



MITSUBISHI ELECTRIC RESEARCH LABORATORIES

Voice Search Without Errors: Impossible Dream?

Kent Wittenburg, VP & Director
wittenburg@merl.com

Mitsubishi Electric Research Laboratories
Cambridge, MA USA

Voice Search
San Diego, CA USA
March 10-12, 2008

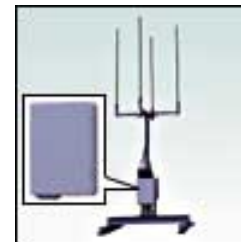
Who are we? Mitsubishi Electric Corporation

- Mitsubishi Electric (MELCO)

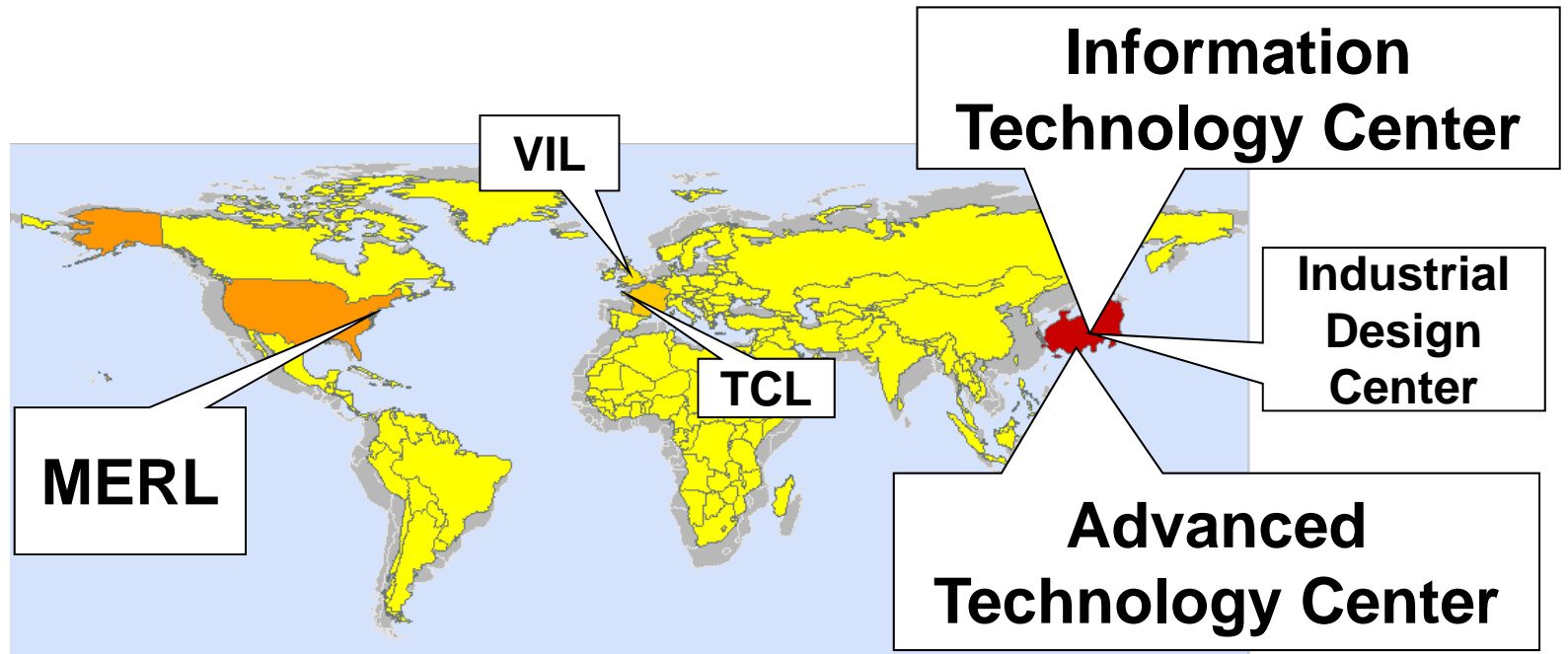
- Energy & Electric
- Industrial Automation
- Information & Communication
- Automotive Electronics
- Electronic Devices for Space, Transportation, and Defense
- Home Appliances

