

Search Vox:

Leveraging Multimodal Refinement
and Partial Knowledge
for Mobile Voice Search

Tim Paek
Microsoft Research

Voice Search 2009

Collaborators at Microsoft Research

- Bo Thiesson

Machine Learning and Applied Statistics (MLAS)

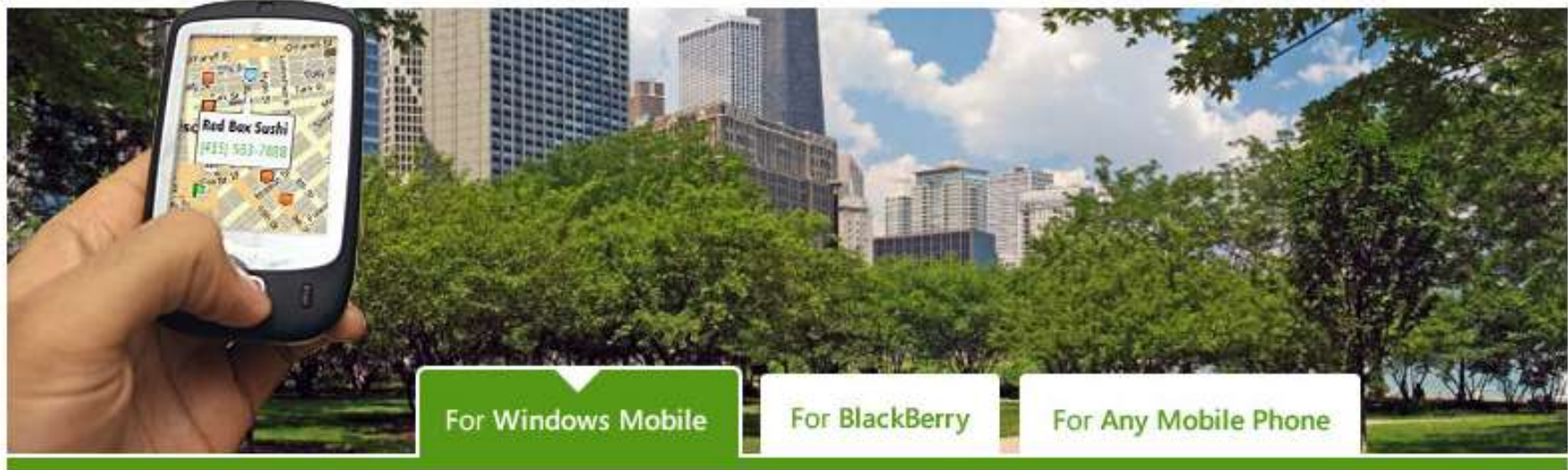
- Y.C. Ju

Speech Research Group (SRG)

- Bongshin Lee

Visualization and Interaction for Business and Entertainment (VIBE)

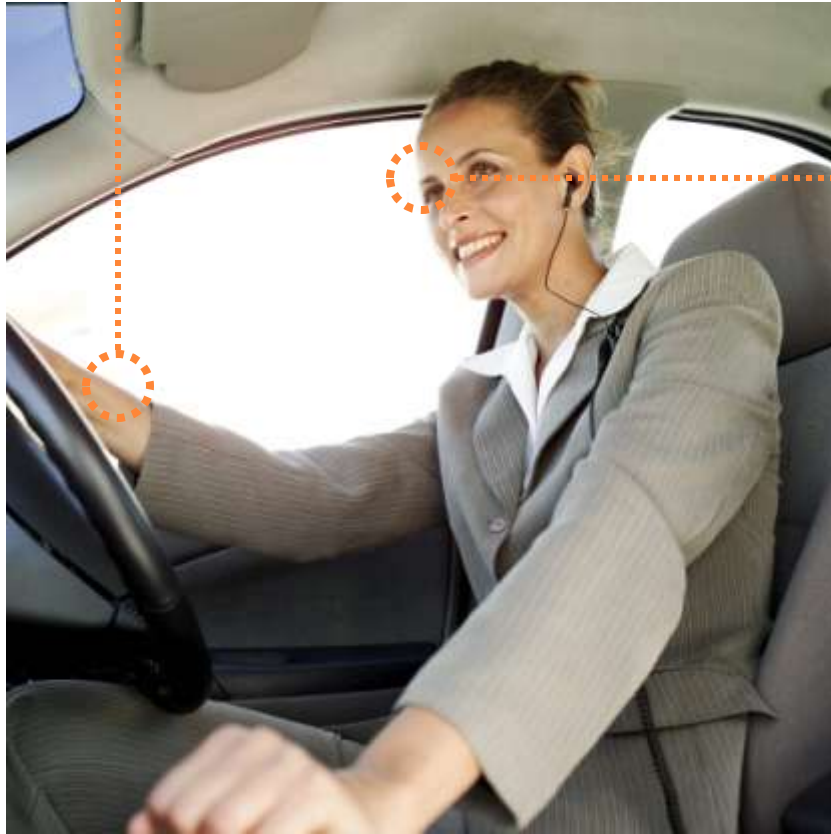




- Speech application for getting on-the-go mobile information from Live Search
- “Just say what you want” for:
 - **411 Directory Assistance (ADA)**
 - Driving Directions
 - Movie Time
 - Local Maps
 - Gas Prices
 - Contacts



Why use speech at all? **Positive** side



Hands-free, eyes-free interaction

Laws prohibiting mobile device use without a “hands-free” kit

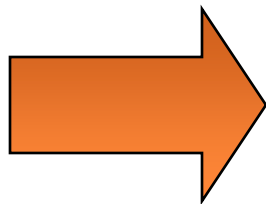


Negative side



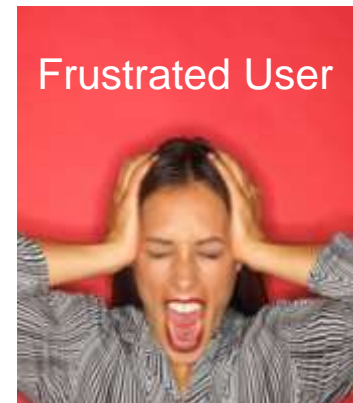
Mobile speech reco is tough!

