

# Fundamentals of user studies

Ed Cutrell

# Why do a user study?

You're designing stuff for people to use.

Wouldn't it be nice to know how they might use it?



# Different times—different goals

- Early on
  - Learn about the domain you're working in
    - Who are your users, what is their context?
  - Get design inspiration
- When you're building
  - Formative studies for iterative testing
  - Fix & understand your artifact, understand how users interact
- Final evaluations
  - Summative studies
  - Does your artifact do what it's supposed to do?
  - Did we achieve our goal? Are we doing a good job?
  - How does it compare to other designs or ways of doing things?

# IRBs & informed consent

- Ethics and research on human ~~subjects~~ participants
- MUST be familiar with the rules and norms of your org for human research—Institutional Review Board
  - Informed consent
  - Privacy considerations
- A few guidelines
  - Respect—remember that ppts are doing YOU a favor
  - Ppts can stop at any time
  - [http://www.hhs.gov/ohrp/irb/irb\\_chapter3.htm](http://www.hhs.gov/ohrp/irb/irb_chapter3.htm)

# Kinds of user studies

- Field studies



- Surveys

- Discount usability

- Controlled studies

**What are you trying to do?**

# Surveys

- Wide net, easy to deploy
- Targeted distribution
- Design guidance
- Evaluation of deployed systems
- Potentially biased
  - Selection bias of who responds
  - Phrasing of questions



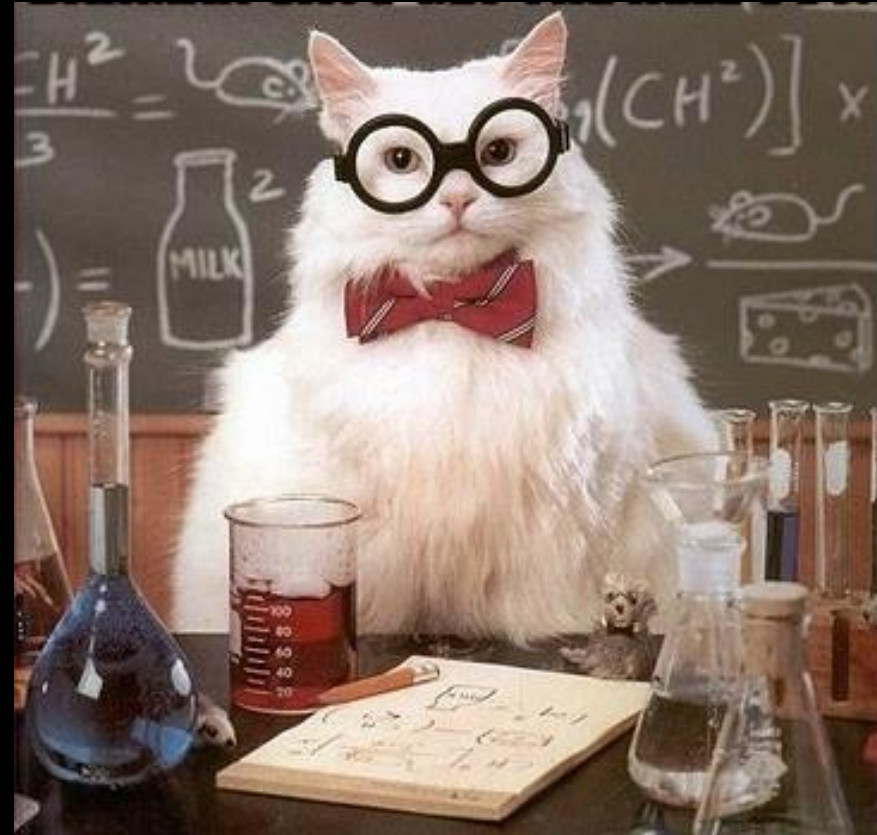
# Discount usability

- Think-aloud protocol
  - Lo-fi prototypes
  - Heuristic evaluation
  - Cognitive walk-through
- 
- Small N, iterative, no stats
  - Identifies big problems (80/20)
  - Bad for correlating personal diffs
  - Bad for understanding *why*
  - Something of an art



# Controlled studies

- Know what you're asking
- Piloting (!)
- Qualitative info
- Quantitative info
- Statistics



# What is your question?

- DON'T go fishing unless you're willing to go home empty-handed
- Controlled studies == hypothesis testing
- How will you interpret the outcome?

