Input Modalities

Wei Dai

Input Modalities

“In the context of human–computer interaction, a modality is the classification of **a single independent channel** of sensory input/output between a computer and a human”

- Wikipedia

Hand-based

- Keyboard
- Touch screen
- Mouse
- Joystick

Speech

- Speech-based assistive technologies
- Siri, Cortana, etc.
- Speech recognition in smartphone keyboards

Question:
What are other input modalities?

Put That There (1979)

PixelTone: A Multimodal Interface for Image Editing (2013)

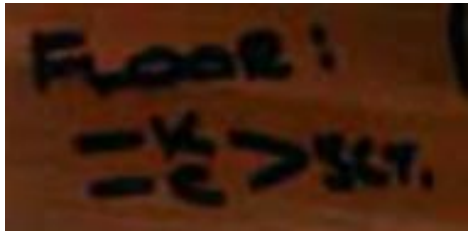
PixelTone

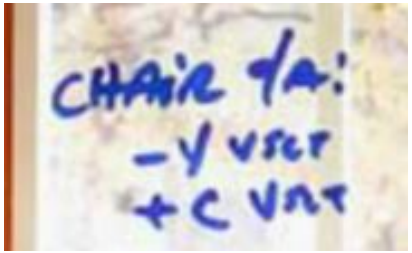
Image editing:

- Action: What filter / transformation do I want to apply?
- Action parameters: how “much” should a transformation be applied?
- Object: Where in the image do I want to apply this?

“Change the color of Sarah’s shirt”

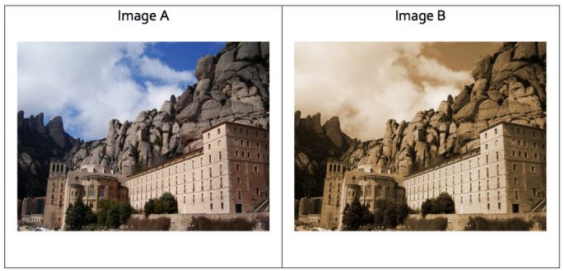
Natural Language In Photo Editing

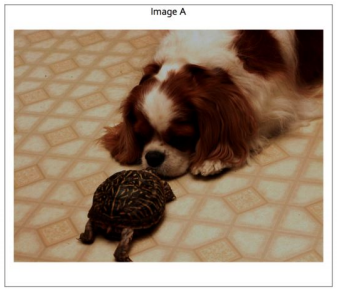




Study 1: Novice Editing Behavior

- Amazon Mechanical Turk
- 10 individual images (how to improve the image)
- 14 image pairs (how to transform image from A to B)
- 10 responses for each task (240 responses overall)
- 211 valid responses from 35 unique Turkers were collected