- Acoustic setting poor
- Microphone quality poor
- Users over-compensate for acoustic noise (Lombard effect)



Recognition Errors Galore

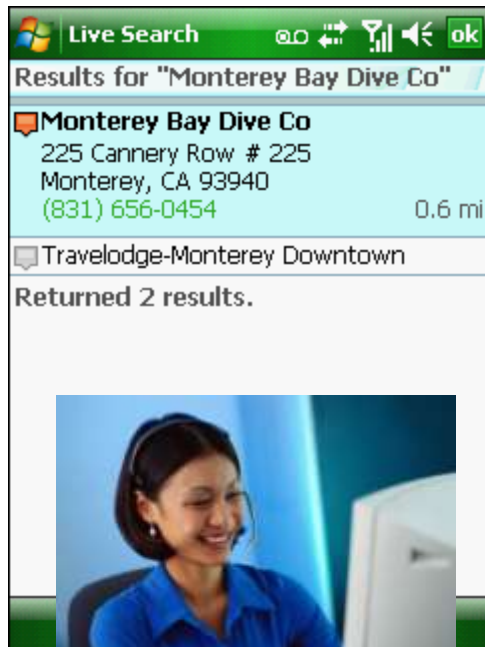
=



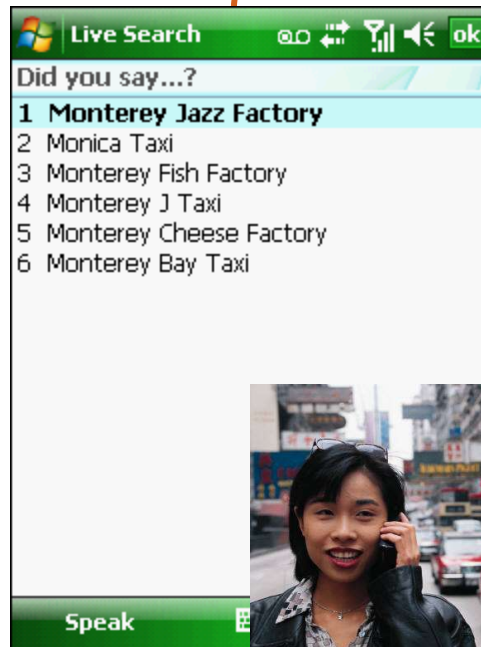
Current error handling

Monterey Bay Dive Company

N-Best List



Good recognition



Poor recognition



Can we leverage other modalities?

- Two modalities are better than one



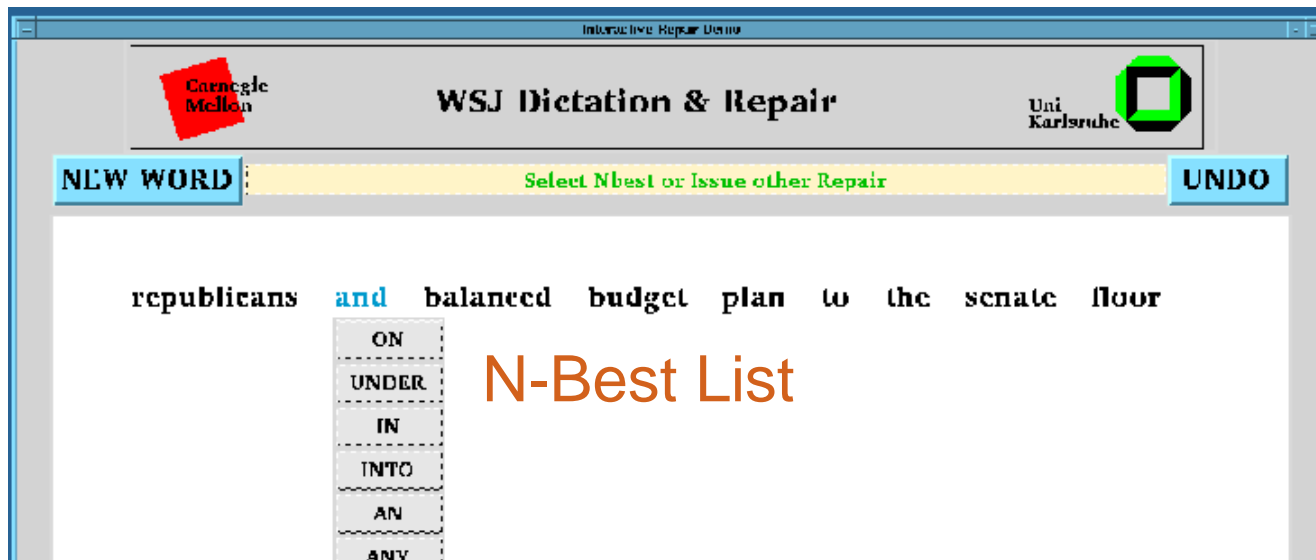
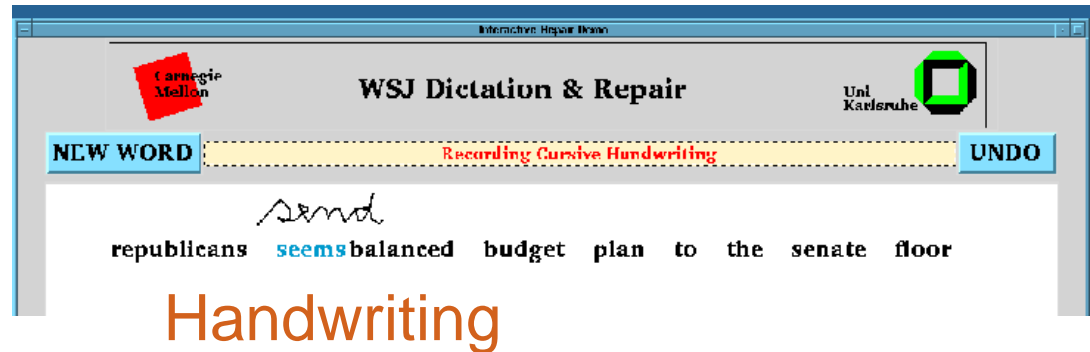
(QuickSet: Oviatt et al., 94, 99, 2000)