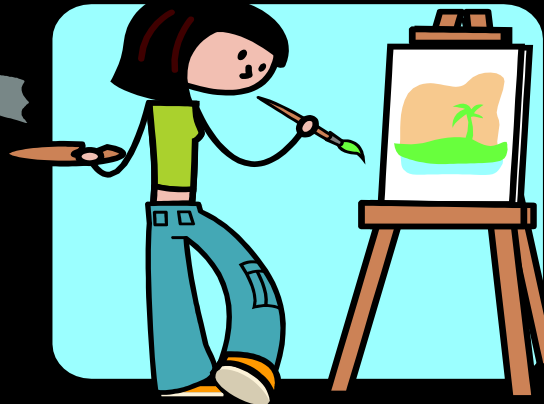
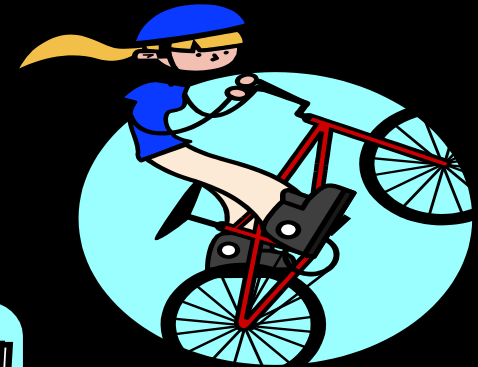
# Participant Profile

- What types people do you want to participate?
  - All the same?



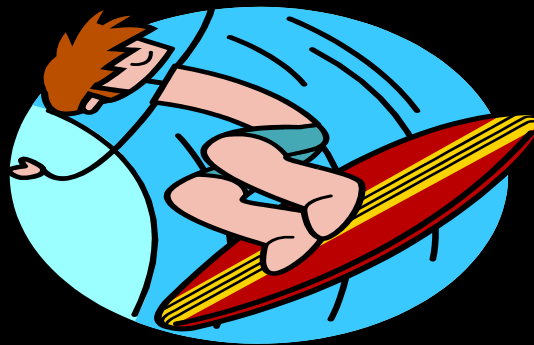
# Participant Profile

- What types people do you want to participate?
  - All the same?
  - All different?



# Participant Profile

- What types people do you want to participate?
  - All the same?
  - All different?
  - People with extreme characteristics vs. “normal” people



# Participant Profile

- What types people do you want to participate?
  - All the same?
  - All different?
  - People with extreme characteristics vs. “normal” people
- Consider
  - Age
  - Gender
  - Technology experience
  - ....

# Participant Profile (2)

- How many?
  - This can be a difficult question
  - Between or within subjects
  - What claims you are making
  - What is feasible
  - Some will drop out!



# Pilot, pilot, pilot!

Piloting will:

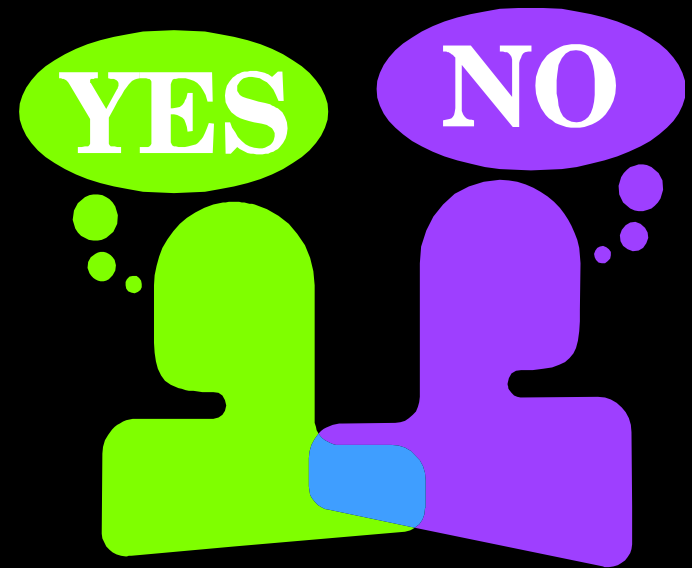
- Uncover system problems
- Uncover experimental design problems
- Uncover problems with materials

