

→ (VoiceXML)

Design of Voice Automated System

With Java and Voice Extensible Markup Language

Bhanu S. Tanty

Triangle Java User Group

June 15, 2009

Before beginning...

- Thanks for the opportunity to share
- Feel free to interrupt with Questions/Comments
- This will hopefully be of interest to you!

Introduction

- What is VoiceXML
- Architecture

Element Set

- Overview of VoiceXML tags

Trijug Voice Automated System

- Overview

Tools

- Most popular tools
- Demo of BeVocal Café

What is VoiceXML?

XML-based markup language for creating distributed voice applications.
Enables phone access to internet hosted content

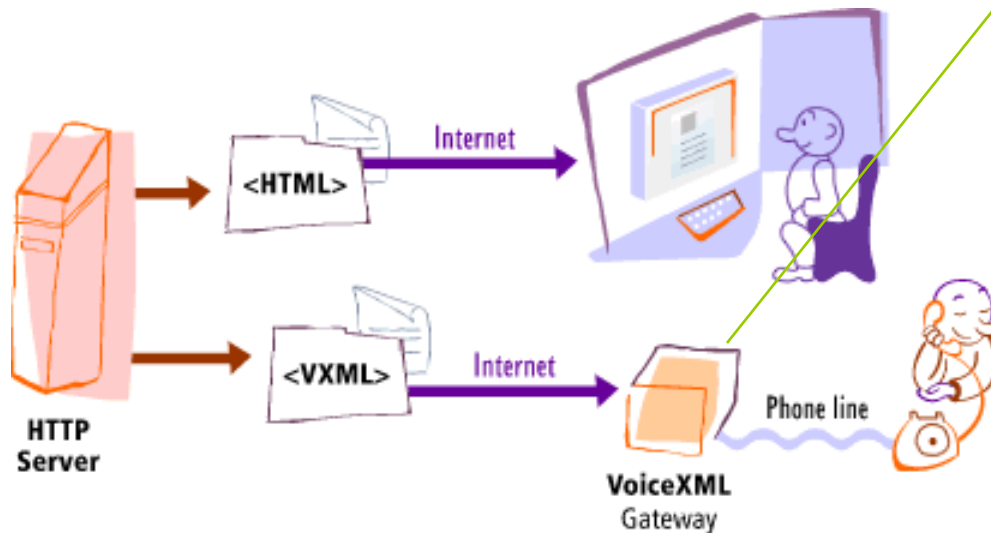
Enabling Technology:
Automated Speech
Recognition (**ASR**)

Your application can **speak to the user via synthesized speech** or by pre-recorded audio files. Your software can **receive input from the user via speech** or by the tones from their telephone keypad

Enabling Technology:
Text to Speech (**TTS**)

Network Architecture

The user interacts with a Web site over the phone using a **VoiceXML Browser**, which is hosted on a **Gateway**. Instead of rendering and interpreting HTML, the VoiceXML Browser renders and interprets VoiceXML.



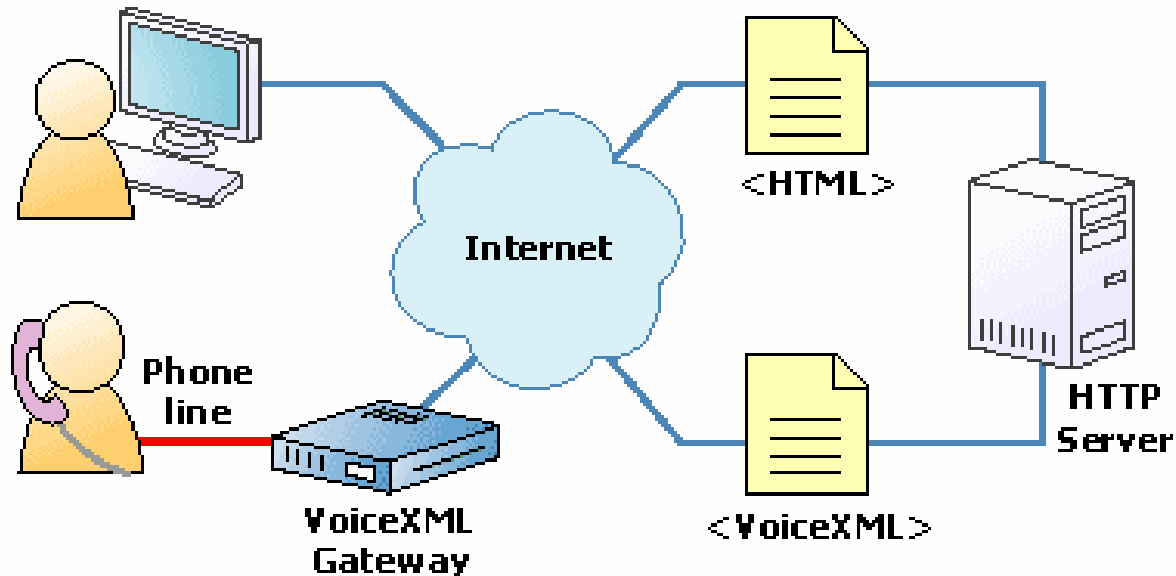
The Gateway is the key bridge technology, responsible for:

