A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India

Neil Patel, Sheetal Agarwal, Nitendra Rajput,
Amit Nanavati, Paresh Dave, Tapan S. Parikh

IBM
dsc Development Support Centre
UC Berkeley School of Information

The Case for Voice Interaction

Asking a question on Avaaj Otalo

AO: <tune> Welcome to Avaaj Otalo!
You can get to the information by saying a single word.
To ask a question, say ‘question’.
To listen to announcements, say ‘announcements’.
To listen to the radio program, say ‘radio’.

User: I want to ask a question.

AO: Sorry, I didn’t understand. I can only understand single words.
Do you want to ask a question – yes or no?

User: Yes.

AO: OK, you want to ask a question.
To ask a question about agriculture, say ‘agriculture’.
To ask about animal husbandry, say ‘animal’.

Key design choice: input modality

- Application requirements
  - Inexperienced/low literacy users
  - Learnable without training
- Speech is natural
  - But speech recognition requires lots of data

Small- vocabulary, isolated word speech interfaces

Tamil Market

- 27-word vocabulary
- 18 speakers' training data
- 98% accuracy

[Plauché et. al. 2006]

The Experiment

VS.

SPEECH

TOUCHTONE

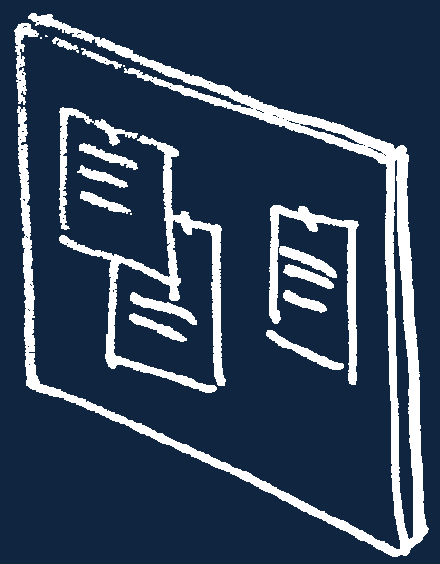
- Previous work for technical professionals in U.S.
- This study: low-literacy, inexperienced users

Sketches courtesy of Bill Verplank

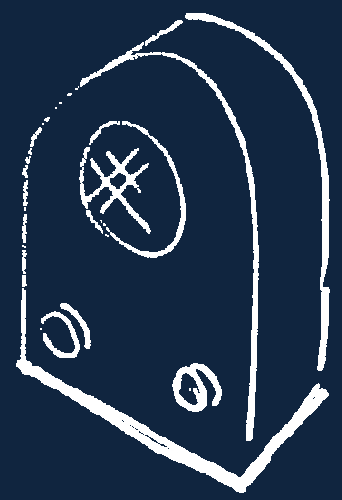
Methodology and Participants

- 45 participants, two treatments, between-subjects
- Small-scale farmers (median 10 acres)
- Native Gujarati, no English
- No experience with voice interfaces
- 73% less than 8th grade education;
  87% never used a PC

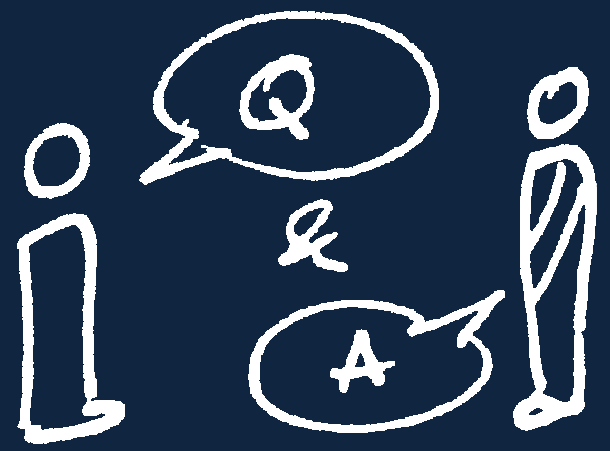
Application Features



ANNOUNCEMENTS



RADIO ARCHIVE



QUESTION AND ANSWER

Sketches courtesy of Bill Verplank

Tasks

1. Listen to announcements (1 step)
   - sequential, 30-60 second audio snippets
2. Listen to a radio program (2 steps)
3. Record a question (9 steps)
   - Categorize question (4 steps)
   - Record question (2 steps)
   - Provide personal contact information (3 steps)

