Robust User Interface for Mobile Voice-Enabled Applications

Mark Anikst

mtanikst@gmail.com

Applications & Platforms

- Network-based client applications
- Wireless mobile devices: handhelds/PDAs, tablet PCs
- Multimedia presentation: Text, audio, images, video
- Multimodal interface: Graphic, Voice and Digital UI
- UI workflow: directed/mixed initiative dialogs

UI Modalities

- VUI using microphone, speaker
- GUI using keypad, stylus & touch-screen, display
- DUI using scanner, digital camera, RFID, printer, etc.

Application Input

- Command/Query/Response (overlapping)
  - GUI: button clicks, text entry, list item selection
  - VUI: verbal commands, queries, responses
  - DUI: alphanumeric data entry
- Audio/Image/Video/
  - VUI: audio recording of notes/messages
  - DUI: image/video capture

Application Output

- Prompt/Info (overlapping)
  - GUI: text (can be in table/list format)
  - VUI: spoken summary of text displayed via GUI
- Audio/Image/Video/Print
  - GUI: image display, video playback
  - VUI: playback of recorded notes/messages
  - DUI: mobile printing

Multimodal UI

- GUI provides full control of application functionality (non-efficient, likely degraded)
- VUI, DUI compliment GUI and provide efficient alternatives for some input/output
- VUI efficiently supports “hands/eyes-free” use scenarios
- VUI “barge-in”: audio output can be interrupted by a speech input/button click

Disable/Enable/Adjust VUI

- MUTE/TALK button to disable/enable speech input
- QUIET/LOUD button to disable/enable audio output (incl. speech)
- “Speak quieter/louder/faster/slower” commands for speech audio output

Noisy Background

- Headset with noise-canceling microphone & speaker
- Noise models integrated into speech recognition
- Rejection, confirmation/error-correction (noise vs. speech confidence thresholds)

Speaker Dialects, Accents

- Run-time adaptation of speech recognizer
- User-specific voice profiles (stored in the network)
- User-native language/dialect specific speech recognition (coupled with TTS and GUI)

Phonetic Confusability of Speech Input

- Restricting to phonetically-distinct alternatives
- “Tight” SR grammars for each dialog state
- Confirmation/error-correction

Alphanumeric Data Input

- Filling fixed-format slots by speech (digits, letters)
- Using letter name recognition
- Using a spelling alphabet (English: “Alpha”-“Zulu”, Spanish: “Alicia”-“Zaragoza”)
- Alternative input using keypad, scanner, RFID

Noise-Degraded Intelligibility of Speech Output

- Speaking numbers as strings of digits
- Speaking letters using a spelling alphabet
- Alternatively, viewing GUI display of speech output (as text/semantic tags)

Conclusion

- Robust UI for mobile voice-enabled applications = VUI augmented with GUI & DUI
- VUI supports a subset of application functionality
- VUI optimized for “hands/eyes-free” segments of user operations

Robust User Interface for Mobile Voice-Enabled Applications

Questions?