UI of the day: Gesture-controlled TV

Technology:
Canesta depth camera (think Kinect)

UI:
Kicker Studio, SF
kickerstudio.com/canesta.html
Assignments

1. Due Today: Group Video Prototype
2. Due Friday: Individual Twitter App
3. New Assignment: Test Low-Fi Prototype with 3 users. You have 1 week – make it short and sweet
Plan until Spring Break

1. Today: Threads & Designing Usability Studies
2. Wednesday: Statistics & Analyzing Study Data
3. Monday: Midterm Review
4. Wednesday: In-class Midterm
Midterm on 3/16

In class. 75 minutes.
Closed book & notes.
If you are registered with the DSP office and have special needs, we need to see your letter by this Wednesday, 3/10, 1pm to make accommodations.
Threading in User Interfaces
(based on John Canny’ CS160 F10 Lecture)
What is a thread?

A **thread** is a **partial virtual machine**.

Each thread has its own stack (and local variables) but shares its heap with other threads in the same application.

Threads can be independently scheduled by the OS/VM.

```plaintext
for (i=0; i<n; i++)
{
    tmp = A[i];
    A[i] = B[i];
    B[i] = tmp;
}
```
Threads vs. Processes

A **process** is a **complete virtual machine** with its own stack and heap.

Threads share memory – processes don’t.

Threads can communicate through shared memory, processes need other mechanisms (IPC = inter-process communication).
Pros and Cons

Why use threads?
Useful model of concurrent execution, both on single processors (time-division multiplexing) and on multi processor/multi-core systems. Threads are relatively cheap to create, versatile because of shared memory.

Why wouldn’t one use threads?
Complicated programming model. Multithreaded programming is one of the biggest productivity killers of all time. (locks, semaphores, monitors, mutexes, signals, spawn, fork, join, …)
“After a long and careful analysis the results are clear: 11 out of 10 people can't handle threads.”

—Todd Hoff
Why use multithreading for UIs?

Interactive programs need to respond **quickly** to user input. Direct manipulation assumes that objects onscreen respond to user’s touch/cursor.
Why use multithreading for UIs?

Not all code can complete quickly inside an event handler. Examples?

- Network access
- File and Database IO
- Simulation

We need to decouple code for long-running computations from code for event handling and screen updates!
btnStart.setOnClickListener(new OnClickListener() {
    public void onClick(View arg0) {
        // start long computation
        Thread.sleep(60000);
        // update UI when done
        txtResult.setText("Done.");
    }
});
Event Dispatch Loop

Event Queue
- Queue of input events

Event Loop (runs in dedicated thread)
- Remove next event from queue
- Determine event type
- Find proper component(s)
- Invoke callbacks on components
- Repeat, or wait until event arrives

Component
- Invoked callback method:
  ...
  ...
- Update application state
- Request repaint, if needed

Mouse moved \((t_0, x, y)\)

Long-running operation
Stopped the event loop!
btnStart.setOnClickListener(new OnClickListener() {
    public void onClick(View v) {
        new Thread(new Runnable() {
            public void run() {
                // start long computation
                Thread.sleep(10000);
                // update UI when done
                txtResult.setText("Done.");
            }
        }).start(); // start new thread
    }
});
android.view.ViewRoot$CalledFromWrongThreadException:

Only the original thread that created a view hierarchy can touch its views.
Event Dispatch Loop

Event Queue
• Queue of input events

Mouse moved \((t_0,x,y)\)

Event Loop (runs in dedicated thread)
• Remove next event from queue
• Determine event type
• Find proper component(s)
• Invoke callbacks on components
• Repeat, or wait until event arrives

Launch Thread
...Compute...
Update app state

Component
• Invoked callback method
• Update application state
• Request repaint, if needed
Updating the UI from another thread

All common UI frameworks have a single UI thread. You are only allowed to modify the UI from the main thread.

Two fundamental rules:
Do not block the UI thread
Background threads they must not modify the UI.