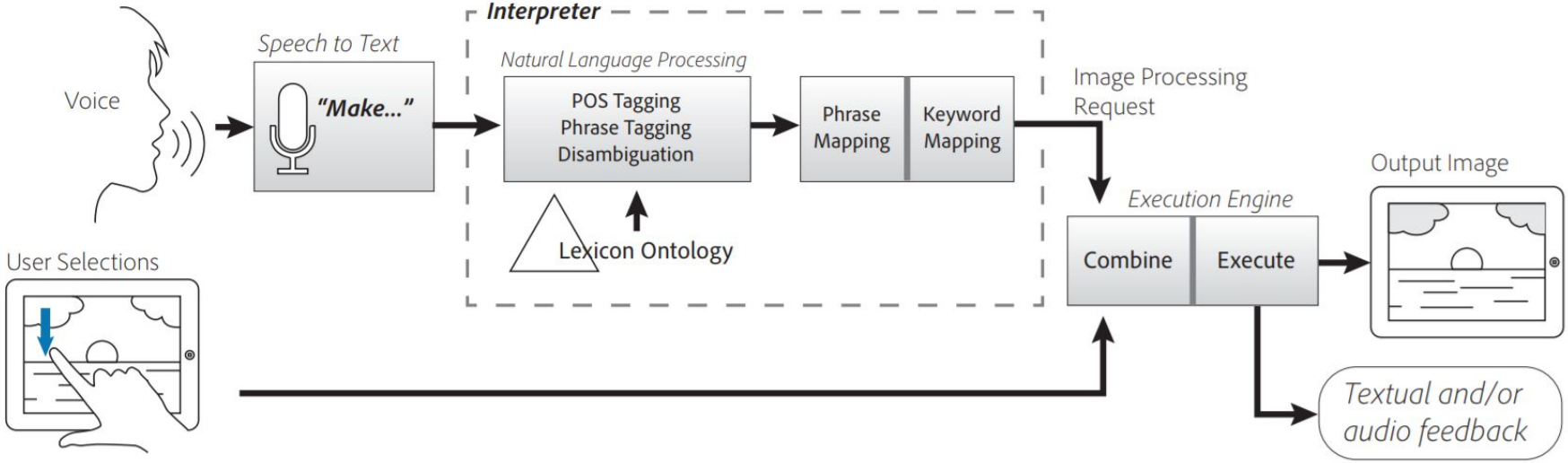


Key finding: “while users had a common language for describing changes, discoverability of the lexicon terms was difficult without guidance.”

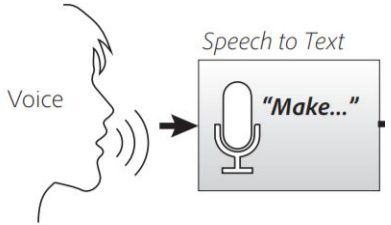
Both **imperative** and **declarative** phrases were used.

Imperative: “Make this brighter.”
Declarative: “This is too dark.”

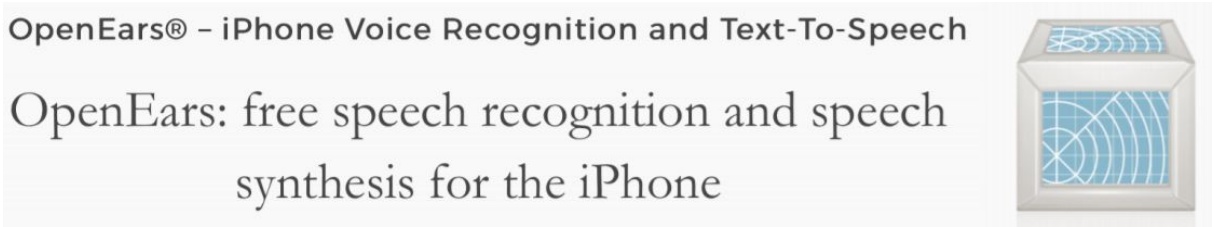
Architecture of PixelTone



Speech Recognition

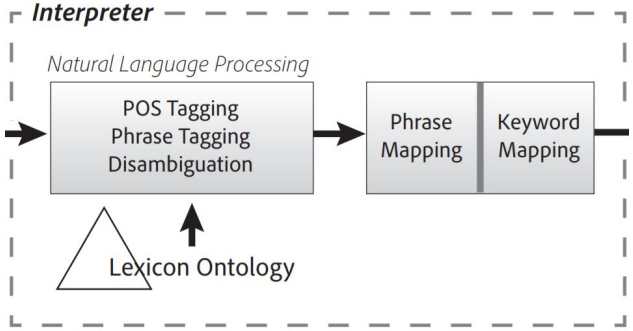


- Local recognition



- Remote recognition

Speech Interpretation



"Make the shadows on the left slightly brighter"

VX	NX	AX
	NNS NN	VB JJ

Word level: Penn Treebank tags

"slightly" - Adverb (VB), "brighter" - Adjective (JJ)

Phrase level: Verb (VX), Noun (NX), Adjective (AX)

"Make" - VX, "the shadows on the left" - NX, "slightly brighter" - AX

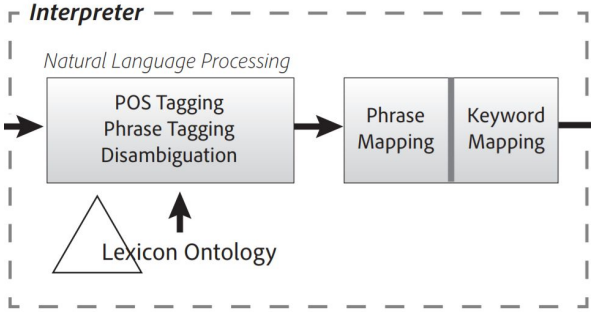
1. CC	Coordinating conjunction	19. PRP$	Possessive pronoun
2. CD	Cardinal number	20. RB	Adverb
3. DT	Determiner	21. RBR	Adverb, comparative
4. EX	Existential *there*	22. RBS	Adverb, superlative
5. FW	Foreign word	23. RP	Particle
6. IN	Preposition or subordinating conjunction	24. SYM	Symbol
7. JJ	Adjective	25. TO	*to*
8. JJR	Adjective, comparative	26. UH	Interjection
9. JJS	Adjective, superlative	27. VB	Verb, base form
10. LS	List item marker	28. VBD	Verb, past tense
11. MD	Modal	29. VBG	Verb, gerund or present participle
12. NN	Noun, singular or mass	30. VBN	Verb, past participle
13. NNS	Noun, plural	31. VBP	Verb, non-3rd person singular present
14. NNP	Proper noun, singular	32. VBZ	Verb, 3rd person singular present
15. NNPS	Proper noun, plural	33. WDT	Wh-determiner
16. PDT	Predeterminer	34. WP	Wh-pronoun
17. POS	Possessive ending	35. WP$	Possessive wh-pronoun
18. PRP	Personal pronoun	36. WRB	Wh-adverb

