Mobile Voice Standards: Integrating HTML5 and Speech

Deborah Dahl, Conversational Technologies
dahl@conversational-technologies.com
Daniel Burnett, Voxeo
Michael Johnston, AT&T
Mobile Voice Conference
March 12-14, 2012
San Francisco
Problem

- Speech is a compelling user interface for web applications, especially mobile applications.
- Currently, speech application development requires significant expertise that most Web developers lack.
However

• There are many web developers who are interested in developing speech-enabled applications
• Integrating HTML5 with speech in an easy-to-use and interoperable way could enable web developers to create speech applications
• The World Wide Web Consortium (W3C) launched an effort to explore how to do this – the HTML-Speech Incubator Group
The W3C HTML-Speech Incubator Group

• Chairs: Dan Burnett, Voxeo, Mike Bodell, Microsoft, Dave Burke, Google
• Members represent browser vendors as well as speech companies
  • Voxeo
  • Microsoft
  • Openstream
  • Google
  • AT&T
  • Mozilla
  • Nuance
  • Everspeech
  • Conversational Technologies
  • German Research Center for Artificial Intelligence
Goal: Determine the Feasibility of Integrating Speech Technology in HTML5

• Leverage the capabilities of both speech and HTML (e.g., DOM)
• Provide a high-quality, browser-independent speech/multimodal experience
• Avoid unnecessary standards fragmentation or overlap
• Support both initial exploratory efforts at speech applications as well as robust, enterprise-quality applications
• Tasks
  – Collect and review use cases, requirements, and HTML change requests
  – Present a consolidated summary as a final report, published December 6, 2011
Proposals Include

- JavaScript API for controlling speech recognizers and synthesizers
- Declarative markup that can be included in an HTML5 page
- Protocol for communication between a web page and a remote speech service
JavaScript API for Speech Recognition

- A SpeechReco object with
  - attributes for setting common speech recognition attributes such as
    - Grammars
    - Maximum number of nbest results
    - Language
    - Confidence threshold
    - Timeouts
    - Endpointing
    - Extensions for supporting vendor-specific parameters
  - Methods to control recognition, such as start, stop, abort
  - Events such as “onaudiostart”, “onaudioend” and “result”
JavaScript API: Results

• SpeechInputResult event
  – result (a SpeechInputResult object)
  – transcript
  – confidence
  – interpretation

• SpeechInputResult object
  – EMMA representations of the result, length of the nbest list

There are many other features in the API – see the final report for details!
function speechClick() {
    var sr = new SpeechReco(); // Build grammars from scratch
    sr.grammars = new SpeechGrammarList();
    sr.grammars.addFromUri("http://example.org/topChoices.srgs", 1.5);
    sr.grammars.addFromUri("builtin:input?type=text", 0.5); // Say what happens on a match
    sr.onresult = function(event) {
        var q = document.getElementById('q');
        q.value = event.result.item(0).interpretation;
        var f = document.getElementById('f');
        f.submit();
    };
}
JavaScript API for TTS

- A TTS object with attributes serviceURI, text, and lastMark
- Includes attributes common to all media elements, such as autoplay, preload, etc.
- Supports SSML
Markup

• Markup-based control is also proposed
  – `<reco>` element
  – `<tts>` element
Protocol

• Goal is to enable a web application to utilize the same network-based speech resources regardless of the browser used to render the application
• Enables HTML user agents and applications to make interoperable use of network-based speech service providers,
• Enables applications to use the service providers of their choice
• A sub-protocol of WebSockets
• Borrows some ideas from MRCP v2
Next Steps

• The final report is not a standard, at this point it’s a proposal only
• Many group members are interested in continuing work towards official standardization
• Options for standardization are actively under discussion
  – A new Working Group
  – A Community Group
  – Add work to an existing Working Group, for example, Voice Browser
More Information and Feedback

• Incubator Group home page:
  – http://www.w3.org/2005/Incubator/htmlspeech/

• The HTML-Speech Incubator Group Final Report
  – http://www.w3.org/2005/Incubator/htmlspeech/XGR-htmlspeech-20111206/

• Mailing list for feedback and suggestions
  – public-xg-htmlspeech@w3.org
Editors

- Michael Bodell, Microsoft
- Björn Bringert, Google
- Robert Brown, Microsoft
- Daniel C. Burnett, Voxeo
- Deborah Dahl, W3C Invited Expert
- Dan Druta, AT&T
- Patrick Ehlen, AT&T
- Charles Hemphill, EverSpeech
- Michael Johnston, AT&T
- Olli Pettay, Mozilla
- Satish Sampath, Google
- Marc Schröder, German Research Center for Artificial Intelligence (DFKI) GmbH
- Glen Shires, Google
- Raj Tumuluri, Openstream
- Milan Young, Nuance