Solution: When worker thread completes, request update back in the UI thread.
Almost all GUI frameworks offer some convenient mechanism to notify the main thread from another thread. Android has at least three such mechanisms:

1. Call `View.post(Runnable)` from worker thread
2. Subclass `AsyncTask` – creates threads behind the scenes
3. Send messages in one thread with `Handler.sendMessage()` – message is received in another thread (like IPC)
Handler.sendMessage Example

**Main thread**
- Handle event
- Handle event btn.OnClick()
- Handle event
- Handle event
- handleMessage()
- update GUI

**Helper thread**
- Start new thread
- Message queue
- Long computation
- sendMessage(“done”)
public class ThreadDemo extends Activity {
    final Handler handler = new Handler() {
        public void handleMessage (Message msg) {
            // update UI
            txtResult.setText((String)msg.obj);
        }
    };
}
public void onClick(View arg0) {
    new Thread(new Runnable() {
        public void run() {
            // long computation…
            Message msg = new Message();
            msg.obj = "Done."
            handler.sendMessage(msg);
        }
    }).start();
}
CS160 Thread Demo

Start Long Computation

Result:

Do Something Else
Usability Testing Methods
Iterative Design

Design
- Brainstorming
- Task analysis
- Contextual inquiry

Prototype
- Low-fi, paper

Evaluate
- Low-fi testing,
- Qualitative eval
- Quantitative eval
## Genres of assessment

<table>
<thead>
<tr>
<th>Genre</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Automated</strong></td>
<td>Usability measures computed by software</td>
</tr>
<tr>
<td><strong>Inspection</strong></td>
<td>Based on skills, and experience of evaluators</td>
</tr>
<tr>
<td><strong>Formal</strong></td>
<td>Models and formulas to calculate measures</td>
</tr>
<tr>
<td><strong>Empirical</strong></td>
<td>Usability assessed by testing with real users</td>
</tr>
</tbody>
</table>
Empirical Testing is Costly

User studies are very expensive – you need to schedule (and normally pay) many subjects.

User studies may take many hours of the evaluation team’s time.

A user test can easily cost $10k’s
“Discount Usability” Techniques

Cheap
No special labs or equipment needed
The more careful you are, the better it gets

Fast
On order of 1 day to apply
(Standard usability testing may take a week)

Easy to use
Can be taught in 2-4 hours
“Discount Usability” Techniques

Heuristic Evaluation
Assess interface based on a predetermined list of criteria

Cognitive Walkthroughs
Put yourself in the shoes of a user
Like a code walkthrough

Other, non-inspection techniques are on the rise
e.g., online remote experiments with Mechanical Turk
Cognitive Walkthrough

Given an interface prototype or specification, need:
• A detailed task with a concrete goal, ideally motivated by a scenario
• Action sequences for user to complete the task

Ask the following questions for each step:
• Will the users know what to do?
• Will the user notice that the correct action is available?
• Will the user interpret the application feedback correctly?

Record: what would cause problems, and why?

From: Preece, Rogers, Sharp – Interaction Design
Empirical Assessment: Qualitative

Qualitative: What we’ve been doing so far

**Contextual Inquiry**: try to understand user’s tasks and conceptual model

**Usability Studies**: look for critical incidents in interface

Qualitative methods help us:
Understand what is going on
Look for problems
Roughly evaluate usability of interface
Empirical: Quantitative Studies

Quantitative
- Use to reliably measure some aspect of interface
- Compare two or more designs on a measurable aspect
- Contribute to theory of Human-Computer Interaction

Approaches
- Collect and analyze user events that occur in natural use
- Controlled experiments

Examples of measures
- Time to complete a task,
- Average number of errors on a task,
- Users’ ratings of an interface*

*You could argue that users’ perception of speed, error rates etc is more important than their actual values
Comparison

**Qualitative studies**
Faster, less expensive → esp. useful in early stages of design cycle

**Quantitative studies**
Reliable, repeatable result → scientific method
Best studies produce generalizable results
You will conduct a **qualitative** study

We don’t have enough time or subjects for quantitative studies

But you should do a little quantitative analysis

What are your measures?

Compute summary statistics (mean, stdev)

Do you have independent, dependent, and control variables?
Designing Controlled Experiments
Steps in Designing an Experiment

1. State a lucid, testable hypothesis
2. Identify variables (independent, dependent, control, random)
3. Design the experimental protocol
4. Choose user population
5. Apply for human subjects protocol review
6. Run pilot studies
7. Run the experiment
8. Perform statistical analysis
9. Draw conclusions
Example: Bubble Cursor

http://husk.eecs.berkeley.edu/projects/bubble
Lucid, Testable Hypothesis

H1: Users will acquire targets faster with the Bubble cursor (their movement time will be lower).

H2: Users will have a lower error rate with the Bubble cursor.

