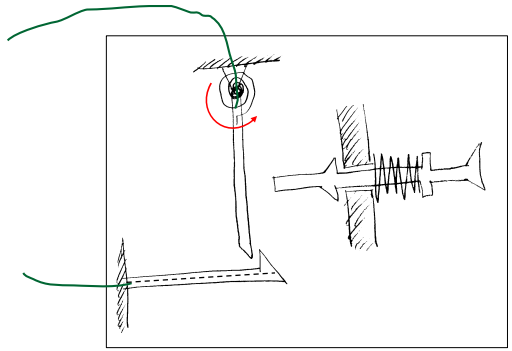
Natural Interaction

Randall Davis
and numerous collaborators
MIT CSAIL

Consider This Device...



A Vision of HCI

- Make interacting with computation as cognitively effortless as face-to-face human interaction *can* be
- Consider how you interact with a computer
  - unidirectional
  - narrow set of modalities (type, click)
  - must be exhaustive, complete, unambiguous …
  - insanely literal-minded

Interaction Should Be

- *Non-distracting*
  - Cognitively effortless
- *Collaborative, mixed initiative*
  - Dialogue vastly reduces demands on speaker
- *Multimodal*
  - modality *opportunistic*
  - modality *agnostic*
  - modality *symmetric*

What Will It Require?

- Intelligence
  - Why is it so much easier to interact with (most) people than with computers?
  - Natural interaction is knowledge-based
    - knowledge of how people draw, talk, gesture, converse,…
    - knowledge of the domain

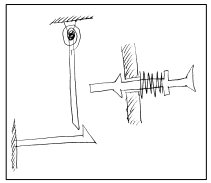
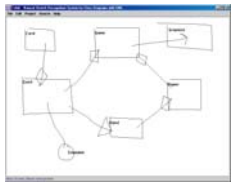
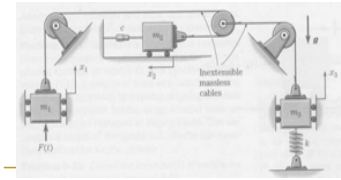
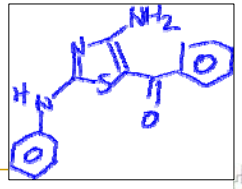
What Will It Require?

- Multimodal interaction
  - sketch
  - talk
  - gesture
- *Symmetric* multimodal interaction

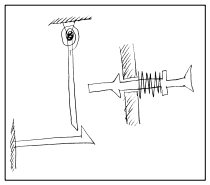
Why Do We Need This?

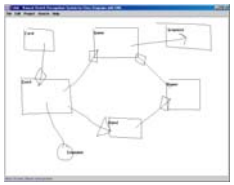
- We're surrounded by computation
  - Ubiquitous computing (1991)
- *Computation is surround by us*.
  - Computation lives in *our* world
  - ➔ it needs to understand human interaction
  - ➔ it needs to be part of our social context
  - A new meaning for *social computing*

