# Input Modalities

Kaiyi Zhang

Nov 5, 2015

# Input modalities

- "a modality is the classification of a single independent channel of sensory input/output between a computer and a human"<sup>[1]</sup>
- human-computer and computer-human modalities

#### Input modality

[1] Karray, Fakhreddine; Alemzadeh, Milad; Saleh, Jamil Abou; Arab, Mo Nours (March 2008). "Human-Computer Interaction: Overview on State of the Art" (PDF). International Journal on Smart Sensing and Intelligent Systems 1 (1).

# **Multiple Modalities**

- Equivalence
- Specialization
- Redundancy
- Complimentarity
- Transfer
- Concurrency

#### PixelTone: A Multimodal Interface for Image Editing

Gierad Laput, Mira Dontcheva, Gregg Wilensky, Walter Chang, Aseem Agarwala, Jason Linder, and Eytan Adar

# Learning goals

- Dive into how the combination of speech and direct manipulation benefits visual tasks like photo editing
- Understand the technical ideas of PixelTone
- Discuss the potential drawbacks and limitations

#### PixelTone: A Multimodal Interface for Image Editing

Motivation:

- 1. The language of image processing can be difficult
- 2. Portable devices make complex interaction even more challenging
- "Visual tasks benefit from a combination of speech and direct manipulation interfaces"

#### Demo: PixelTone

#### PixelTone

A Multimodal Interface for Image Editing



#### University of Michigan

Gierad Laput Eytan Adar



Gierad Laput Mira Dontcheva Gregg Wilensky Walter Chang Aseem Agarwala Jason Linder

#### How does a multimodal interface help?

Think about using speech or direct manipulation alone...

- When you know what you want to do but don't know how to achieve it
- When you know how to do but don't know what to do

# The combination of speech and direct manipulation can:

- Permit flexible use of input modes, including alternation and complementary integration of modes (Oviatt & Olsen, 1994; Oviatt, DeAngeli, &Kuhn, 1997).
- Support enhanced error avoidance and ease of error resolution (Oviatt&vanGent, 1996).
- Satisfy higher levels of user preference (Oviatt, Cohen, &Wang, 1994; Oviatt& Olsen, 1994).

# Discussion

Can you think of some tasks other than photo editing that can benefit from a combination of speech and direct manipulation? How can they benefit? (groups of 2-3)



# Activity: pilot experiment

In groups of 2-3, direct your partners to transform image A into image B (B1 to B6). Take turns to play the role of "director". Write down your commands.

# PixelTone

- Speech recognition
- Interpreter
- Execution engine



# Two-tiered speech recognition

"make the shadows on the left slightly brighter"

- Local speech recognition
- Remote speech recognition

# Interpreter

"make the shadows on the left slightly brighter"

- Two-level tag hierarchy for parsing phrases
- phrase level Verb, Noun, and Adjective

make  $\rightarrow$  VX | the shadows on the left  $\rightarrow$  NX | slightly brighter  $\rightarrow$  AX

*slightly*  $\rightarrow$  Adverb (VB) | *brighter*  $\rightarrow$  Adjective (JJR)

# Penn Treebank tags

- 1. CC Coordinating conjunction
- 2. CD Cardinal number
- 3. DT Determiner
- 4. EX Existential there
- 5. FW Foreign word
- 6. IN Preposition or subordinating conjunction
- 7. JJ Adjective
- 8. JJR Adjective, comparative
- 9. JJS Adjective, superlative
- 10.LS List item marker
- 11.MD Modal
- 12.NN Noun, singular or mass
- 13.NNS Noun, plural
- 14.NNP Proper noun, singular
- 15.NNPSProper noun, plural
- 16.PDT Predeterminer
- 17.POS Possessive ending
- 18.PRP Personal pronoun