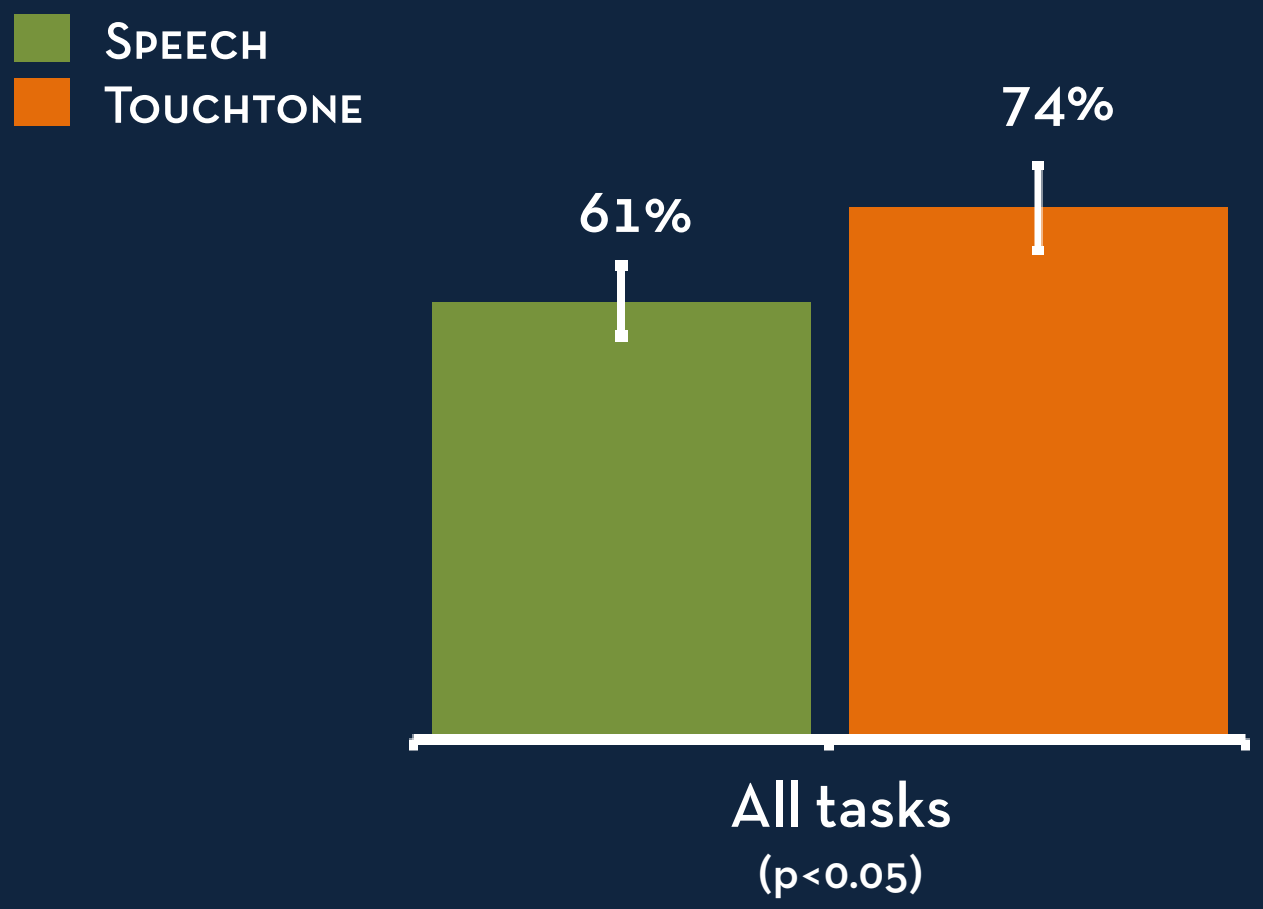
Speech Recognition Accuracy

- Our method: cross-language transfer
  - Apply unmodified acoustic model using transliterated vocabulary
- Accuracy: 94% (commercial systems: ~98%)
- Alternative: model adaptation