**SAVE YOUR HIDE!**



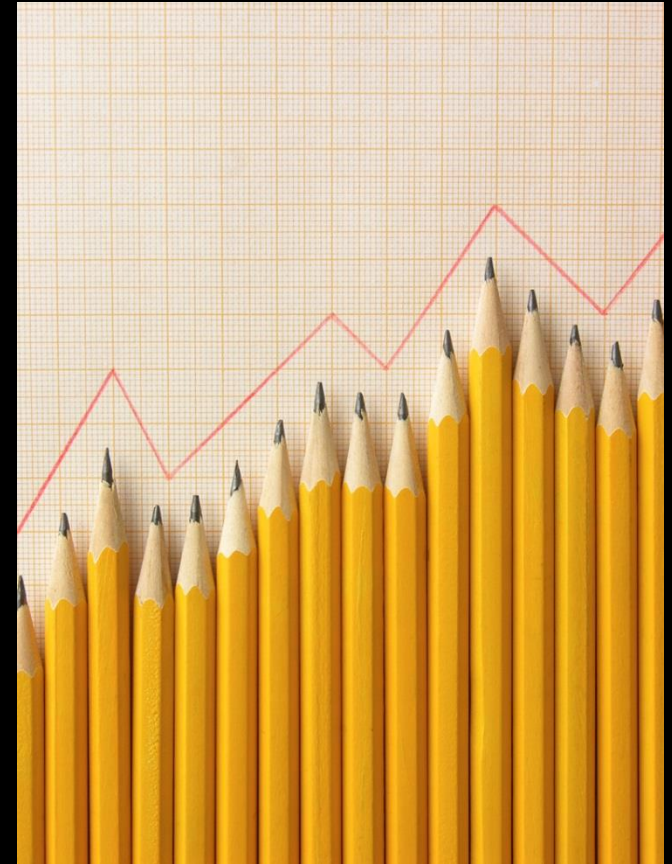
# Qualitative data

- Opinions
  - Elicited and volunteered
- Your observations
- Frustration and delight
- Likert scales & ratings  
(Scale of agreement 1-5)



# Quantitative data

- Timing
- Errors
- Task completion
- Mortality
- Modeling
- Usage
  - (What and how often?)



# Field studies

- Ethnography
  - Understand who you're designing for
- *In vivo vs. in vitro* evaluation
  - Things are different in the real world
- Deployment
  - When the cat's away...
- Instrumentation & logging
  - What can you use?



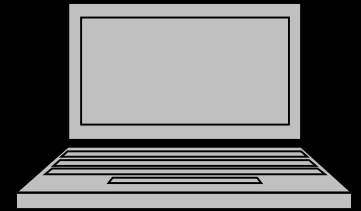
# Field studies

**Slides stolen from AJ Brush!**



# Formative

- Learn about a new domain
- Observe use of existing technology
- Methods
  - Interviews
  - Ethnography
  - Logging



# Evaluation of Artifact

- Have a prototype that you want to evaluate
- Research Question (again)
  - Participant profile
  - Length of deployment/study
  - Appropriate data to collect
  - Where can you skimp during the study....
  - What absolutely has to work

There are very few right decisions,  
instead decisions you need to justify



# Data Collection

- How will you understand if you answered your research question?
- Quantitative Data (logs)
- Qualitative Data (interviews, surveys, etc)

**TRIANGULATE** between  
multiple sources

# Logging

- You must have a plan going in about how you will use the log data
  - Risk of forgetting to log something important
  - Logging too much can create an analysis nightmare
- Make a list of questions you expect to answer with log data
  - How many times did they upload a photo?
  - How many days did they use your prototype?

# Qualitative Data

- Surveys
  - Pre-survey
  - Post Condition
  - Post-survey
- Experience Sampling Methodology
  - Small set of questions
  - Event triggered, random....
- Diaries
- Interviews (semi-structured, structured)
- Observation

# Analyzing Qualitative Data

- Affinity Diagramming
- Coding of Comments
- Inter-rater reliability



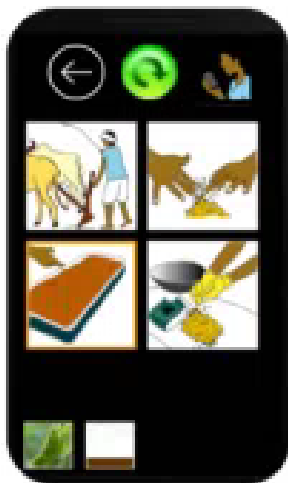
# What if it doesn't work?

