Comparing Speech and Keyboard Text Entry for Short Messages on Touchscreen Phones

Sherry Ruan[1], Jacob O. Wobbrock[2], Kenny Liou[3], Andrew Ng[13], James A. Landay[1]

[1] Stanford University

[2] Information School, University of Washington

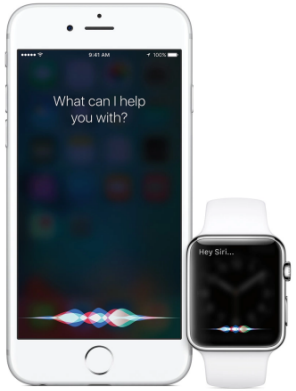
[3] Baidu Research

Background and Motivation

Background

- Speech UIs have not enjoyed widespread use despite decades of research
- Recent advances in speech recognition due to deep learning
- Popular speech-based assistants have started to attract interest

Baidu Duer



Apple Siri

Amazon Echo

Google Home

Motivation

- Previous ineffective speech-based off-desktop text entry methods
- Today’s speech recognition systems have the potential to be suitable for general-purpose text entry
- Unknown how state-of-the-art speech-based dictation & miniature touch screen keyboards compare

Related work was mostly dated

- The speed of speech input was far inferior to keyboard input

  13.6 WPM vs. 32.5 WPM (Karat et al. 1999)

- The accuracy of speech input was far inferior to keyboard input

  33-44% speech error rates (Price et al. 2004)

- Hardly any research results comparing the performance of Mandarin speech and typing input methods

Contribution

Contribution

- First evaluation of a *state-of-the-art* deep learning-based speech recognition system against a *state-of-the-art* touch-based keyboard for mobile text entry
- Made for two languages
- Novel speech-based measures
- Insights for improving interaction design for speech-based text entry

Experiment Design

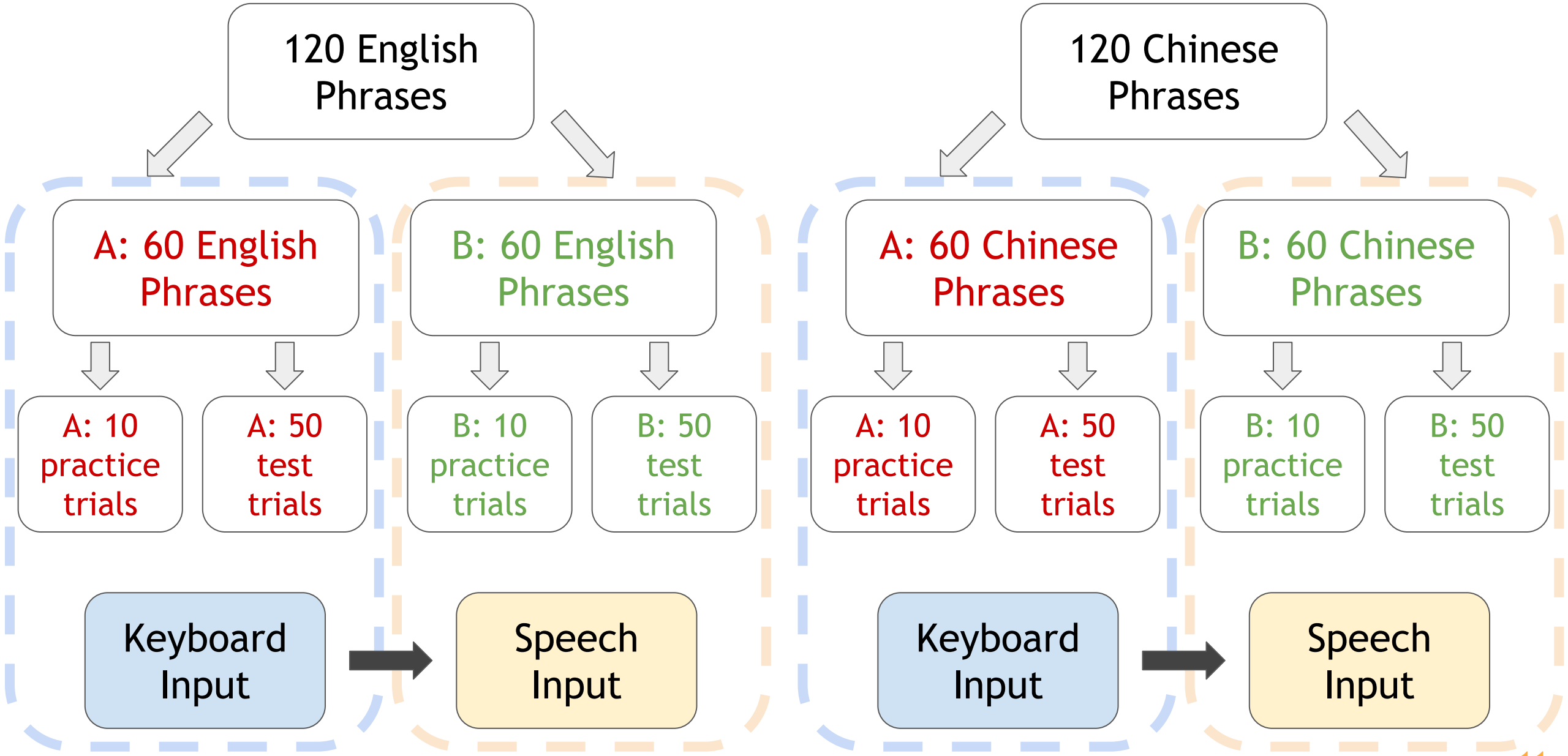
2×2 mixed factorial design

Factors	Options	Within or Between
Input Method	Keyboard	Within-Subjects
	Speech	
Language	English	Between-Subjects
	Mandarin Chinese	