- ~\$35B annual revenues

- *NOT Mitsubishi Motors, Mitsubishi Heavy Industries, Mitsubishi Bank,....*



Who are we? MELCO Corporate R&D

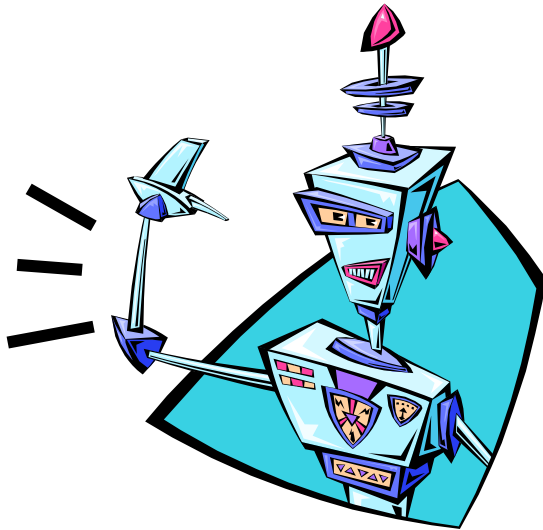


- MERL is the North American arm of Melco Corporate R&D
 - Cambridge Research Lab founded 1991
 - Completed merger in 2003 of Murray Hill Lab (DTV & Communications) and Horizon Systems Lab (Computer Systems)
- MERL's Mission
 - Generate key intellectual property in areas of importance to Melco
 - Significantly impact Melco products with innovative technology
 - Set industry standards for key business areas

Best way to handle errors in dialogs?

- Don't allow them!

1. Conversational dialogs

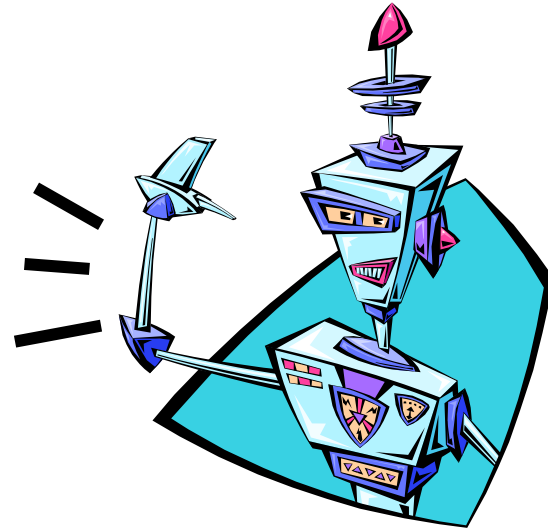


2. Speech-in List-out dialogs



Conversational Dialog Paradigm

- Based on human-human communication
- Speech must be disambiguated
- Highly structured, multi-state dialogs
- Grammars are state-dependent
- Users must stay within the sublanguage allowed in each state
- Else: errors -> error recovery