Phrase Mapping

“Make the shadows on the left slightly brighter”

NNS NN	VB JJ
VX	NX AX

Image operation: “brighter” -> BRIGHTEN
Mask: “shadows” and “left” -> SHADOW & LEFT
Parameters: “slightly” -> SLIGHT



Keyword Mapping

If parsing fails:
Keyword scan in bag-of-word model

“Left shadow brighten”

{“left”, “shadow”, “brighten”}

Available operations


Original


Exposure


Auto-Color


Brighten


Darken


Black & White


Posterize


Soft Focus


Contrast


Hue


Vibrance


Saturation


Blur


Vignette


Sepia


Warmth


Coolness


Green Tint


Magenta Tint


Sharpen

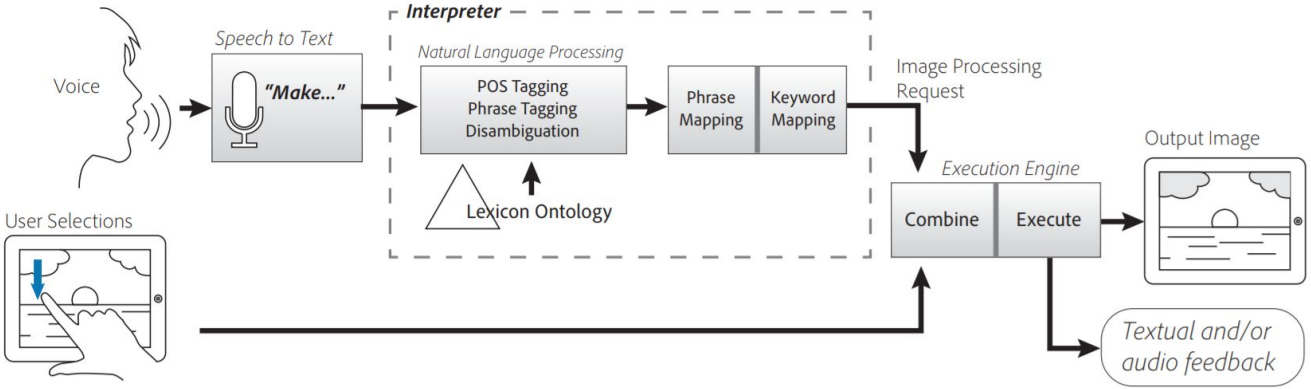

Pixellate


Vintage


Lomo

PixelTone


a
b
c
"This is a shirt"
"Change the color of the shirt"


Interpreter
Voice
Speech to Text
"Make..."
Natural Language Processing
POS Tagging
Phrase Tagging
Disambiguation
Lexicon Ontology
Phrase Mapping
Keyword Mapping
Image Processing Request
Execution Engine
Combine
Execute
Output Image
User Selections
Textual and/or audio feedback

On PixelTone

“Adding speech modality to photo editing has many benefits. One is that it **reduces task switching** by allowing the user to stay focused on the part of the photo they're working on while issuing verbal commands to the AI to apply certain filters.” - Heitor Schueroff

Improving PixalTone

“One way to improve this technique is to increase the image understanding capability. Many recent works on computer vision like image recognition can greatly improve the techniques. This will help with the annotating phase, users do not need to clearly specify the region of the graphs. More advanced techniques can apply learning capability to automatically adjust the image to be ``good'' enough for medium level users, like google photos” -Bingyu Shen

Discussion

- How would you improve PixelTone by utilizing more of speech / auditory modality?

