Voice Browser Working Group

Jim Larson
Co-chair, Voice Browser Working Group
Mission

• To **prepare and review documents** related to voice browsers
  – Voice browsers allow people to access the Web using speech synthesis, pre-recorded audio, and speech recognition.

• To serve as a **coordination body** with existing industry groups working on related specifications.

• To serve as a **pool of experts** on Voice Browsers, some of which will participate in the other W3C working groups relevant to Voice Browsers.
W3C Speech Interface Framework

Grammar

Semantic Interpretation

Speech Synthesis

VoiceXML 2.0

Call Control

Other

Other
<menu>
   <prompt>
      Welcome to Ajax Travel. Do you want to fly to
      <emphasis>
      New York
      </emphasis>
      or
      <emphasis>
      Washington
      </emphasis>
   </prompt>
   <choice next="http://www.NY...">
      <one of>
         <item> New York </item>
         <item> Big Apple
               <tag> New York </tag>
         </item>
      </one of>
   </choice>
   <choice next="http://www.Wash...">
      <one of>
         <item> Washington </item>
         <item> The Capital
               <tag> Washington </tag>
         </item>
      </one of>
   </choice>
</menu>
Status of W3C Speech Interface Languages

- **Recommendation**
- **Proposed Recommendation**
- **Candidate Recommendation**
- **Last Call Working Draft**
- **Working Draft**
- **Requirements**

<table>
<thead>
<tr>
<th>Voice XML 2.0</th>
<th>Grammar</th>
<th>TTS</th>
<th>Semantic Interpretation</th>
<th>Call Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>2Q03</td>
<td>1Q03</td>
<td>3Q03</td>
<td>4Q03</td>
<td>4Q03</td>
</tr>
<tr>
<td>1Q03</td>
<td>3Q03</td>
<td>3Q03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1Q03</td>
<td>1Q03</td>
<td>1Q03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3Q03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Follow-on to VoiceXML 2.0
backup
Motivation for Speech Applications

- There are about 10 times as many telephones as connected PCs
- Cell phones usage is growing dramatically
- Speaking and listening are the natural usage modes for phones
What about HTML?

• What is missing from HTML?
  – Tapered prompts
  – Grammar specifying alternative words that the user can speak in response to questions
  – Instructions to the text-to-speech synthesizer about how to say words and phrases

• Adding these capabilities would complicate HTML, a language designed for visual UIs
A variety of dialog styles

- Traditional system-directed dialogs for novice users
- Mixed initiative dialogs for experienced users
- Novice users smoothly become experienced users at their own pace