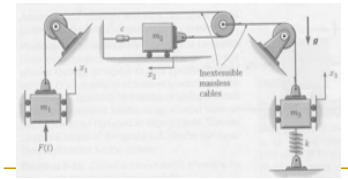
Sketches are Ubiquitous

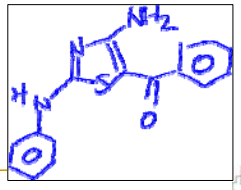





Sketches are *Dead*

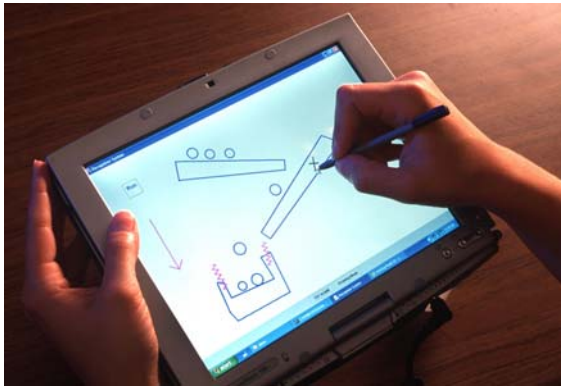






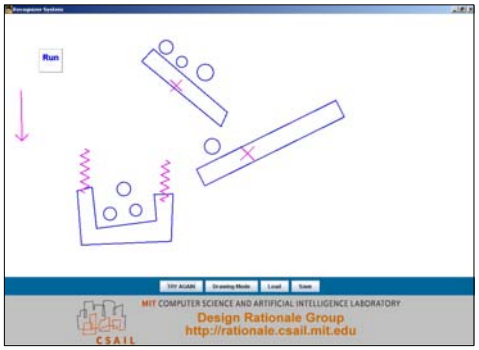


Magic Paper



Christine Alvarado

Sketching Physics

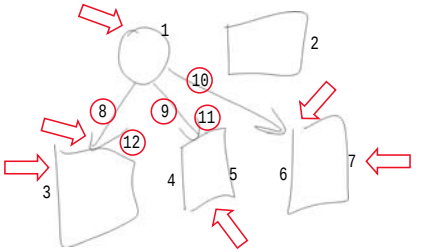


(Why) is this *interesting*?

Sketch Understanding: What's Interesting

- Pragmatic answer
  - Sketching is ubiquitous
- Visionary Answer
  - Electronic textbooks: What if every figure…
- Research answer
  - Sketch understanding as signal interpretation

Sketch Understanding: What's *Difficult?*



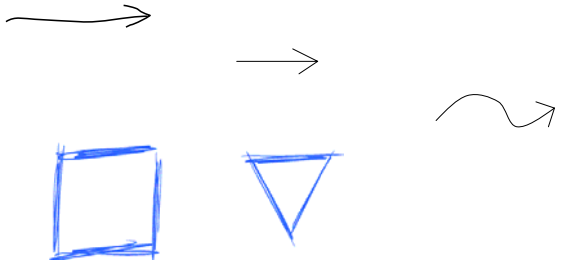
- Task is incremental
- Signal is noisy
- Styles vary
- Segmentation is difficult
- The signal is 2-d, non-chronological

Research Issues: Four Representations

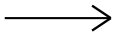
- Spatial/Image: what it looks like
- Temporal: how it's drawn
- Conceptual: how it's defined
- Semantic: what it means

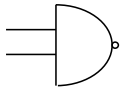
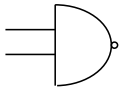
Four Representations

- Spatial/Image



Four Representations

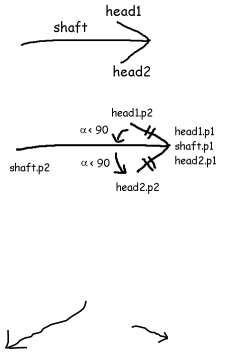
- Temporal 
  - shaft
  - head (barb to shaft)
  - head (shaft to barb)

Four Representations

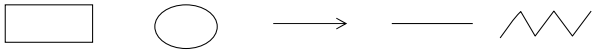
- Conceptual

```
components
 Line shaft
 Line head1
 Line head2
constraints
 equal head1 head2
 coincident head1.p1 head2.p1
 coincident head1.p1 shaft.p1
 acuteMeet head1 shaft
 acuteMeet shaft head2
```

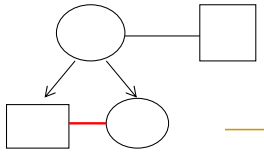


Four Representations

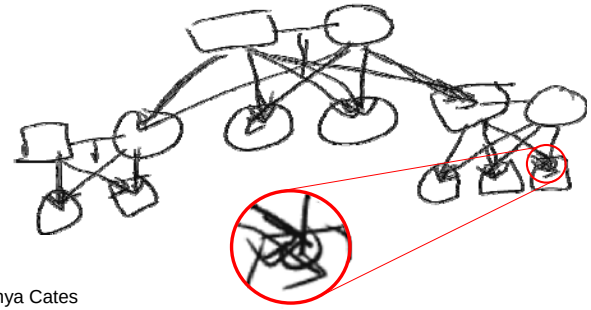
- Semantics
  - Family trees
    - Lexicon