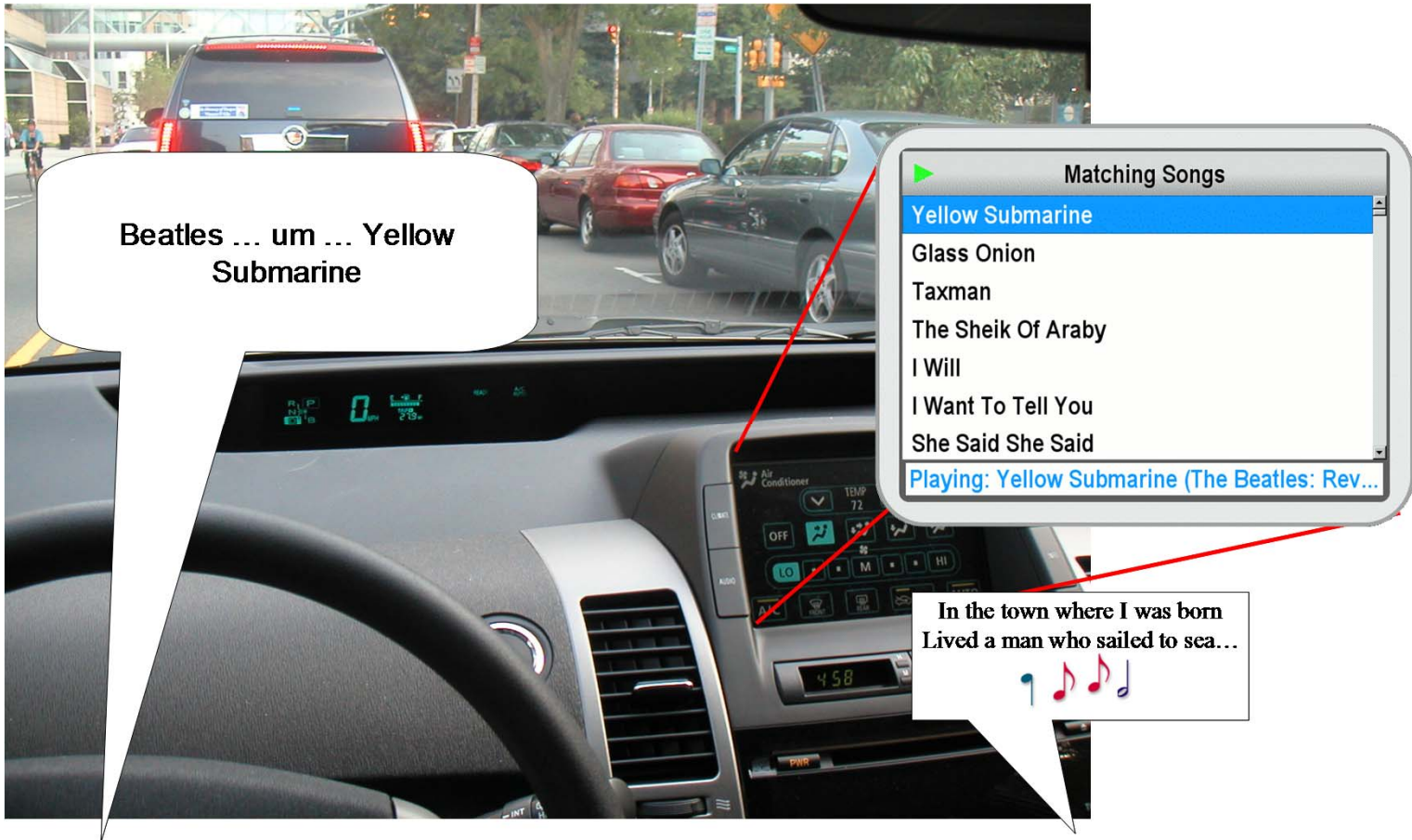


Speech-in List-out Paradigm

- Voice search (Google with your voice)
- Results (good, bad, or ugly) are returned
- Speech is never disambiguated
- No vocabulary or grammar to learn
- Robust to noise and out-of-vocabulary input
- No errors
 - Just try again OR
 - Navigate from there

