

**IBM Pervasive Computing**

November 2002



## **Extending an Enterprise with IBM WebSphere Voice Application Access**

by **Eddie Epstein**  
**WebSphere Voice System Architect**

---

**Contents**

---

**Why Voice Access is Needed 3**

*Voice access as a competitive advantage 3*

*Voice access as a personal requirement 3*

**Why the time for Voice Access is Now 3**

*Impact of VoiceXML 4*

**Building on WebSphere Portal Server 4**

**IBM WebSphere Voice Application Access 5**

**Voice Portals versus Visual Portals 7**

**The Voice Application Interaction Model 9**

**Roadmap for Future Functions 9**

*Support of 3rd Party Speech Technologies 9*

*VoiceXML 2.0 10*

*Multimodal Applications 10*

*Mixed Initiative Application Design 10*

**Summary 10**

## **Executive Summary**

Many business enterprises are providing mobile computing access for their employees, customers, and business partners. They understand the need for their employees to have access to office services such as e-mail and calendars, as well as current business information. They also realize a strong competitive advantage in giving their partners and customers access to the information they need, when and where they need it.

Voice access to computers has become a preferred interface. The availability of high quality automated speech recognition and speech synthesis technologies, combined with lower cost and higher performance hardware, make automated voice access feasible for most applications. What is particularly important is that most applications can be multichannel in nature: providing voice access in addition to the traditional visual interface.

Tight integration between voice and visual access technologies requires a combined deployment infrastructure. Delivering voice access with a positive return on investment (ROI) calls for a technology that reuses existing infrastructure and organizational skills. To succeed, the right support structure is essential. This includes:

- *Code sharing between voice and visual applications*
- *Common application development tools and a common technical approach that allows the same skill set to support both types of applications*
- *Where possible, packaged solutions that deliver functions from Day One*
- *Integrated tools to enable a development team to mobilize existing enterprise assets, including tools and code samples for developers to use in building a solution*
- *Easy-to-deploy solutions based on open-standard technologies to protect technology investments*

This white paper is intended for solution architects and developers who want to provide voice access for their mobile employees and customers, as well as those already developing interactive voice response solutions. This information can assist those who want to move to open standards-based application development, better integrate with visual self-service applications, and increase their overall organizational efficiency. It introduces *IBM WebSphere® Voice Application Access* and its capabilities. It also describes the issues involved in creating voice applications.

---

## Highlights

---

*Companies that offer automated voice access to their employees, business partners and customers benefit from an immediate competitive advantage*

*Some disabled employees can't use visual interface devices. For them, voice access is a fundamental requirement to doing their job*

*Recent breakthroughs in the use of concatenative text-to-speech technology have resulted in voice quality comparable to human speech*

*Providing voice access is much more than just voice-enabling a visual interface*

## Why Voice Access is Needed

### Voice access as a competitive advantage

Many mobile employees spend a considerable amount of time in cars or in other venues where a voice telephone (wired or cellular) is the only viable means of communication -- and the only way to access remote information sources. As self-service access to business applications becomes essential to more and more jobs, automated voice access becomes a key requirement. It is estimated that around half of cellular phone calls originate from automobiles. For a large segment of the professional workforce, the mobile phone has opened up hours of weekly commuting time for productive business purposes. Companies that offer telephone access derive a competitive advantage over those that do not.

While new mobile computing devices offer remote access, their small visual displays and limited input capabilities often result in a frustrating and tedious experience. For example, the selection of items from a long list or menu is much more efficient by voice, simplifying actions such as finding a name in an address book, selecting a date on a calendar, or finding a note with a specific subject line.

### Voice access as a personal requirement

Some employees with disabilities are not able to use visual interface devices, and others can't use input devices dependent on fine hand control. For these individuals, voice access is more than a competitive advantage, it is a fundamental requirement to doing their job.

Providing voice access is much more than just voice-enabling a visual interface; it requires a basic redesign of an application for conversational interaction.