  - Linear transformations based on GMM parameters
  - Requires some speech in target language

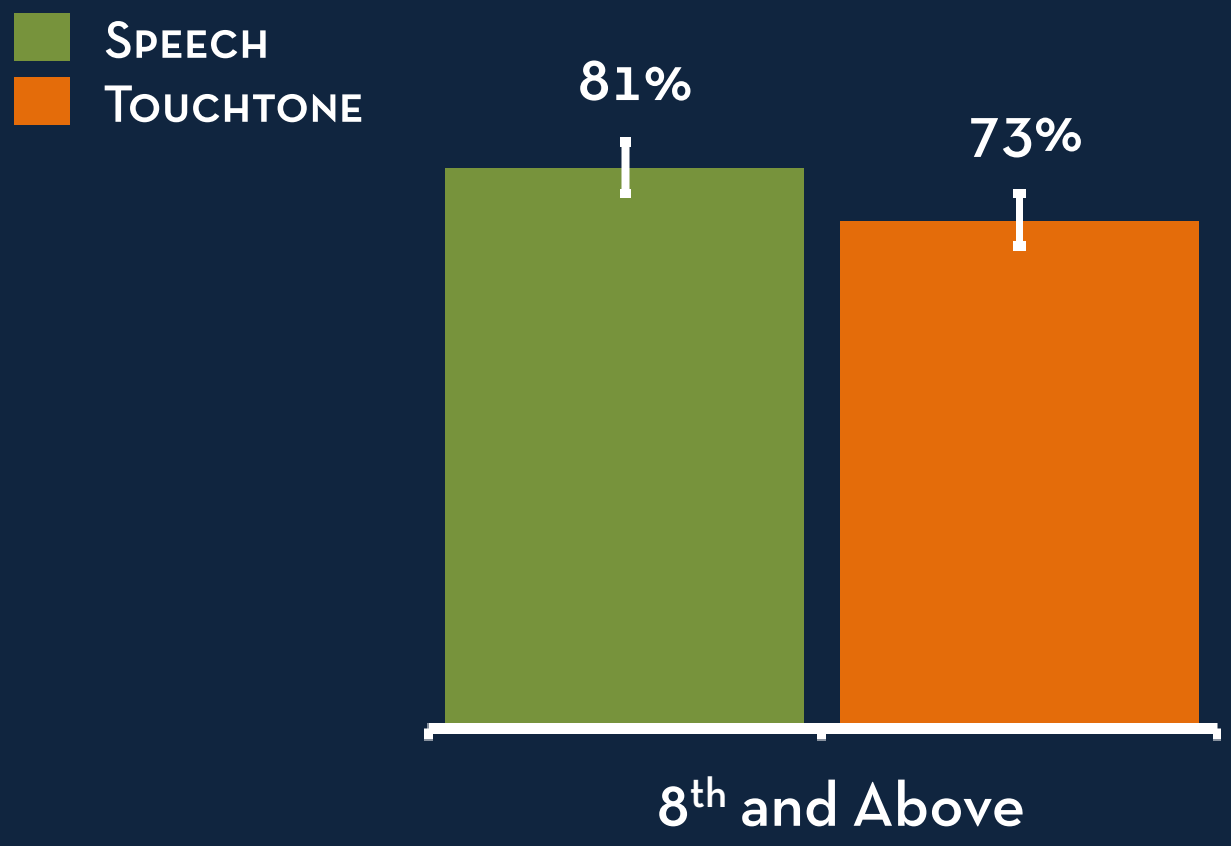
Testing Environment

	OFFICE	VILLAGE
PARTICIPANTS	38 (half speech, half touchtone)	7 (half speech, half touchtone; all women)

