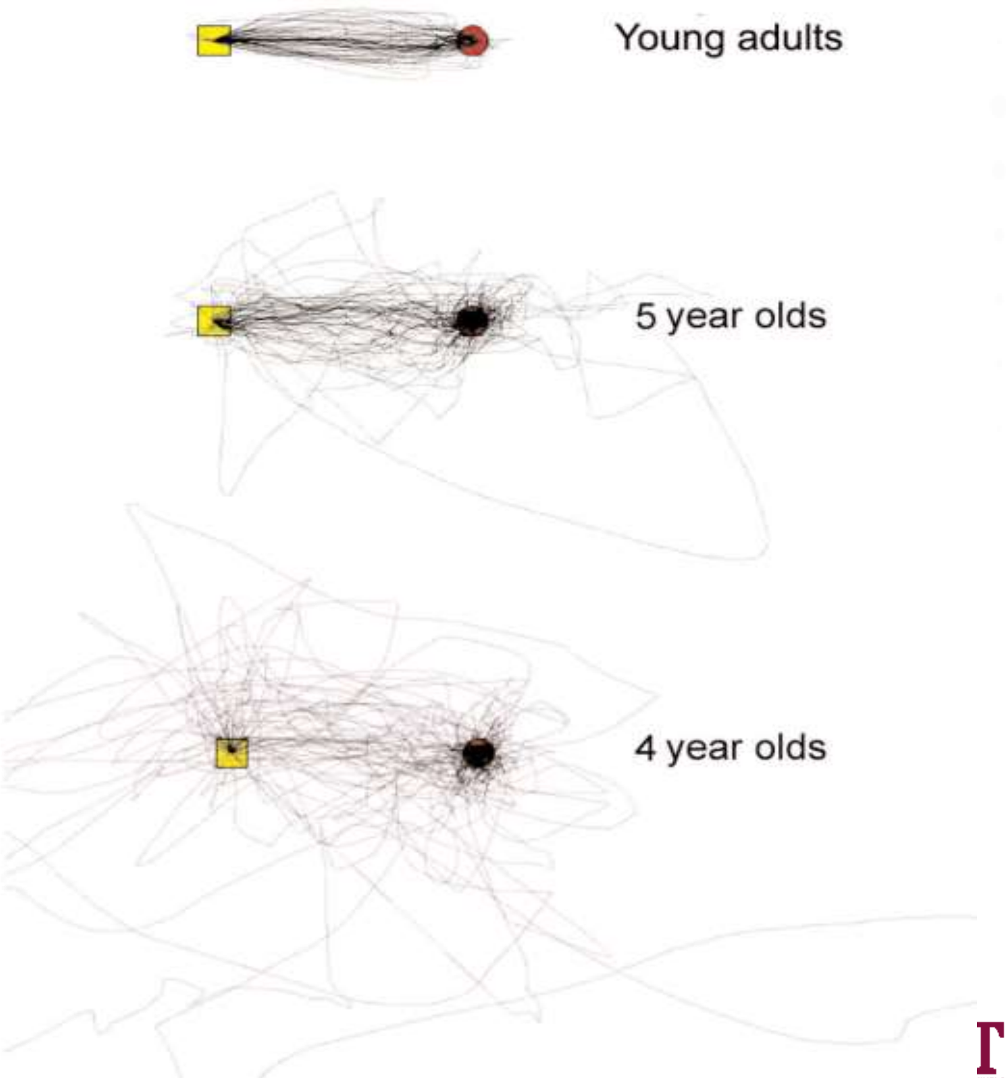
Task-Related Organisation

"The primary goal for menu, form-fillin, and dialog-box designers is to create a sensible, comprehensible, memorable, and convenient organization relevant to the user's task."