## Why the time for Voice Access is Now

A major stumbling block for the voice interface has been the unnatural and difficult-to-understand nature of computer generated voices. Recent breakthroughs in the use of concatenative text-to-speech technology has eliminated this limitation and resulted in voice quality comparable to human speech. Speech recognition accuracy has also continued to improve, so that millions of people daily use their voice to "dial" phone numbers by saying a person's name, manage their investment portfolios, and access weather information, sports scores and other information. In addition to technology improvements, the steady refinement of conversational dialogue design has

## Extending an Enterprise with IBM WebSphere Voice Application Access

### Voice Access as a Competitive Advantage

---

#### Highlights

---

*VoiceXML is an open standards-based voice application design protocol that is supported by all major speech technology suppliers*

*VoiceXML enables WebSphere Voice Application Access to integrate voice interface capabilities*

*Portals serve as simple unified access points to Web applications*

resulted in a much more efficient and pleasant user experience than was provided by earlier voice activated systems.

Advances in hardware have also made it possible to deploy automated support for large numbers of simultaneous callers without large capital investments. In particular, the cost of CPU processing power, memory, and telephony interface cards have been falling by the rule of Moore's Law.

#### **Impact of VoiceXML**

The last critical piece to fall into place has been the availability of *VoiceXML*, an open standards-based voice application design protocol that is supported by all major speech technology suppliers. This standard was designed to allow voice applications to run on all enterprise-quality computer hardware and operating system platforms. Companies can be sure that their investment in a *VoiceXML* application infrastructure won't lock them into a single supplier for critical system components.

Voice application development had traditionally required a variety of skills, knowledge and programming techniques, including:

- *Specific Integrated Voice Response (IVR) application development environments*
- *Interfacing between specific IVR environment and middleware applications*
- *Using speech recognition and speech synthesis technologies*
- *Conversational design*
- *Middleware design*

VoiceXML was introduced specifically to eliminate the need for proprietary IVR application design environments, to automatically provide the integration to middleware using the view-and-form based model of Web application design, and to create a standardized interface to speech recognition and speech synthesis technologies. VoiceXML enables WebSphere Voice Application Access to integrate voice interface capabilities in the same way WebSphere Portal Server applications are built on HTML and WML. These protocols provide a modular application design environment with common components sharable across all access modalities.

### **Building on WebSphere Portal Server**

Today, portals serve as a simple, unified access point to Web applications. Portals provide a runtime

## Extending an Enterprise with IBM WebSphere Voice Application Access

### Voice Access as a Competitive Advantage

---

#### Highlights

---

*IBM WebSphere Voice Application Access combines the modular application design of IBM WebSphere Portal Server with VoiceXML to add voice access to WebSphere Portal Server*

*By building on VoiceXML, the growing community of voice application developers can leverage the voice application access platform*

*VoiceXML customers can choose from leading speech recognition and text-to-speech offerings*

platform and tools that give a consistent presentation view across multiple pages, navigation control to access applications, and personalized selection and customization of content. The *WebSphere Portal Server* infrastructure (<http://www.ibm.com/software/webservers/portal/>) accomplishes this by providing functions that:

- *Provide access to information across a spectrum of users, devices, and customization options.*
- *Integrate and automate business processes*
- *Build, connect and manage applications*

Pervasive portal offerings are part of a new generation of applications designed to obtain information and execute transactions from a variety of remote access devices.

