HCI: Empirical Research Case Study

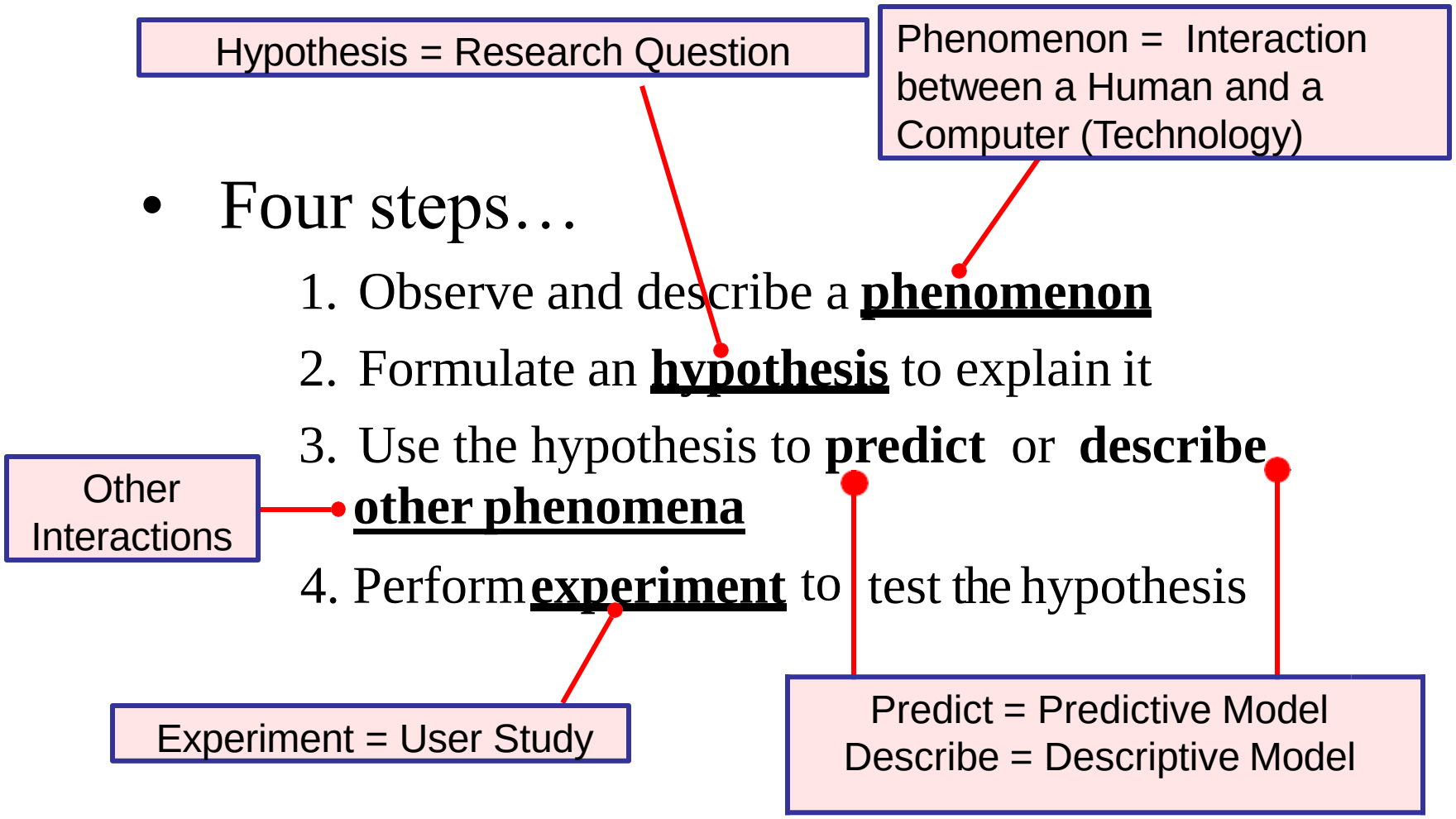
Learning Objective

- In the previous lectures, we discussed different empirical research methods involved in HCI
- We got introduced to several concepts such as testable question formulation, experiment design, data collection and statistical analysis of data
- In this lecture, we shall consider a case study

