

# VoiceXML-Applications for E-Commerce and E-Learning

## Projektleiter

Prof. Dr. Dr. h.c.mult.  
Peter J. A. Reusch

## Zeitraum

2001–2004

## Mitarbeiter

Bastian Stoll,  
Daniel Studnik

## Förderung

Fachhochschule  
Dortmund,  
Forschungssemester

## Kontakt

Prof. Dr. Dr. h.c.mult.  
Peter J. A. Reusch,  
Fachhochschule  
Dortmund  
Fachbereich Wirtschaft,  
Emil-Figge-Str. 44,  
44227 Dortmund

## Telefon:

(0231) 755-4909

## E-Mail:

peter.reusch@  
fh-dortmund.de

## Abstract

VoiceXML<sup>1</sup> is a language of the W3C<sup>2</sup> to create voice-user interfaces, particularly for the telephone. It uses speech recognition and touchtone (DTMF keypad) for input, and pre-recorded audio and text-to-speech synthesis (TTS) for output. The text-to-speech synthesis feature of advanced VoiceXML tools like WebSphere<sup>3</sup> opens new perspectives for e-commerce and e-learning. We are no longer restricted to pre-recorded audio but can bring any text to the ear of the user – a user that could be visually impaired and needs a voice channel to communicate – or a user who can read but who prefers to listen. VoiceXML-applications have been implemented by the authors to support e-commerce (selection of commodities from catalogues) and user guides for mobile phones. New contributions to e-learning are considered.

## 1. INTRODUCTION

### Voice-enabled applications

Voice-enabled applications are available in several areas, for example:

### Information providers

Voice response facilities are used for various kinds of information over the phone: time, weather, horoscopes, lottery results, sports events, news, exchange rates, and so on.

### Financial institutions

A bank can let its customers access their account balances, obtain information on interest rates and mortgages, calculate loan payments, or transfer funds, all using voice response applications. An application can also call customers to inquire about transactions such as renewing a Certificate of Deposit. Using a voice response application, brokerage firms can make current stock prices, quotations, and portfolio balances available over the telephone. Clients can perform complex transactions without the intervention of a broker. When a broker's advice is required, the application can transfer the call.

### Educational institutions

A voice response application can provide information about class schedules, availability, and course content. Students can register using the telephone, and the application that handles the registration process can also update the database containing enrollment information. A voice response application can call students to inform them of schedule changes

or openings in a class for which enrollment had been closed.

Voice-enabled applications will be improved and expanded significantly in the future, because there is a growing demand – and new technologies are available, that support the integration of voice-enabled functions in more and more applications.

In the past voice-enabled applications were based upon pre-recorded audio. New information models and speech-synthesis tools will boost the development of voice-enabled applications.

This kind of transformation allows new voice-enabled applications and supports visually impaired users significantly for example by offering user guides – an application described below.

## VoiceXML

The Voice Extensible Markup Language VoiceXML is an XML-based language of the W3C to create voice-user interfaces, the latest version is available at: <http://www.w3.org/TR/voicexml20/VoiceXML> is designed to create audio dialogs. Its major goal is to bring the advantages of Web-based development and content delivery to interactive voice response applications

VoiceXML describes the human-machine interaction provided by voice response systems, which includes:

- Output of synthesized speech (text-to-speech).
- Output of audio files.
- Recognition of spoken input.
- Recognition of DTMF<sup>4</sup> input.
- Recording of spoken input.
- Telephony features such as call transfer and disconnect.

The XML root element of a VoiceXML file is `<xml>`, which is mainly a container for dialogs. There are two types of dialogs:

forms and menus. Forms present information and gather input; menus offer choices of what to do next.

Figure 1 shows part of the definition of a VoiceXML menu according to the XML Schema Language XSD, available at <http://www.w3.org/TR/voicexml20/vxml.xsd>

1)<http://www.w3.org/TR/2004/WD-voicexml21-20040323/>

2)<http://www.w3.org/>

3)<http://www.-306.ibm.com/software/websphere/>

4) DTMF (Dual Tone Multi-Frequency) Touch-tone or push-button dialing. Pushing a button on a telephone keypad generates a sound that is a combination of two tones, one high frequency and the other low frequency

```

<xsd:element name="menu">
  <xsd:complexType mixed="true">
    <xsd:choice minOccurs="0"
      maxOccurs="unbounded">
      <xsd:group ref="audio" />
      <xsd:group ref="choice" />
      <xsd:group ref="event.handler" />
      <xsd:group ref="property" />
      <xsd:group ref="prompt" />
    </xsd:choice>
    <xsd:attribute name="id" type="xsd:ID" ?>
    <xsd:attributeGroup
      ref="GrammarScope.attrib" ?>
    <xsd:attribute name="dtmf"
      type="Boolean.datatype"
      default="false" />
    ..</xsd:complexType>
  </xsd:element>