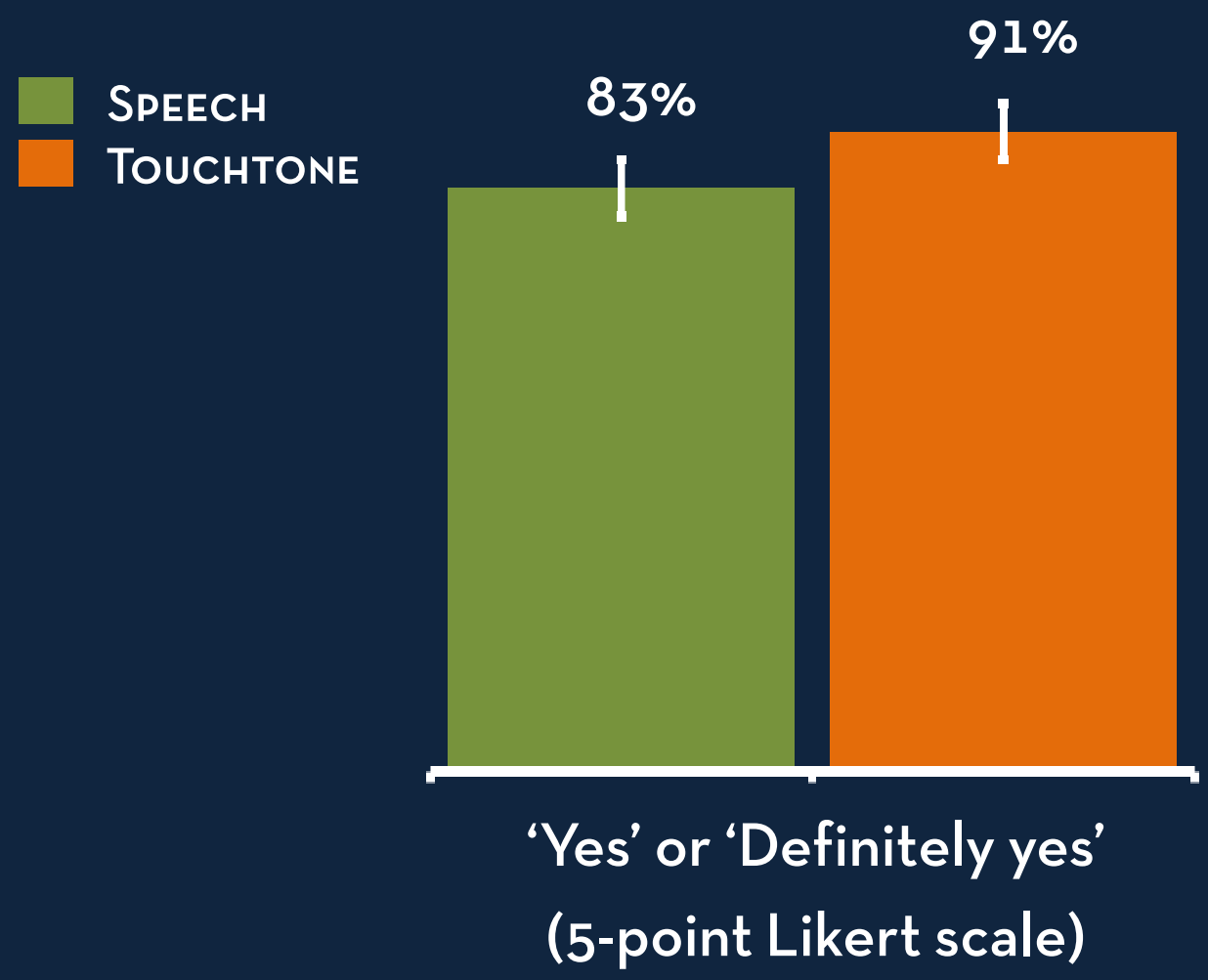
Overall task completion: touchtone higher than speech



