Technologies, Tools, and Standards for Multimodal Application Development

Deborah Dahl
Principal, Conversational Technologies
Mobile Voice Conference 2014
San Francisco
I want to develop a multimodal application!
….Now what?

• What devices do I want my application to run on?
• What technologies does my application need?
• What development tools are available?
• What standards do I need to know about?
Devices: The Olden Days
Devices: Now

• Smartphones: Android, iPhone, Windows Phone, Blackberry
• Wearables (Xowi, Google Glass, Pebble)
• Tablets
• Car
• TV
• Appliances
• Smart Environments (home, office, public places)
• Things
Technologies

Input
• Speech recognition
• Natural language understanding
• Dialog management
• Touch
• Sensors (GPS, accelerometer, proximity sensor)
• Camera
• Handwriting
• Haptics

Output
• One or more displays
• Audio (audio files or TTS)
• Video
• Vibration
Example: Smartphone with speech recognition

Needs
• Speech recognition
• GUI

Other possibilities
• Camera input (recognizing face, objects, gestures...)
• Audio output (TTS or audio files)
• Sensors
Speech Recognition Options

- Google Android
- Google Chrome
- CMU PocketSphinx.js
- Nuance NDEV
- ATT Developer
- OpenEars
- Sphinx
- iSpeech
Consider requirements

- Offline or cloud
- Languages
- What language models are available?
- Are grammars or SLM’s needed
- Cost
- Platform-independence
- Development complexity
Natural Language Understanding

• Component technologies for various tasks
  – Translation (Bing, Google)
  – Part of speech tagging
  – Parsing
  – Named entity recognition

• Open source
  – Stanford NLP tools
  – OpenNLP

• All in one -- text in meaning out
  – WIT ([https://wit.ai/](https://wit.ai/)) API for natural language
Dialog Tools

• OpenDial (open source dialog manager)
  – https://code.google.com/p/opendial/
  – Java-based software for developing robust and adaptive dialog applications using XML rules

• Pandorabots (API)
  – http://www.pandorabots.com/
  – AIML (Artificial Intelligence Markup Language) based framework for

• State Chart XML (SCXML)
GUI Options

- Native iOS, Android, Windows
- PhoneGap (HTML5)
- Appcelerator Titanium
- Web Browser (HTML5)
Tool Tradeoffs: All-in-one vs. Component Tools

- In a mashup world, it’s not so important that all the components of a multimodal application come from the same place.
- There are more options and more flexibility if the application can use components from different vendors.
- But this makes applications more complex.
All in one

• Openstream
  – platform-independent
  – standards-based
  – Cue-me browser and Cue-me Studio

• MIT App Inventor
  – Android
  – Very simple to use
  – Limited capabilities
Example: MIT App Inventor

- A simple app development environment for Android
- Includes speech recognition, typical HTML elements, typical sensors, media capture and playback and many more
- http://appinventor.mit.edu
MIT App Inventor Screen
Standards

• Organizing interaction
  – MMI Architecture
  – SCXML

• Moving information around
  – WebRTC
  – WebAudio
  – WebSockets
  – HTTP
  – Ajax

• Managing Speech
  – SRGS, SISR, SSML

• Representing user input and system output
  – EMMA
  – EmotionML
  – InkML
Many Options

- Must carefully weigh requirements of application
- This area developing very rapidly
- New options are continually becoming available