

Multimodal Interfaces

Shiri Azenkot

May 29, 2013

LNG 575

Multimodal Interface Papers

Oviatt. 2012.

Multimodal Interfaces

Feng et al. 2011.

Speech and Multimodal Interaction in Mobile Search

Oviatt, Gent. 1996.

Error Resolution During Multimodal Human-Computer Interaction

Paek et al. 2008.

**Search Vox: Leveraging Multimodal Refinement and
Partial Knowledge for Voice Search**

Suhm et al. 2001.

Multimodal Error Correction for Speech User Interfaces

Multimodal Interface Papers

Suhm et al. 2001.

Multimodal Error Correction for Speech User Interfaces

Paek et al. 2008.

Search Vox: Leveraging Multimodal Refinement and Partial Knowledge for Voice Search

Feng et al. 2011.

Speech and Multimodal Interaction in Mobile Search

Oviatt, Gent. 1996.

Error Resolution During Multimodal Human-Computer Interaction

Oviatt. 2012.

Multimodal Interfaces

What is a
multimodal Interface?



3G 9:41 AM

Cancel Back at school Send

To: Melissa Chevalier

Cc/Bcc; From: zgosler55@me.com

Subject: Back at school

|

Sent from my iPhone



Done

3G 9:41 AM

Cancel Back at school Send

To: Melissa Chevalier

Cc/Bcc; From: zgosler55@me.com

Subject: Back at school

●●●|

Sent from my iPhone

Q W E R T Y U I O P
A S D F G H J K L
↑ Z X C V B N M ↵
123 microphone space return