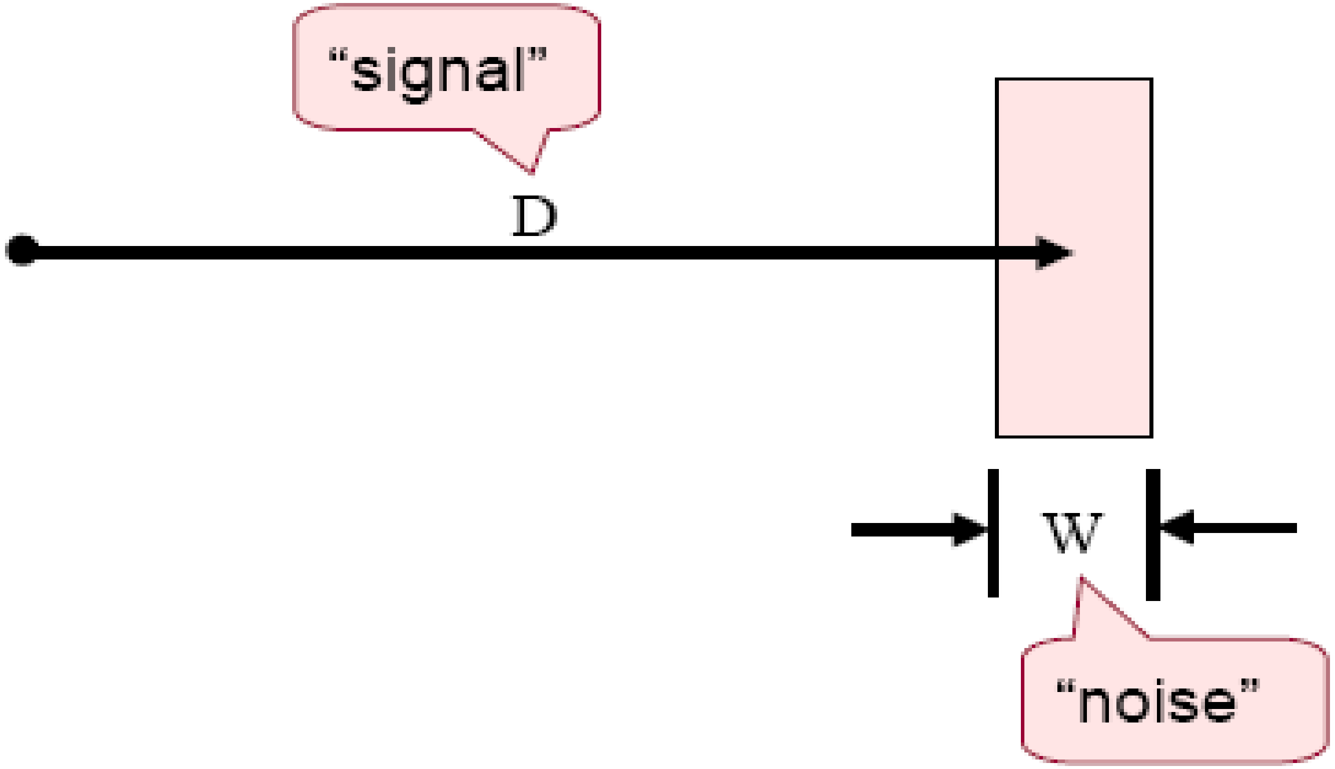


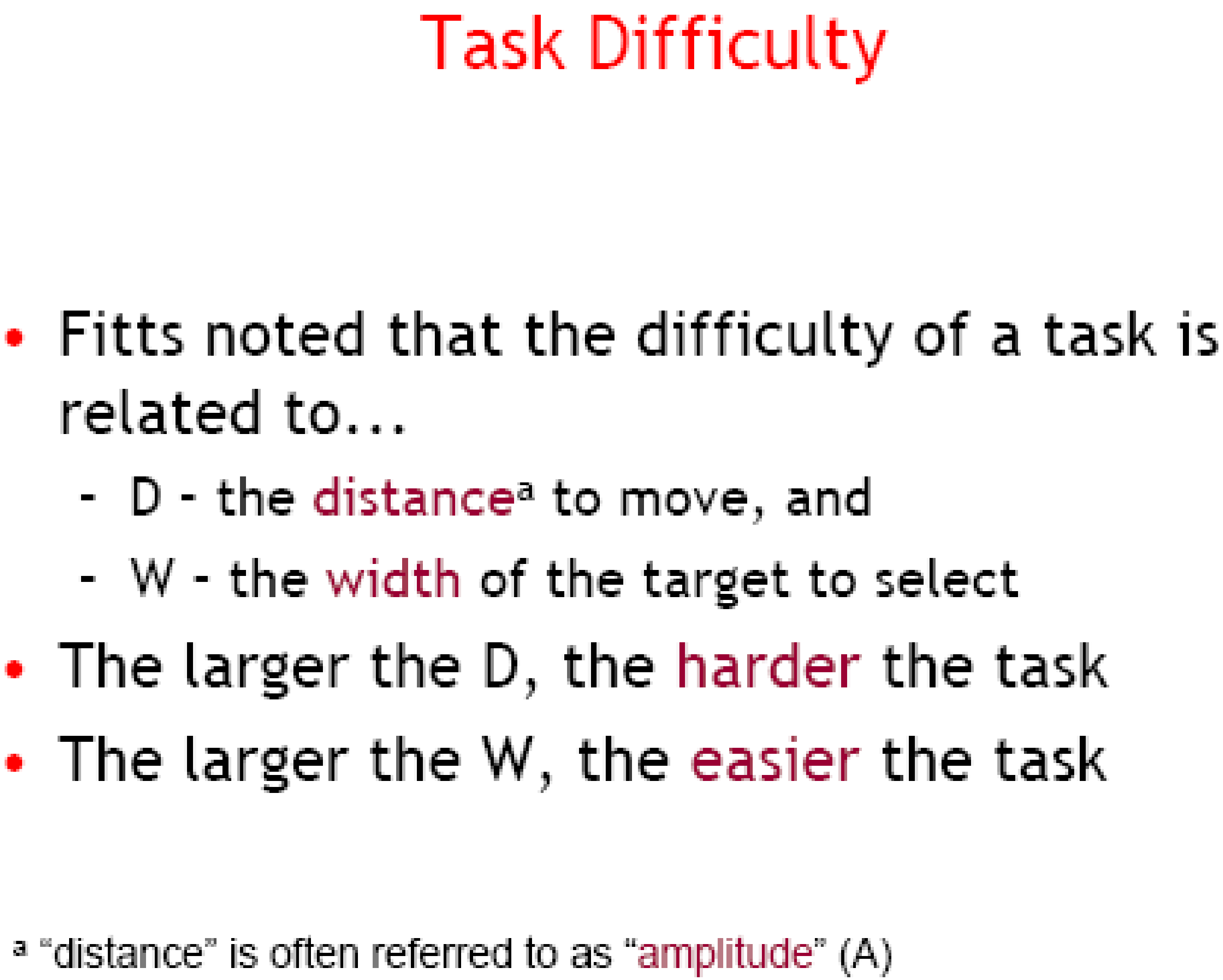
Fitts' Law

- Index of difficulty = $\log_2$ (D / W+1)
  - D = distance between buttons
  - W = target size
- Time to point = C1 + C2 (index of difficulty)
- C1 and C2 and constants that depend on the device
  - C1 = start/stop time in seconds
  - C2 = speed of the device
- Index of difficulty is $\log_2$ (2*8/1) = $\log_2$ (16) = 4 bits



Fitts' Analogy





Task Difficulty

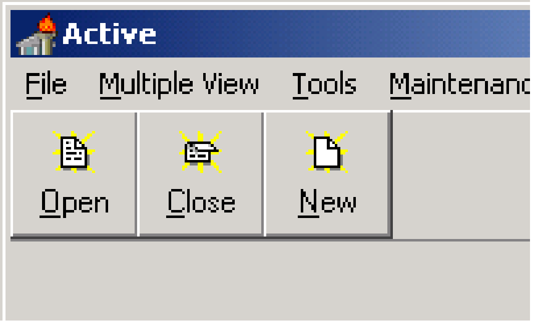
- Fitts noted that the difficulty of a task is related to…
  - D - the distance[a] to move, and
  - W - the width of the target to select
- The larger the D, the harder the task
- The larger the W, the easier the task