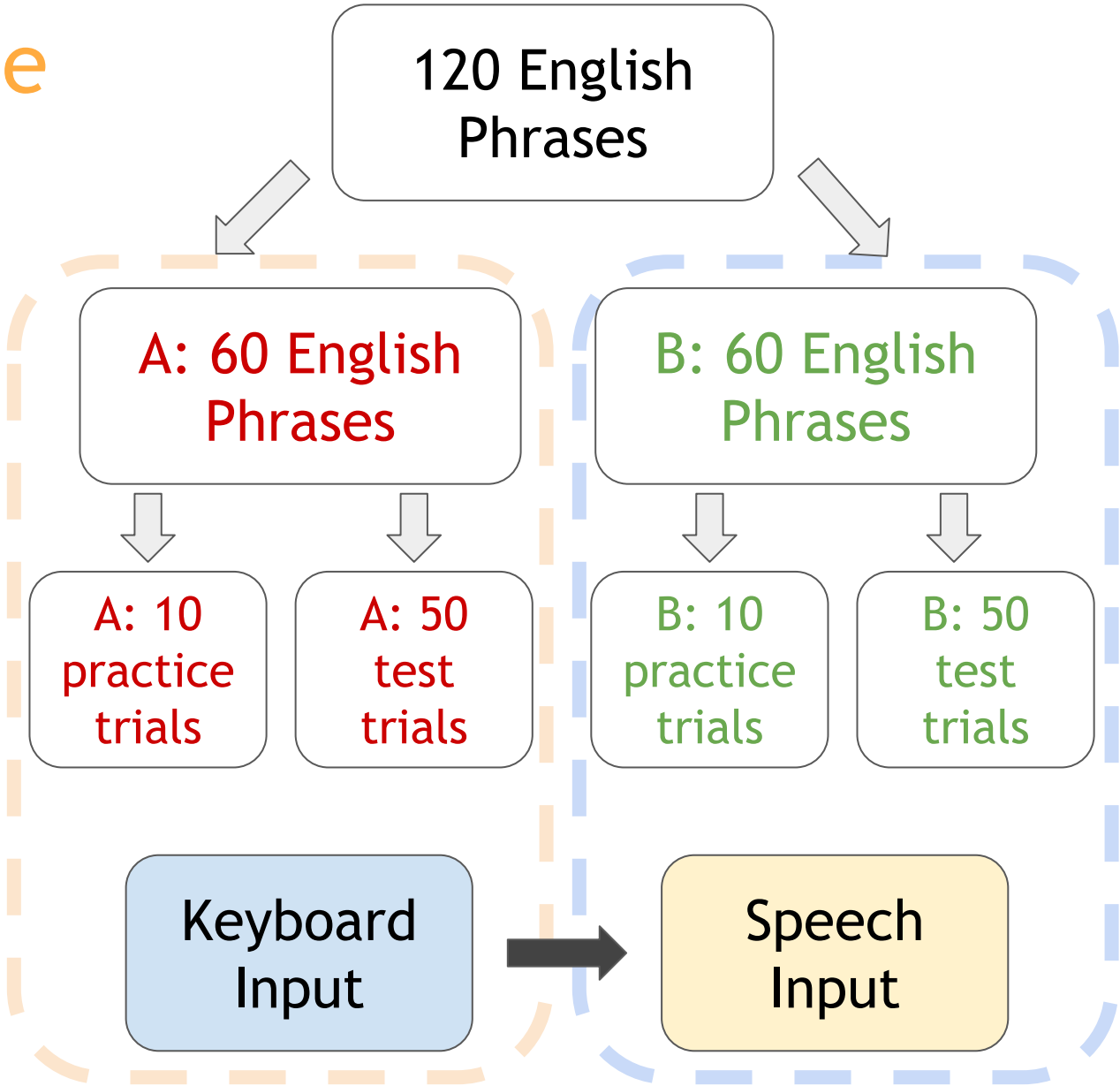
Procedure

- Each participant entered text only in their native language, **English or Mandarin Chinese**, with each text entry method (**keyboard or speech**)
- The order of which input method to use first was *counterbalanced*
- Participants transcribed 60 phrases drawn from one of the two phrase sets (**A or B**)
- A or B was assigned in each session period in *alternating* order
- 10 practice trials before the test, which consisted of 50 testing trials

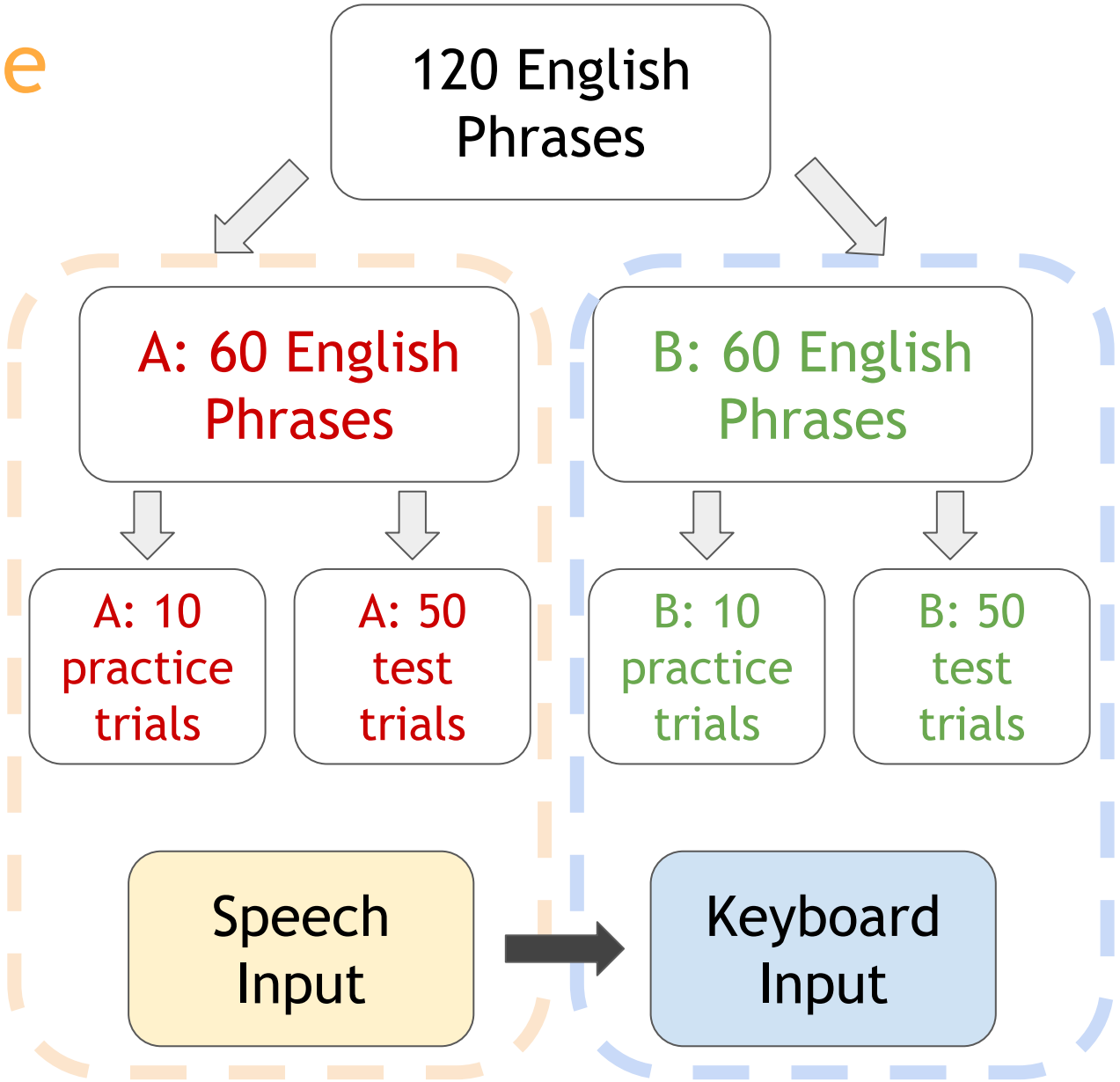
Procedure


120 English Phrases
A: 60 English Phrases
B: 60 English Phrases
A: 10 practice trials
A: 50 test trials
B: 10 practice trials
B: 50 test trials
Keyboard Input
Speech Input
120 Chinese Phrases
A: 60 Chinese Phrases
B: 60 Chinese Phrases
A: 10 practice trials
A: 50 test trials
B: 10 practice trials
B: 50 test trials
Keyboard Input
Speech Input