3G 9:41 AM

Cancel Back at school Send

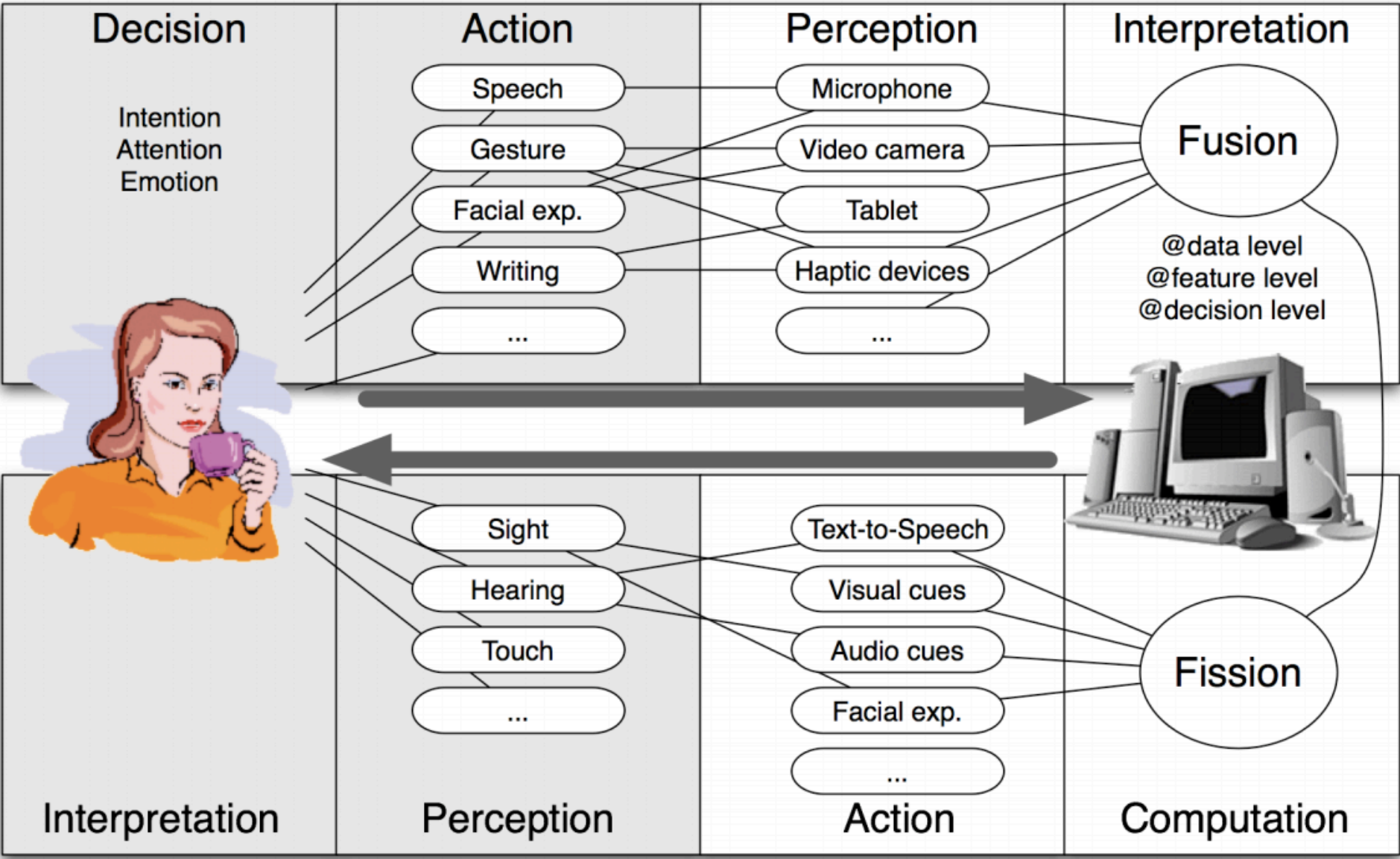
To: Melissa Chevalier

Cc/Bcc; From: zgosler55@me.com

Subject: Back at school

I'm back at school now and most of my new professors.

Q W E R T Y U
A S D F G H
↑ Z X C V B N M ↵
123 microphone space return



Suhm et al. 2001.

Multimodal Error Correction for Speech User Interfaces

The Problem

- Speech recognition systems are limited
- Correcting errors is challenging
“repair problem”

