

Created
2004-11-25

Updated
2006-10-18

Presentation

Multimodality with XHTML+Voice

An XML application:

Introduction and demonstration of X+V

Alexandre Alapetite

- Informatics engineer (France)
- PhD student at Risø National Laboratory (Denmark)

Multimodal interfaces (1/2)

Definition

- The user can choose between several modalities

Modalities

- Screen, Sounds, Voice synthesis, ...
- Mouse, Touch screen, Keyboard, Voice input, ...

Synchronisation

Multimodal interfaces (2/2)

☛ Advantages

- Robustness
- Flexibility

☛ Drawbacks

- Complexity

☛ How to implement?

- SALT (Speech Application Language Tags)
- **XHTML+Voice**
- ...

What is XHTML+Voice?

✂ For voice interaction on Web pages

✂ Actors: IBM, Opera, Access

– <http://www.ibm.com/pvc/multimodal>

✂ Required notions:

– XHTML

• HTML

• XML

– VoiceXML

– Javascript (ECMAScript)

– CSS

Multimodality with XHTML+Voice

✦ Live demonstration of X+V...

- <http://alexandre.alapetite.net/phd-risoe/mxml/>

HTML

HyperText Markup Language (1995)

- ✦ Used in current 'normal' Web pages
- ✦ Has some semantic

```
<html>  
...  
<h1>This is a title</h1>  
<p>This is a paragraph</p>  
...  
</html>
```

XML

Extensible Markup Language (1998)

- ✖ No native semantic
- ✖ But can be used to fusion several vocabularies within the same document (Namespaces)
- ✖ RDF: OWL, ontologies

XML Namespaces example

```
<foaf:Person>
  <foaf:name>Hans Hansen</foaf:name>
  <foaf:birthday>1970-01-15</foaf:birthday>
  <foaf:based_near>
    <geo:Point xml:lang="da-DK">
      <dc:title>København, Danmark</dc:title>
      <geo:lat>55.7013</geo:lat>
      <geo:long>13.1221</geo:long>
    </geo:Point>
  </foaf:based_near>
  <lang:masters>da</lang:masters>
  <lang:masters>en</lang:masters>
  <lang:speaks>fr</lang:speaks>
</foaf:Person>
```


XHTML

Extensible HTML (2000)

✂ XML document with the ‘meaning’ of HTML

✂ Current standard for Web pages

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
xml:lang="en-GB">
...
<h1>This is a title</h1>
<p>This is a paragraph</p>
...
</html>
```

Which XHTML?

🌟 XHTML 1.0

- Transitional
- **Strict**

🌟 XHTML Basic (Base for modularisation)

🌟 **XHTML 1.1** (Module-based XHTML)

- **Voice**
- MathML, SVG, etc.

🌟 XHTML 2.0 (Future...)

VoiceXML (2000)

- ✚ XML document to design mixed initiative voice interaction systems
 - Speech synthesis, audio
 - Voice recognition, DTMF key input

✚ Telephony

```
<audio src="sound.wav"/>
<prompt>Would you like coffee, tea, or milk?</prompt>
<grammar src="drink.grxml" type="application/srgs+xml"/>
<catch>Please say coffee, tea or milk</catch>
<filled>
  <vxml:if cond="x > 1">
    <submit next="http://mysite/next"/>
  </vxml:if>
  <vxml:else/>
    <submit next="http://mysite/next"/>
  </vxml:if>
</filled>
```

VoiceXML grammars

☛ Two main types

- ABNF: Augmented Backus-Naur Form

- **JSGF:**

- Java API Speech Grammar Format

- JSpeech Grammar Format <http://www.w3.org/TR/jsgf/>

- Yacc: Yet Another Compiler Compiler

- XML

- SRGS: Speech Recognition Grammar Specification

☛ Inline / external grammar

JSGF

Java Speech API Grammar Format (1998)

JSpeech Grammar Format (2000)

```
<vxml:field name="voice_field_news" xv:id="voice_news" modal="true">
  <vxml:grammar><![CDATA[
    #JSGF V1.0;
    grammar menu;
    public <menu> = ([I would like]|[I'd like]|[I want]) to
                    (hear|listen|read) [to] <news> news [please]
                    {$.voice_field_news = $news;} ;
    <news> = BBC | New-York Times {$="newyorktimes";} | Guardian ;
  ]]></vxml:grammar>
  <vxml:prompt>What news to you want to read?</vxml:prompt>
  <vxml:catch event="nomatch noinput help">
    For example, say: I would like to hear BBC news.
    You can say you would like to hear BBC, New-York Times or Guardian.
  </vxml:catch>
  <vxml:filled>
    You will now listen to <vxml:value expr="voice_field_news"/>
  </vxml:filled>
</vxml:field>
```

XHTML+Voice (2001)

- ✂ VoiceXML embedded into an XHTML document
- ✂ That is to say HTML and Voice vocabularies in the same XML document
- ✂ Compatibility with HTML browsers
- ✂ Optimal if combined with other technologies
 - CSS, Javascript, ...

```
<html:p>This is a paragraph</html:p>
```

```
<vxml:prompt>Hello! I am the computer</vxml:prompt>
```

CSS

Cascading Style Sheets (1996)

✦ For the presentation of SGML: HTML, XML

✦ Also for audio

– CSS 2

- Aural style sheets

– CSS 3

- Speech Module

```
h1 {  
  azimuth:far-right;  
  color:red;  
  elevation:above;  
  pause-after:500ms;  
  pause-before:500ms;  
  richness:75;  
  speech-rate:x-slow;  
  stress:75;  
  voice-family: "Paul",male;  
  volume:loud;  
}
```

ECMAScript (1998)

- ✂ Standard Javascript, client side
- ✂ Interaction between VoiceXML and Javascript
 - Variables
 - Events

```
<script type="text/javascript" id="test">  
  function test()  
  {  
    if (document.getElementById('newyorktimes').checked)  
      alert("You have chosen to read the New-York Times");  
  }  
</script>
```

```
<ev:listener ev:event="vxmldone"  
             ev:observer="voice_field_news" ev:handler="#test" />
```


HTTP

Hypertext Transfer Protocol (1996)

✂ Text protocol

✂ HTTP headers: before the document

```
GET /xhtml_voicexml_test.vxml HTTP/1.1
Host: test.server.net
Accept: application/xhtml+xml,application/xhtml+xml,text/html;q=0.9
```

```
-----
HTTP/1.x 200 OK
Date: Mon, 07 Mar 2005 10:47:52 GMT
Server: Apache
Vary: Accept
Content-Length: 3647
Content-Type: application/xhtml+xml; charset=UTF-8
```

```
<?xml version="1.0" encoding="UTF-8"?>
<html>...
```

X+V and Internet technologies

- ✚ XHTML+Voice is text-based
- ✚ All the XML enabled tools can handle it (ex: XSL)
- ✚ Can be generated like normal dynamic Web pages with PHP, ASP, JSP, CGI, etc. server side
- ✚ The data for voice interaction can come from a database
- ✚ Client side programmable with Javascript
- ✚ The voice style can be configured with normal CSS

Move to X+V

- ✦ It is possible to add voice interaction to existing Web pages
- ✦ It does not modify the way classic users can use the page (mouse, keyboard, screen, speakers)
- ✦ Targeted to mobile devices, but will probably be used on desktop computers also
 - Since Opera browser 7.55 (2004) / 8.0 (2005) / 8.5+

Questions?

☛ Thank you for your attention

☛ Questions?

☛ Try by yourself!

☛ Alexandre Alapetite

- <http://alexandre.alapetite.net/phd-risoe/mxml/>