Combining Voice & Visual in Conversational Apps

Balaji Narayana: Openstream, Inc.
http://www.Openstream.com
balaji@Openstream.com
Multimodal is not new....

- Google introduced voice modality in Version 2 of Android® - several years back
- ...evolved into... a voice modality for aiding web search...
- Fast forward.... Google Now! is a conversational assistant that’s been literally running our lives with several features and utility apps; Assistant maintains conversational and lookup context
... but in-application usage is more complex...
“natural interaction” is beyond just using speech

✓ Managing context beyond speech
✓ Disambiguation via other modalities
✓ Simultaneous multimodality
✓ Mixed-mode, mixed initiative interaction
✓ Driving & deep-linking into the application navigation
And so it follows....

Conversational Applications (interpreted as Natural Interaction Applications) are beyond just Speech and Visual, and incorporate other modalities like gesture, handwriting, keystrokes, camera, and situational context parameters to drive interaction.

Creating an application means merging and working with all these sentient IOs that we take for granted while having a “conversation”.
Wow! If we have to manage all these contexts and modalities to build conversational applications...

.. We need a State Machine that is extensible and malleable when new contexts come in...

... and knows enough about the application context to drive a conversational flow.

Utopia? PMM2?
Segue: W3C® Multimodal Interaction Working Group

• https://www.w3.org/2002/mmi/

Charter
The primary goal of this group is to develop W3C Recommendations that enable multimodal interaction with various devices including desktop PCs, mobile phones and less traditional platforms such as cars and intelligent home environments including digital TVs/connected TVs. The standards should be scalable to enable richer capabilities for subsequent generations of multimodal devices.

Users will be able to provide input via speech, handwriting, motion or keystrokes, with output presented via displays, pre-recorded and synthetic speech, audio, and tactile mechanisms such as mobile phone vibrators and Braille strips. Application developers will be able to provide an effective user interface for whichever modes the user selects. To encourage rapid adoption, the same content can be designed for use on both old and new devices.
W3C® MMI: A slew of standards to enable multimodal application development

• Multimodal-Architecture
• Discovery and Registration of Multimodal Mobile Components
• EMMA – common data exchange format for multimodal annotations
• inkML – XML markup for digital ink traces
• emotionML
Interaction Management is the linchpin

HTML “apps” run in the GUI Modality Component... which can talk to other components via Interaction Manager

IM is the linchpin for Multimodal, Context-Aware interaction

© Openstream, Inc.
Openstream’s MMI based offering: Cue-me™

- SC-XML based IM runs on device
- Placeholder for transformation or synchronization logic for inter-component communication
- Abstracts capabilities as against API or specific features
  - Capabilities (components) provide services
- Framework for context-aware architecture

(C) Openstream, Inc. Prepared for MVC 2016
Example IM Interaction with Voice

<!- User clicked on a link with id “micrButton” -->
<go on="click" from="x-html" node="micrButton" >
<send event="startRecognition" to="x-voice" />
</go>

<go on="RecoResult" from="x-voice" node="event.name"> 
<send event="execute" to="x-html" data="event.value" />
</go>

GUI <x.html> Component

Voice <x.voice> Component

JS { validate ID; Post to web server}
Architectural Advantages

• Clear separation of concerns
• Event driven architecture for context based interaction management
• Trigger to reco can happen from anywhere in the system – controlled by IM... and not the application
• Mix and match reco engines and strategies
• EMMA based annotation for common understanding of modality confidence and parameters
• Control visual navigation flow based on context
• Built completely on W3C® Open Standards and Markups!
Cue-me™: Deploy and manage app and modality resources using a secure managed server offering

- Dynamically manage SC-XML
- Manage modality components and security
- Manage application resources

(C) Openstream, Inc. Prepared for MVC 2016
Development Methodology

Normal Web Dev Practice

Author HTML(GUI) App with service/data sync stubs

Author and Connect HTML app with services and data sync

Author interaction markup for Cue-me™

Test App in Cue-me™ Simulator

Generate Cue-me™ App Package

(C) Openstream, Inc. Prepared for MVC 2016
Preparing for the Postapp Era

The shape of a “world beyond apps” is emerging where mobile apps will be joined by many alternative approaches to providing functions and experiences for mobile users, and rewards for developers and service providers.

“Don’t assume that if the question involves the word “mobile” the answer involves an app. There are many alternatives to traditional apps (as discussed above), and apps can play new roles as components of bigger systems. Ask what the best way to deliver value to a mobile user is, because the answer might be a VPA.......” – Nick Jones, Gartner

EVA’s objective is to deliver a Virtual Assistant foundation for the Enterprise Digital Strategy;

EVA is built on Cue-me™ and is a transformative strategy to take the Enterprise from an “app” model into an Intelligent Assistant Architecture.
EVA Solution Components

**EVA Client**
Secure, multi-modal software Virtual Assistant; Choreographs context for mobile app execution – hybrid or otherwise

**EVA Server: Security, Secure Notifications, Management**
Unified management platform for versioning, deployment, security, notification, and analytics; scalable, distributed, server architecture

**EVA Context/AI Engine**
Server platform for context management, Natural Language Processing and related AI technologies

**Integration Proxy (mBaaS)**
Mobile-ready services, and service optimization

**Studio and Designer**
Design, build, test, debug and deploy Virtual Assistant applications using tooling infrastructure. Import existing

(C) Openstream 2015; Confidential; Under Strict NDA
Danke schön – Questions?

Balaji Narayana

balaji@Openstream.com
http://www.Openstream.com