    - Grammar

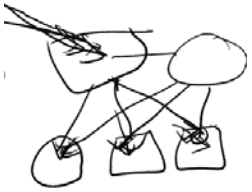


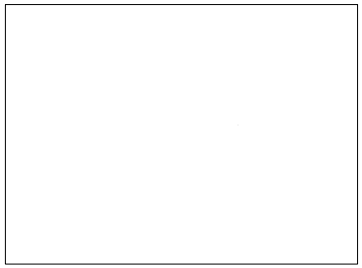
Confusion And Complexity Are Often Representation Specific



Sonya Cates

Confusion And Complexity Are Often Representation Specific



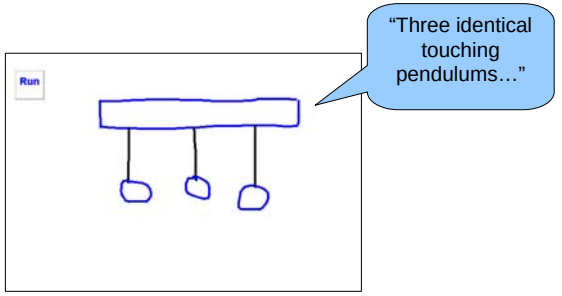


"Look" At It From Multiple Views

- Group spatially related elements
  - We (tend to) separate objects
- Group temporally related elements
  - We (usually) finish one idea before starting another
  - We (typically) pause between ideas, not in the middle

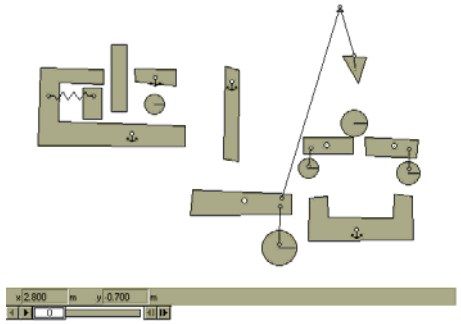
Sketching Isn't Enough

Multi-Modal Interaction: Sketching and Speech

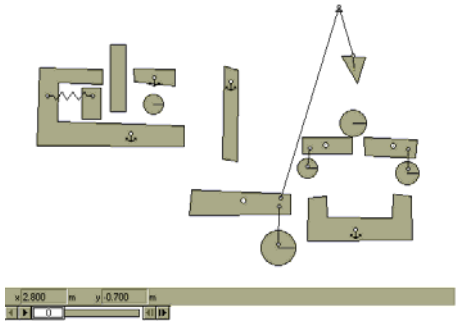


• Graphical vs verbal communication:
What's best said, what's best sketched?

Sketching and Speech Isn't Enough

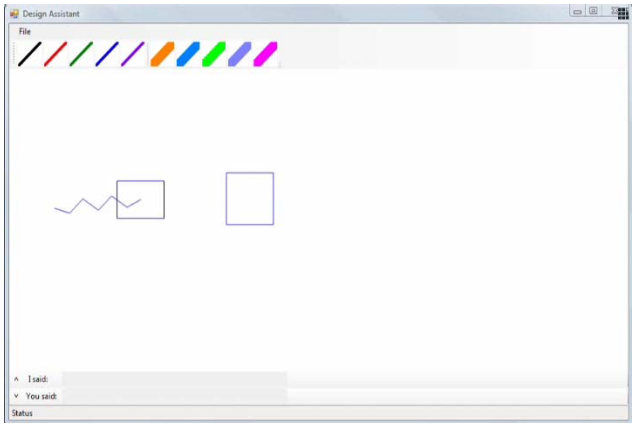


Sketching and Speech Isn't Enough



Symmetric multimodal interaction

- Interaction is *much* easier as a dialogue.
- We can sketch, talk, and gesture.
- Why can't the system?



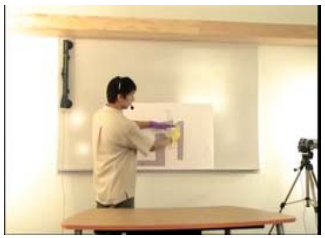
Aaron Adler

Multi-Modal Interaction: Speech and Gesture

- Gesture interpretation depends on linguistic context.
- No adequate representation of individual gestures
  - diverse, fluid, idiosyncratic

Jacob Eisenstein

Key Idea



"**This thing** clicks back…" **0:46** "…**it** goes out and clicks back…" **1:10**

***Patterns* in gesture predict *patterns* in language**

Another Example?

