Beyond the Form Interpretation Algorithm

Standards-based flexible dialog management for conversational voice and multi-modal applications

Rahul Akolkar, IBM Research
Motivation

Browser-based rich speech interaction

• VoiceXML 2.1
  – Form Interpretation Algorithm
  – Dynamic server-side generation and frameworks

• VoiceXML 3.0 requirements
  – Richer dialog management in browser
  – Multiple modalities in browser
  – Align voice and web

• Sample usecases
  – Visit form items in reverse document order
  – XHTML + VoiceXML
Recursive MVC

- Transition controllers at multiple levels of dialog granularity, localized control
  - Form
  - Document
  - Session
  - Application
- DOM Events
  - Well-defined integrations
  - Opportunistic (application level)
- Backwards compatible (VoiceXML 3.0 and 2.1)
Transition Controllers, e.g. SCXML

```
<v3:form
    <scxml:scxml initial="...">
      <!-- Controller -->
    </scxml:scxml>
  </v3:form>

DOM events as glue:

- Initiate execution of a VoiceXML field

```scxml:send target="#vfield" type="DOM" event="DOMActivate"/>

- Notification on filling fields / slots

```scxml:transition event="filled.vfield" cond="vfield == 'foo'"
    target="statefoo"/>
```
Compound Documents, e.g. X+V

```xml
<form action="myForm.do" method="post">
   <input id="myInput" ev:event="onfocus" ev:handler="#voice"/>
   <!-- Rest of the XHTML form -->
</form>

<v3:field id="myField" ev:event="filled" ev:handler="#sync">
   <v3:grammar src="gram.grxml" type="application/srgs+xml"/>
   <v3:prompt>
      What is the value of the input?
   </v3:prompt>
</v3:field>

<script xml:id="voice" type="application/ecmascript">
   document.myField.dispatchEvent('DOMActivate');
</script>

<script xml:id="sync" type="application/ecmascript">
   document.myInput.value = document.myField.value;
</script>
```
Summary

• Alignment of web and voice programming model
  – Increasing complexity may be entrusted to the browser
  – Transition controllers at multiple levels of dialog granularity
  – Flexible network boundary

• Well-defined as well as opportunistic compositions of speech with other modalities
  – DOM Events

• Work being done in the W3C Voice Browser Working Group, discussed at F2F this week
  – VoiceXML 3.0
  – SCXML 1.0