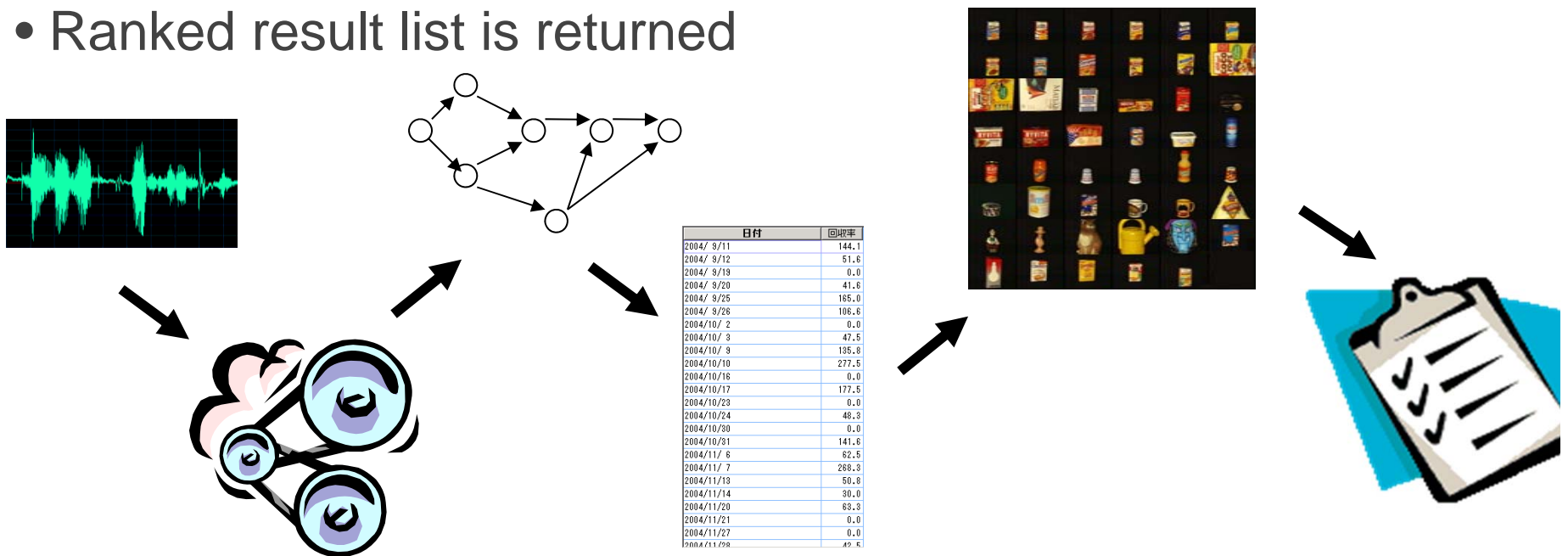


Demo of MERL's "SpeakPod" SpokenQuery for Digital Music Collections



How does it work?*

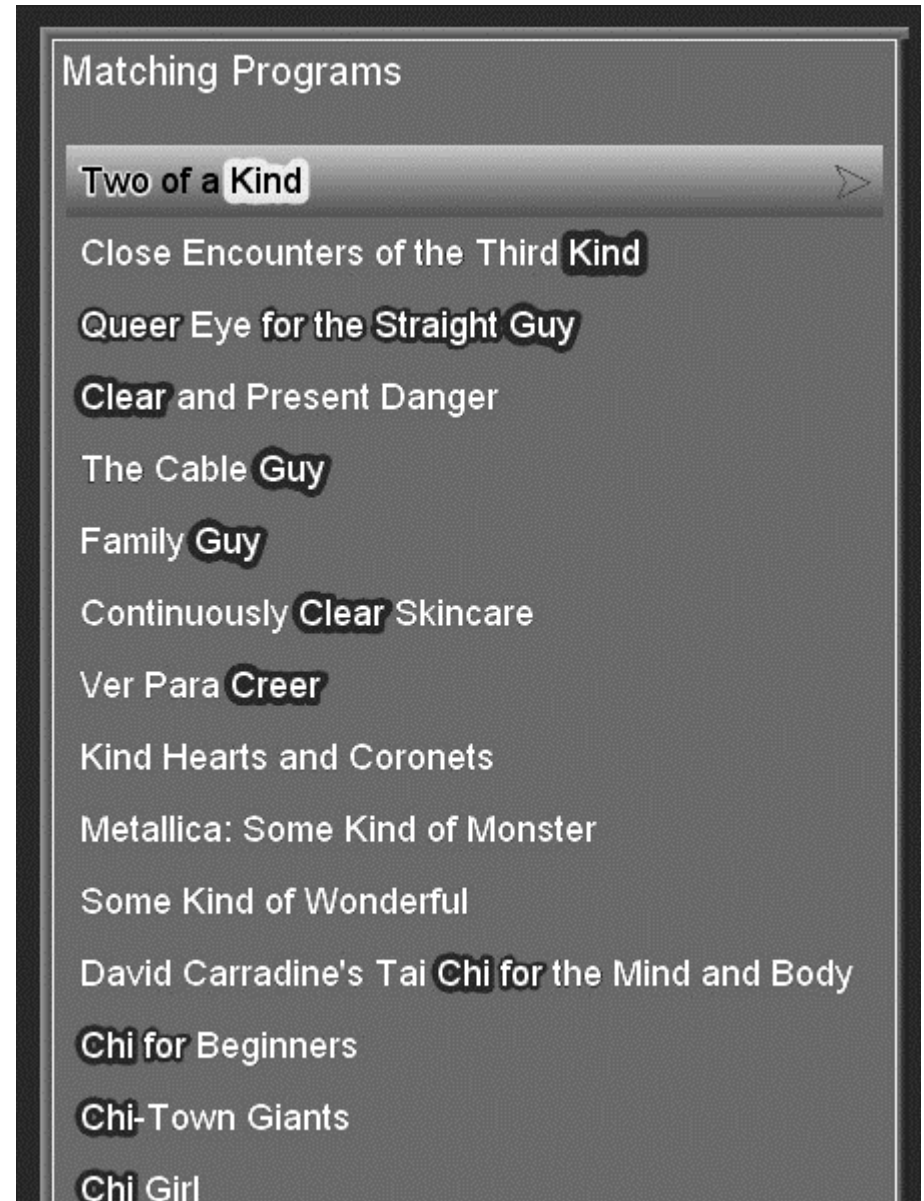
- Speech is fed to a recognition engine
- Yields recognition lattice
- Vector of words/bigrams/trigrams with probabilities is extracted
- Query vector is matched against database
- Ranked result list is returned



*Patented method

What is the hitch?

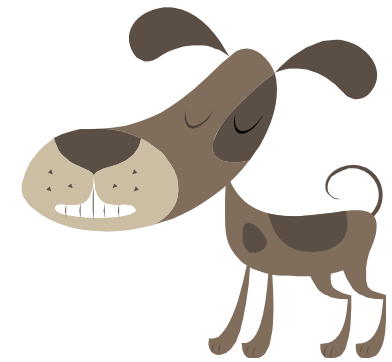
- Intermediate feedback is difficult
- Users must not expect the system to “understand” them
- Are just the results enough?
- Will users just “try again”?



Development Advantages

- Streamlined interface development
 - Deep, stateful voice dialogs can be difficult to specify, design, test, and debug
 - Interfaces (both spoken and visual) kept flat and minimal; less development time
- Streamlined application development
 - Much data (e.g. POI, music, product catalogs) already indexed in a suitable form
 - No need to construct, test, and maintain
 - grammars
 - dialogs
- Easily extensible to new domains

Parting thought:
What makes a dialog “natural”?





Acknowledgements

- Bent Schmidt-Nielsen, Bhiksha Raj, Bret Harsham, Garrett Weinberg, Peter Wolf, Joe Woelfel, and Hugh Secker-Walker for conception & development of SpokenQuery

Further Information

- <http://www.merl.com/projects/SpokenQuery>
- See Garrett Weinberg at this conference for more demos