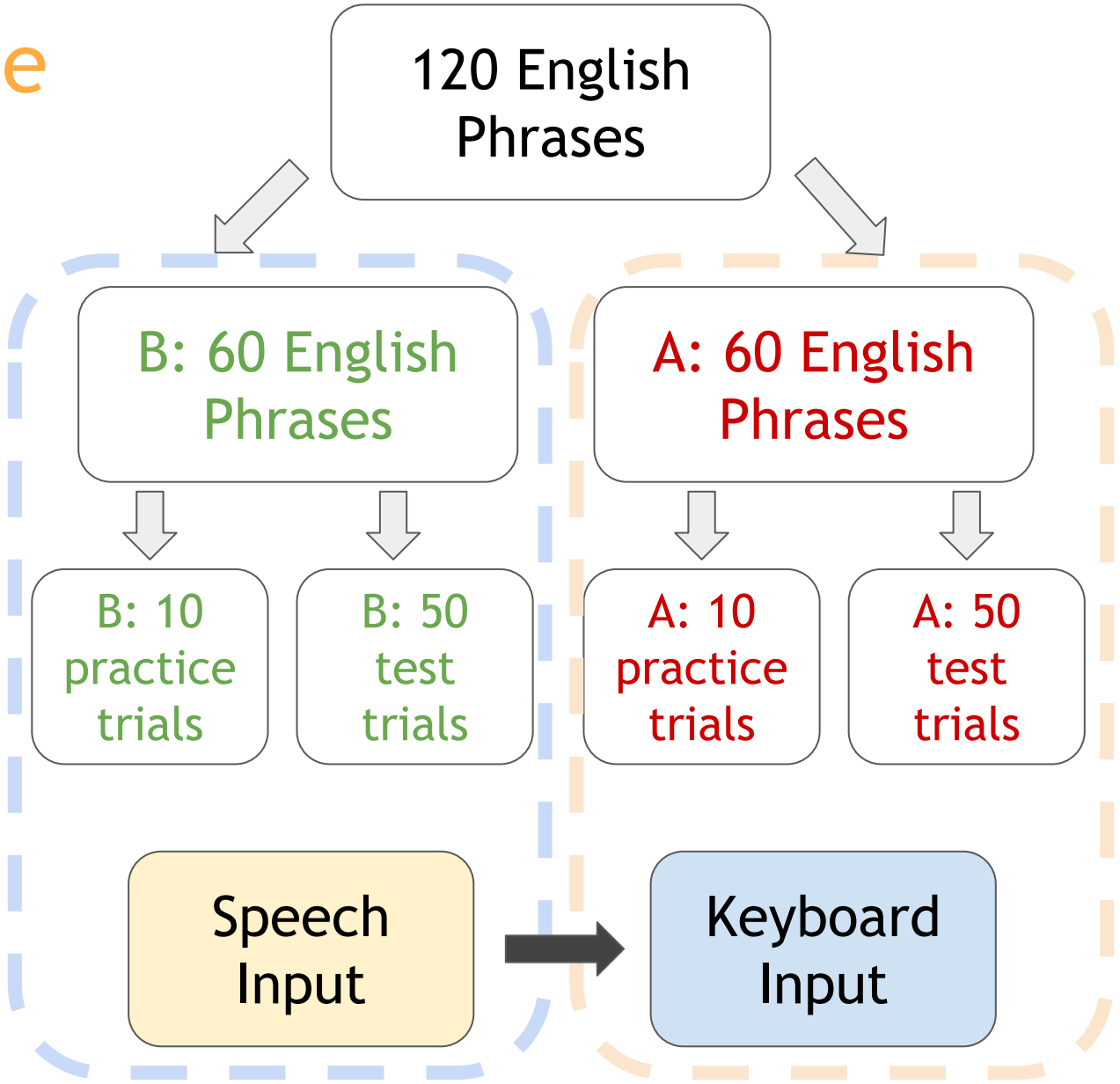
Counterbalance



Counterbalance

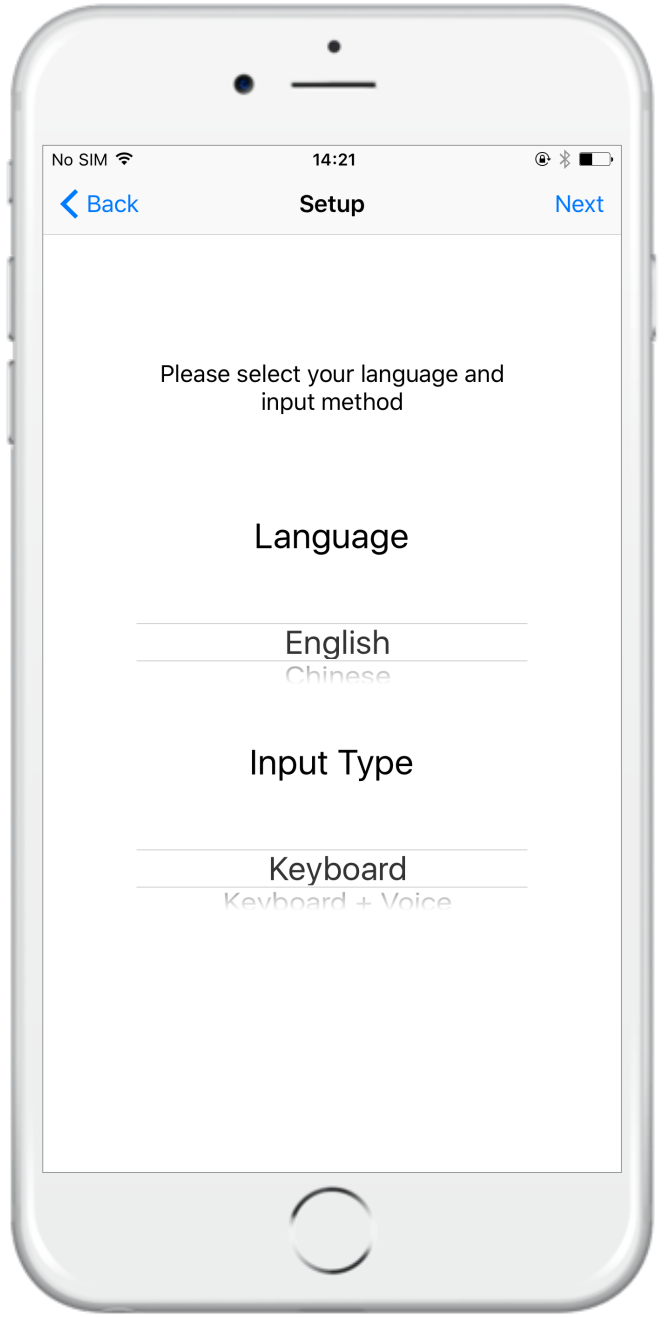


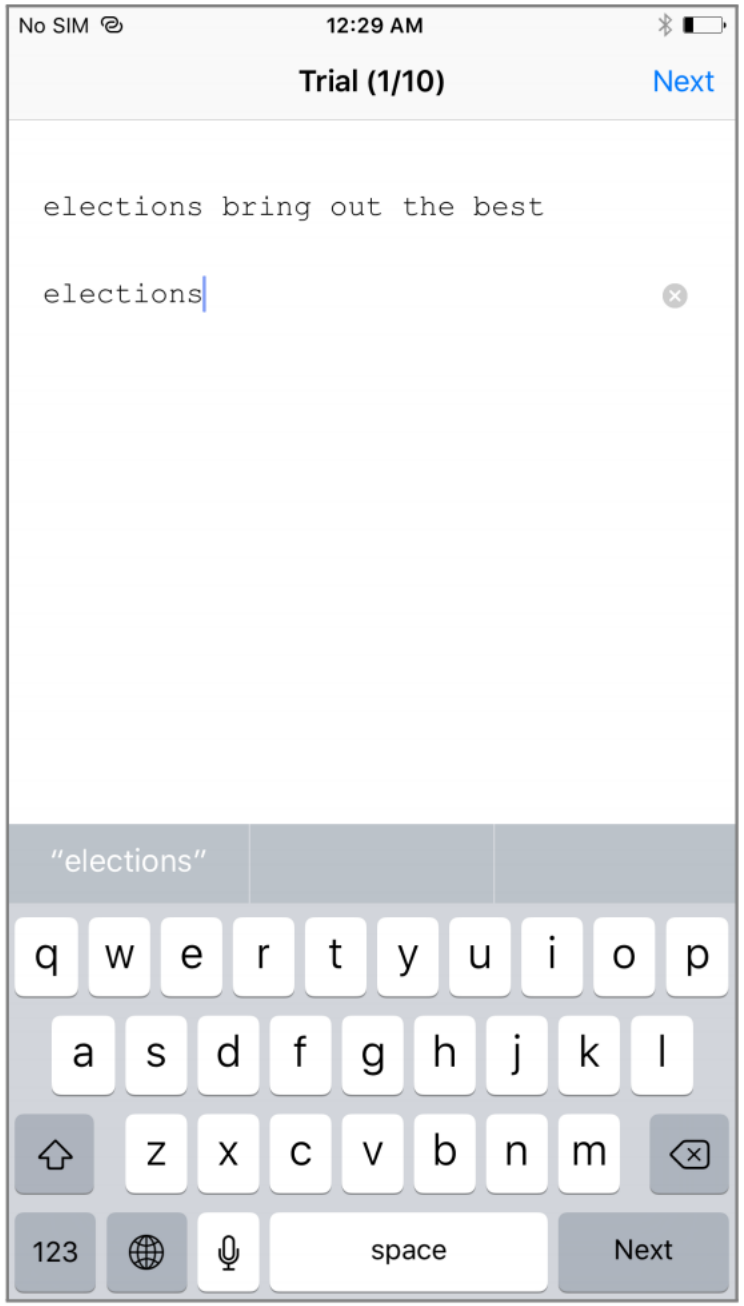
Counterbalance



Apparatus

- Developed a custom experiment test-bed app using Swift 2 and Xcode 7
- The app was connected to Baidu's speech recognition system: Deep Speech 2
- Transcription tasks with two input interfaces: keyboard or speech
- All participants used the same iPhone 6 Plus