Multimodal Dictation System

Dictation recognizer +
multimodal error correction



Audio Control

Reset All

Acoustic Adaptation

Quit Demo

Speech Recognition: better



faster

Carnegie
Mellon

Multimodal Listening Typewriter

Uni
Karlsruhe

Add Word

Insert at Cursor, Substitute Selected Words or Delete

Undo

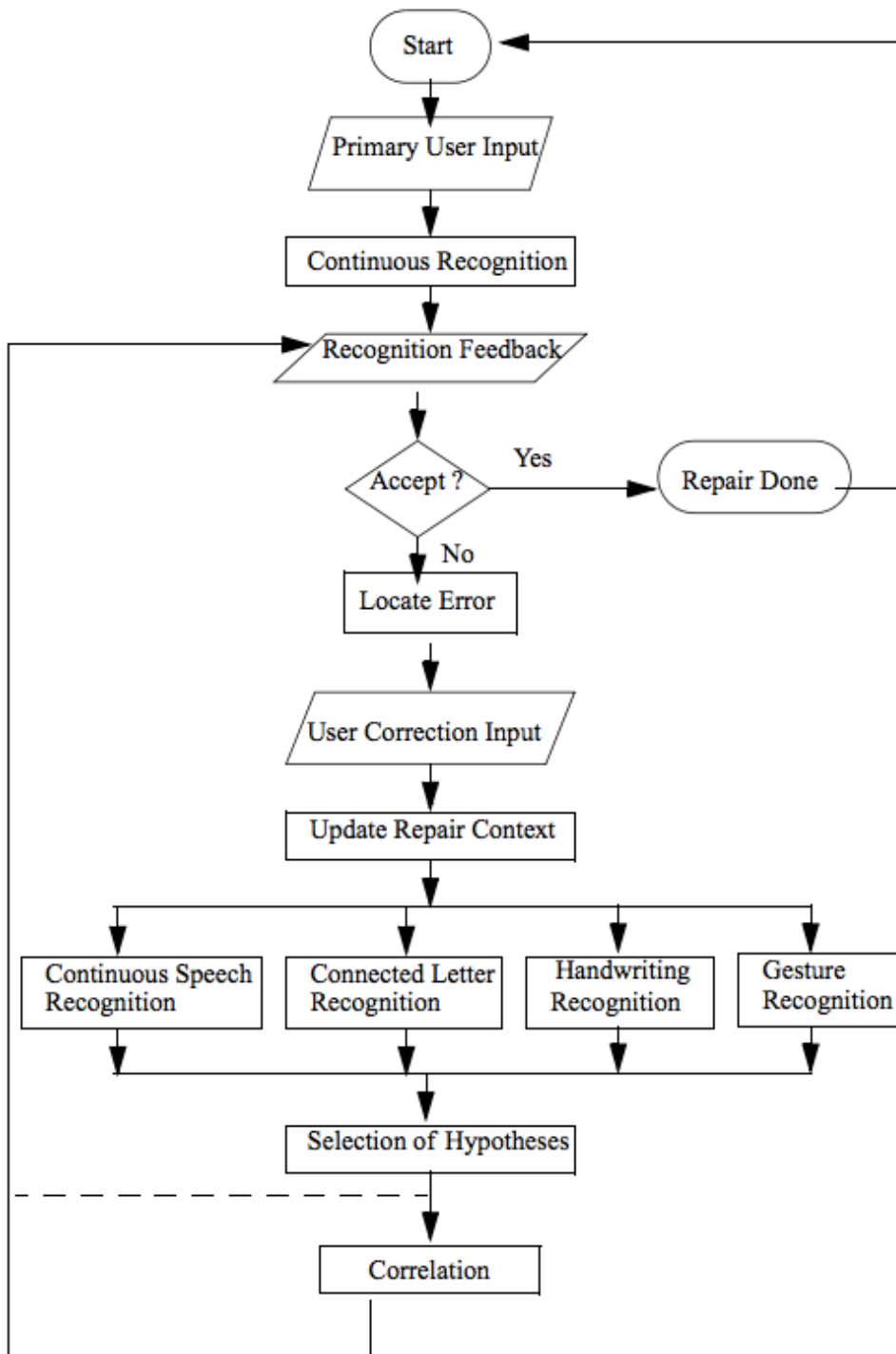
**The multimodal listening typewriter
allows to input and correct text using
continuous speech, spelling, and gestures**