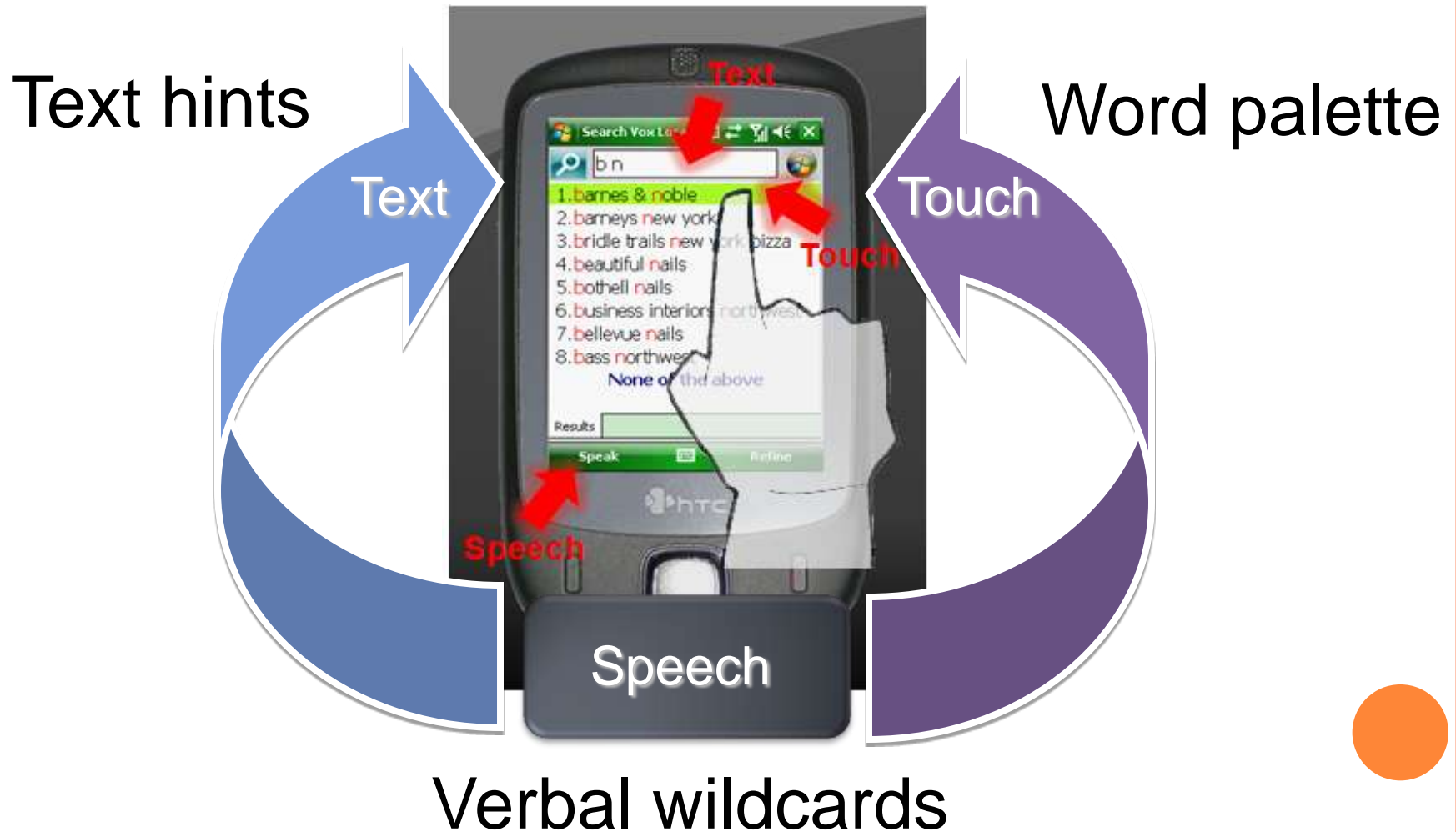
(MATCH: Johnston et al., 2001, 2002)



Multimodal dictation correction (Suhm et al., 99, 2001)