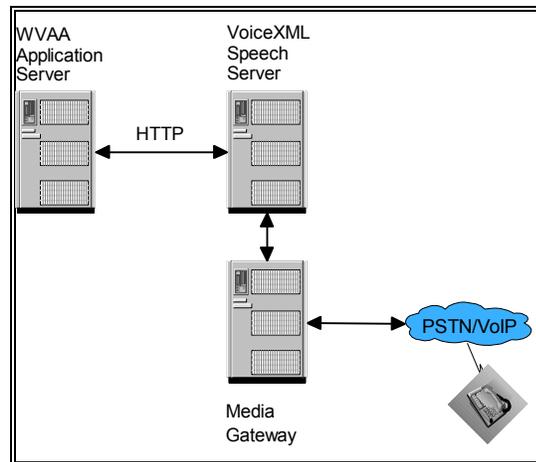


Figure 1 -- Speech Application Components

In addition, the portal platform is ideal for supporting both voice and visual access through a common personalization store-and-shared business logic.

## IBM WebSphere Voice Application Access

Most existing automated voice solutions have been created using proprietary voice application environments combined with custom interfaces to back-end business logic and data. These custom interfaces are difficult to integrate with traditional GUI Web access solutions. IBM WebSphere Voice Application Access combines the modular application design paradigm of IBM WebSphere Portal Server with VoiceXML (<http://www.w3.org/Voice/>) to add voice access to the other modalities supported by WebSphere Portal Server. By building on VoiceXML, not only is the growing community of voice application developers able to directly leverage the voice application access platform, but platform customers should be able to choose between leading speech recognition and text-to-speech offerings.

---

Highlights

---

*The VoiceXML browser works with ASR and TTS engines to interpret spoken input and generate voice output*

*In order to interpret voice input, ASR engines use active vocabularies that identify all the words which can be recognized*

*Tools facilitate building grammars and pronunciations, and create good quality conversational call flows*

Figure 1 shows the main components of an application implementation using IBM WebSphere Voice Application Access. As with traditional visual interfaces, application logic and the generation of presentation markup is done in the application server middleware. VoiceXML markup is delivered to a speech server stack including a VoiceXML browser and underlying automatic speech recognition (ASR) and text-to-speech (TTS) technologies. A media gateway such as IBM WebSphere Voice Response is required to provide connectivity with the telephone network.

### IBM WebSphere Voice Application Access Runtime Components

More detail on the application and speech technology software is shown in Figure 2. Individual portlets deliver VoiceXML markup to the Voice Aggregator, which creates complete VoiceXML documents including support for a global main menu. Markup is sent to a compliant VoiceXML browser using standard HTTP connectivity. The VoiceXML browser works with ASR and TTS engines to interpret spoken input and generate voice output. The browser can also accept DTMF (telephone keypad) as input and use prerecorded audio

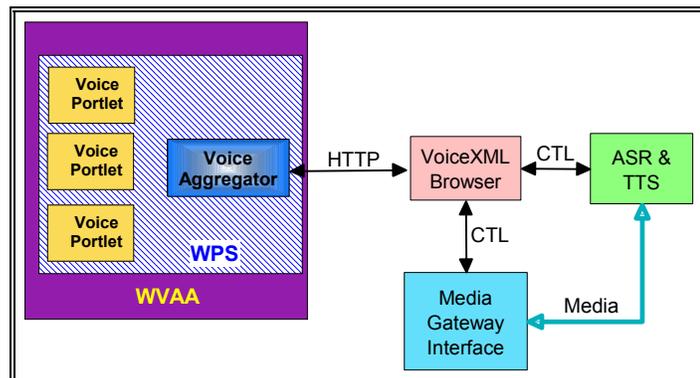


Figure 2 -- Application and Speech Software

files for output.

In order to interpret voice input, ASR engines use active vocabularies that identify recognizable words. These vocabularies also specify allowable word sequences; this combination of vocabulary and specific word ordering is called a speech recognition *grammar*. Each word in a grammar is represented by a spelling, but it is actually the word's pronunciation that is used by the ASR engine. Although both ASR and TTS speech technologies have large dictionaries of word pronunciations, applications will often use words or abbreviations outside the dictionary that require the definition of new pronunciations.