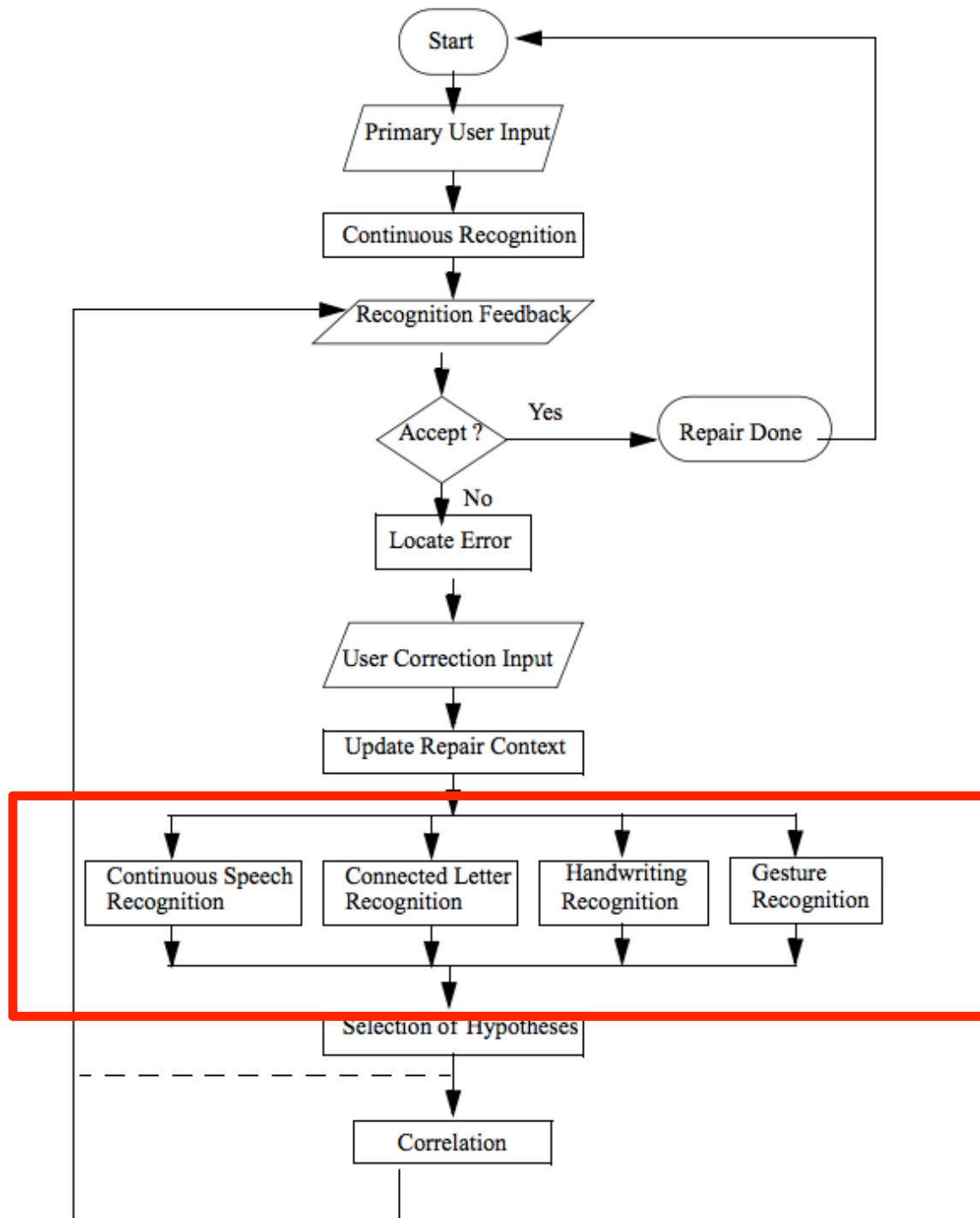
handwriting

Select Next Error

Dictate/Respeak

Spell





Gesture Recognition

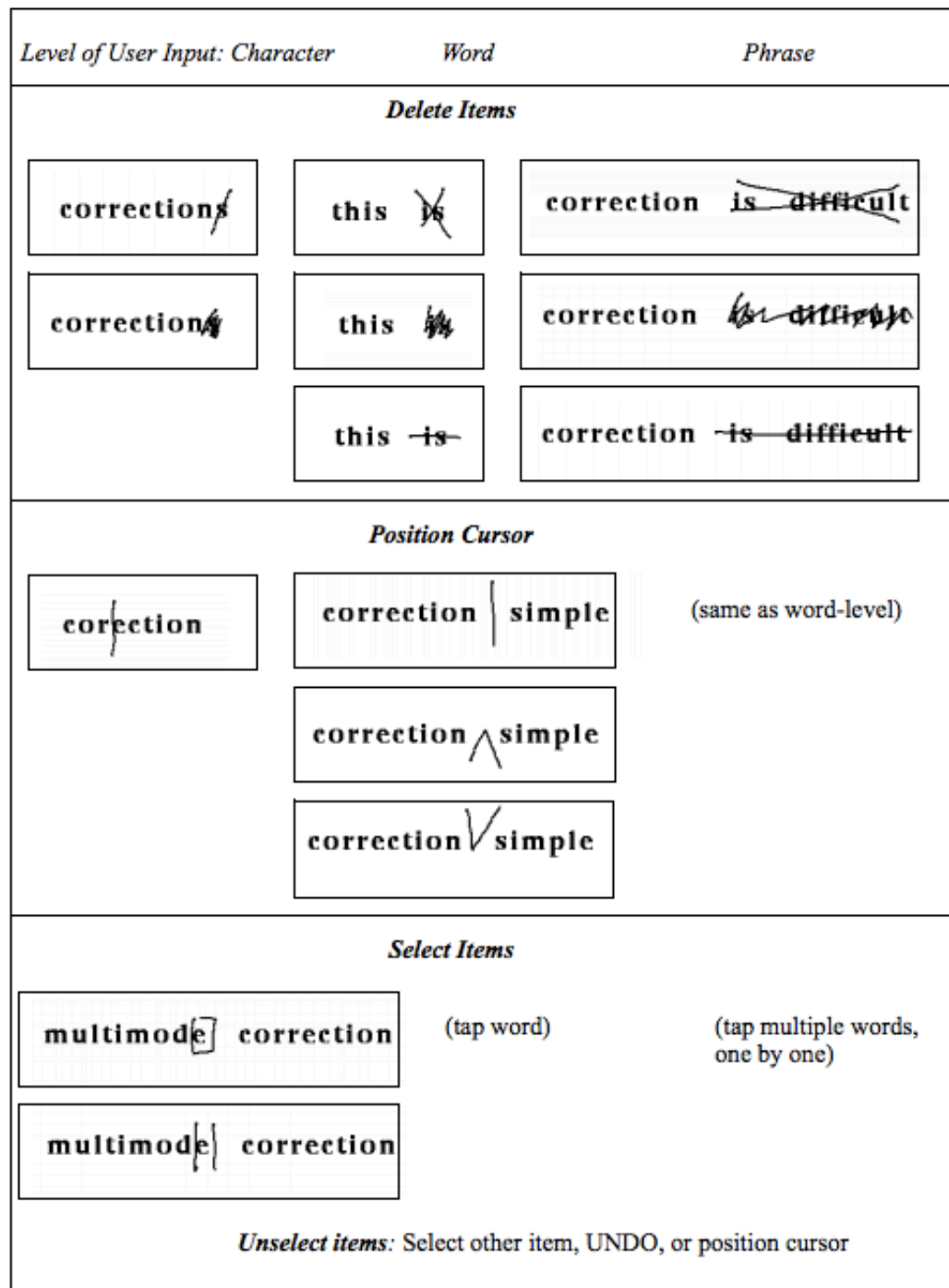


Fig. 4. Editing using pen-based gestures.

Evaluation

- Hypotheses:
 - Multimodal correction is faster
 - Detecting potential errors automatically speeds up correction process
 - Users prefer fastest modality

Evaluation

- Tasks: participants read newspapers
- Factors:
 - Correction modality
- Measures:
 - Input rate (WPM)
 - System response time
 - Recognition accuracy
 - Correction accuracy
 - **Correction speed**

Correction accuracy for modality

Modality	Correction Accuracy
Choose from alternatives	24%
Respeaking	35%
Handwriting	75%
Spelling	82%
Typing	87%

Use of different modalities

- - -

Modality	Usage Frequency
Spelling	0.14
Choose from alternatives	0.21
Respeaking	0.28
Handwriting	0.35

Paek et al. 2008.