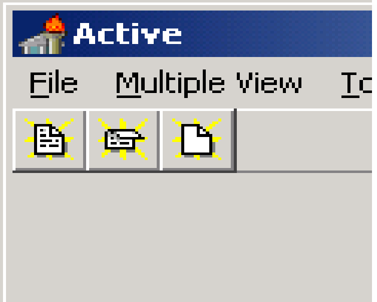
[a] "distance" is often referred to as "amplitude" (A)

Fitts in Practice

- Microsoft Toolbars allow you to either keep or remove the labels under Toolbar buttons

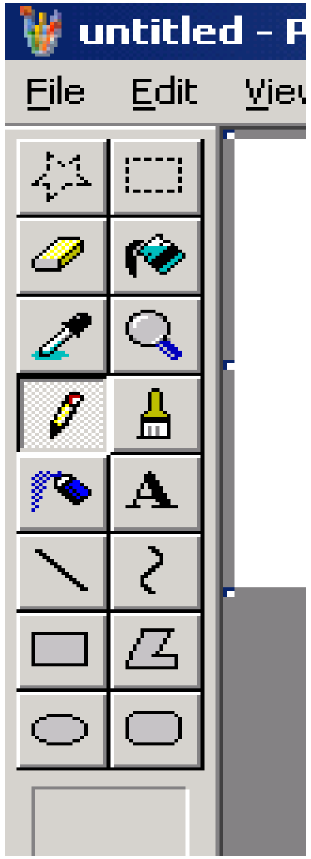
- According to Fitts' Law, which is more efficient?





Source: http://www.asktog.com/columns/022DesignedToGiveFitts.html

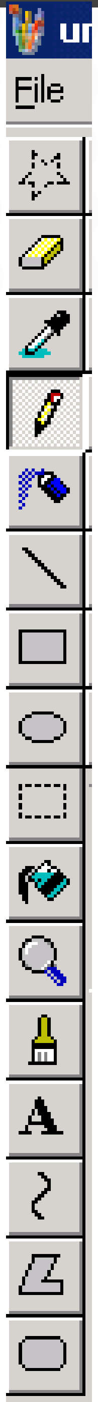
Fitts in Practice



- You have a toolbar with 16 icons, each with dimensions of 16x16

- Without moving the array from the left edge of the screen, or changing the size of the icons, how can you make this more efficient?

Adapted from Hearst, Irani



Fitts in Practice

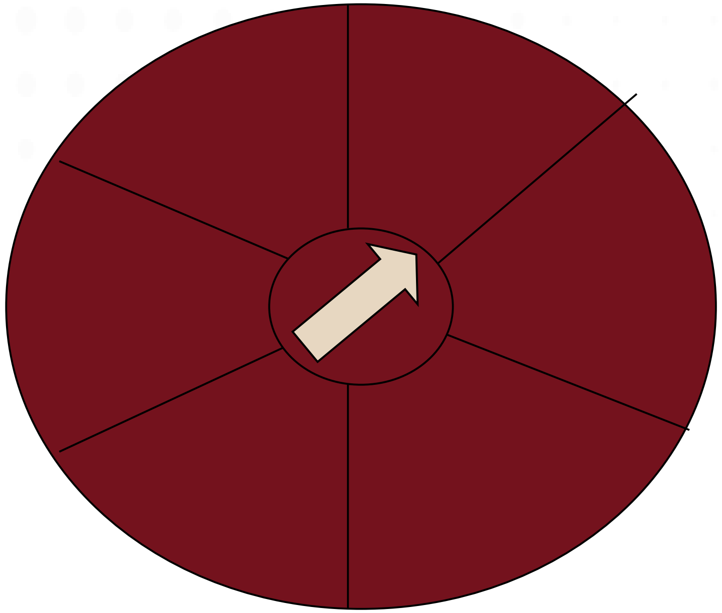
Answer:

- Line up all 16 icons on the left hand edge of the screen
- Make sure that each button can be activated up the last pixel on the left hand edge
- Why? Because you cannot move your mouse off of the screen, the effective width s is infinite