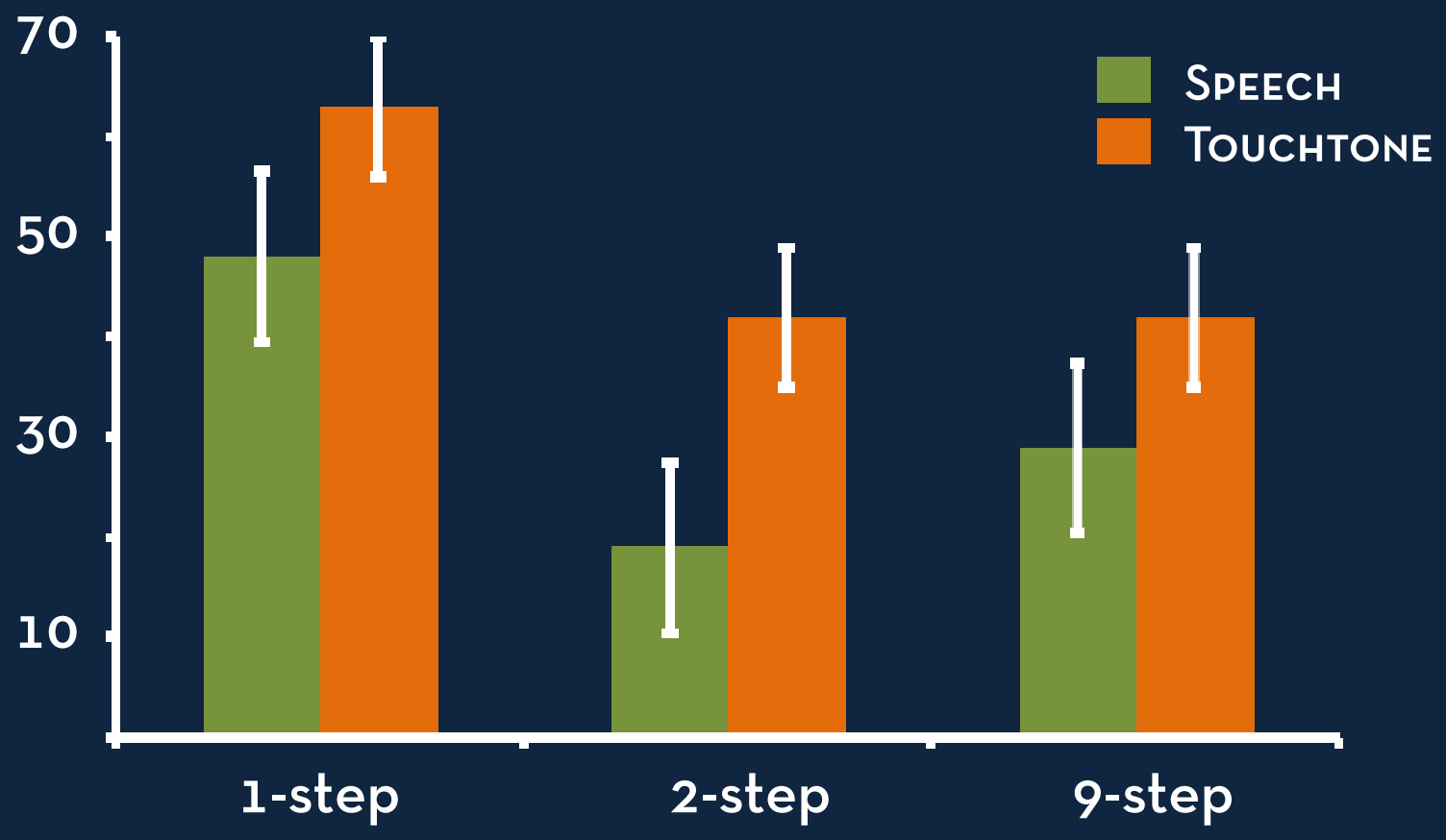
Speech: slightly higher task completion with most educated users



Overall user satisfaction: comparable



Percentage of tasks rated "difficult" or "very difficult"



Across all tasks: 49% (speech) vs. 30% (touch), $p<0.05$

Why was speech less successful?