**Search Vox: Leveraging Multimodal
Refinement and Partial Knowledge
for Voice Search**

The Problem

Voice search is common

Challenges of voice search on mobile devices

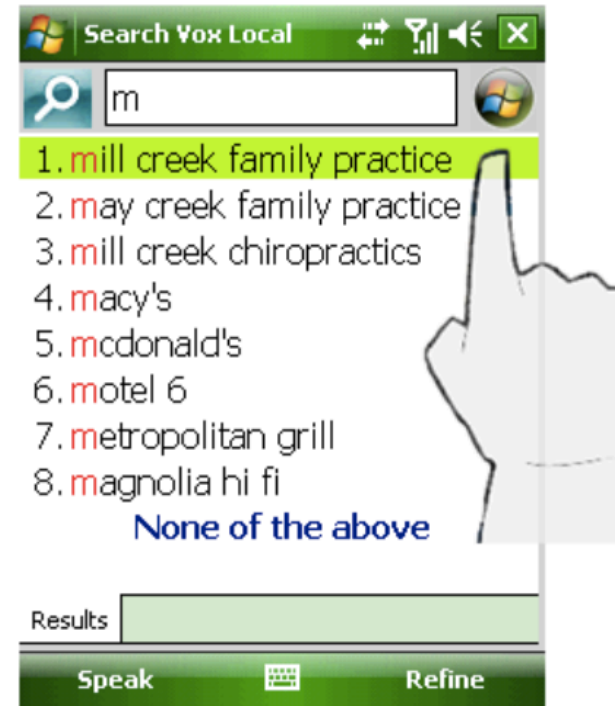
1. Noisy environments
2. User compensate in unhelpful ways



Text

Touch

Speech



Text hints



Search Vox Local

black * restaurant

1. black angus restaurant
2. mexican restaurant
3. claim jumper restaurant
4. canlis restaurant
5. arnies restaurant
6. cactus restaurant
7. azteca restaurant
8. canyons restaurant

None of the above

Results

Speak Refine

Search Vox Local

black angus restauran

1. black angus restaurant
2. mexican restaurant
3. claim jumper restaurant
4. canlis restaurant
5. arnies restaurant
6. cactus restaurant
7. azteca restaurant
8. canyons restaurant

None of the above

Results

Speak Refine



Using partial knowledge



Search Vox Local

b* angus

1. black angus restaurant
2. black angus motor inn
3. black angus steakhouse
4. b & b auto parts
5. b c surf & sport
6. c b richard ellis
7. j b factory carpets
8. b & j roofing