Case Study

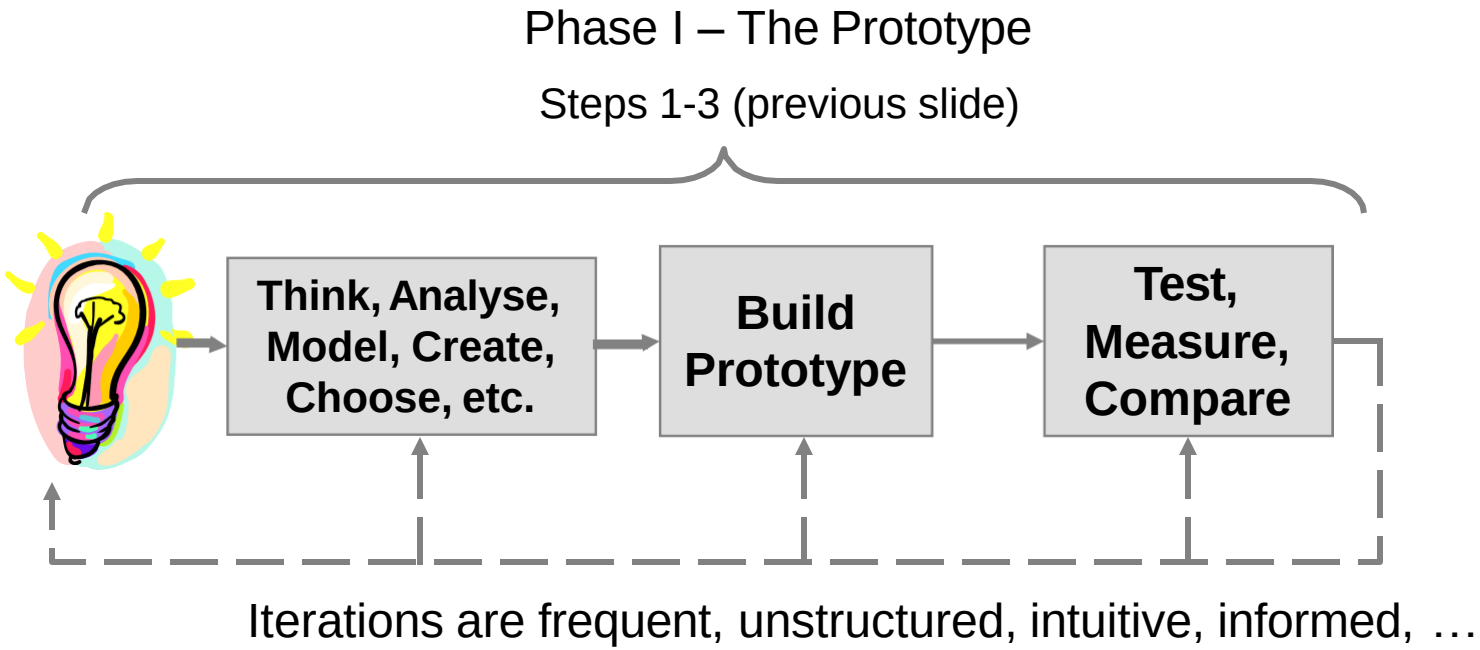
- Suppose, we want to study the application of eye tracking technology to text entry (i.e., typing through eye gaze). Let us initiate an empirical inquiry to explore the performance limits and capabilities of various feedback modalities for keys in on-screen keyboards used with eye typing
  - Suppose four feedback modalities are considered by us, namely: Audio only [A], Click+Visual [C], Speech+Visual [S], Visual only [V]

Steps in Empirical Research (Classical View)



