Speech Interface Implementation for XML Browser

Aki Teppo & Petri Vuorimaa
Telecommunications Software and Multimedia Laboratory

Petri.Vuorimaa@hut.fi
http://www.tml.hut.fi/~pv/
Agenda

• Introduction
• X-Smiles XML Browser
• VoiceXML Implementation
• Movie Service Example
• Conclusions
Introduction

- Web content is becoming more popular in different kinds of handheld devices
- Since the display size is often limited different kinds of multimodal user interfaces are an interesting alternative
- XML and - especially - VoiceXML are the most promising markup languages
- In this paper, we present how VoiceXML can be used in X-Smiles XML browser
X-Smiles History

- The XML browser was started as a student software project 1998
  - X-Smiles SMIL-browser
- Support for XSL stylesheet and XML parser was improved during summer 1999
- XSL Formatting Objects, Scalable Vector Graphics, XForms, and Streaming were added during 2000
- Released as open source (www.x-smiles.org) 2001
Some X-Smiles features

- XSL Formatting Objects (XSL FO)
- Synchronized Multimedia Integration Language (SMIL) and streaming
- Scalable Vector Graphics (SVG)
- XForms
- XML Messaging
- Session Initiation Protocol (SIP) client
- Specific Graphical User Interfaces (GUIs)
VoiceXML Implementation

- A special Markup Language Functional Component (MLFC) was made for VoiceXML
- In addition, a separate VoiceXML interpreter was created
- Public domain components were used for text to speech conversion and speech recognition
- Java Speech API was used to connect the components together
VoiceXML Interpreter

- The VoiceXML Interpreter translates the XML content into suitable actions for the underlying speech engines
- We implemented only part of the VoiceXML specification
- Prompt and menu are most important features
Text to Speech Engine

• We used the Festival Text to Speech engine
• Due to a license problem, we had to implement our own Java Speech API for the Festival
Speech Recognition Unit

• We used the Sphinx Automatic Speech Recognition (ASR) library as the speech recognition unit
• The ASR server runs on a separate Linux server
• Dynamic grammars are not supported
Movie Service Example

• We used a movie service as a demonstration service
• The user can browse available movies and get information about them
• Parts of the information is rendered using text to speech engine
• Speech can be used for navigation
XML Sample Data

<movie name="Star Wars" id="star">
  <information>
    When the opening scroll of Star Wars mentions "a galaxy far, far away," it might unwittingly refer to the '70s, a time when "the force" went hand in hand with "the Fonz," and hokeyness ran unchecked.
  </information>
  <picture file="sw.jpg"/>
</movie>
XSL Transformation

<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:template match="/">
    <vxml version="1.0">
      <xsl:apply-templates select="movies"/>
    </vxml>
  </xsl:template>
  <xsl:template match="movies">
    <!-- Creates the main menu -->
  </xsl:template>
</xsl:stylesheet>
VoiceXML Main Menu

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE vxml SYSTEM "voicexml1-0.dtd">
<vxml version="1.0">
  <menu>
    <prompt>Welcome to current movies</prompt>
    <prompt>Select one of:<enumerate/></prompt>
    <choice next="pulp.fo">Pulp Fiction</choice>
    <choice next="fifth.fo">Fifth Element</choice>
    <choice next="star.fo">Star Wars</choice>
    <choice next="sound.fo">Sound of Music</choice>
  </menu>
</vxml>
```
Main Menu

Current Movies

Now in theater:
- Pulp Fiction
- Fifth Element
- Star Wars
- Psycho
- Mulan
- Sound of Music

Ready
Movie Information

Information:
Quentin Tarantino's award-winning homage to dime-store novels is presented in a collector's edition featuring director's commentary and never-before-seen additional footage, including an entire scene with Travolta and mob moll Thurman.

Back
VoiceXML Dialog

**Browser:** Welcome to current movies! Select one of: Pulp Fiction, Fifth Element, Star Wars, Sound Of Music.

**User:** Pulp Fiction

**Browser:** Pulp Fiction – Information – Quentin Tarantino’s award-winning homage to dime-store novels is presented in a collector’s . . . Please select one of: Back

**User:** Back

**Browser:** Welcome to current movies! . . .
Results

• The demonstration run well on Intel Celeron 450 MHz computer with 128 Mbytes of memory
• It did not work well with Intel Pentium II 300 MHz computer with 64 Mbytes of memory
• The text to speech engine started in few seconds, while the speech recognition engine started in about ten seconds after opening a page
Conclusions

• VoiceXML is convenient tool to implement speech based web applications
• XSL Transformations can be used to convert XML based information to VoiceXML
• Integration of VoiceXML to XML browser is possible, but consumes a lot of resources
• Commercial use requires further optimization