```

Fig. 1 – XML Schema Definition of a VoiceXML Menu

VoiceXML is a key used to transfer text to speech or database entries to speech in a very flexible way. Any text with its index of contents can be transferred from the document file into forms and menus of VoiceXML files that can be read out by text-to-speech synthesis tools like WebSphere.

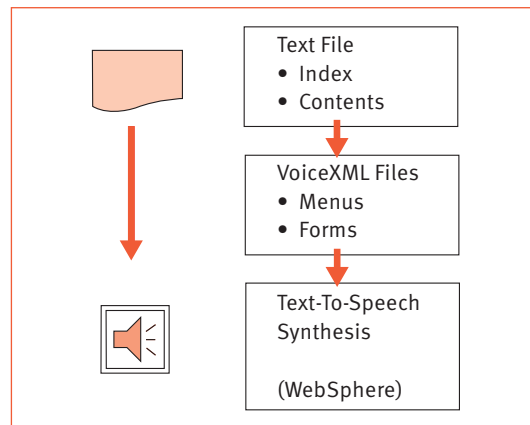


Fig. 2 – Text to Speech Transformation through VoiceXML

The following figure shows the core architecture of VoiceXML applications.

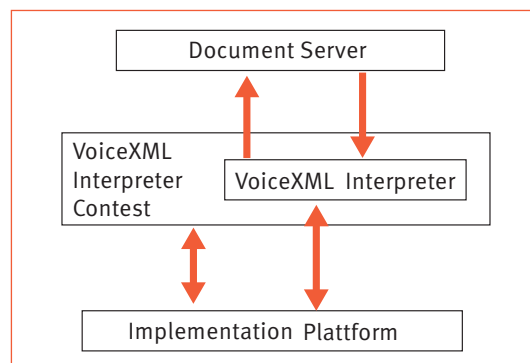


Fig. 3 – Architectural Model of VoiceXML

A document server processes requests from a client application, the VoiceXML Interpreter, through the VoiceXML interpreter context. The server produces VoiceXML documents in reply, which are processed by the VoiceXML Interpreter. The VoiceXML interpreter context may monitor user inputs in parallel with the VoiceXML interpreter. The implementation platform is controlled by the VoiceXML interpreter context and by the VoiceXML interpreter.<sup>5</sup>

### 3. VoiceXML-based USER GUIDE FOR A MOBILE PHONE

All kinds of user guides can be transferred into voice-enabled applications. Here a voice-enabled user guide for the Siemens SL55 mobile phone will be described.

The following figure shows the mobile phone Siemens SL55. The user guide is available in a printed version attached to the product, a user guide is also available on web sites<sup>6</sup> and can be downloaded.<sup>7</sup> It is no problem to study this user guide – 73 pages – if you can read. If you are visually impaired, you need others that can help – or a tool that can transform the document to speech.



Fig. 4 – Siemens SL55 Mobile Phone

The authors have implemented a voice-enabled user guide for this mobile phone. VoiceXML vxml-files include the information and menus that allow the user to navigate through the user guide.

The following figure shows the files of the user guide, starting with the vxml-file of the main menu – main\_menu.vxml – and other vxml-files with specific information on selected items. For all vxml-files there are corresponding grammar files.

5) <http://www.voicexml.org/specs/VoiceXML-100.pdf>

6) <http://communications.siemens.com/cds/frontdoor/0,2241,hq>

7) [http://communications.siemens.com/repository/169/16990/s155/userguide\\_3\\_comaucacnh-kieisrllibmanzpaksgzatwthaeu-kusy\\_e\\_eng.pdf](http://communications.siemens.com/repository/169/16990/s155/userguide_3_comaucacnh-kieisrllibmanzpaksgzatwthaeu-kusy_e_eng.pdf)

