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 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.

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)
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.

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

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

```xml
<?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 or byebye -->
        <grammar>goodbye | byebye </grammar>
    </link>
    <form id="welcome" scope="document">
        <prompt>Welcome to CU Weather. Today is \[date\] \</prompt>
        <goto next="#type" />
    </block>
    <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">
        <block>
            <prompt>Hong Kong’s current weather condition is… \</prompt>
        </block>
    </form>
    <form id="world" scope="document">
        <field namelist="other_city">
            <!--the list of city names is obtained from the grammar file eng_city.gram for filling in the variable other_city. -->
            <grammar src="eng_city.gram"/>
        </field>
        <prompt>Please say the city name \</prompt>
        .......
    </form>
    <form id="exit">
        <block>
            <prompt>Thank you for calling.\</prompt>
            <exit expr="success" />
        </block>
    </form>
</vxml>
```

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:

- **OpenVXI**
- **Dialogic** (dll)
- **Xerces** (dll)
- **XHTML**
- **XML library**
- **Prompt** (interface)
- **Task manager** (interface)

Figure 3. Architecture of the bilingual CU Voice Browser.

The CU Voice Browser supports bilingual human-computer conversations in VXML. It includes the following components:

- **Internet Module** (interface)
- **XML Parser** (interface)
- **XHTML**
- **Prompt** (interface)
- **Task manager** (interface)
- **Speechify** (Eng TTS)
- **CURSBB** (Can ASR)
- **CU VOCAL** (Eng ASR)
- **Dialogue** (dll)

III - 770
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 Speakify [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. Real-time weather information is sourced from the Hong Kong Observatory websites. 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: 欢迎致电中文大学天气热线，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，定係世界個城市嘅天氣 |
| (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: 上海既天氣而來係…… (translation: The weather information for Shanghai is...) |
| H: Change language. |
| C: 請選擇語言，請講廣東話或者英文 |
| Please select language. Please say Cantonese or English |

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

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

The Hong Kong Observatory websites provide weather information in both English and Chinese for both local and world weather (www.hko.gov.hk).
that can be used to fill in the variable other_city. -->
<prompt xml:lang="zh-HK">請講出城市名稱</prompt>
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>

that can be used to fill in the variable other_city. -->
<grammar src="eng_city.gram"/>