Adapted from Hearst, Irani

Pie Menus

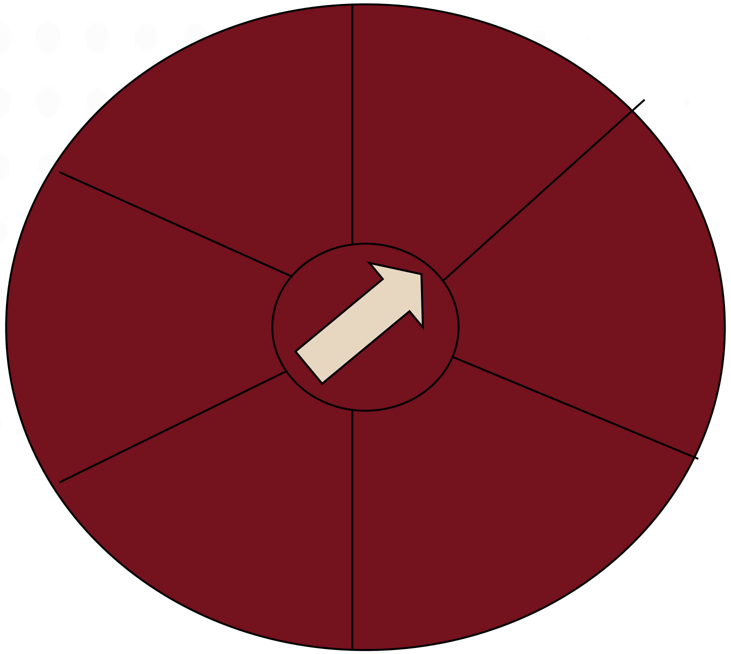
- A circular pop-up menu
  - no bounds on selection area
    - basically only angle counts
    - do want a “dead area” at center
  - What are Fitts’ law properties?



Pie Menus

- A circular pop-up menu
  - no bounds on selection area
    - basically only angle counts
    - do want a “dead area” at center
  - Fitts’ law properties:
    - minimum distance to travel
    - minimum required accuracy
    - very fast

- Why don’t we see these much?
    - Just not known
    - Harder to implement
      - particularly drawing labels
      - but there are variations that are easier
    - Don’t scale past a few items
      - No hierarchy



