Bilingual Chinese/English Voice Browsing based on a VoiceXML Platform

Helen Meng, Yuk-Chi Li*, Tien-Ying Fung*, Kon-Fan Low*,*
Ka-Fai Chow, Tin-Hang Lo*, Man-Cheuk Ho*and P. C. Ching***

***Human-Computer Communications Laboratory,**
Department of Systems Engineering and Engineering Management,
****Digital Signal Processing Laboratory,**
Department of Electronic Engineering,
The Chinese University of Hong Kong,
Hong Kong SAR, China

{hmmeng, ycli, tyfung, kflow, kfchow, thlo, mcho@se.cuhk.edu.hk, pcching@ee.cuhk.edu.hk}

Abstract

We report on the development of English-Chinese bilingual speech applications on a VXML platform. VXML support displayless voice browsing of Web content. We have developed the CU Voice Browser based on OpenVXI 2.0. We have also integrated the voice browser with the OpenSpeech Recognizer (for English speech recognition), CU RSBB (for Chinese speech recognition), Speechify (for English speech synthesis) and CU VOCAL (for Chinese speech synthesis in order to support *bilingual* voice browsing. The CU Voice Browser includes an attribute for identifying the appropriate language for speech input/output, thereby invoking the appropriate speech engine for recognition / synthesis. We have developed two bilingual sample applications – CU Weather and CU News. This paper provides the associated VXML documents that specify these bilingual dialogs for browsing weather and news information respectively.

1. Introduction

Universal accessibility to the Web content offers the convenience of information access at anytime, from anywhere, by anyone and with any device. Web browsing by the Internet population is no longer restricted to desktop personal computers, but is expanding to mobile client devices of diverse form factors such as personal digital assistants (PDAs), WAP phones, smart phones as well as the conventional (displayless) telephone. A recent effort in our laboratory attempts to develop a platform that supports universal accessibility. We refer to this platform as the AOPA (Author, Once Present Anywhere) platform.[1] AOPA is illustrated in Figure 1. We follow standards from the W3C[2] and adopt the Extensible Markup Language (XML) for content specification and the Extensible Stylesheet Language family (XSL) for presentation specification. We also developed the CU Transcoder that can automatically transcode existing Web content written in HTML to XML, in order to be hosted in the unified content repository. XSL automatically transforms the content to suitable layouts for different client devices with different form factors (e.g. medium to small display screens or displayless devices). More specifically, Web browsers for desktop PCs render documents in HTML, mobile mini-browsers for PDAs render documents in HTML3.2, WAP browsers render documents in WML and (displayless) voice browsers render documents in the Voice Extensible Markup Language (VXML) [1]. In each case, the XSL stylesheets automatically repurpose the documents for layouts that suit information visualization using the various client devices.

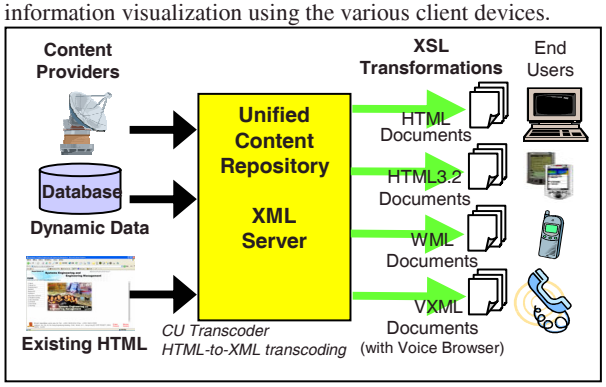

Figure 1. The AOPA software platform aims to support universal accessibility.

This paper focuses on displayless voice browsing of Web content through the telephone channel. The content has been transcoded and transformed into VXML documents and these can be accessed by a voice gateway that connects to the Internet and the public switch telephone network. The voice browser is a software that runs on the voice gateway and can interpret VXML documents. A VXML document specifies a human-computer dialog – human speech input is supported by speech recognition; and computer speech output is supported by speech synthesis/digitized audio. These core speech engines (for recognition and synthesis) are invoked by the voice browser. In addition, the voice browser can also handle touch-tone (telephone keypad) input. Figure 2 illustrates voice browsing of Web content.