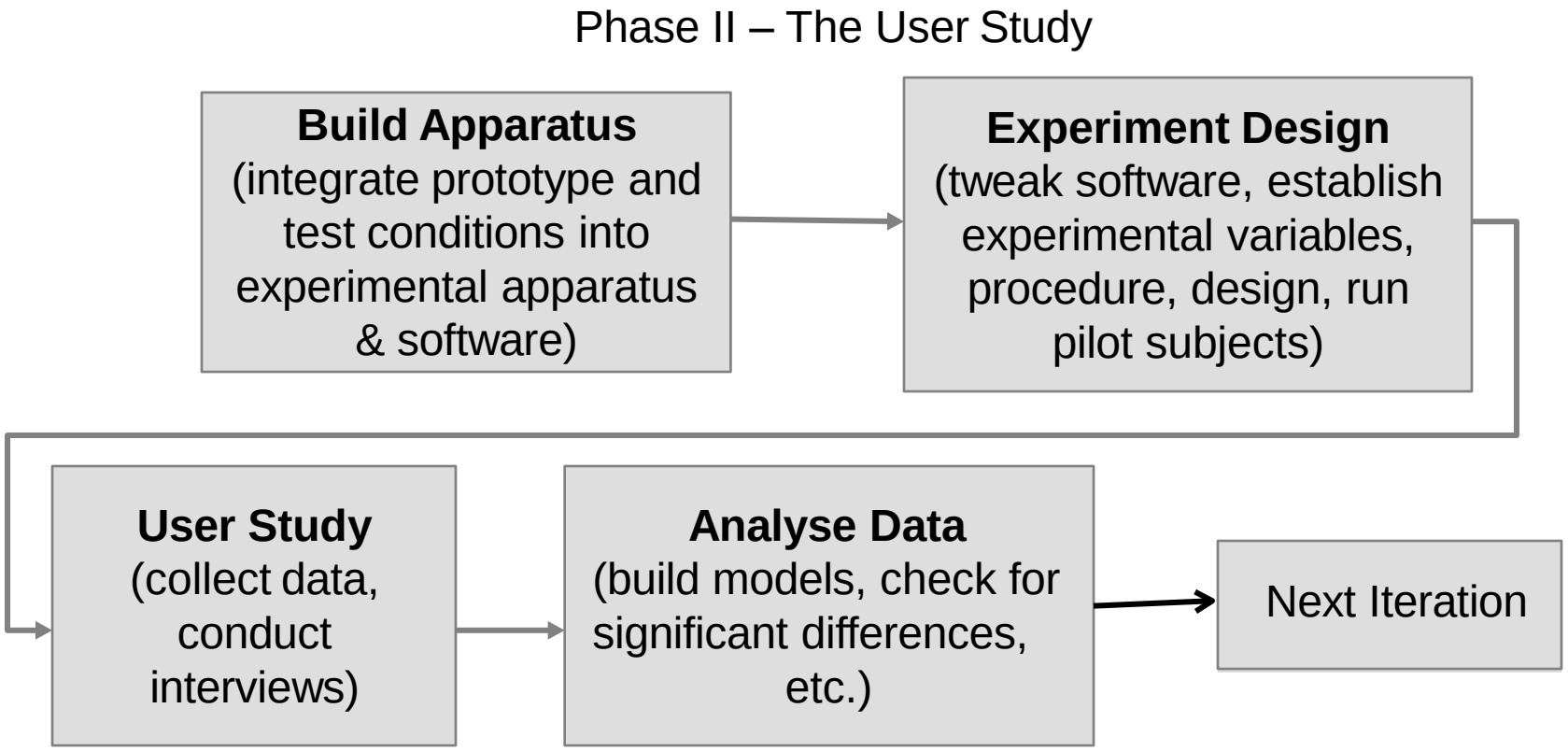
- Four steps…
  1. Observe and describe a **phenomenon**
  2. Formulate an **hypothesis** to explain it
  3. Use the hypothesis to **predict** or **describe other phenomena**
  4. Perform **experiment** to test the hypothesis

Steps in Empirical Research (Practical View)



Research questions “take shape” (I.e., certain measurable aspects of the interaction suggest “test conditions”, and “tasks” for empirical inquiry

Steps in Empirical Research (Practical View)



The User Study

- Describe the participants employed for our study
  - Thirteen, volunteers, recruited from university campus, age, gender, computer experience, eye tracking/typing experience
- Apparatus
  - Describe hardware and software, etc.

The User Study

- Experiment Design
  - We decided to have a 4 x 4 repeated measures design
  - There are two independent variables (factors) with four levels each
    - Feedback modality (with the levels A, C, S, V)
    - The participants were asked to enter blocks of text at a time and four such blocks were there for each participant. So, “block” is a factor with four levels 1, 2, 3, 4

The User Study

- Experiment Design
  - We have identified dependent variables (measures)
    - Speed of text entry (in “words per minutes”)
    - Accuracy of text entry (in “percentage of characters in error”)
    - Key selection activity (in “keystrokes per character”)
    - Also… responses to “broad” questions
  - Order of Conditions
    - Feedback modality order differed for each participant (using a Latin Square Method)

The User Study

- Procedure for Data Collection
  - We first explained to the participants the general objectives of the experiment
  - Then the eye tracking apparatus was calibrated
  - The participants were put through some practice trials for familiarization
  - Afterwards, let us begin data collection