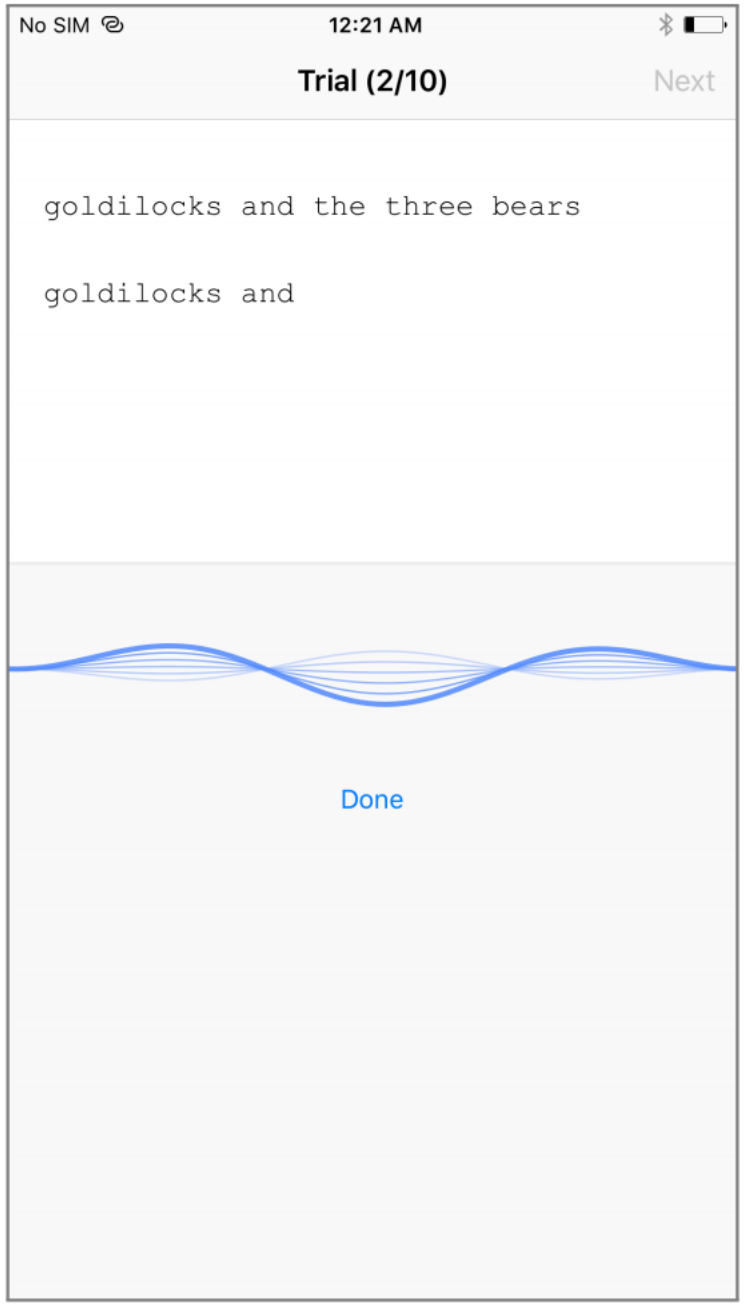


(Left)

Keyboard input interface with English QWERTY keyboard

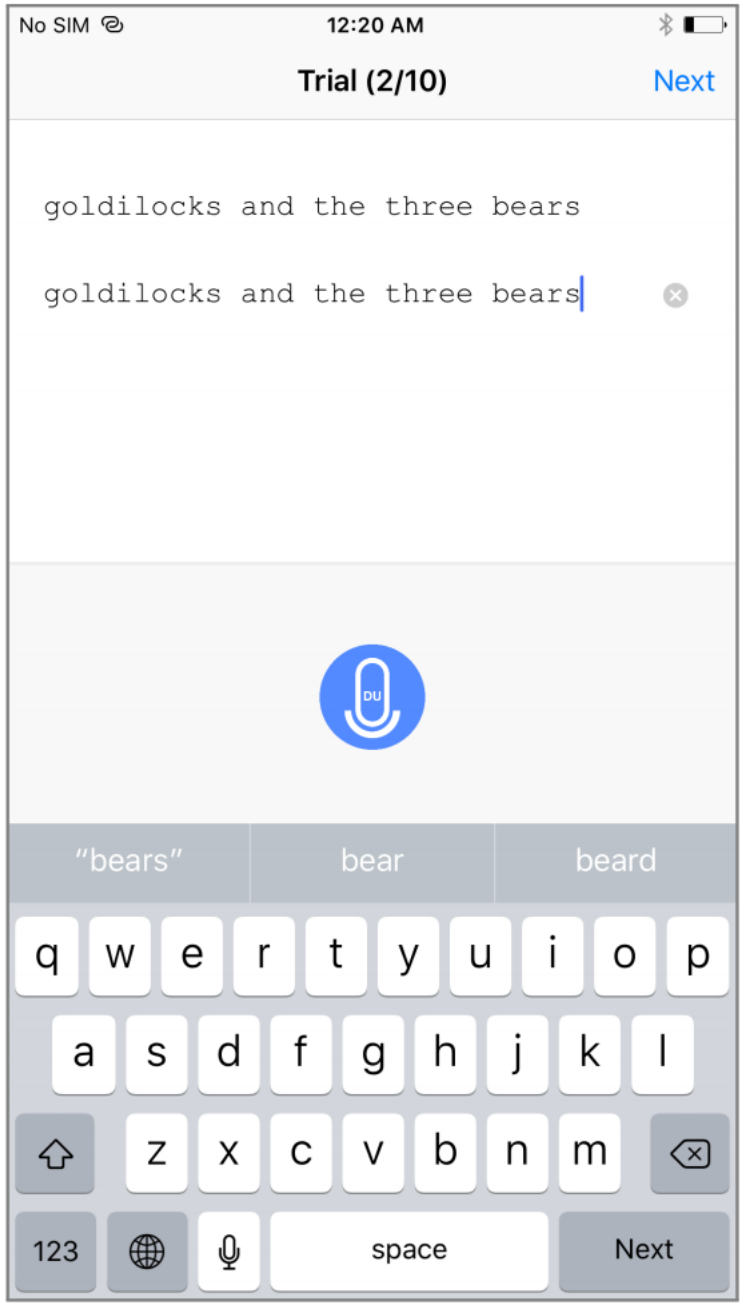
(Right)

Keyboard input interface with pinyin QWERTY keyboard



(Left)

English speech input interface: user is speaking



(Right)