Existing voice browsers generally support a single, primary language, e.g. English, for voice browsing. However, since English and Chinese are the two predominant languages in Hong Kong, the objective of this work is to incorporate VXML into the AOPA platform to support *bilingual voice browsing*. Bilinguality refers to English and Chinese in textual form; or English and Cantonese in spoken form. Cantonese is the predominant Chinese dialect used in Hong Kong and is also a major dialect of Chinese.

[1] www.se.cuhk.edu.hk/~aopa
[2] World Wide Web Consortium (www.w3.org)

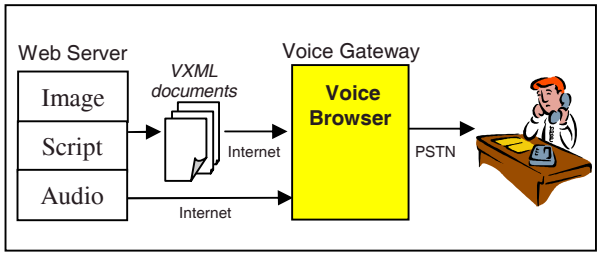

Figure 2. Browsing Web content by voice.
(Reference: www.voicexml.org)

2. Background – Example of a Monolingual VXML Document

This section presents a simple example of an English dialog in the weather domain. The dialog flow is presented in Table 1a and the associated VXML document is presented in Table 1b.

C:	Welcome to CU Weather. Today is 22/11/2003. Please select Hong Kong weather or world weather.
H:	World weather
C:	Please say the city name.
H:	Shanghai
C:	The current weather conditions for Shanghai is…

Table 1a. A human-computer dialog in the weather domain (C:Computer, H:Human).

```
<?xml version="1.0"?>
<vxml version= "1.0">
   <link next="#exit">
<!--This is a link at the VXML level and so the grammar
remains active throughout the document. The user can exit
the dialog by saying goodbye of byebye -->
      <grammar>goodbye | byebye </grammar>
   </link>

   <form id="welcome" scope="document">
     <block>
<!-- The [date] information below is automatically filled in
as XSL transformation takes place to generate the VXML
document from XML content -->
      <prompt> Welcome to CU Weather. Today is [date]
      </prompt>
      <goto next="#type" />
     </block>
   </form>

   <menu id="type" scope="document">
      <prompt> Please select Hong Kong weather or world
      weather </prompt>
      <choice next="#local">
<!--Jumps down to the form with id local below -->
         <grammar>hong kong weather | hong kong<grammar>
      </choice>
      <choice next="#world">
<!--Jumps down to the form with id world -->
         <grammar>world weather | world<grammar>
      </choice>
   </menu>

   <form id="local" scope="document">
<!-- form with id local -->
      <block>
         <prompt> Hong Kong's current weather condition is…
         </prompt>
         <goto next="#type" />
      </block>
   </form>

   <form id="world" scope="document">
<!-- form with id world -->
      <field namelist="other_city">
      <grammar src="eng_city.gram"/>
<!--the list of city names is obtained from the grammar file
eng_city.gram for filling in the variable other_city. -->
      <prompt> Please say the city name </prompt>
<!-- Based on the filled variable, the system retrieves the
relevant weather information. -->
      …….
   </form>

   <form id="exit">
      <block>
         <prompt> Thank you for calling.</prompt>
         <exit expr="success" />
       </block>
      </form>
</vxml>
```

Table 1b: The VXML document that specifies the English dialog in Table 1a. Explanations are boldfaced in the table.