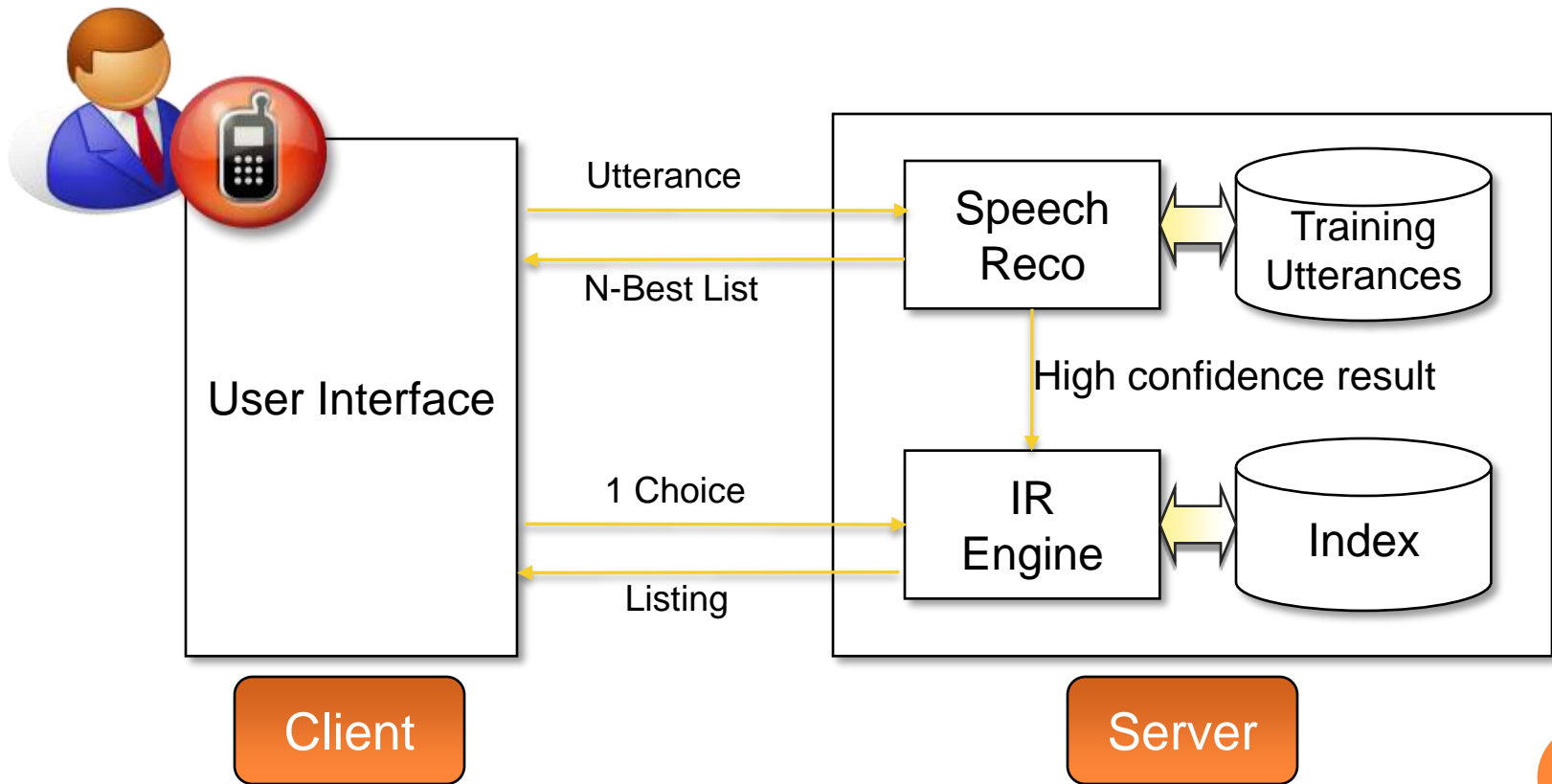
Leveraging Multimodal Refinement



Search Vox Demo

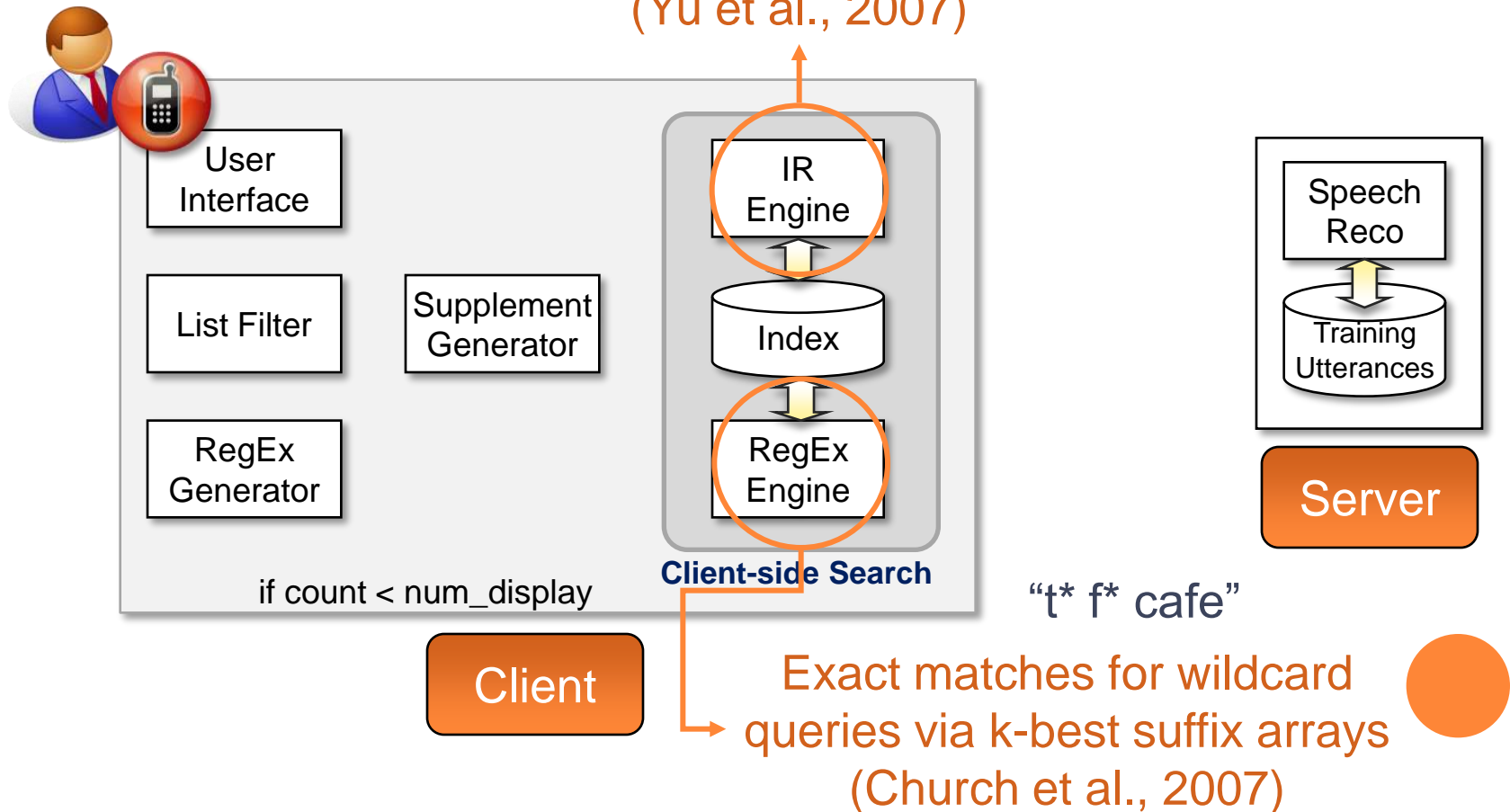


Typical voice search architecture