English speech input interface: user finishes speaking

Enabled all the standard iOS text entry features

	English Keyboard	Pinyin Keyboard	Speech
Layout	Qwerty	Qwerty	NA
Spell Check	Yes	Always outputs valid Mandarin characters	Already eliminates all invalid words
Auto Correct	Yes		
Word Completion	Yes	Yes	NA

Text entry phrase set

- From a standard text entry phrase set (MacKenzie & Soukoreff 2003)
- Representative of everyday English / Mandarin including proper nouns
- Excluded punctuation marks and capital letters except for “I”

Language	Phrase Length	Mean	Stdev
English	16 - 37 chars	26.8	4.3
Chinese	3 - 14 chars	7.7	2.2

Text entry phrase set

- From a standard text entry phrase set (MacKenzie & Soukoreff 2003)
- Representative of everyday English / Mandarin including proper nouns
- Excluded punctuation marks and capital letters except for “I”

the dow jones index has risen
fish are jumping
goldilocks and the three bears
mystery of the lost lagoon

Participants

- A total of 48 university students
- Gender balanced
- Age ranging from 19 to 32 years old
- Familiar with an English Qwerty (or a Mandarin Pinyin Qwerty) keyboard

	American English	Mandarin Chinese
Female	12	12
Male	12	12

Results

Entry rate formula

$$Words\ per\ Minute = \frac{|T| - 1}{t} \times 60 \times \frac{1}{L}$$

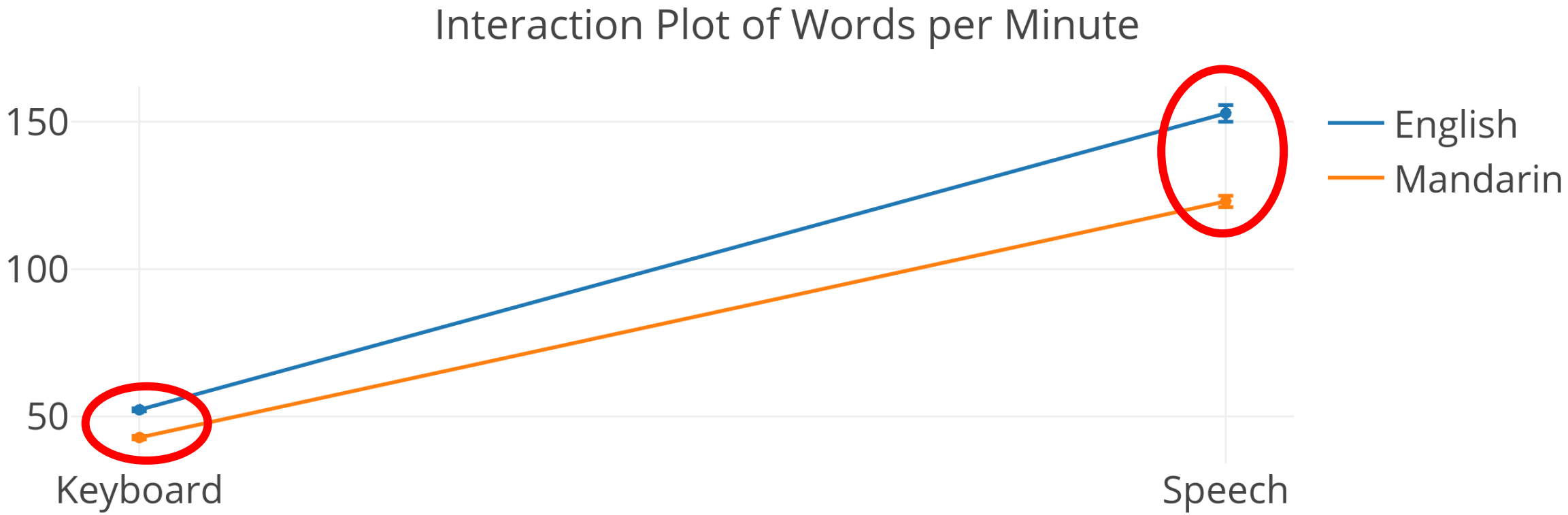
T: the transcribed string

t: time in seconds

L: the average word length in characters

(5 for English and 1.5 for Mandarin Chinese)

Speech was 2.9x faster than typing for both English and Mandarin Chinese



Higher is faster (better). Error bars represent +/-1 standard error

Classify characters in the input stream

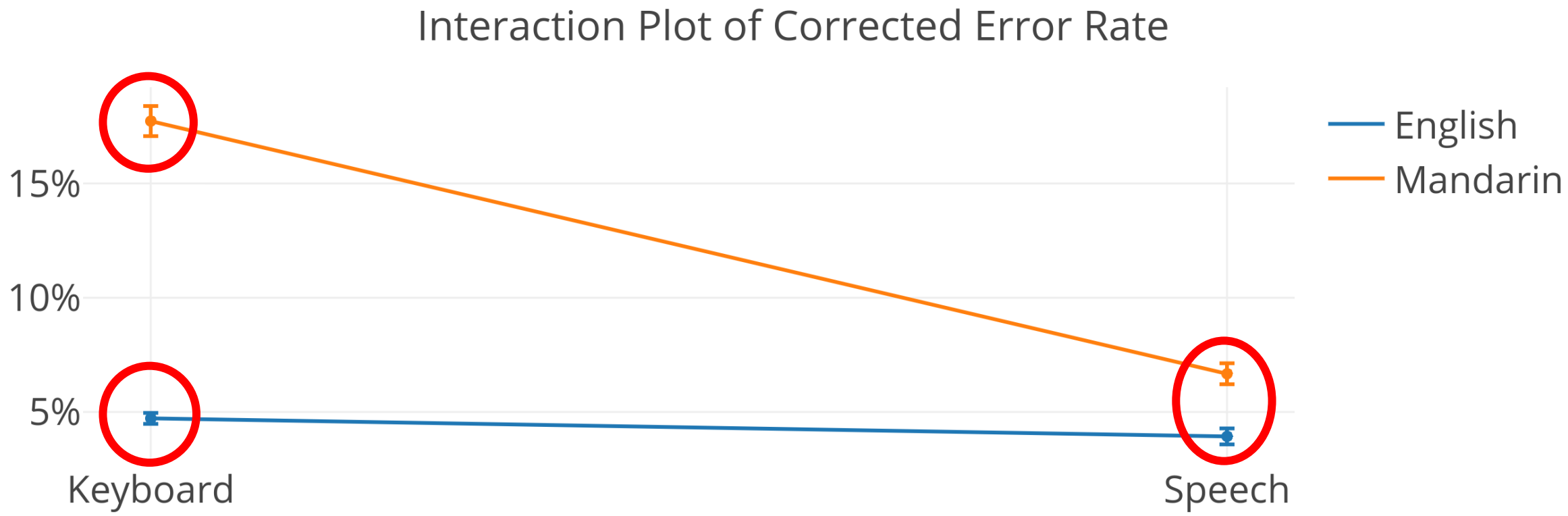
- Correct (C)
- Incorrect-not-fixed (INF)
- Incorrect-fixed (IF)
- Fixes (F)

