Multimodal Search in Mobile Devices

March 11, 2008

Thomas Schalk
Agenda

• Multimodal user interfaces
• Interactions of voice, vision, touch
• Portable navigation
• Live search
• Embedded versus off-board
• Future mobile device trends
Multimodal User Interfaces

- Multiple interaction modes
  - Speaking, listening, vision, and touch
- Saying what you want
- Seeing what you asked for
- Hearing what you asked for
- Buttons, keyboards, knobs, stylus pens

- Mixing a visual display with speech is key
- HMI principle: ability for vehicle driver to glance
Speech: Latency and Accuracy

System Latency and Accuracy are both critical for user satisfaction. Off-board access latency needs to be managed by user interface.
Portable Navigation: TomTom Destination Entry

- Must initialize with tap
- Optional dialog mode
- Speech is very effective

- N-best lists are critical
- Confirm by tap or voice
TomTom Street Entry

- Conditional on city
- Critical speech function
- Visual n-best listing

- Speech or tap
- Tap to confirm
- Tap may be easier
TomTom Routing Display

- Turn-by-turn directions by selectable TTS
- Street names are not spoken, just how and when to turn
- Display not needed while driving

HMI Principle: Glancing rule to minimize driver distraction
Microsoft Live Search Multimodal Interface

Wing Device
Asking for Restaurants

“I want Italian”

Choices List

Selection
Live Search Routing Information

Location

Text Turn-by-turn

Route Map

Detail Menu Map Menu

Route Summary

Depart on Freeport Pkwy (S) (East)

Take Ramp (RIGHT) onto I-635 [Lyndon B. Johnson Fwy (S)]
At exit 31, take Ramp (LEFT) onto I-635
Road name changes to Local road(s)
Turn LEFT (North) onto N MacArthur Blvd
Arrive at Michael’s Italian Kitchen, 8140 N MacArthur Blvd

Total: 4.4 mi (7min)

Detail Menu Map Menu

Depart on Freeport Pkwy (S) (East)
Embedded Versus Off-board

• Embedded Solutions
  – ASR and data within the mobile
  – Limited computing resources, but reasonable
  – Limited usability and adaptation
  – Direct audio input
  – Tap-to-talk or push-to-talk

• Off-board Solutions
  – ASR is server-based and away from device
  – Unlimited computing resources
  – Unlimited usability and adaptation
  – Audio communicated via data channel
  – Push-and-hold audio capture and transmit
Off-board Search Functional Architecture

Users -> Voice Server
- Speech Recognition
- TTS & Audio Files
- Voice Browser
- Audio Interface

Operator -> ATX Back-end Infrastructure
- Intelligent Whisper

Call Center
- GPS
- Application Server
  - Voice XML
  - Java
- IP
  - VoiceXML
- Search DB
- Usability Monitoring

Users to Call Center:
- Voice Browser
- TTS & Audio Files
- Audio Interface

Voice Server to Application Server:
- Voice XML
- Java

Operator to Users:
- Intelligent Whisper
CONCLUSION: Challenges and Trends

- Thin mobile device audio quality
- Off-board latency prevents dialog
- Agent-assisted speech recognition
- Improved speech recognition
- Multimodal UI usability improvements
- Portability trends over next 5 years
- Mobile device HMI controlled by vehicle
  - Key trend for future navigation and infotainment
Thank You