- Single-word input awkward
- Recognition errors
  - 67% to 42% drop with 1+ errors
- Touchtone benefited from simple, linear tasks

Comparing speech vs. touchtone studies

	TECHNOLOGY	LITERACY	TRAINING	RESULT; USER PREF
IBM staff, U.S *[Lee and Lai, 2005]*	Natural language *~80% accuracy*	High	None	Touchtone; prefer speech
Hospital staff, Botswana *[Sharma et. al., 2009]*	Wizard of Oz *100% accuracy*	Low	Yes	No sig. diff.; prefer touchtone
Community health workers, Pakistan *[Sherwani et. al., 2009]*	Cross-language transfer 93% accuracy	Low	Yes	Speech; no preference
THIS STUDY				
Farmers in India	Cross-language transfer 94% accuracy	Low	None	Touchtone; no preference

Comparing speech vs. touchtone studies

	TECHNOLOGY	LITERACY	TRAINING	RESULT; USER PREF
IBM staff, U.S *[Lee and Lai, 2005]*	Natural language *~80% accuracy*	High	None	Touchtone; prefer speech
Hospital staff, Botswana *[Sharma et. al., 2009]*	Wizard of Oz *100% accuracy*	Low	Yes	No sig. diff.; prefer touchtone
Community health workers, Pakistan *[Sherwani et. al., 2009]*	Cross-language transfer 93% accuracy	Low	Yes	Speech; no preference
THIS STUDY Farmers in India	Cross-language transfer 94% accuracy	Low	None	Touchtone; no preference

Comparing speech vs. touchtone studies

	TECHNOLOGY	LITERACY	TRAINING	RESULT; USER PREF
IBM staff, U.S *[Lee and Lai, 2005]*	Natural language *~80% accuracy*	High	None	Touchtone; prefer speech
Hospital staff, Botswana *[Sharma et. al., 2009]*	Wizard of Oz *100% accuracy*	Low	Yes	No sig. diff.; prefer touchtone
Community health workers, Pakistan *[Sherwani et. al., 2009]*	Cross-language transfer 93% accuracy	Low	Yes	Speech; no preference
THIS STUDY Farmers in India	Cross-language transfer 94% accuracy	Low	None	Touchtone; no preference

Comparing speech vs. touchtone studies

	TECHNOLOGY	LITERACY	TRAINING	RESULT; USER PREF
IBM staff, U.S *[Lee and Lai, 2005]*	Natural language *~80% accuracy*	High	None	Touchtone; prefer speech
Hospital staff, Botswana *[Sharma et. al., 2009]*	Wizard of Oz *100% accuracy*	Low	Yes	No sig. diff.; prefer touchtone
Community health workers, Pakistan *[Sherwani et. al., 2009]*	Cross-language transfer 93% accuracy	Low	Yes	Speech; no preference
THIS STUDY Farmers in India	Cross-language transfer 94% accuracy	Low	None	Touchtone; no preference

Current Status - Pilot

- Live with 50 farmers; over 3500 hits/month
- 70% of calls in first month used touchtone



Thanks to...

- Our partners
  - Development Support Center, Gujarat, India
  - IBM India Research Laboratory
- Our funding sources
  - Stanford SOE
  - IBM India Research Laboratory
- The farmers of Gujarat!
- My advisors: Scott Klemmer and Tapan Parikh

http://hci.stanford.edu/research/otalo/