3. CU Voice Browser

The CU Voice Browser (see Figure 3) is developed to support bilingual human-computer dialogs in VXML. The Browser is implemented based on OpenVXI 2.0 [2], a voice browser developed by SpeechWorks and hosted at CMU. OpenVXI contains a task manager (including the VXML interpreter) and five component interfaces. The interfaces are:

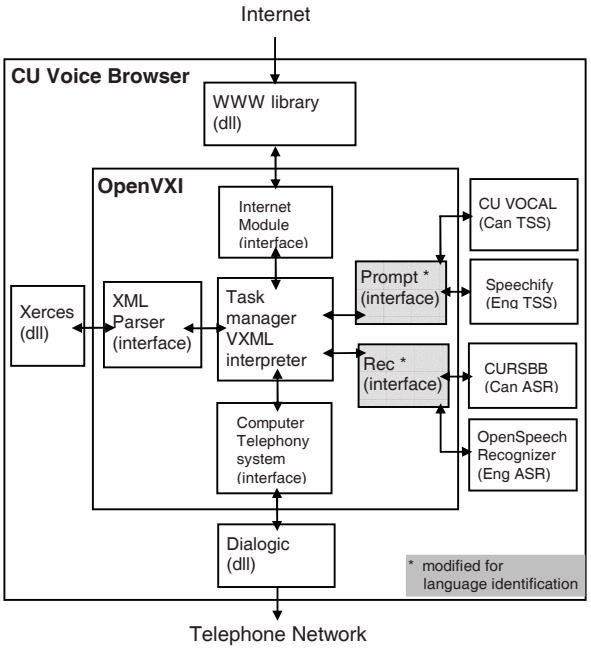

Figure 3. Architecture of the bilingual CU Voice Browser.

(1) an Internet module to fetch VXML documents via HTTP; (2) an XML parser to parse a VXML document into a DOM

object; (3) a computer telephony module for telephone speech I/O; (4) a Prompt interface to pass the VXML prompt text to the corresponding text-to-speech engines (we use CU Vocal [3,4] for Cantonese and Speechify [5] for English); (5) a Rec interface to pass the VXML grammar to the corresponding speech recognition engines (we use CU RSBB [6] for Cantonese and OpenSpeech Recognizer [7] for English). The Dialogic component is a modularized computer telephony interface to handle phone calls. As indicated by the shading in Figure 3, major enhancements in the CU Voice Browser reside in the Prompt and Rec interfaces for bilingual speech recognition and speech synthesis. The CU Voice Browser sets the default language to Cantonese. It also interprets the value of the xml:lang attribute in VXML – the values "en-US" indicates English and "zh-HK" indicates Chinese. Hence the CU Voice Browser can invoke the appropriate (English versus Cantonese) text-to-speech engine as well as to activate the appropriate grammar for speech recognition. If *both* grammars are activated, the Cantonese recognizer will by default be applied first to the speech input. However, if the Cantonese recognizer cannot search for any suitable recognition hypothesis in its lattice, it will reject the speech input and then the CU Voice Browser will pass the speech input to the English recognizer.

4. CU Weather – A Bilingual Application Implemented with VXML

This section extends the example provided in Section 2 to illustrate the bilingual CU Weather application that is implemented with VXML. CU Weather is a research prototype that is available for experimentation.[3] Real-time weather information is sourced from the Hong Kong Observatory websites.[4] The dialog flow is presented in Table 2a and associated VXML is presented in Table 2b. It provides menus for language selection and change, as well as grammar files that list city names in English and Chinese (see Table 3).

C:	歡迎致電中文大學天氣熱線, 今日係 22/11/2003 Welcome to CU Weather, today is 22/11/2003
C:	請選擇語言, 請講廣東話或者英文 Please select language. Please say Cantonese or English
H:	廣東話 *(translation: Cantonese)*
C:	請問你想查詢本地嘅天氣,定係世界城市嘅天氣 *(translation: Would you like local weather or world weather information?)*
H:	世界城市 *(translation: World weather.)*
C:	請講出城市名稱 *(translation: Please say the city name.)*
H:	上海 *(translation: Shanghai)*
C:	上海既天氣而家係…… *(translation: The weather information for Shanghai is...)*
H:	Change language.
C:	請選擇語言, 請講廣東話或英文 Please select language. Please say Cantonese or English

Table 2a. A *bilingual* human-computer dialog implemented with VXML in the CU Weather system.