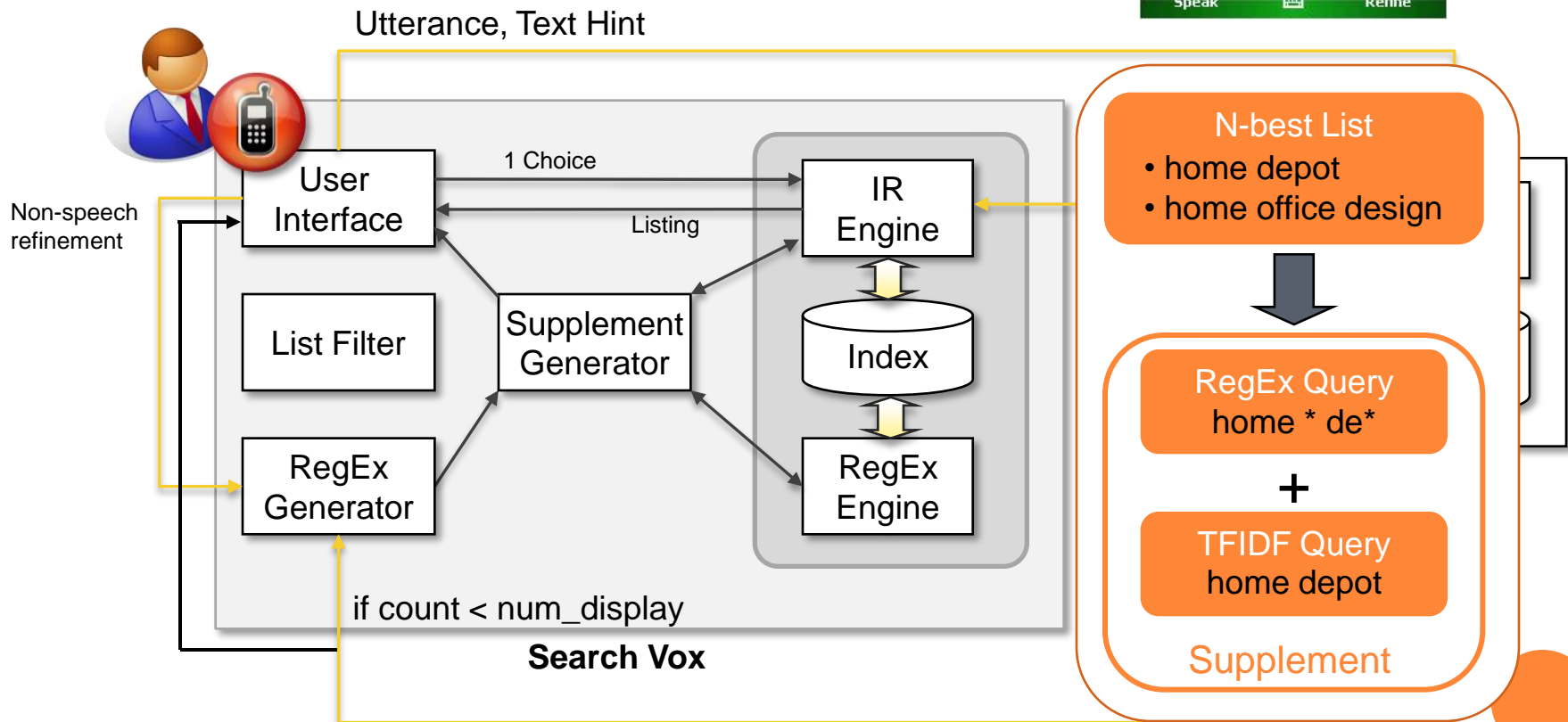


Search Vox architecture

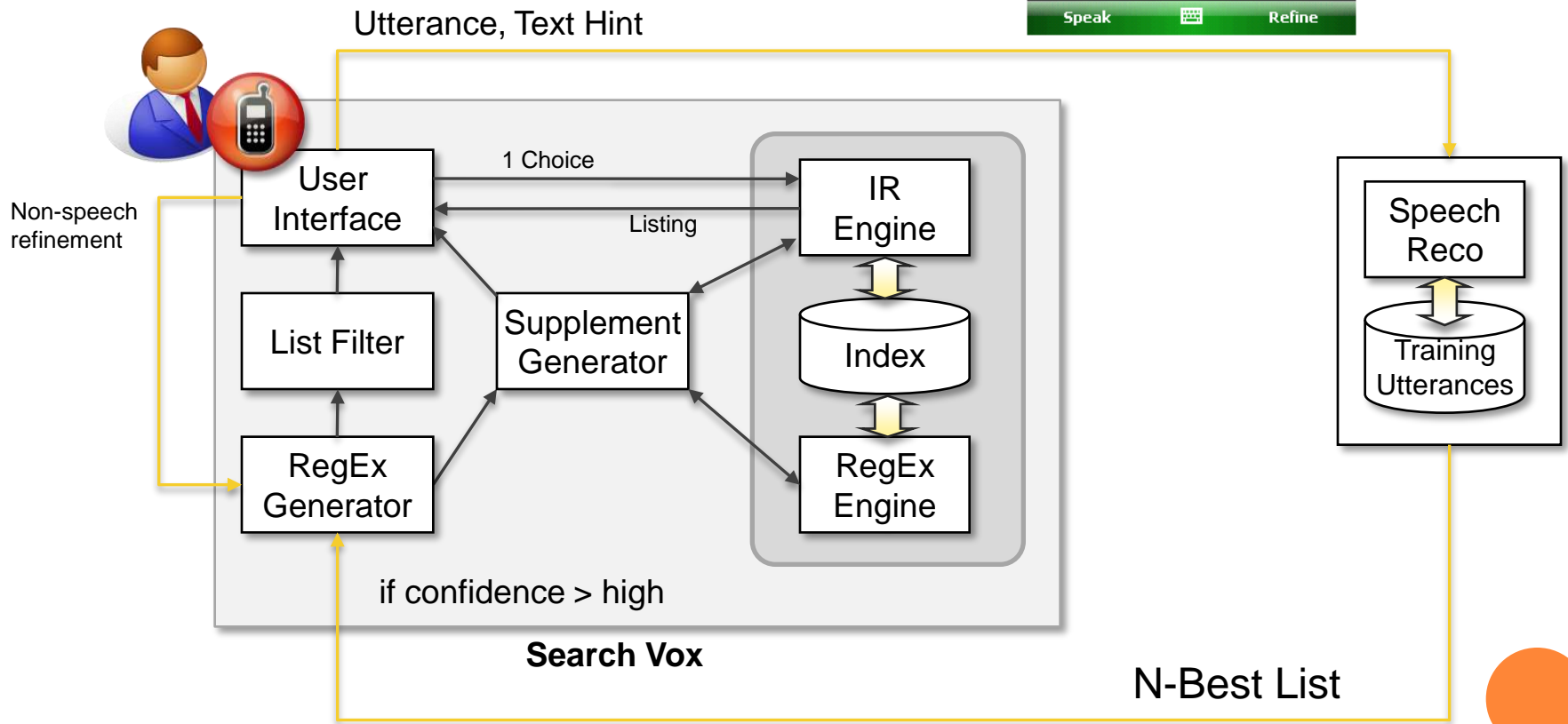
Approximate matches
TFIDF for voice search
(Yu et al., 2007)