- VoiceXML Browser
- ASR Resource
- TTS Resource
- Telephony Resource
- Audio Resource
- TCP/IP Resource

→ (VoiceXML)

Network Architecture

- Trijug Phone : **206-462-3588**



Document Architecture

A VoiceXML document can contain static information, or it can be generated dynamically from data stored in an enterprise database using the same type of server-side logic (CGI scripts, ASPs, JSPs etc.) that is used to generate dynamic HTML documents

Analogous to HTML

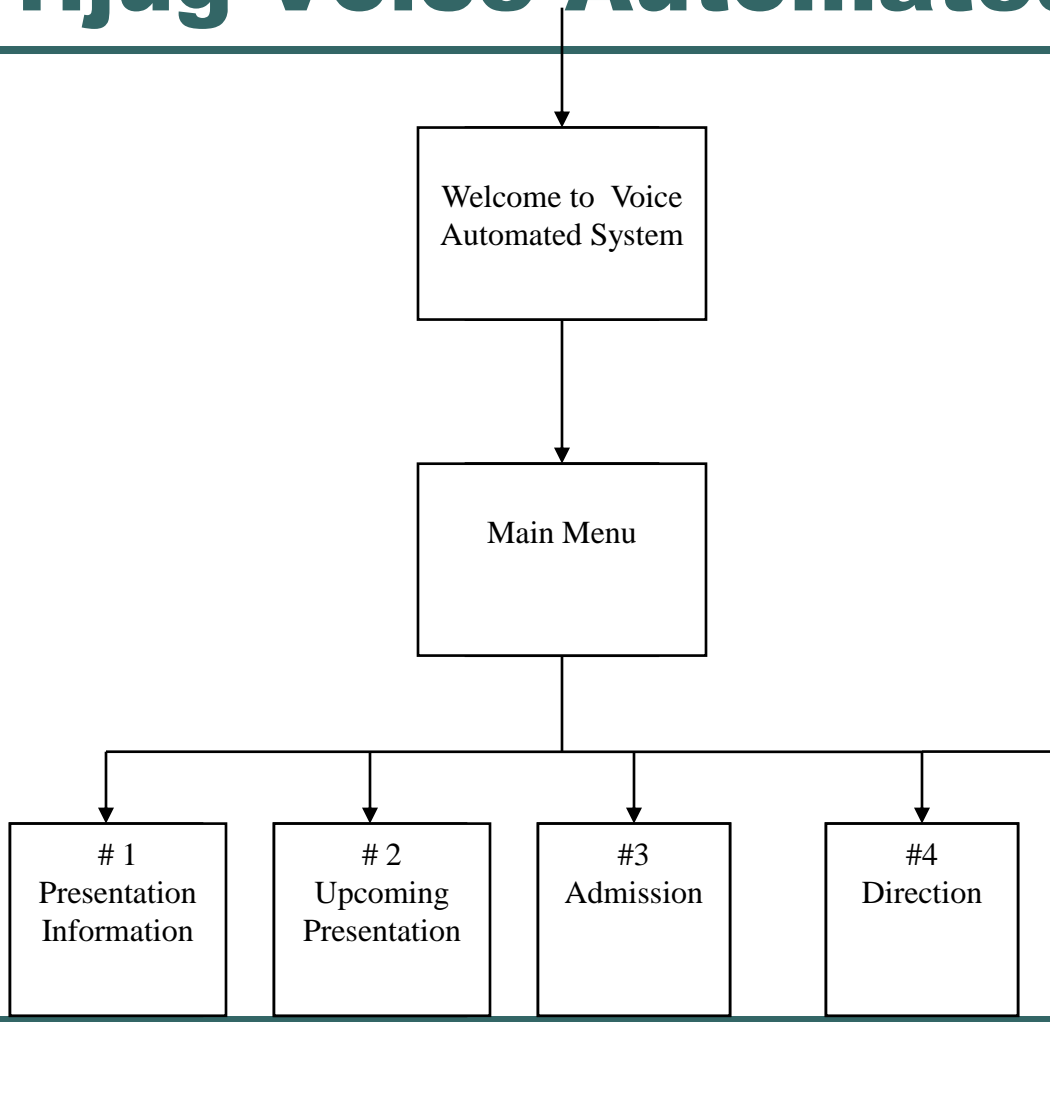
- Dialog \equiv HTML Page
- Supports ECMAScript/Javascript
- Interface with Back-end Infrastructure (CGI etc)

47 VoiceXML Tags

- <form>, <field>, <submit>
- <prompt>
- <noinput>
- <nomatch>
- <help>
- <filled>
- <grammar>
- ...