$$\textit{Uncorrected Error Rate} = \frac{INF}{C+INF+F}$$

$$\textit{Corrected Error Rate} = \frac{IF}{C+INF+F}$$

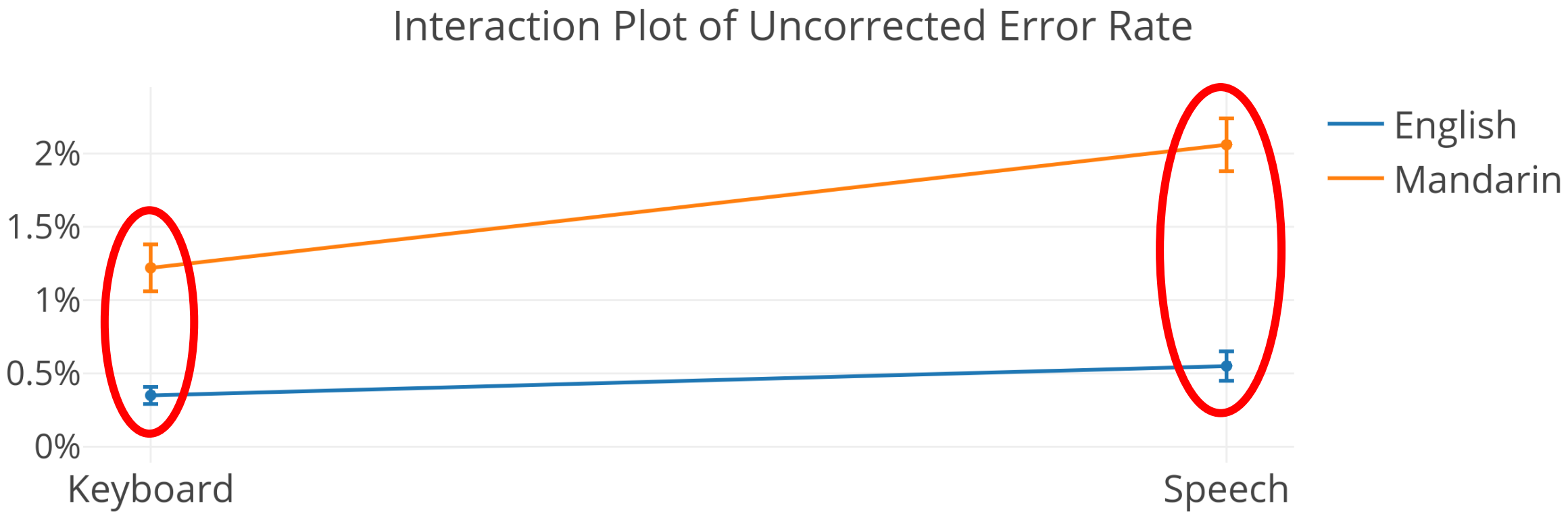
(Soukoreff & MacKenzie 2003)

Speech was much more accurate than keyboard for errors made and fixed during entry



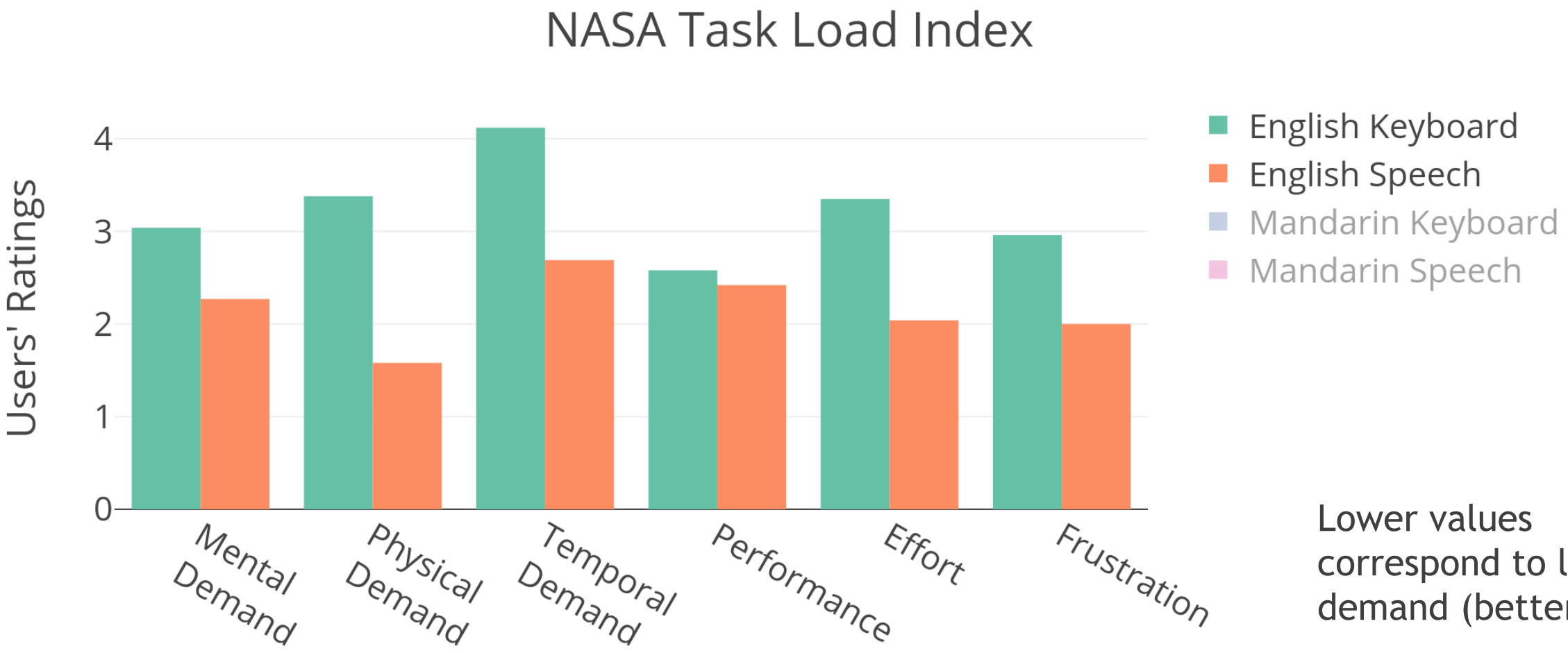
Lower is more accurate (better). Error bars represent +/-1 standard error

Speech left slightly more errors in the transcribed text than the keyboard did



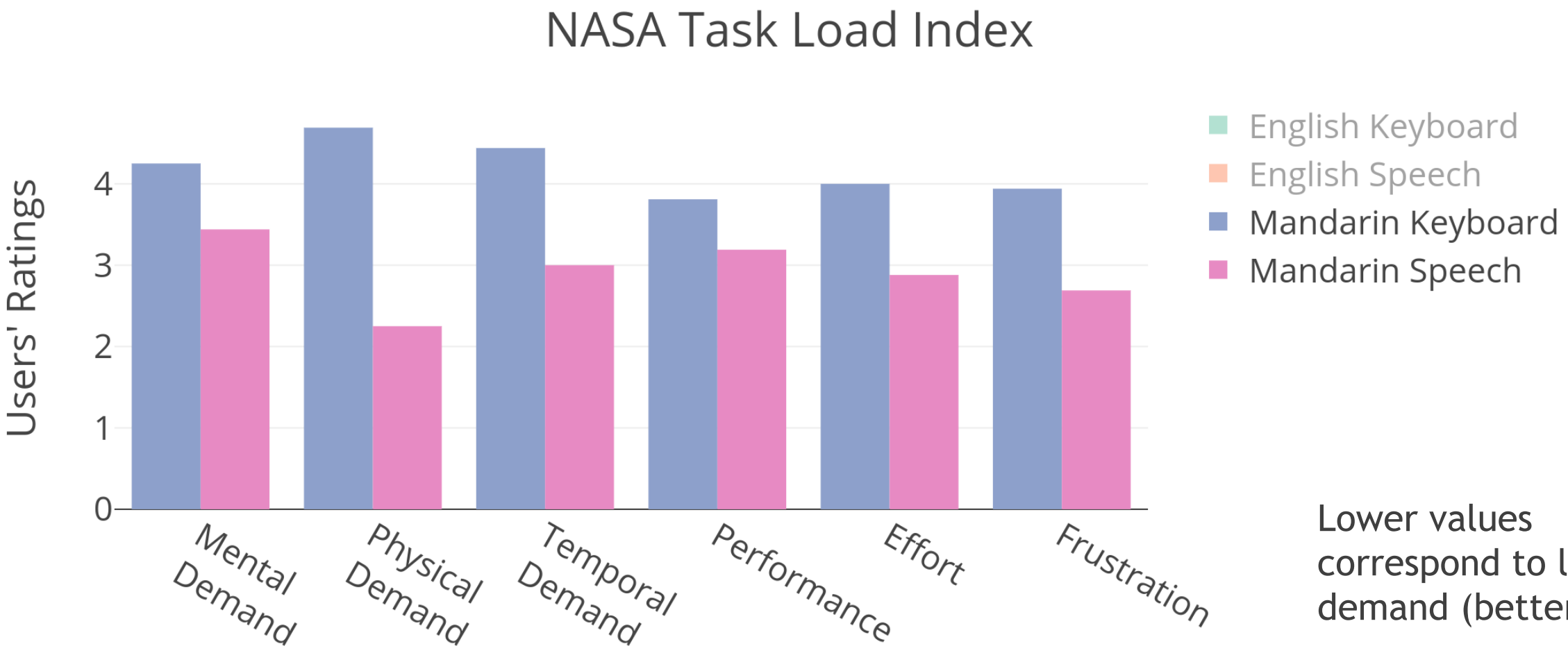
Lower is more accurate (better). Error bars represent +/-1 standard error

