

# Multimodal Dialogue

## Intelligent User Interfaces

Professor Charles Rich  
Computer Science Department  
[rich@wpi.edu](mailto:rich@wpi.edu)

# Readings

---

- Chai et al, A Probabilistic Approach to Reference Resolution, IUI'04
- Horchani et al, A Platform for Output Dialogic Strategies in Natural Multimodal Dialogue Systems, IUI'07

# “Put That There”

---



[http://www.futureofthebook.org/discussions/2006/02/put\\_that\\_there.html](http://www.futureofthebook.org/discussions/2006/02/put_that_there.html)

MIT Architecture Machine Lab, 1980

# Basic Concepts

---

## 1. Modalities:

- speech
- gesture
- gaze
- touch
- vision
- graphics
- smell?

## 2. As input to system (e.g., reference resolution)

## 3. As output (e.g., presentation planning)



# Common Issues

---

## 1. Ambiguity

- what is the person pointing at?
- what does this word mean in this context?
- etc.

## 2. “Fusion” (combining info from multiple modalities)

- to reduce ambiguity (e.g., head nod and “yes”)
- for convenience (e.g., “put that there”)



# Ten Myths of Multimodal Interaction

---

[Sharon Oviatt, CACM, Nov. 1999](#)

*Myth #1:*

If you build a multimodal system, user will interact multimodally.

- users like to interact multimodally, but they don't always do so
- often intermix unimodal and multimodal



# Ten Myths of Multimodal Interaction

---

## *Myth #2:*

Speech and pointing is the dominant multimodal integration pattern.

- studies show “speak and point” only 14% of spontaneous multimodal utterances
- also facial expression, touch, manipulation, etc.

# Ten Myths of Multimodal Interaction

---

*Myth #3:*

Multimodal input involves simultaneous signals.

- estimated only 25% of “speak and point” commands overlap temporally.
- gesture often precedes speech
- i.e., synchronized but not simultaneous

# Ten Myths of Multimodal Interaction

---

## *Myth #4:*

Speech is the primary input mode in any multimodal system that includes it.

- spatial inputs often necessary (e.g., pen input)
- speech degrades badly in noisy environments

# Ten Myths of Multimodal Interaction

---

## *Myth #5:*

Multimodal language does not differ linguistically from unimodal language.

- often much simpler language use
- cannot just reuse same grammars, etc.

# Ten Myths of Multimodal Interaction

---

*Myth #6:*

Multimodal integration involves redundancy of content between modes.

- complementarity is more common
- e.g., “put that there”



# Ten Myths of Multimodal Interaction

---

## *Myth #7:*

Individual error-prone recognition technologies combine multimodally to produce even greater unreliability.

- usually the opposite (i.e., disambiguation)
- e.g., speech and handwriting

# Ten Myths of Multimodal Interaction

---

## *Myth #8:*

All users' multimodal commands are integrated in uniform ways.

- different users have different patterns, e.g., some prefer pen
- systems need to adapt to user differences

# Ten Myths of Multimodal Interaction

---

## *Myth #9:*

Different input modes are capable of transmitting comparable content.

- admirable desire for “universal” interfaces ignores real differences in modalities
- e.g., need to design a speech interface differently from the bottom up

# Ten Myths of Multimodal Interaction

---

## *Myth #10:*

Enhanced efficiency is the main advantage of multimodal systems.

- sometimes more important to reduce errors than to speed up interaction
- e.g., 10% speed increase reported for pen plus speech vs. speech alone
- e.g., 36-50% reduction in task-critical errors

# Readings

---

- Chai et al, A Probabilistic Approach to Reference Resolution, IUI'04
  - multimodal *input*
  - data fusion
- Horchani et al, A Platform for Output Dialogic Strategies in Natural Multimodal Dialogue Systems, IUI'07
  - multimodal output
  - very “methodological”