- There are many ways a field study can fail
  - Technical Problems
  - They don't like your artifact
- Ensure success!
  - Figure out what has to work and skimp other places
  - Comparison between prototypes
  - Pilot studies

Example design & study of  
an ICTD project:

VideoKheti

# VideoKheti



Microsoft  
**Research**

# Started by talking with DG

- What are the problems they face?
  - Browsing video on pico projectors
  - Videos only available by mediator
  - Low literacy

# Decide on goals & parameters

- Design probe:
  - Don't worry too much about current state of infrastructure or cost
  - What would a solution look like in 5 years?
  - Inject a prototype into context (farming village) to see how people respond
    - Is this a promising direction?
    - What works and what doesn't?

# Field visits

- First visit: travel to MP, talk to farmers, mediators, & DG staff
- Second visit: collect speech data
- Third visit: user study & evaluation



# User study in the field

- 20 participants from Rajgargh, MP:
  - 8 women, 12 men
  - 10 low-literate (up to 5<sup>th</sup> grade)
- 2 conditions:
  - Speech and touch with graphics and audio (STGA)
  - Touch with graphics and audio (TGA)



Source: commons.wikimedia.org



# Task

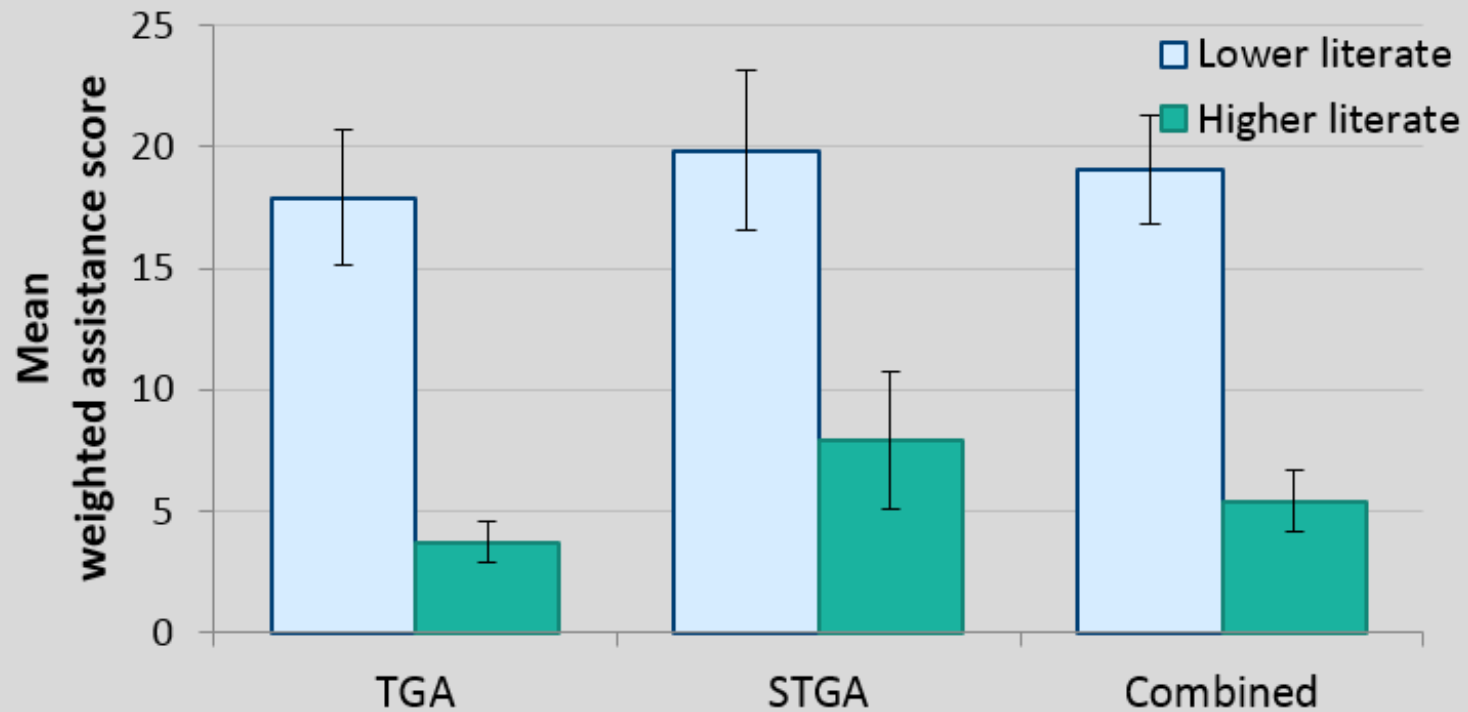
- 4 scenarios, 1 completed with the experimenter

*“This year you chose to cultivate maize. You need to treat the seed before you can sow. Find a video that will explain how to treat the seed using an organic method.”*
- Measuring the assistance provided:
  - encouragement, spoken reminder, hand-holding
  - assistance score: # enc. + 2 \* # spok. rem. + 3 \* hand-h.