[3] Callers can dial +852.2603.7884 to experiment with the CU Weather system.

[4] The Hong Kong Observatory websites provide weather information in both English and Chinese for both local and world weather (www.hko.gov.hk).

```
<?xml version="1.0"?>
<vxml version="1.0">
  <link next="#exit">
<!--This is a link at the VXML level and so the grammar
remains active throughout the document. The user can exit
the dialog by saying goodbye, byebye,再見 or 拜拜.
Recognition grammars for both languages are activated
since xml:lang is set to both "en-US" and "zh-HK". -->
    <grammar xml:lang="en-US">
      goodbye | byebye
    </grammar>
    <grammar xml:lang="zh-HK">
      再見|拜拜
    </grammar>
  </link>

  <link next="#chlang">
<!-- Add a link to allow the user to change language at any
time by saying change language or 轉語言-->
    <grammar xml:lang="en-US">change language
    </grammar>
    <grammar xml:lang="zh-HK ">轉語言</grammar>
  </link>

  <form id="welcome" scope="document">
<!-- The xml:lang attribute attached with the prompt tag
specifies the synthesizer to invoke for the delimited text-->
    <block>
      <prompt xml:lang="zh-HK">歡迎致電中文大學天氣
       熱線, 今日係 22/11/2003</prompt>
      <prompt xml:lang="en-US"> Welcome to CU Weather.
      Today is 22/11/2003 </prompt>
<!-- Jumps down to the menu with id chlang below -->
      <goto next="#chlang"/>
    </block>
  </form>

  <menu id="chlang" scope="document">
<!-- menu that supports language selection, id is chlang -->
    <prompt xml:lang="zh-HK">
    請選擇語言,請講廣東話或者英文</prompt>
    <prompt xml:lang="en-US"> Please select language.
    Please say Cantonese or English</prompt>
      <choice next="#chitype">
        <!-- two ways of referring to "Cantonese" -->
        <grammar xml:lang="zh-HK"> 廣東話 | 廣府話
        </grammar></choice>
      <choice next="#engtype">
        <!-- two ways of referring to "English" -->
        <grammar xml:lang="zh-HK">英文 | 英語
        </grammar></choice>
       <choice next="#chitype">
         <grammar xml:lang="en-US">Cantonese
         </grammar></choice>
        <choice next="#engtype">
          <grammar xml:lang="en-US">English
          </grammar></choice>
  </menu>
    …….
  <form id="chiworld" scope="document">
      <field namelist="cityname">
      <grammar src="chi_city.gram"/>
<!-- grammar file chi_city.gram lists city names in Chinese
```

```
that can be used to fill in the variable other_city. -->
        <prompt xml:lang="zh-HK">請講出城市名稱
        </prompt>
<!-- Based on the filled variable, the system retrieves the
relevant weather information. -->
            ……
  </form>

  <form id="engworld" scope="document">
        <field namelist="cityname">
          <grammar src="eng_city.gram"/>
<!-- grammar file eng_city.gram lists city names in English
that can be used to fill in the variable other_city. -->
            <prompt xml:lang="en-US">Please say the city
             name</prompt>
               ……
  </form>
     …….
</vxml>
```

Table 2b: VXML document that specifies the bilingual CU Weather dialog in Table 2a. Explanations are boldfaced.

Except from eng_city.gram
Beijing | Changsha | Los Angeles | New York | San Francisco | Montreal | Tokyo | Sapporo | London | Amsterdam |.....
Except from chi_city.gram
北京 | 長沙 | 洛杉磯 | 紐約 | 三藩市 | 蒙特利爾 | 東京 | 札幌 | 倫敦 | 阿姆斯特丹 |.....