None of the above

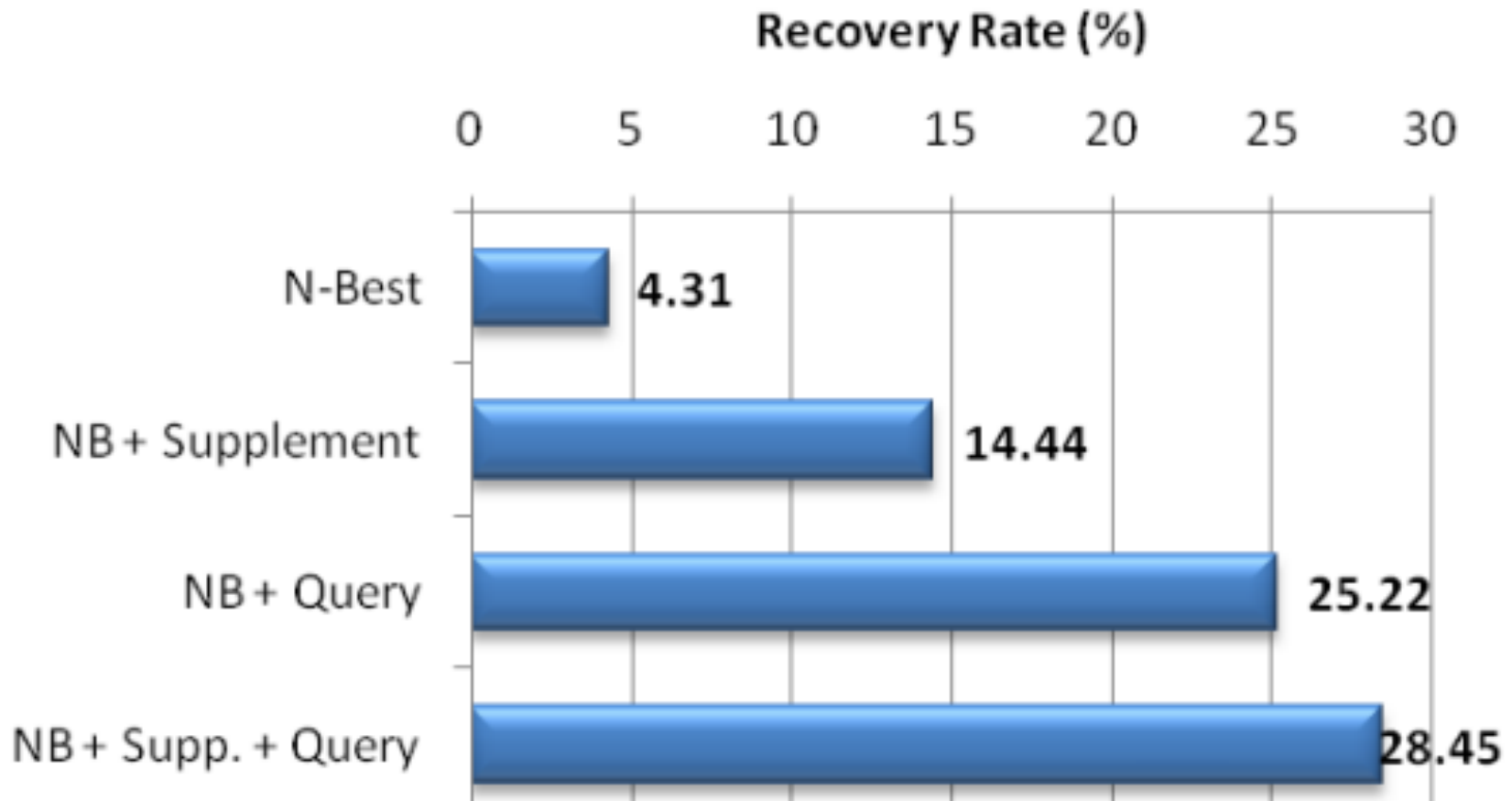
Results

Speak Refine

Evaluation

- Simulation using DB of voice search utterances
- Experiments with cases where *ASR failed*

Recovery Rate with Multimodal Word Palette



Feng et al. 2011.