Other hypotheses?
Experiment Design

**Testable hypothesis**
Precise statement of expected outcome

**Independent variables (factors)**
Attributes we manipulate/vary in each condition
Levels – values for independent variables

**Dependent variables (response variables)**
Outcome of experiment (measurements)
Usually measure user performance
Experiment Design

**Control variables**
- Attributes that will be fixed throughout experiment
- Confound – attribute that varied and was not accounted for

Problem: Confound rather than IV could have caused change in DVs
Confounds make it difficult/impossible to draw conclusions

**Random variables**
- Attributes that are randomly sampled
- Increases generalizability
Variable Types

Nominal: categories with labels, no order

Ordinal: categories with rank order

Continuous: interval (w/o zero point), ratio (w/ zero point)
Common Metrics in HCI

Performance metrics:
- Task success (binary or multi-level)
- Task completion time
- Errors (slips, mistakes) per task
- Efficiency (cognitive & physical effort)
- Learnability

Satisfaction metrics:
- Self-report on ease of use, frustration, etc.
Performance Metric: Errors
Performance Metric: Lostness

Smith 1996:
N: # of different pages visited
S: # of total pages visited, incl. revisits
R: minimum # of pages to accomplish task

Lostness = \sqrt{(\frac{N}{S}-1)^2+\left(\frac{R}{N}-1\right)^2}
Satisfaction Metric: Likert Scales

Respondents rate their level of agreement to a statement

Likert data is ordinal, not continuous (matters for analysis)!

“Overall, I am satisfied with the ease of completing the tasks in this scenario”

1: Strongly Disagree
2: Disagree
3: Neither agree nor disagree
4: Agree
5: Strongly agree
Variables for the Bubble Cursor

- Independent variables
- Dependent variables
- Control variables
- Random variables
Variables

Independent variables
Cursor type (bubble, normal, area?)
Target Distance
Target Width (Effective vs. Actual?)

Dependent variables
Movement Time
Error Rate
User Satisfaction

Control variables
Color scheme, input device, screen size

Random variables
Location, environment, Attributes of subjects
Age, gender, handedness, …

Conducting studies online vs. in person strongly influences which variables are controlled and which are random.
Goals

**Internal validity**
Manipulation of IV is cause of change in DV
Requires eliminating confounding variables (turn them into IVs or RVs)
Requires that experiment is replicable

**External validity**
Results are generalizable to other experimental settings

*Ecological validity* – results generalizable to real-world settings

**Confidence in results**
Statistics
Experimental Protocol

What is the task? (must reflect hypothesis!)
What are all the combinations of conditions?
How often to repeat each combination of conditions?
Between subjects or within subjects
Avoid bias (instructions, ordering, …)
Number of Conditions

Consider all combinations to isolate effects of each IV (factorial design)

\[(3 \text{ cursor types}) \times (3 \text{ distances}) \times (3 \text{ widths}) = 27 \text{ combinations}\]

Adding levels or factors can yield lots of combinations!
Reducing Num. of Conditions

Vary only one independent variable
leaving others fixed

Problem: ?
Reducing Num. of Conditions

Vary only one independent variable leaving others fixed

Problem: Will miss effects of interactions
Other Reduction Strategies

Run a few independent variables at a time
If strong effect, include variable in future studies
Otherwise pick fixed control value for it

Fractional factorial design
Procedures for choosing subset of independent variables to vary in each experiment
Choosing Subjects

Pick balanced sample reflecting intended user population

Novices, experts
Age group
Sex
....

Example

12 non-colorblind right-handed adults (male & female)

Population group can also be an IV or a controlled variable

What is the disadvantage of making population a controlled var?
Between Subjects Design

Wilma and Betty use one interface  Dino and Fred use the other
Within Subjects Design

Everyone uses both interfaces
Between vs. Within Subjects

Between subjects
Each participant uses one condition
+/- Participants cannot compare conditions
+ Can collect more data for a given condition
- Need more participants

Within subjects
All participants try all conditions
+ Compare one person across conditions to isolate effects of individual diffs
+ Requires fewer participants
- Fatigue effects
- Bias due to ordering/learning effects
Within Subjects: Ordering Effects

In within-subjects designs ordering of conditions is a variable that can confound results
Why?

Turn it into a random variable
Randomize order of conditions across subjects
Counterbalancing (ensure all orderings are covered)
Latin square (partial counterbalancing)
...
Run the Experiment

Always pilot it first!
Reveals unexpected problems
Can’t change experiment design after starting it

Always follow same steps – use a checklist

Get consent from subjects

Debrief subjects afterwards
For Wed: Bubble Cursor Experiment

Participate in Bubble Cursor Experiment for Wednesday’s class!