Speech input was perceived as having a lower workload than keyboard input for English



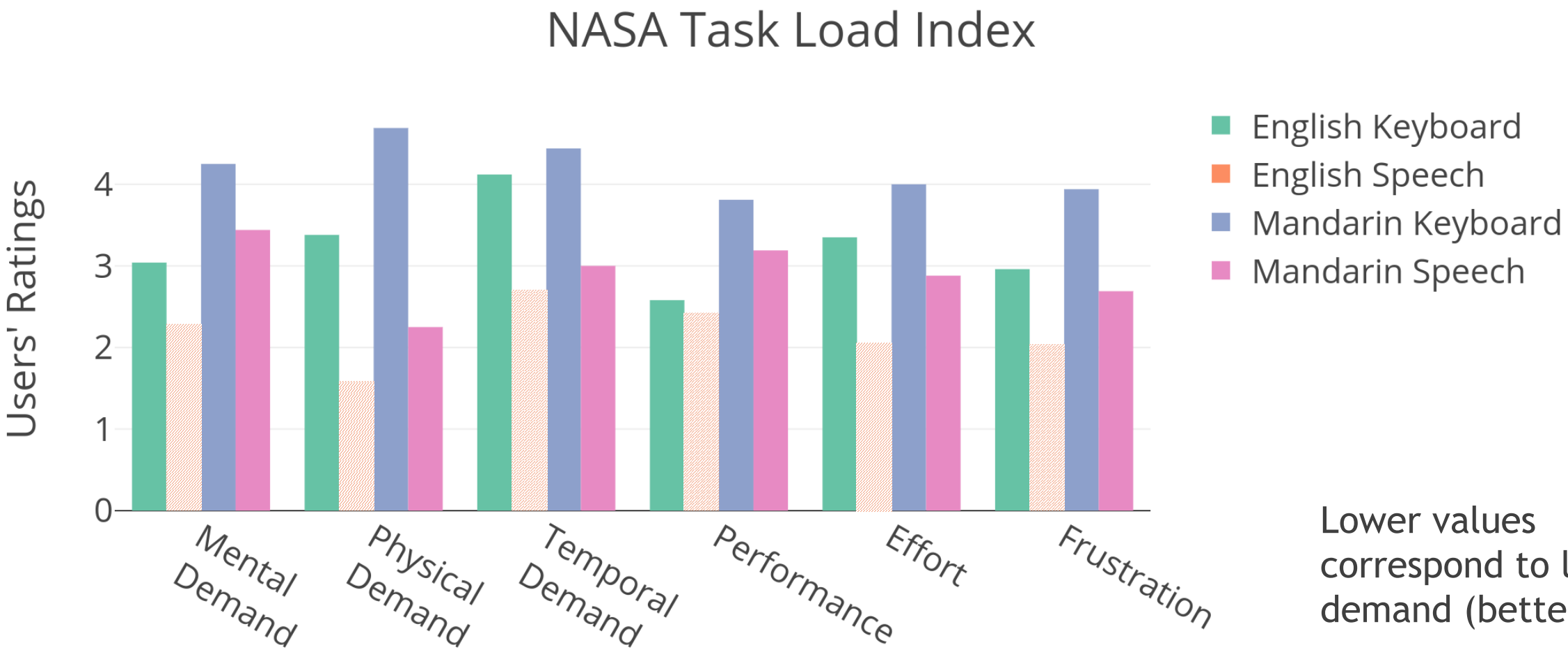
Lower values correspond to less demand (better)

Speech input was perceived as having a lower workload than keyboard input for Mandarin Chinese



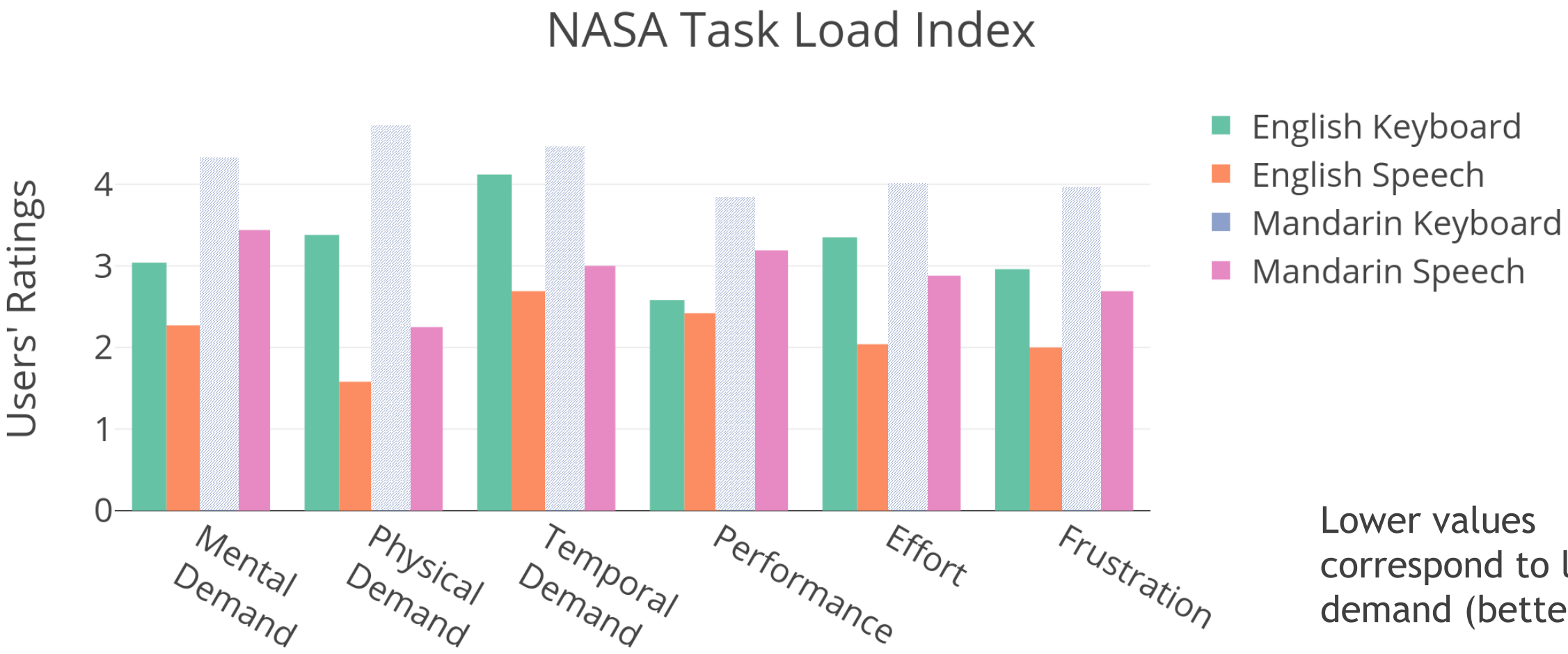
Lower values correspond to less demand (better)