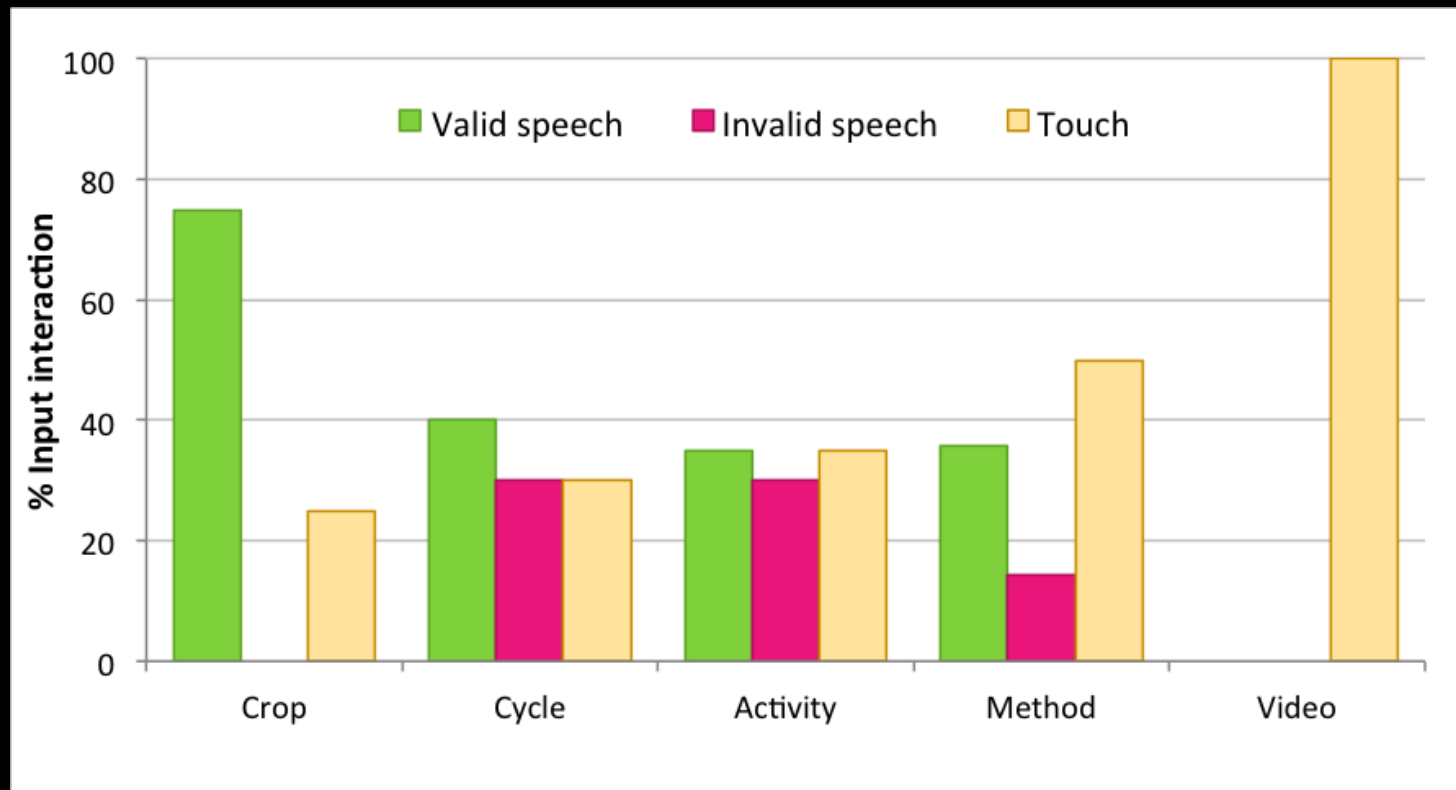
# Results: global performance

- All participants could complete all three tasks
- Low-literate users needed more assistance