Word palette control flow



Text hint control flow



Verbal wildcards

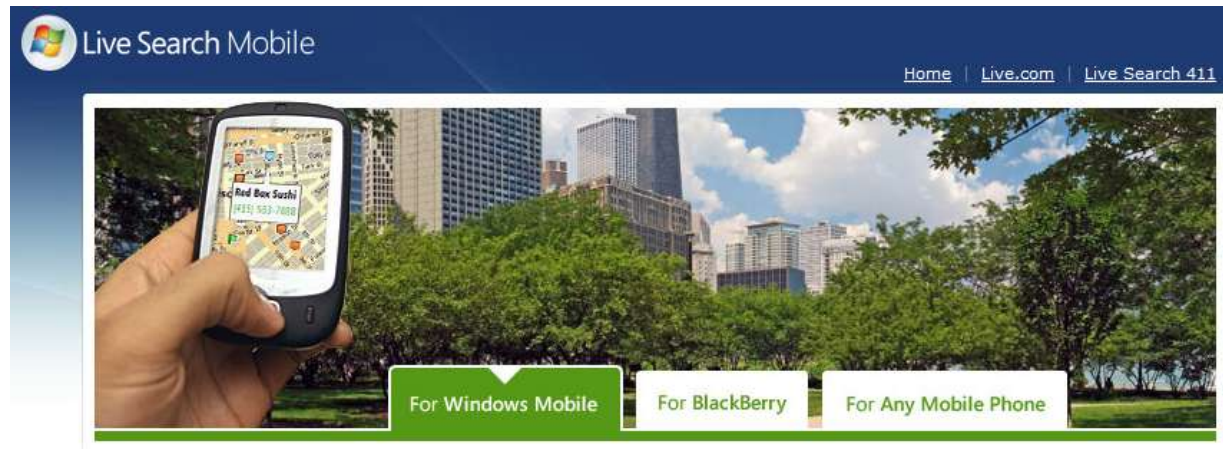
(Paek & Ju, Interspeech 2008)



- Changed language model training sentences
- For all listings with more than 1 word
 - Replaced each word with “*” and “*i*-*” where *i* = first letter of the word
 - If words > 2, we focused on replacing interior words
 - Added duplicate listings to preserve counts for priors
- Semantic rules converted “*” to “something” during inverse text normalization to avoid conflicts with real “something” listings
- Wildcard queries then submitted to RegEx Engine



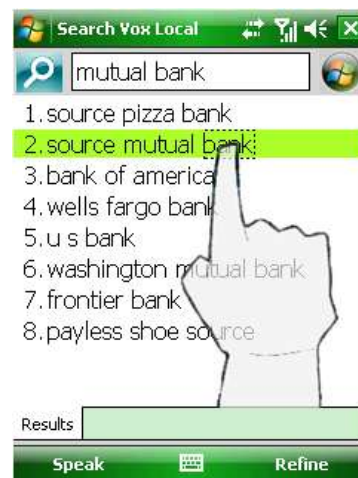
Evaluated multimodal refinement



- Obtained transcribed 2317 utterances for **Microsoft Live Search for Mobile** for which we knew desired listing
- In 20% of the data, transcription did not appear as a choice in the n-best list → **Failure Cases**
- Evaluated how many of the failure cases we could recover; i.e., **Recovery Rate = Relative reduction**