The User Study

- Procedure for Data Collection
  - Phrases of text presented to the participants by experimental software
  - Participants instructed to enter phrases “as quickly and accurately as possible”
  - Five phrases were entered by the participants per block
  - Total number of phrases entered in this experiment is found to be 13 x 4 x 4 x 5 = 1040

Experiment Replication

- The description of the experimental methodology (i.e., participants, participant selection, apparatus, design, procedure) must be sufficient to allow the experiment to be replicated by other researchers
  - This is necessary to allow the possibility for the results to be verified or refuted as part of performance evaluation
  - An experiment that cannot be replicated is useless

Data Tables

- The next slide contains example data on text entry speed, recorded in this empirical study (user study)
  - Create a Table to arrange data i.e. Data Table
  - From the Data Table, calculate other quantities such as grand mean = 6.96 wpm
  - The Data Table also allows us to make salient observations (for example, 4^{th} block speed for best condition was…)

Data Tables

Factors and Levels

Speed																	
	A	A	A	A	C	C	C	C	S	S	S	S	V	V	V	V	
Participant	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	Mean
1	6.17	7.19	7.04	7.09	6.76	7.40	7.54	7.94	6.44	6.17	7.84	6.81	5.20	6.29	7.39	7.63	6.93
2	6.71	7.25	7.05	7.15	7.73	7.57	8.04	7.26	7.00	6.75	7.68	7.46	7.50	7.07	7.32	7.06	7.29
3	6.80	6.65	7.62	7.98	6.61	7.18	7.34	8.19	6.65	7.53	7.09	7.90	5.73	7.24	6.94	7.13	7.16
5	6.30	6.31	7.59	7.38	6.85	7.64	7.58	7.88	7.07	6.43	7.26	7.65	6.75	6.59	6.97	7.72	7.12
7	6.68	6.89	7.32	7.51	7.00	7.81	7.64	7.2						72	7.57	7.20	7.11
8	6.08	6.55	6.83	5.92	7.44	6.93	7.56	6.4						3	7.45	7.16	6.98
9	7.62	7.01	6.60	7.07	6.91	6.81	6.91	7.73	6.50	7.57	7.59	7.80	6.62	7.06	7.16	7.41	7.15
10	5.88	5.71	7.33	7.11	6.66	7.97	7.64	8.15	6.35	7.21	6.56	7.33	5.00	6.97	6.54	6.36	6.80
12	6.89	7.61	7.42	7.88	7.79	8.28	8.20	8.39	6.62	6.87	7.99	8.23	9.57	8.17	7.91	7.09	7.81
13	6.85	6.57	8.14	6.00	5.92	7.89	7.49	6.98	6.05	7.45	5.34	7.46	7.21	6.81	6.80	8.24	6.95
14	5.37	5.56	6.04	6.86	6.20	6.82	7.71	7.76	5.85	6.37	6.74	6.69	5.98	6.43	6.38	5.87	6.41
15	5.51	6.12	6.32	7.00	6.16	6.49	7.21	7.19	5.65	6.52	6.49	7.10	5.31	6.88	6.36	6.93	6.45
16	5.88	7.18	5.95	6.00	4.85	6.98	7.37	6.98	6.88	6.21	4.96	5.34	6.72	7.14	4.96	6.80	6.26
																	6.96

Outlier

Each cell is the mean for five phrases of input

Statistical Analysis of Data

- The data recorded in the Data Table are analyzed statistically to identify (statistically) significant effects
- For example, we may have the following findings:
  - Main effect for Feedback mode significant [$F(3,36)=8.77$, $p<.0005$]
  - Feedback mode by block interaction not significant [$F(9,108)=0.767$, ns]

Data Tables

- Apart from the main tables, other tables are also created, which helps in making more useful observations
- The next slide shows an example of a summary table created from the data on text entry speed