English speech was ranked as the easiest input method across all six categories



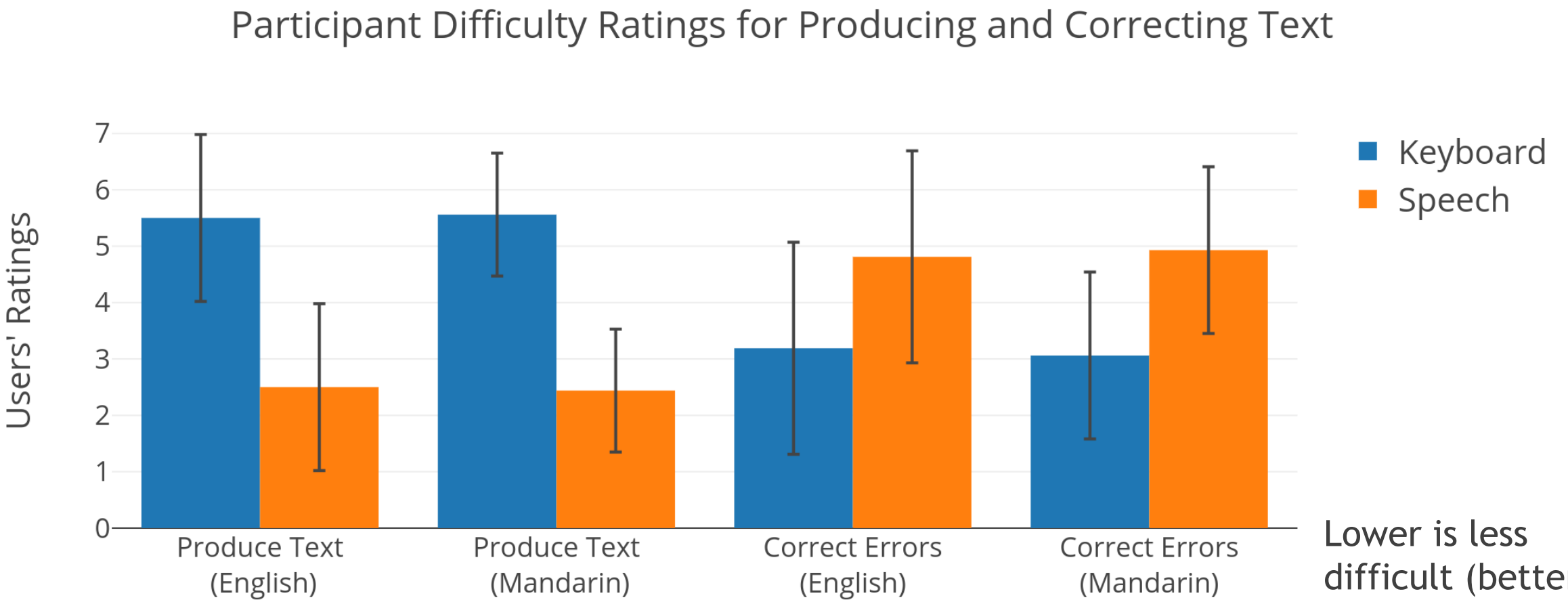
Lower values correspond to less demand (better)

Mandarin Keyboard was ranked as the most difficult input method across all six categories



Lower values correspond to less demand (better)

Speech input is easier to produce text, but keyboard is easier to correct errors with



Qualitative responses

Most users preferred speech due to lower error rate

> “It worked quite smoothly and recognized most of the words I said immediately. It felt very natural once I got used to speaking into the phone.”

> “It is difficult to type an entire sentence using the keyboard, since a typo at the beginning of the sentence could be very hard to correct later.”

Qualitative responses

Some felt more comfortable with keyboard

> “I am comfortable typing on a keyboard. There was less uncertainty about what text was going to be produced. I could correct errors as they were made. That being said, it seemed like I introduced more errors typing than [with] the speech system.”

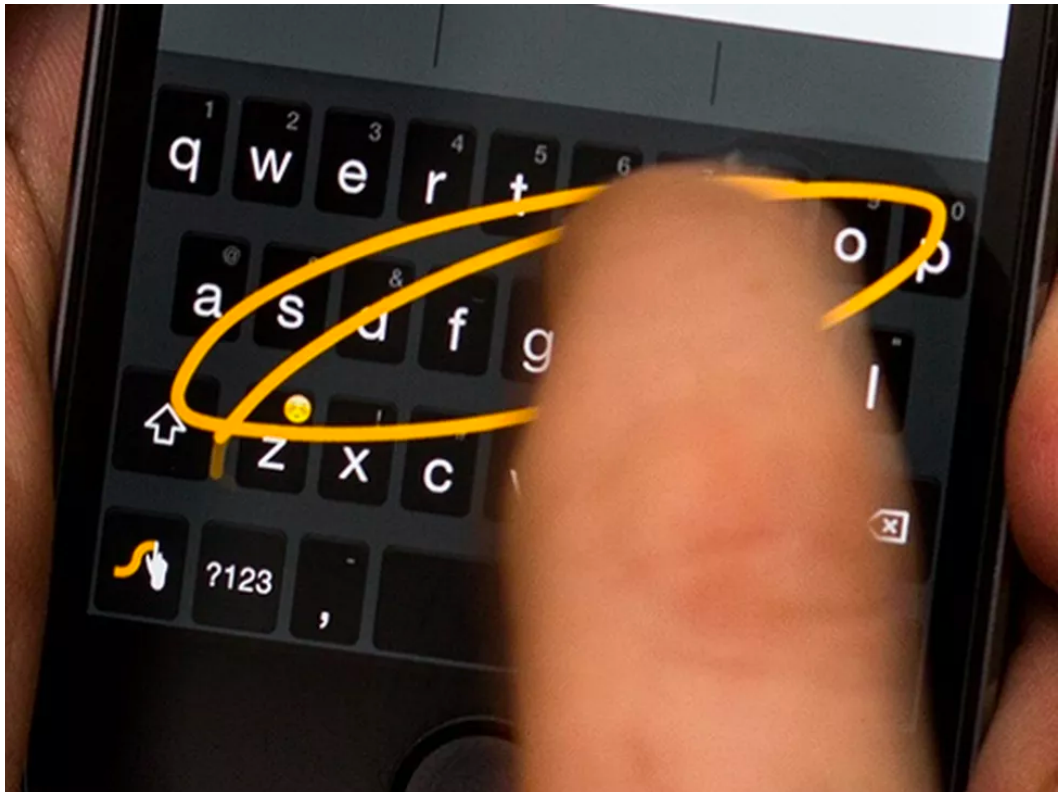
Limitations and Future Work

Limitations

- Only focused on the situation where users were sitting in a relatively quiet environment talking on the phone
- Only tested transcription tasks on short messages without punctuation

Future Work

- Investigation under less-than-ideal real-world settings
- Other keyboards as alternatives to the standard Qwerty keyboards
- Other tasks such as text composition
- Evaluation of embedded speech applications



Conclusion

Conclusion

- The first empirical study demonstrating the **practicality of speech input over keyboard** input on mobile phones
- Speech is **2.9 times faster** than keyboard
- Speech made **fewer errors during entry** and only slightly more after entry

Thanks to

My advisors and coauthors: Jacob O. Wobbrock, Kenny Liou, Andrew Ng, James A. Landay

He Dang of Baidu's Speech Technology group, who had performed a preliminary study that inspired this research project


A significant shift from typing to speech may be imminent and impactful