PixelTone: User Study

Experiment setup

- 14 users
- 8 tasks (transformation)+ 8 tasks (improvement)
- Random assignment to PixelTone / PixelTone w/o Speech
- Difficulty gradually increased
- Each task is scored (1-5)
- User evaluates interface at the end (1-5)

Discussion:
How would you improve this user study?

Findings:

Success rate for both interfaces were identical

- Task score: 4.37 SD=0.31 vs. 4.32 SD=0.45 between multi-model vs. non-speech

(Discussion: how?)

Users preferred the multimodal interface.

- User rating: 4.36, SD=0.50 vs 3.64, SD=0.63 multi-modal vs non-speech

Speech Recognition Statistics

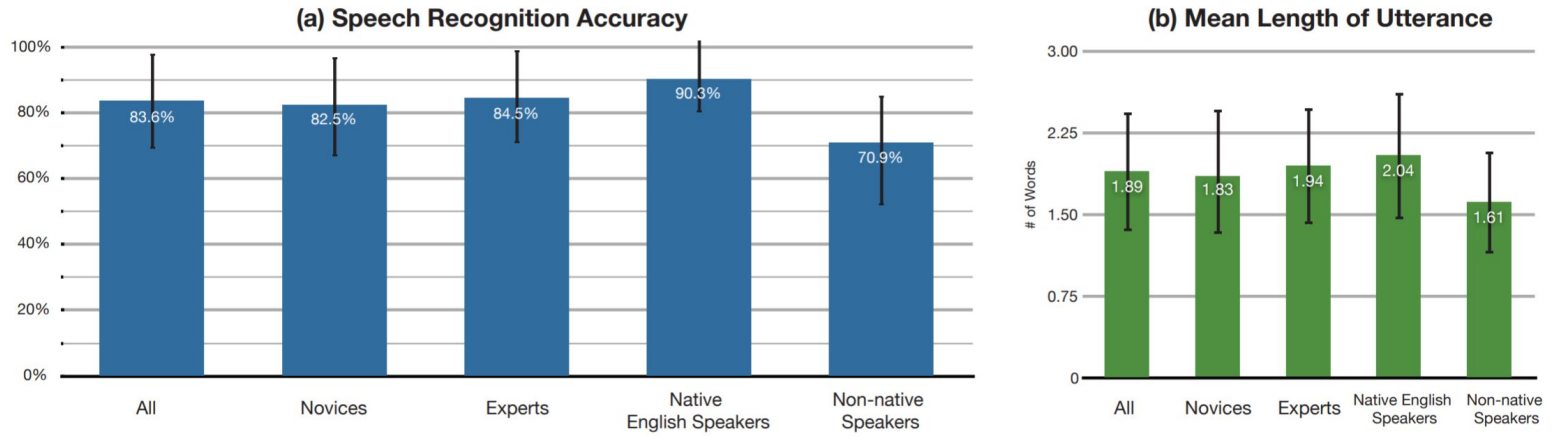


Figure 5. Speech recognition accuracy (a) and mean word length of utterance (b) from a total of 386 utterances across 14 users.

User Study: Qualitative Findings

1. Users use the speech interface when they have a good idea of what they want to do.
2. Users use the gallery mode when they want to explore options and compare different effects.
3. Users use direct manipulation to fine-tune and explore.
4. Non-native English speakers with accents used speech interaction much less.

Discussion

- What other application would or would not benefit from multimodal interfaces? Why?

Gestural Interfaces: A Step Backward In Usability

“When users think they did one thing but actually did something else, they lose their sense of controlling the system because they don’t understand the connection between actions and results.”

Good features

- Visibility
- Feedback
- Consistency (aka standards)
- Non-destructive operations (e.g. Undo)
- Discoverability (All operations can be discovered by systematic exploration of menus.)
- Scalability (The operation should work on all screen sizes, small and large.)
- Reliability (Operations should work. Period. And events should not happen randomly.)

Question:
What are the problems that you encounter in user interfaces today?
Which principle was not satisfied?

Disccusion

Should we define a standard set of gestures for touch-based interfaces? Why or why not?

Gestural Interfaces: A Step Backward In Usability

- Gestural interfaces are inconsistent, undiscoverable, etc.
- No established conventions
- Companies ignorance of findings in HCI literature

iOS

Android

Mac

Windows

Gestural Interfaces: A Step Backward In Usability

Donald A. Norman Jakob Nielsen

2010

- 1st gen iPad
- iOS 4.0
- Android 2.0

Discussion: How has things changed in the past 9 years?