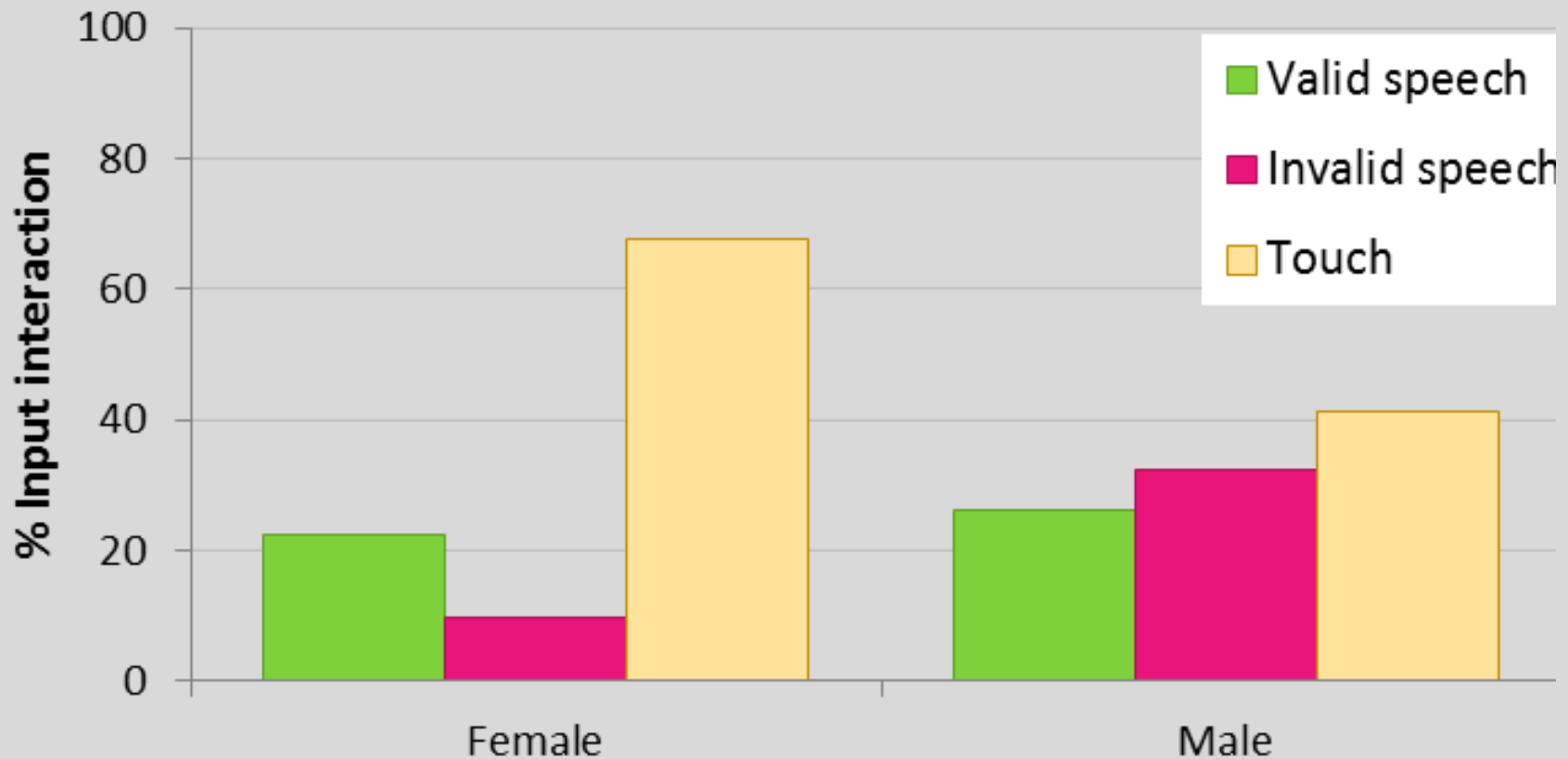
# Results: usage of speech vs. touch

- Speech is mostly used for crop selection, because:
  - Short words, common vocabulary, easy to remember
  - Many choices (2 screens)



# Results: gender differences

- Women use touch more (cultural issues)



# Problems in field evaluation

- Village chaos!
- No 2G coverage—had to use local server
- Huge difference in status between farmers & experimenters: very little frank discussion
- Gender issues: women reluctant to use voice

# Conclusions

- Implemented and evaluated a speech + graphics phone-based system for low-literate users
  - Speech recognition with SALAAM worked well
  - Participants were enthusiastic about using the system...
  - ... but the multimodal interface did not overcome many barriers for low-literate users.
- Problems less related to literacy per se, rather to other skills learned at school. It is not clear that adding speech to the interface was useful.
- Future work: how would a longer exposure to VideoKheti impact the usage by low-literate users?