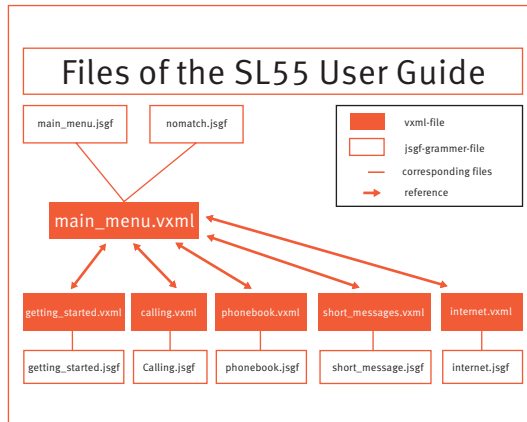


Fig. 5 – Files of the VoiceXML-based User Guide for the Siemens SL55 Mobile Phone

The following figure shows a part of main\_menu.vxml .

```

<form id="selection">
<field name="main_menu">
<grammar src="main_menu.jsgf"></grammar>
<option dtmf="1" value="getting_started">
Getting started</option>
<option dtmf="2" value="Calling">
Calling</option>
<option dtmf="3" value="Phonebook">
Phonebook </option>
<option dtmf="4" value="Short_Messages">
Short Messages </option>
<option dtmf="5" value="Internet"> Internet
</option>
<option dtmf="6" value="Exit"> Exit </option>
<prompt bargein="true">

```

Welcome to the voice-controlled user guide for your mobile phone SL 55.  
If you are familiar with the system you can get direct access to the main menu. Please interrupt me now and tell me, what kind of help you need. Below you will find a menu with 6 choices. In this menu the most important control elements for your mobile phone are explained.  
You are free to switch between the menu items, whenever you like. If you would like to listen to some menu items again, just tell me.  
Now, take your time and listen to the instructions carefully.  
I suggest you listen to the getting started instructions first.  
There you get an overview of your mobile phone.

Fig. 6– Part of main\_menu.vxml

This file includes a text to welcome the user who is listening to the user guide and it includes the menu for the first choices:

- Getting started
- Calling
- Phonebook
- Short Messages
- Internet
- Exit

The menu is also included in the corresponding grammar file

```

#JSGFV1.0 iso-8859-1;
grammar main_menu;
public <main_menu> = Getting started | Phonebook | Calling | Short Messages | Internet | Exit ;

```

Fig. 7 Grammar file main\_menu.jsgf

The VoiceXML user guide was implemented on IBM’s Websphere. The WebSphere Voice Server includes<sup>8</sup>:

- A speech-recognition (ASR) engine (speech-to-text software) that detects and recognizes words that are spoken over a telephone, then passes those words as text to an application.
- A text-to-speech (TTS) engine (voice-synthesis software) that synthesizes speech from application text for play-back over a telephone.
- Concatenative to TTS text.
- Connection to WebSphere Voice Response for AIX.
- Development tools, including support for applications that are written in VoiceXML, Java™ Beans, or C.

The following figure shows the structure of the WebSphere Voice Server.

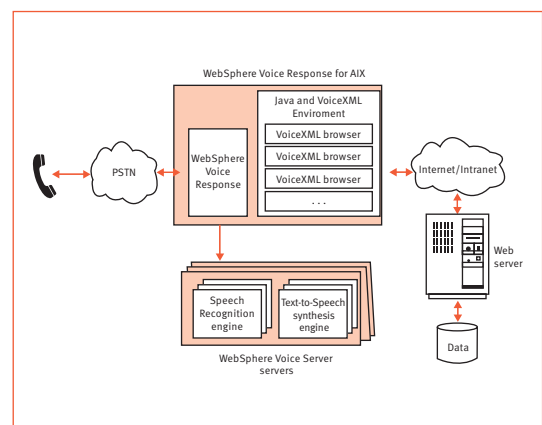


Fig. 8: WebSphere Voice Server<sup>9</sup>

8) [http://www-306.ibm.com/software/pervasive/voice\\_server/about/](http://www-306.ibm.com/software/pervasive/voice_server/about/)

9) <http://publibfp.boulder.ibm.com/epubs/pdf/c3463792.pdf>

And the next figure shows the screen of WebSphere while running the voice-enabled user guide for SL55. The user interaction is done by microphone and loudspeaker.

#### 4. VoiceXML FOR E-LEARNING

Voice-enabled applications can support e-learning in many ways. They can open e-learning systems to visually impaired users. As mentioned before – Fig. 2 – any document can be transferred through VoiceXML to speech-synthesis tools and the user can listen to it.