Table 3 Excerpts from the English and Chinese grammar files to guide recognition in both languages in CU Weather.

5. CU News – Voice Browsing with Changing Browsing News Articles by Voice

CU News is another bilingual system that we have implemented with VXML. While the CU Weather domain involves a human-computer spoken dialog that evolves around a static recognition vocabulary (i.e. around 60 city names in both English and Chinese), the CU News domain calls for recognition of dynamically loaded recognition vocabularies based on the news titles. Table 4 shows the partial XML document of the news content as well as the dialog specified by the corresponding VXML that is generated by XSL transformation. During this transformation process, the <newsTitle> tag in XML will be mapped into the <grammar> tag in VXML. Hence the CU Voice Browser interprets these as keywords that need to be recognized – the Browser's task manager signals to the speech recognizer(s) to extract the keywords, perform pronunciation lookup and incorporate the keywords into reloaded speech grammars. Hereafter the speech recognizer(s) will be able to recognize the (dynamically changing) news titles.

6. Conclusions and Future Work

This paper describes our work in the development of English-Chinese bilingual speech applications on a VXML platform. We have developed the CU Voice Browser (based on OpenVXI 2.0) that can integrate with English speech engines (OpenSpeech Recognizer for speech recognition and Speechify for speech synthesis) and (home-grown) Chinese speech engines (CU RSBB for recognition and CU VOCAL for synthesis). We have also developed two bilingual systems – CU Weather and CU News that can serve as sample applications running on the VXML platform. Future work investigates possible extensions to voice markups [8,9] for recognition/synthesis to increase support for Chinese.

XML document (excerpt):	
<news> <newsTitle>楊利偉函謝港人難忘東方之珠風采 *(translation: news story about the first Chinese astronaut Yang Liwi)*</newsTitle> <newsContent> ….</newsContent> </news> <news> <newsTitle>網中人講身份 電子證書「智」合你 *(translation: news story about smart ID cards)* </newsTitle> <newsContent> ….</newsContent> </news> etc.	
VXML-based dialog:	
C:	請選擇以下新聞標題*(translation: please select from the following news titles)* 楊利偉函謝港人難忘東方之珠風采 *(first news title)* 網中人講身份 電子證書「智」合你*(second news title)…*
H:	楊利偉函謝港人難忘東方之珠風采 *(Explanation: first news title is chosen)*
C:	隨同中國首次載人航天飛行代表團 11 月初訪港的航天員楊利偉… *(Explanation: news content of first news story)*
H:	轉語言 *(translation: change language)*
C:	Please select from the following news titles…

Table 4: News content in XML (above). Dialog from the VXML document generated via XSL transformation (shaded).

7. Acknowledgments

This work is substantially supported by the Innovation and Technology Fund from the Hong Kong SAR Government (project code CUHK ITS/117/01). We thank SpeechWorks International Ltd. (now Scansoft Inc.) for their support in providing OpenVXI, Speechify and the OpenSpeech Recognizer.

8. References

1. Voice eXtensible Markup Language (VoiceXML) version 1.0: http://www.w3.org/TR/2000/NOTE-voicexml-20000505/ (www.voicexml.org)
2. OpenVXI: http://fife.speech.cs.cmu.edu/openvxi/
3. Meng, H. et al. CU VOCAL: Corpus-based Syllable Concatenation for Chinese Speech Synthesis across Domains and Dialects. Proc. of ICSLP 2002.
4. Fung, T. Y. and Meng, H. Concatenating Syllables for Response Generation in Domain-Specific Applications. Proc. of ICASSP 2000.
5. Speechify, Scansoft Inc. http://www.scansoft.com/speechify/
6. CURSBB (Chinese University Recognition Software Building Blocks) http://dsp.ee.cuhk.edu.hk/speech/cursbb/
7. OpenSpeech Recognizer, Scansoft Inc. http://www.scansoft.com/openspeech/recognizer/
8. http://www.w3.org/TR/speech-synthesis/
9. http://www.w3.org/TR/speech-grammar/