Evaluation of Mobile Interfaces

Professor John Canny

Spring 2006

Review

Last time:

- User testing process
- Severity and Cost ratings
- Discount usability methods
- Heuristic evaluation
- HE vs. user testing

Mobile Evaluation

What changes in the mobile space?

- People are moving
- People are distracted, in a hurry
- Its hard to use both hands
- There are often other people around
- …

Mobile Evaluation

But both readings showed that "laboratory" evaluation can still work very well.

"Series 60 Voice Mailbox in Rome" used all the methods we talked about:

- Lo-fi paper prototypes
- Wizard-Of-Oz simulation
- Horizontal and vertical prototypes

But with an Italian moderator and translator

Mobile Evaluation

Kjeldskov and Stage's "New Techniques…" paper compared 6 methods for mobile evaluation.

1. Sitting or standing around a table – same as standard usability studies.

Mobile Evaluation

2. Walkin' in the lab, on a treadmill at constant speed.

Mobile Evaluation

3. Walkin' on a treadmill with varying speed

Mobile Evaluation

4. Walkin' at constant speed on a track that is changing with moving obstacles

Mobile Evaluation

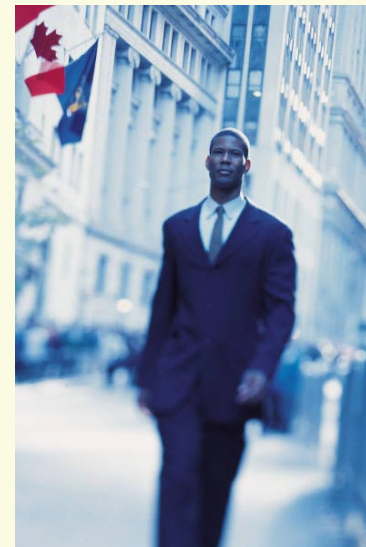5. Walkin' at varying speed on a track that is changing with moving obstacles

Mobile Evaluation

6. Walkin' on a pedestrian street.

Results

Largest number of usability problems were detected in case 1. (users sitting down!)

But, when broken down across severity (critical, serious, cosmetic), the only significant difference was in cosmetic errors.

i.e. sitting users found cosmetic errors what walking users did not.

Workload Results

Task Load Index (TLX): a method from NASA to measure the difficulty of a task.

TLX increased in with the task difficulty (roughly from 1. to 6.), but not significantly.

So did physical effort, and overall workload.

Workload Results

So the elaborate variations on walking in the lab really did make the task harder for users.

Their setups are quite easy to replicate (treadmills, walking obstacle courses).

Other ways to recreate the outdoor experience?

Error Results

Typing error rates were higher on difficult tasks (higher condition numbers), which is a warning against sitting tests.

The task difficulty was also responsible for the higher number of cosmetic errors found in case 1.

Going Mobile

In a laboratory user study we have:

- A facilitator helping the user to think aloud
- Several observers looking for critical incidents
- A Wizard, simulating the computer

How can we replicate this on a mobile device?

Ideas?

Ideas

Facilitator can use the phone for voice contact with the subject –

It is probably important to give the facilitator some context, such as frequent images or video.

Ideas

WoZ: need two-way data transfer from phone keys to dialog elements. Paper wont work, but a Denim-like lo-fi could.

Ideas

Recording: Sound is OK, but video?

Really want to see the user's expression and also their hands near the phone.

Try eyeglass cameras for hands.

Camera on phone for users face – need a video-style phone for this

Break

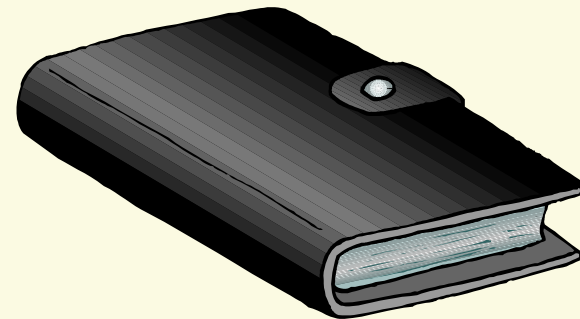# Research in Mobile Evaluation

More advanced methods:

- **ESM (Experience Sampling Method):** A mobile system prompts the user at random times: "What are you doing right now?"

- Diary studies: users record the significant events around them.

Experience Sampling

Method:

- System generates events at random times.
- These events trigger data capture on the mobile device. This could include:
  - * Location or other sensor data (no user input)
 - * A short question for the user, especially multiple choice.

Experience Sampling

Method:

- System generates events at random times.
- These events trigger data capture on the mobile device. This could include:
  * Location or other sensor data (no user input)
  * A short question for the user, especially multiple choice.

Experience Sampling

Even though a tiny fraction of the user's actual behavior is sampled, random sampling gives an unbiased estimate of what they're doing the rest of the time.
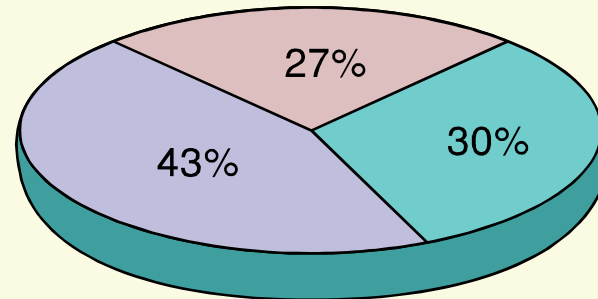
Experience Sampling

Experience sampling originated as a method for studying human behavior, not UI performance.

However, it is quite simple to do, and can provide good information about certain mobile applications.

Experience Sampling Tradeoffs

Pros:

- Unbiased sampling of user behavior.
- Simple to do, typically text-only interaction (you can use SMS for it).

Cons:

- Randomly samples **all** behavior, not system use.
- Very unlikely to catch critical incidents.
- Very little information is recorded, hard to recognize and diagnose problems later.

Experience Sampling Variants

Use system interaction as a trigger:

- Only sample when user is "using" the system, i.e. for up to 30 seconds after the last key press.

- If possible, look for apparent critical incidents and use those as triggers.

- Other ideas?

Diary Studies

In a diary study, the user records significant events by themselves. There are two kinds:

- **Feedback studies**: Users answer predefined questions about each significant event.

- **Elicitation studies**: Users make a recording when the event happens to serve as a memory aid in a later interview.

    * The recording could be a photo, or voice recording, or video.

Diary Study tradeoffs

Feedback studies are hard – users are notoriously unreliable in reporting all events.

They are also bad at judging how often and for how long they do things.

Intel Smart Home Study:

- Participants wear location sensors all day (Ubisense)
- Their behavior is also studied using diary questionnaires.

Elicitation Studies

Have a number of advantages.

Some recent research results:

- **Photos** seem to work better than other media, in terms of recall.

- But photo capture is very intrusive in social situations, and then **audio** is often used.

Still better results are obtained with media capture *and* question-and-answer annotations.

Summary

- Laboratory evaluations of mobile systems can work quite well (but more data needed).
- You can simulate cognitive and physical effects of walking in the lab.
- Experience Sampling and Diary Studies are two methods that support true mobile evaluation.
- Elicitation diary studies are probably the most effective
- Annotated photo capture is the best media for an elicitation study.