An equally important aspect is the opportunity to navigate through the "document" using VoiceXML menus that can be derived from the index of the original document automatically. This means that visually impaired users will be able to access a lot of e-learning applications in the future.

VoiceXML applications can be used through computers at home or in computer labs, but also through the telephone – without any computer on the part of the user. The e-learning application in this case is based upon a server with access through the telephone. In our days of telephone flat rates we can sit anywhere with our telephone and can listen to e-learning contributions. On the other hand voice-enabled applications add new channels for everyone. We have our senses.

We can see, hear, taste, touch and smell. Our ears can support e-learning. Today almost all speech components of e-learning systems are based upon pre-recorded audios. We use audiobooks and "talking" dictionaries. Using VoiceXML we can bring much more to the ear for everyone.

#### 5. CONCLUSION

VoiceXML applications can be established in an efficient way – also in new areas like user guides or e-learning. The applications are very flexible and can be derived from existing documents or data bases. The VoiceXML applications are available through computers or through the telephone, which is a universal means of communication.

#### 6. REFERENCES

- [1] Reusch, Peter J. A.; Reusch, Pascal: Classifications of Products and Services to Support Business Process Engineering and e-Commerce, International Scientific Journal of Computing, Vol. 2, Issue 2, 133–140, 2003.
- [2] Stoll, Bastian: Anwendung von VoiceXML bei sprachbasierter Arbeit mit Katalogen erstellt mit IBM Websphere Studio, Diploma Thesis, Dortmund 2004.
- [3] Studnik, Daniel: Erstellung einer sprachgesteuerten Bedienungsanleitung für ein Mobiltelefon mit der Programmiersprache Voice XML 2.0, Project Thesis, Dortmund 2004.

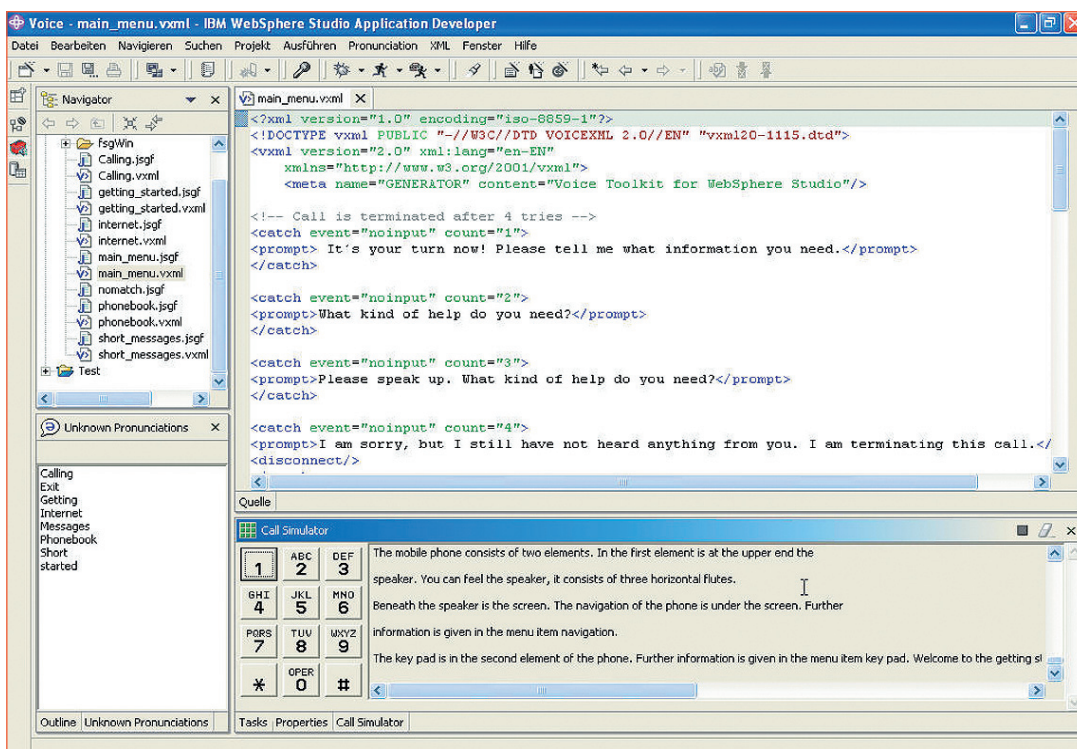


Fig. 9: WebSphere – SL55 User Guide Running