Data Tables

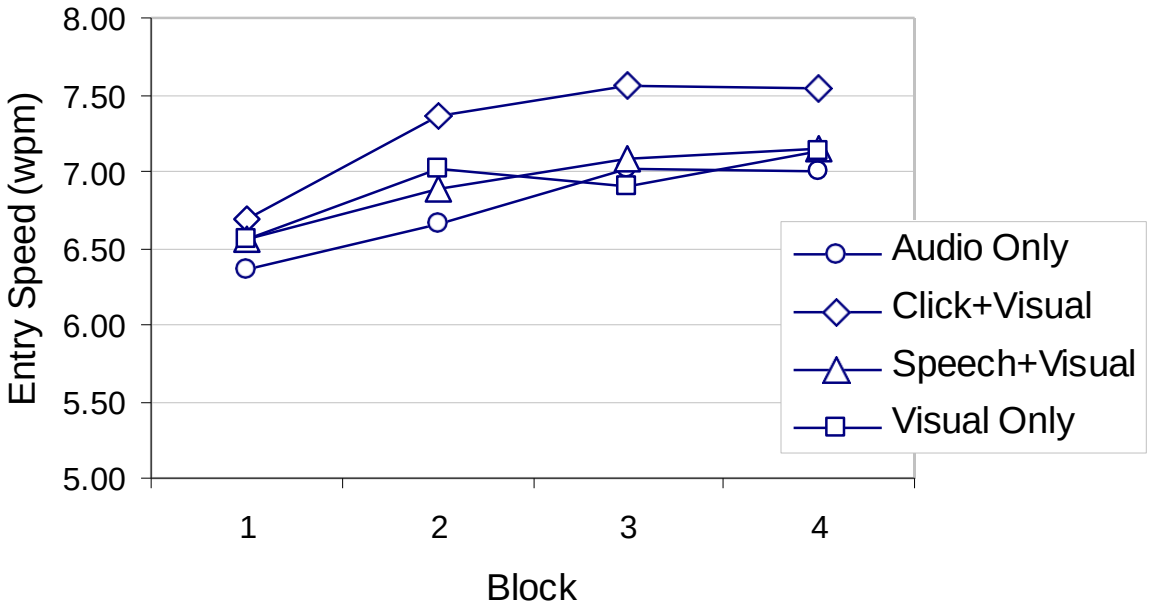
Speed (wpm)					
	Feedback Mode				
Block	Audio Only	Click+Visual	Speech+Visual	Visual Only	Mean
1	6.36	6.68	6.56	6.55	6.54
2	6.66	7.37	6.88	7.02	6.98
3	7.02	7.56	7.09	6.90	7.14
4	7.00	7.55	7.14	7.12	7.20
Mean	6.76	7.29	6.92	6.90	6.97

5.7% faster on 4th block

Each cell is the mean for 13 participants

Charts/Graphs

- Also let us create graphs/charts to visualize findings



The Broad Questions

- Along with the data analysis, an empirical study typically collects the direct feedback from all the participants on “broad” questions
  - For example, all participants can be asked about their preferences, satisfaction levels or even their suggestions for further improvements

The Broad Questions

- In the study, we asked the participants to rank (between 1 to 4) the feedback mode based on their personal preferences
- We obtained the following results:
  - Six of thirteen participants gave a 1st place ranking to the fastest feedback modality

The Broad Questions

- The results obtained is not strong enough to come to any conclusions
  - A reason may be that the differences just weren't large enough for participants to really tell the difference in overall performance

The Broad Questions

- However, we have also made another observation, namely ten of the thirteen participants gave a 1st or 2nd place ranking to the fastest feedback modality
  - This can be treated as a strong indication that better performance yields a better preference rating

What's Missing?

- The case study just described show that the user study involves collection and analysis of usage data as well as participants' feedback
- However, that's not all (it misses an important aspect of empirical research)
  - There is no theoretical account of the phenomena

What's Missing?

- There is no delineation, description, categorization of the known and observed behaviors (…that can form such a theoretical account)
- It is not sufficient to simply observe and conclude, it is also necessary to theorize about the observations (e.g., why the text entry speed is the least in a particular feedback mode)

What's Missing?

- The direct conclusions from observations help us decide an interaction method; a theory about observed behavior can help us do much more
  - Such theories can eliminate need for further investigations as well as can suggest ways for further improvement

Empirical Research in HCI

- Such theories, if found, are another motivation for conducting empirical research in HCI (in fact, many models in HCI have been derived empirically)

Case Study: The Case for a Model

- Is there a “model of interaction” suggested by the observations in the case study?
- Perhaps. Here’s one possibility
  - All gaze point changes were logged as “events”.
    What was the total number of such events?
    Are there categories of such events?
- The identification, labelling, and tabulation of such could form the basis of a model of interaction for eye typing