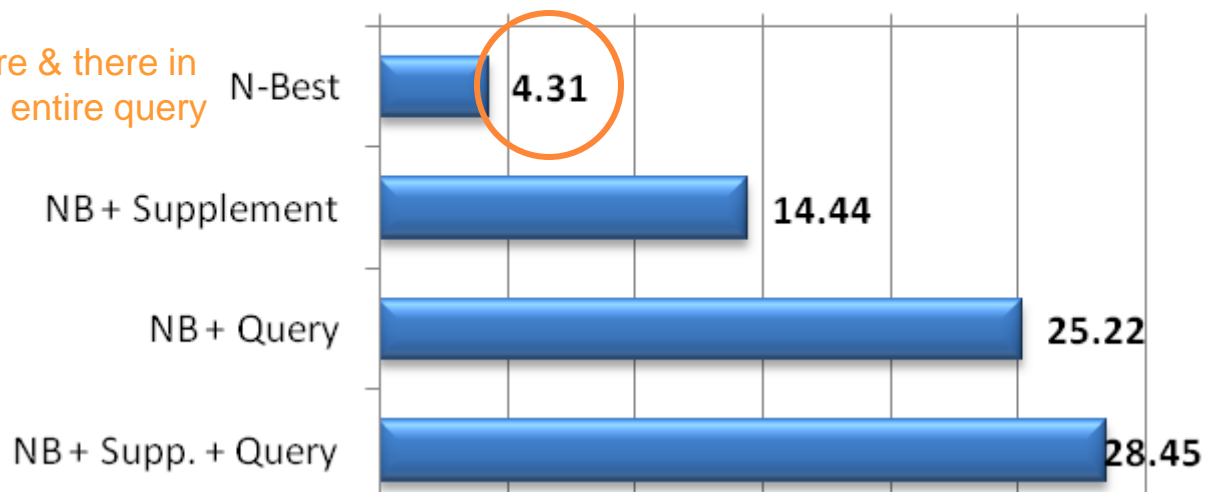
Refinement with word palette



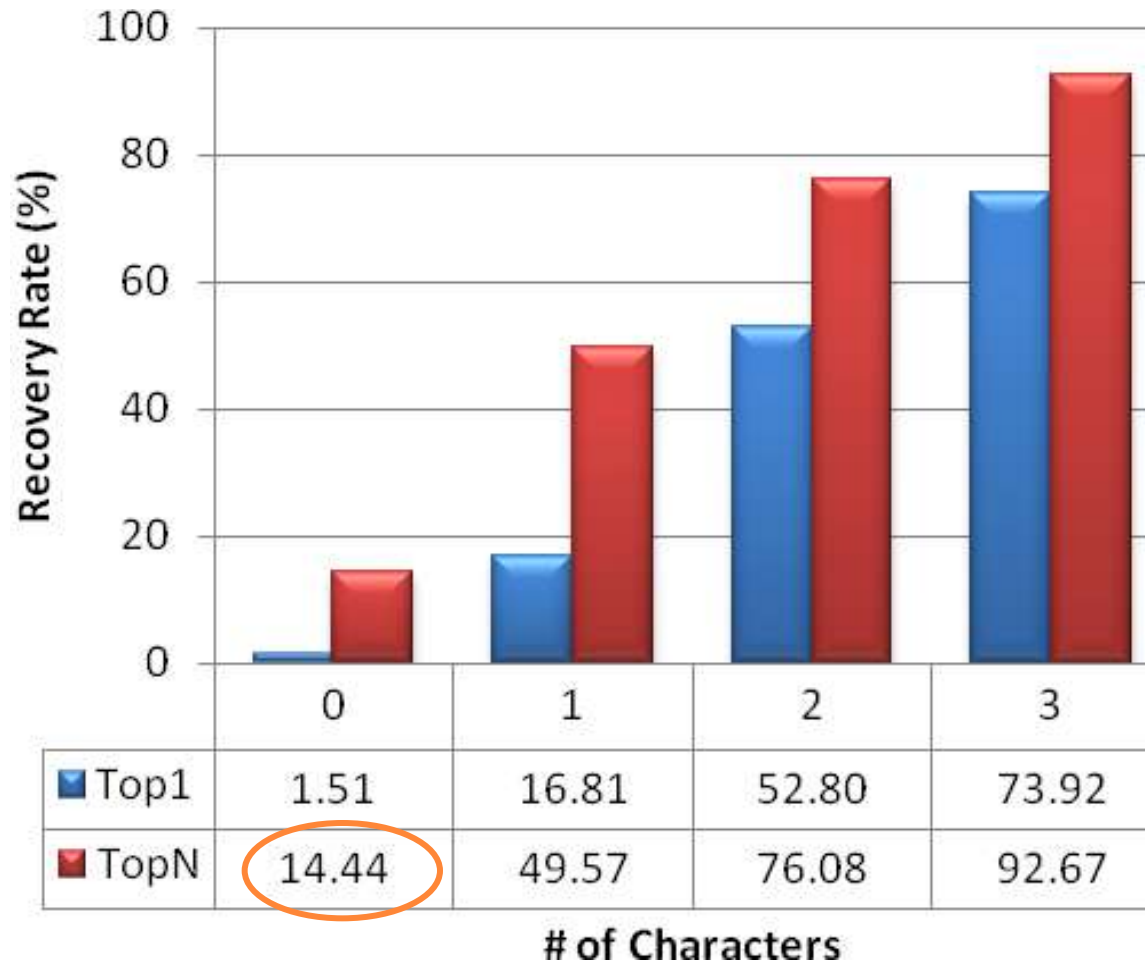
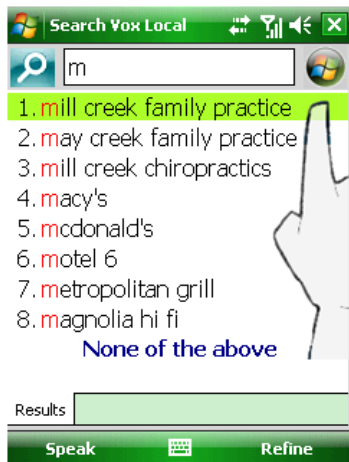
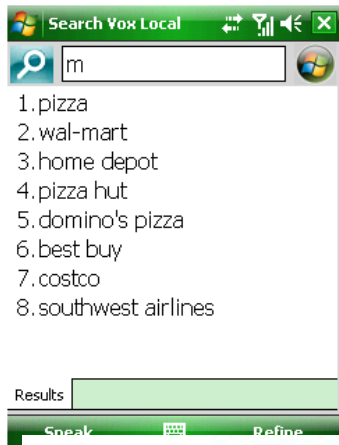
Recovery Rate (%)

0 5 10 15 20 25 30

Using words here & there in result list to form entire query



Refinement with text hints



Guessing top 8 most popular listings was already very good



Evaluation of partial knowledge (Paek & Ju, 2008)

- Language modeling changes cause little damage
 - 1% relative reduction for Top 1, 0.4% for Top N
- Conducted experiment in which users had to recall business listings
 - Recorded listings with both “something” as placeholders and best guess
 - 15 subjects, 178 utterances, 54 comparison cases

	Total (54 Utterances)		
	Top 1	Top N	Error
Guess	16.7%	18.5%	81.5%
Something	33.3%	44.4%	55.6%

31.8% relative reduction
70.6% high / 7.41% low

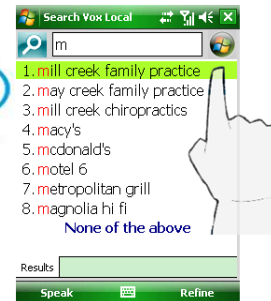
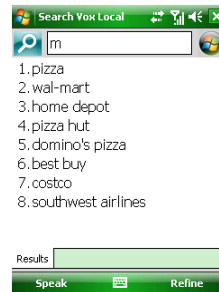


Summary of contributions

- Word palette



- Text hints



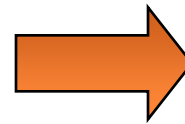
- Verbal wildcards



Future directions

- Text hints during decoding
- Spoken language understanding (SLU)-based recognition of partial knowledge

*“B something bistro ... on Main
... it’s a Italian restaurant”*



SQL Query

- Decoding using multiple utterances (repeats)
- Multimodal fusion
- Incorporating gestures
- User modeling and adaptation