Trijug Voice Automated System

→ (VoiceXML)



→ (VoiceXML)

Trijug Voice Automated System

- Calling TRIJUG Voice Automated System:
- To try out the application, Phone Number : **206-462-3588**
- Once the authentication check is passed, the Voice XML interpreter runs the interactive Trijug Voice System.

→ (Voice)

Welcome Page

- Hello! Welcome to Triangle java user group Automated Voice System.
- The TriJUG is an official Java Users Group based in Research Triangle Park, North Carolina. TriJUG has been established to provide Java related information to its members, to facilitate contacts between its members, and to promote the use of Java technology. TriJUG is a non-profit, vendor-independent organization. You can also visit us any time at [http:// www.trijug.org](http://www.trijug.org).

Main Menu Page

- This is the Main menu.
- For Current Presentation Information, Press 1.
- For information about upcoming presentation, press 2.
- For information about membership press 3.
- For Meeting Direction press 4.
- For information about Board members, press 5.
- To end this call press 6 or say exit at any time

→ (VoiceXML)

TRIJUG VOICE AUTOMATED SYSTEM

Phone Number : 206-462-3588

```
<?xml version="1.0" ?>
<!DOCTYPE vxml PUBLIC "-//BeVocal Inc//VoiceXML 1.0//EN"
"http://cafe.bevocal.com/libraries/dtd/vxml1-0-bevocal.dtd">
<vxml version="1.0">
  <menu accept="approximate">
    <prompt> Welcome. Please choose one of the following
      <enumerate/>
    </prompt>
    <choice next="test1.vxml">one</choice>
    <choice next="test2.vxml">two</choice>
    <choice next="test3.vxml">three</choice>
  </menu>
  <help>I'm sorry. There's no help available here.</help>
  <noinput> I'm sorry. I didn't hear anything.
    <reprompt />
  </noinput>
  <nomatch> I didn't get that.
    <reprompt />
  </nomatch>
</vxml>
```

Grammars specify what will be accepted as proper user input

Two types of grammars: DTMF and Voice

- DTMF Grammars specify what buttons the user can push on a touch-tone phone
- Voice Grammars specify correct spoken input
- Use the `<dtmf>` and `<grammar>` tags, respectively
- Same grammar formats are used for both

There is no standard grammar format – yet.

Vendor-specific formats

- Nuance's Grammar Specification Language (GSL)
- Java Speech Grammar Format (JSGF)
- Augmented BNF (ABNF)
- BNF
- XML Grammar Format

VXML 2.0 – standard based on JSGF

Example - Grammar

External:

```
<grammar type="application/x-jsgf" src="my.gram" />
```

Inline:

```
<grammar type="application/x-jsgf" >  
  <![CDATA[<  
    one | two | three  
  ]]>  
</grammar>
```

VoiceXML Tools & Deployment

● Development Tools

- Voxeo Designer
- HeyAnita FreeSpeech DT
- IBM WebSphere Studio

● Libraries: grammars, audio files, and reusable code.

- Tellme
- BeVocal
- HeyAnita

Online Tools

- BeVocal Cafe
(<http://cafe.bevocal.com>)
- Tellme Studio
(<https://studio.tellme.com/>)
- Voxeo
(<http://evolution.voxeo.com/>)
- Voice genie
<http://developer.voicegenie.com/>

Deployment Options

- Developer Accounts: Free. Good for small applications
- ASP Model: Development and hosting are outsourced
- Gateway Model: Development done in-house, but hosting is outsourced
- In-House Model: Development and hosting done in-house

Summary

Pros

- Simple to use
- Extending the Web to Phone
- Reuses existing back-end infrastructure

Cons

- Voice is not always a good interface – large data sets
- Speech Recognition is not always correct
- Standards are still emerging

- Voice access to web content over the phone allows for virtually *anytime, anywhere* access to information.
- It removes the need for a *client* computer, thereby allowing the deployment of cheaper mobile applications than applications based on technologies such as WAP, i-mode etc.
- It is well suited for applications involving small data sets – Flight Info. etc