GOMS

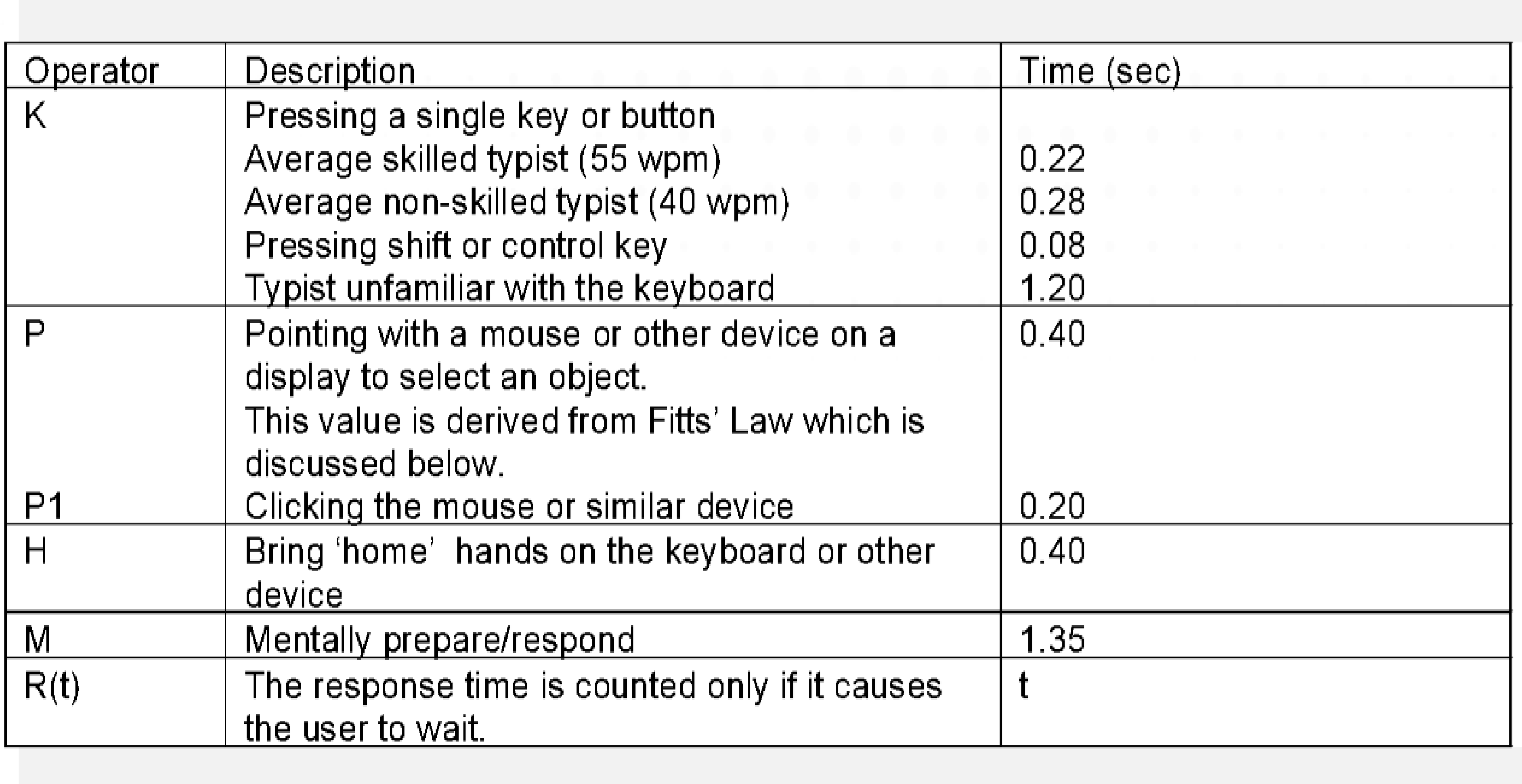
- **Goals** - the state the user wants to achieve e.g., find a website.
- **Operators** - the cognitive processes & physical actions needed to attain the goals, e.g., decide which search engine to use.
- **Methods** - the procedures for accomplishing the goals, e.g., drag mouse over field, type in keywords, press the go button.
- **Selection rules** - decide which method to select when there is more than one.

Keystroke Level Model (KLM)

- GOMS has also been developed to provide a quantitative model - the keystroke level model.

- The keystroke model allows predictions to be made about how long it takes an expert user to perform a task.

Response Times for Keystroke Level Operators (Card et al., 1983)



Operator	Description	Time (sec)
K	Pressing a single key or button Average skilled typist (55 wpm) Average non-skilled typist (40 wpm) Pressing shift or control key Typist unfamiliar with the keyboard	 0.22 0.28 0.08 1.20
P	Pointing with a mouse or other device on a display to select an object. This value is derived from Fitts' Law which is discussed below.	0.40
P1	Clicking the mouse or similar device	0.20
H	Bring 'home' hands on the keyboard or other device	0.40
M	Mentally prepare/respond	1.35
R(t)	The response time is counted only if it causes the user to wait.	t

Summary

- Inspections can be used to evaluate requirements, mockups, functional prototypes, or systems
- User testing and heuristic evaluation may reveal different usability problems
- Design guidelines can be used to develop heuristics
- Web Content Accessibly Guidelines (WCAG) provide guidance for evaluating the accessibility of websites
- Walkthroughs are a fine-grained focused method for evaluating small parts of a prototype or product

Summary (Cont.)

- Analytics involves collecting data about people's activity on a website or product to see which parts are used

- A/B testing is a form of large-scale experiment

- Fitts' Law can be used to predict expert, error-free performance for clearly defined tasks with limited key presses, for example, to evaluate keypress sequences for handheld devices and the position of objects on a small screen