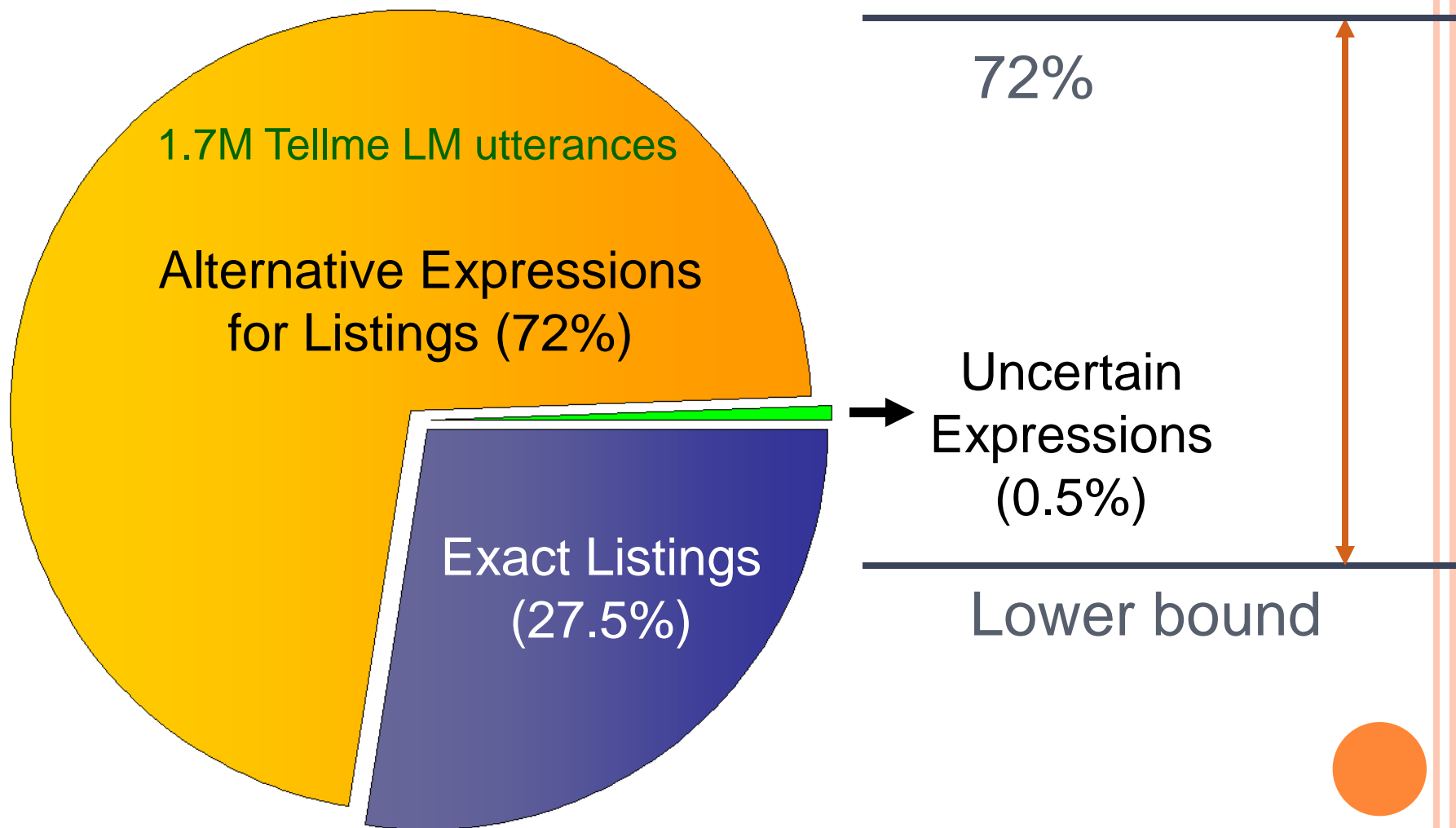


Questions?

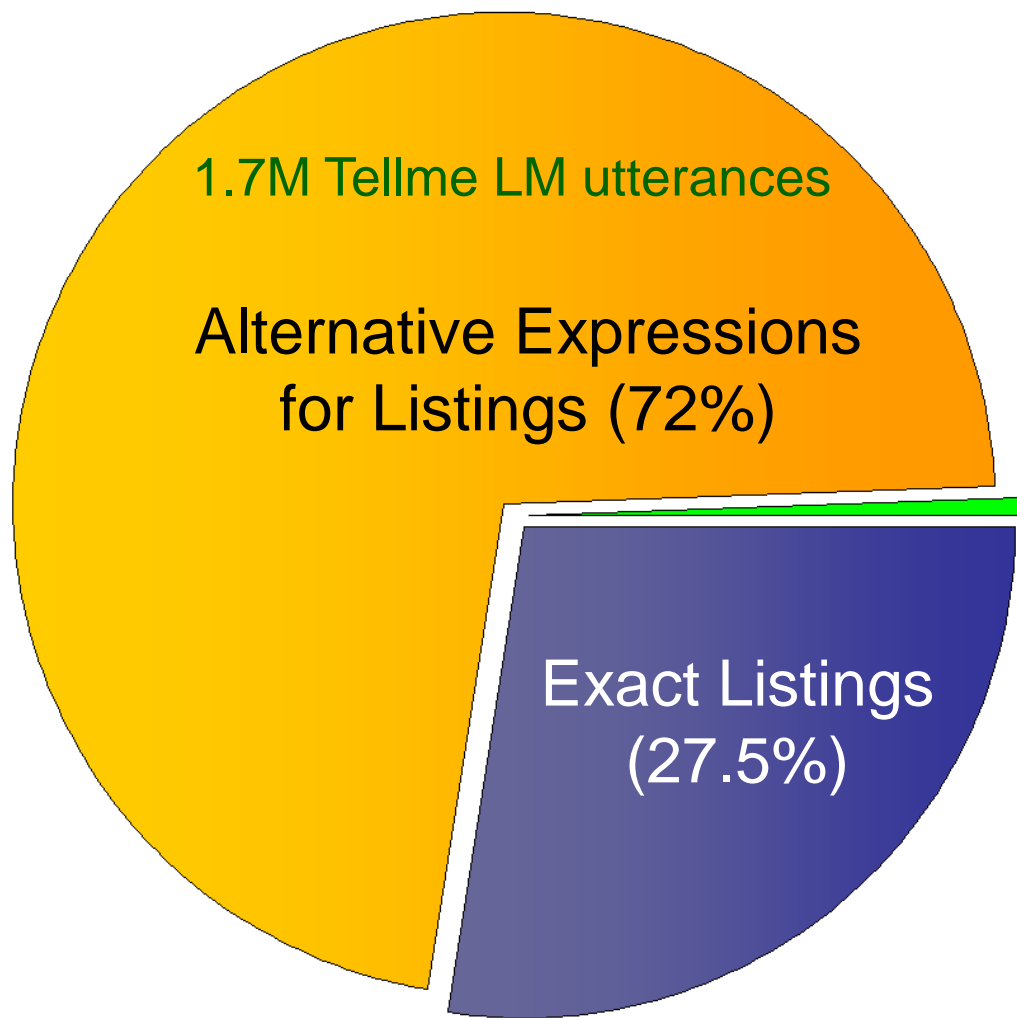
Thank you for your
attention!



How often do uncertain expressions occur in voice search?



How often do uncertain expressions occur in voice search?



Trade-off Issue

Uncertain Expressions (0.5%)

We should not lose more than 1% in overall ADA to handle uncertain expressions



Classified 5 types of uncertain expressions

1. Uncertainty about **specific words** in the name
“U J C United Jewish Something”
2. Uncertainty about **which of several specific words** belong in the name
“Adirondack Steak And Grill Or Grill And Steak”
3. Uncertainty about the **specific address**
“Paper House On Seventy Something In Amsterdam”
4. **General** uncertainty about the entire query
“Jillian's Game Palace Or Something”
5. Uncertainty about the name with **supplementary category** information
“Travel Trailer It's A Trailer I Don't Know They Sell Trailers ...”