Bubble Cursor Online Experiment
UC Berkeley CS160

In this short experiment, you will be asked to click on a sequence of targets on screen. You will do this both with a regular mouse cursor and with a different type of cursor that dynamically expands in size to always select the closest target. This Bubble Cursor is inspired by works at CHI 2005.[1]

Watch the instructions below.

First, you will be asked to click on the red targets - quickly but accurately. Click on a few of them.
Managing Study Participants
The Participants’ Standpoint

Testing is a distressing experience
Pressure to perform
Feeling of inadequacy
Looking like a fool in front of your peers, your boss, ...

(from “Paper Prototyping” by Snyder)
The Three Belmont Principles

Respect for Persons
Have a meaningful consent process: give information, and let prospective subjects freely chose to participate

Beneficience
Minimize the risk of harm to subjects, maximize potential benefits

Justice
Use fair procedures to select subjects (balance burdens & benefits)

To ensure adherence to principles, most schools require Institutional Review Board approval of research involving human subjects.
1971 Experiment by Phil Zimbardo at Stanford
24 Participants – half prisoners, half guards ($15 a day)
Basement of Stanford Psychology bldg turned into mock prison
Guards given batons, military style uniform, mirror glasses,…
Prisoners wore smocks (no underwear), thong sandals, pantyhose caps

Experiment quickly got out of hand
Prisoners suffered and accepted sadistic treatment
Prison became unsanitary/inhospitable
Prisoner riot put down with use of fire extinguishers
Guards volunteered to work extra hours

Zimbardo terminated experiment early
Grad student Christina Maslach objected to experiment
Important to check protocol with ethics review boards

[from Wikipedia]
Ethics

Was it useful?
“…that’s the most valuable kind of information that you can have - and that certainly a society needs it” (Zimbardo)

Was it ethical?
Could we have gathered this knowledge by other means?

http://www.prisonexp.org/slide-42.htm
“In 2001, a faculty member from the business school of a major university designed a study to see how restaurants would respond to complaints from putative customers. As part of the project, the researcher sent letters to restaurants falsely claiming that he and/or his wife had suffered food poisoning that ruined their anniversary celebration. The letters disclaimed any intention of contacting regulatory agencies and stated that the only intent was to convey to the owner what had occurred "in anticipation that you will respond accordingly." Restaurant owners were understandably upset and some employees lost their jobs before it was revealed that the letter was a hoax.”
Beneficience: Example

**MERL DiamondTouch:**
User capacitively coupled to table through seating pad. No danger for normal users, but possibly increased risk for participants with pacemakers. Inform subjects in consent!

Privacy: having control over the extent, timing, and circumstances of sharing oneself with others.

Confidentiality: the treatment of information that an individual has disclosed with the expectation that it will not be divulged.

Examples where privacy could be violated or confidentiality may be breached in HCI studies?
Treating Subjects With Respect

Follow human subject protocols
- Individual test results will be kept confidential
- Users can stop the test at any time
- Users are aware (and understand) the monitoring technique(s)
- Their performance will not have implications on their life
- Records will be made anonymous

Use standard informed consent form
- Especially for quantitative tests
- Be aware of legal requirements
Conducting the Experiment

**Before the experiment**
- Have them read and sign the consent form
- Explain the goal of the experiment in a way accessible to users
- Be careful about the demand characteristic (Participants biased towards experimenter’s hypothesis)
- Answer questions

**During the experiment**
- Stay neutral
- Never indicate displeasure with users' performance

**After the experiment**
- Debrief users (Inform users about the goal of the experiment)
- Answer any questions they have
Managing Subjects

Don’t waste users’ time
- Use pilot tests to debug experiments, questionnaires, etc…
- Have everything ready before users show up

Make users comfortable
- Keep a relaxed atmosphere
- Allow for breaks
- Pace tasks correctly
- Stop the test if it becomes too unpleasant
If you want to learn more…

Online human subjects certification courses:  
E.g., http://phrp.nihtraining.com/users/login.php

The Belmont Report: Ethical Principles and Guidelines for the protection of human subjects of research

1979 Government report that describes the basic ethical principles that should underly the conduct of research involving human subjects

http://ohsr.od.nih.gov/guidelines/belmont.html
For Wednesday

Reading: Martin, Chapter 12.

No reading response required – participate in experiment instead!
We’ll analyze your data in class.