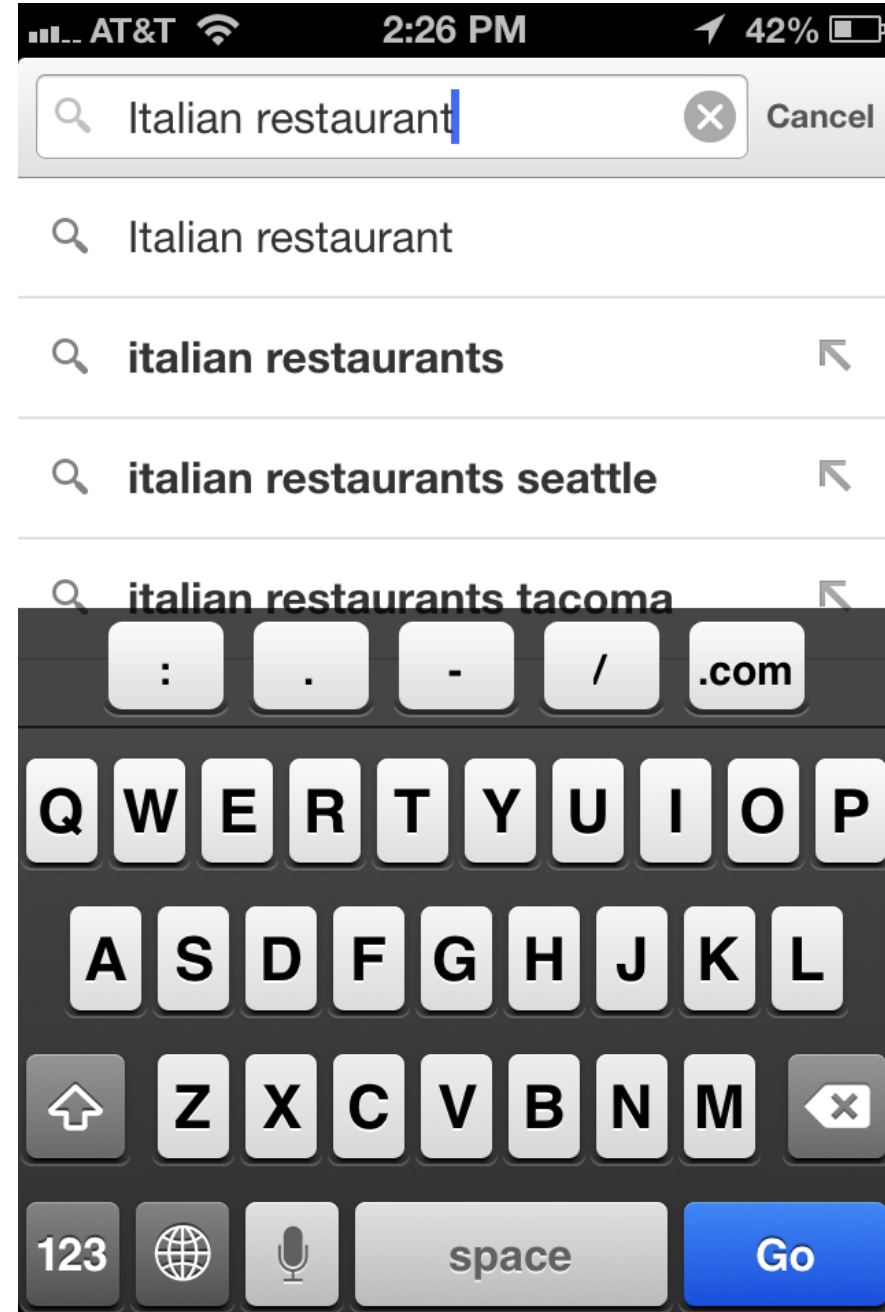
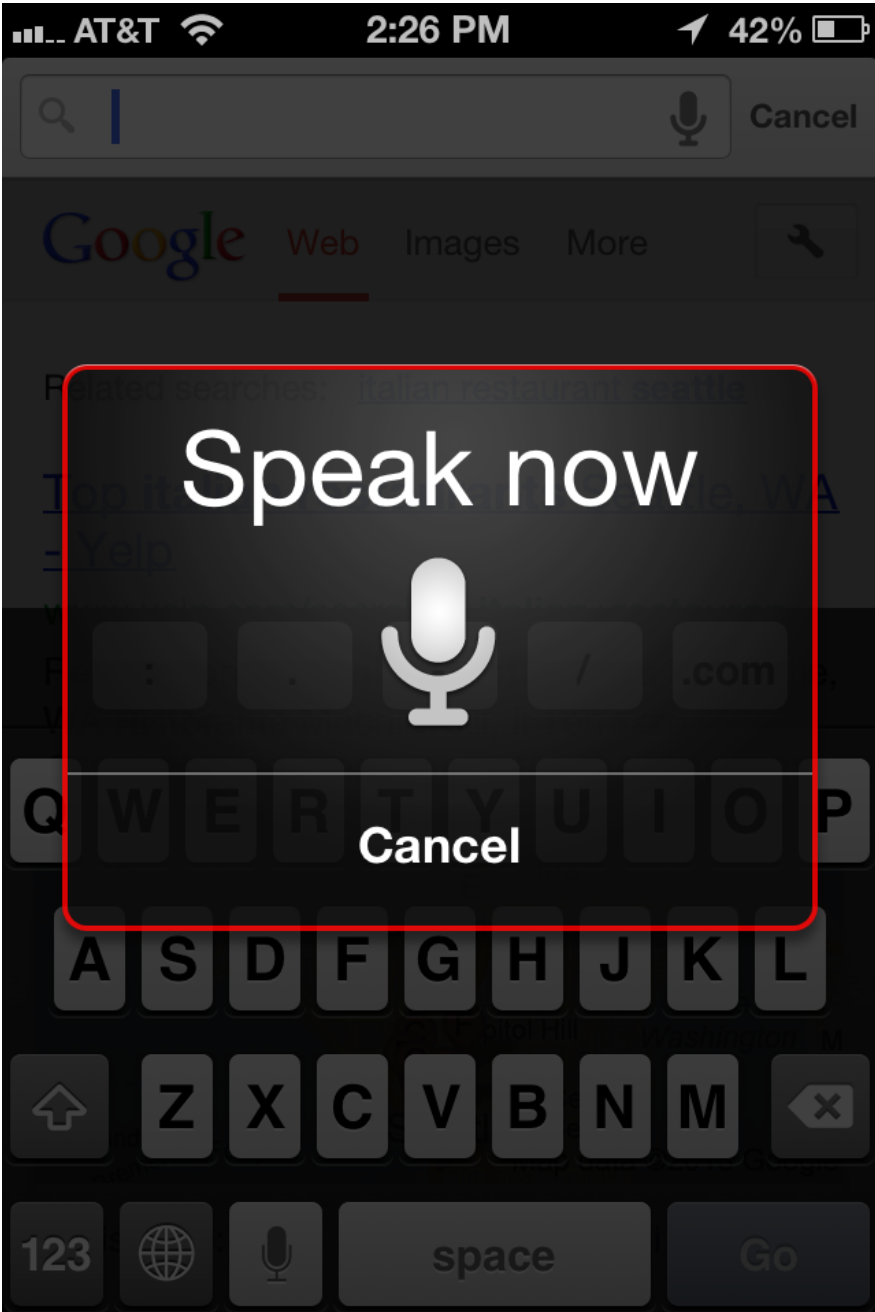
Speech and Multimodal Interaction in Mobile Search

Challenges for Multimodal Interfaces for Voice Search

- Multimodal microphone activation
- Multimodal confirmation and error correction
- Multimodal input and integration

Speak4It





Oviatt, Gent. 1996.

Error Resolution During Multimodal Human-Computer Interaction

Question

How do users use different modalities in multimodal error correction?

Question

How do users use different modalities in multimodal error correction?

Hypothesis:

Users will switch modalities and use different words making corrections. They are more likely to do this as their attempts to correct errors fail.