- 19.PRP\$ Possessive pronoun
- 20. RB Adverb
- 21. RBR Adverb, comparative
- 22. RBS Adverb, superlative
- 23. RP Particle
- 24. SYM Symbol
- 25. TO to
- 26. UH Interjection
- 27. VB Verb, base form
- 28. VBD Verb, past tense
- 29. VBG Verb, gerund or present participle
- 30. VBN Verb, past participle
- 31. VBP Verb, non-3rd person singular present
- 32. VBZ Verb, 3rd person singular present
- 33. WDT Wh-determiner
- 34. WP Wh-pronoun
- 35. WP\$ Possessive wh-pronoun
- 36. WRB Wh-adverb

# Interpreter

"make the shadows on the left slightly brighter"MNSNNVXNXNXAX

Image processing request

*Image operation: "brighter"—>* BRIGHTEN

*Mask:* "shadows" "left" —> SHADOW & LEFT

*Parameters:* "slightly" —> SLIGHT

# Discussion

Check out the commands you just wrote on the paper. Discuss within your same group:

What are the operations, masks and parameters in your command? Define the rules for your *phrase templates* so that your command can be mapped into components of an image processing request.

# Execution engine

- Process command
- Combine with direct manipulation
- Localize
- Blend multiple masks



#### **Evaluation results**

Quantitative Results :

Success rate, users preference, number and complexity of utterances, speech engine accuracy

Qualitative Results :

Speech interface, gallery mode, direct manipulation Non-native English speakers with accents

### From your commentaries

"Implications have been made using only 14 participants in the study, as the number is far too low to generalize. Another problem is that they mention that 13 out of the 14 participants preferred the multimodal interaction but they never mention as to why the 14th participant did not prefer it. Was he a non-native English speaker? Did he find the software too overwhelming?" — Pallavi Agarwal

# From your commentaries

"As the paper noted, there wasn't a significant difference in effectiveness between the single mode interface and the multimodal interface. Is this something that industry would invest in mass-producing? In my opinion, I don't think we're quite at that state yet." — Jesse Qin

"Ultimately where the word falls short (and this is acknowledged by the authors) is in the failure for them to make the language of image editing less ambiguous." — Steven Rick

#### Discussion

In groups of 2-3, discuss other criteria (quantitative & qualitative) that were not been evaluated but you believe are important to address in the evaluation phase.

# Gestural Interfaces: A Step Backward In Usability

Donald A. Norman, Jakob Nielsen

# Learning goals

- Why are we having trouble in usability?
- Understand how gestural interfaces violate interaction design principles
- Discuss how to solve these problems

### Gestural Interfaces: A Step Backward In Usability

Why is it a step backward?

# Gestural Interfaces: A Step Backward In Usability

Principles of interaction design:

Visibility

Feedback

Consistency

Discoverability

Scalability

Reliability

Non-destructive operations





1

**<** Inbox (687)

 $\sim \sim$ 

Executive Vice Chancellor - Acad... To: All Academics Staff and Stu... more...

Warren College Provost November 4, 2015 at 2:08 PM

UC SAN DIEGO CAMPUS NOTICE University of California, San Diego

OFFICE OF THE EXECUTIVE VICE CHANCELLOR -

ACADEMIC AFFAIRS



# Gestural Interfaces: A Step Backward In Usability

Principles of interaction design:

Visibility

Feedback

Consistency

Discoverability

Scalability

Reliability

Non-destructive operations

# Discussion

 How well does PixelTone address these principles of interaction design?

Visibility (signifiers), feedback, consistency, nondestructive operations (undo), discoverability, reliability.

![](_page_31_Picture_0.jpeg)

![](_page_32_Picture_0.jpeg)

#### The promise of gestural interfaces

- A pleasure to use and to see
- New displays promise to revolutionize media
- Add a welcome feeling of activity

#### Discussion

Instead of going back to the previous page within the application, the permanently visible back button on Android phones move the user out the of the application, which is beyond users' expectation. As a designer, how would you solve this problem?