~~"Similar gestures imply similar content"~~

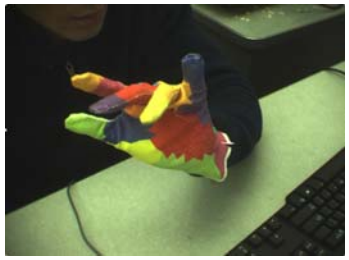
Similar *meaningful* gestures imply similar content.

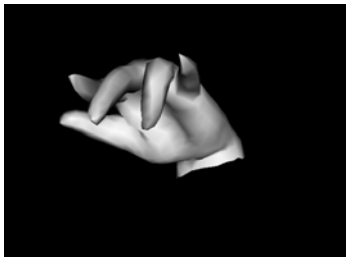
Gestural Salience

- Can viewers distinguish communicative gestures from other hand movements consistently and robustly?
- Yes: (Kendon 1978).

- Salience model learned, unsupervised, from labeled linguistic data

Gesture Understanding





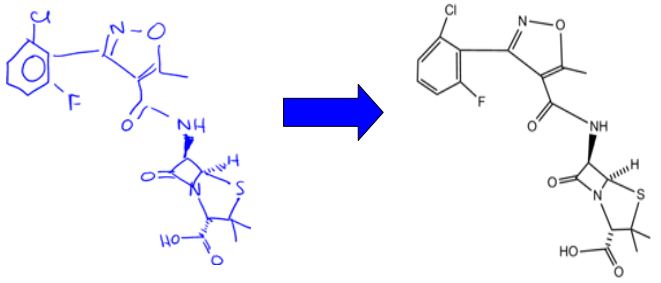
Rob Wang

Gestures on a Tabletop

Ying Yin

Body Tracking

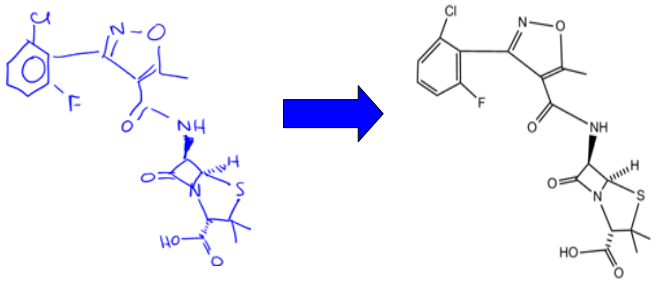
Sketching Chemistry

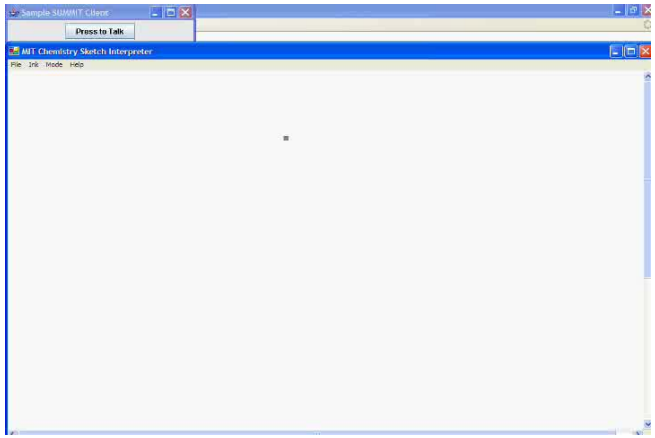


Tom Ouyang

Demo

Multimodal Interaction in Chemistry





Tom Ouyang, Chih-yu Chao

The Vision, Redux

- A smart digital whiteboard, tablet computer, wall
- Offering interaction that is
  - Non-distracting, cognitively effortless
  - Collaborative, mixed initiative
  - Multimodal
    - modality opportunistic
    - modality symmetric
    - modality agnostic
- Because it
  - knows a lot, about interaction, the domain, …