Study

- **Participants:** 20 native English speakers
- **Apparatus:** car rental, conference registration assistant system.
 - pen + speech input
 - Simulated speech rec w/errors
 - “Spiral depth”: 1 – 6
- **Procedure:** participants completed transactions as accurately as possible.

Prototype system

System asks: where would you like to pick up your car?

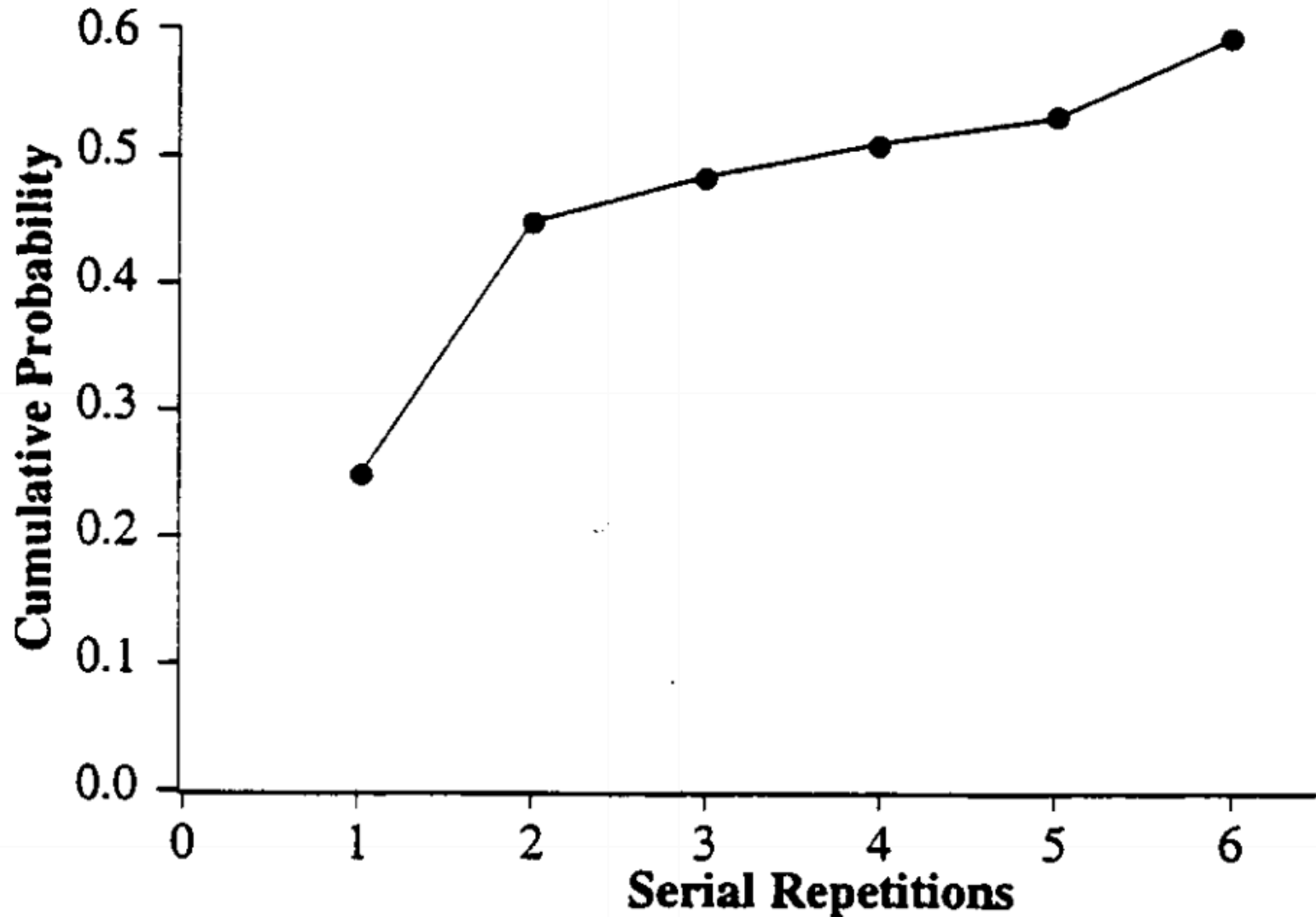
Users speaks: “san francisco”

System displays simulated error: “???”

Results

- **Modality preference** – Speech over written
- **Simultaneous use of modes**
Barely – only 0.7% of all words
- **Modality alternatives**
Increased when people corrected errors
Peaked at spiral depth 5 (40% probability)
- **Lexical alternation**
Probability fluctuated

Probability of spiral depth and lexical alteration



Oviatt. 2012.

Multimodal Interfaces

Goals of multimodal interfaces

- Accommodate
 - User skill & preference
 - Different situations
- Increase efficiency
- Better error handling