## Extending an Enterprise with IBM WebSphere Voice Application Access

### Voice Access as a Competitive Advantage

---

#### Highlights

---

*WebSphere Portal Server and WebSphere Application Server form a modular, multichannel application development environment*

*The WPS Personalization facility is one of many important features*

*Voice Aggregator is a runtime module that exposes voice applications and implements voice navigation menus*

*The Toolkit creates and debugs voice aggregation, voice portlets and customizes WVS speech technologies*

*Interfaces for WebSphere Voice Response and other selected connection environments are also included*

*Prepackaged portlet applications are part of the system install or available by download*

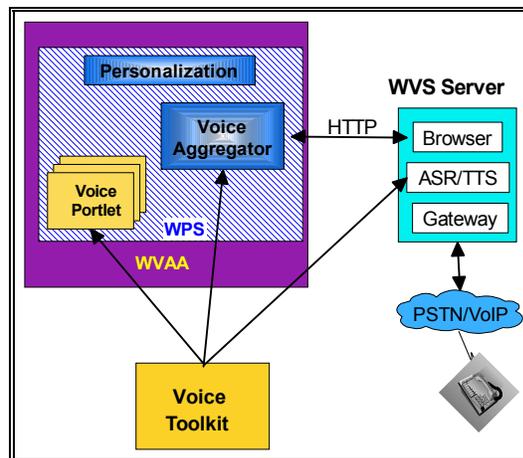


Figure 3 -- The full package. PSTN connection hardware is not included.

Tools are needed to facilitate building grammars and pronunciations, as well as to create good quality conversational call flows.

Figure 3 is a system view of voice application access with server, runtime, application and tool modules. This full package is made up of the following components:

**WebSphere Portal Server** and WebSphere Application Server form a modular, multichannel application development environment. The WPS Personalization facility shown is only one of many important features.

**Voice Aggregator** is the runtime module that exposes the voice applications enabled for the current user and implements the voice menu needed to navigate to a specific application.

**Voice Application Access Toolkit** are the tools for creating and debugging voice aggregation, voice portlets and customizing WVS speech technologies.

**WebSphere Voice Server** provides the VoiceXML browser and underlying speech technologies. Interfaces for WebSphere Voice Response and other selected connection environments are also included.

**Prepackaged portlet applications.** Functional and sample voice application portlets are included in the system installation or available by download. The key functional portlets are:

- Lotus Notes R5 access to e-mail, calendar and contact information
- Microsoft Exchange 2000 access

Both of these portlets rely on the tight security features built into WebSphere Portal framework and extended by Voice Application Access for voice login.

Sample portlet code will also be included that provides a travel example with flight arrival and departure information.

---

#### Highlights

---

*Most people follow spoken dialogs more easily than written instruction*

*Prompts lead users through a conversation step by step*

*People's spoken input is channeled to match up with the computer's recognition strengths*

*Voice interfaces designed for telephony access have evolved significantly other the past few years*

Other portlets and sample code to be available in future releases or available by download include:

- *Stock quote sample to query the current trading price*
- *Weather sample to get a local forecast from a nationwide forecasting service*

#### Voice Portals versus Visual Portals

A voice interface -- such as those provided with IBM WebSphere Voice Application Access -- has significant advantages over visual applications in a portal. Graphical user interfaces (GUIs) tend to have a large amount of text on every screen that can saturate the user. Most people follow spoken dialogs more easily than written instruction. Perhaps the biggest advantage is dialogue focus -- prompts lead users through a conversation step-by-step.

On the other hand, in natural conversations people answer even simple questions in a large variety of ways, often outside the scope of the question. For example, they may answer a question and then explain their answer. Designing automated systems to be able to "understand" most of these arbitrary inputs would generally be quite complex and impractical. Consequently, it is important to channel people's spoken input to match the computer's voice recognition strengths.

Voice interfaces designed for telephony access have evolved significantly other the past few years based on the experiences of many application deployments. Some of the most important things learned include:

- *Conversational flow must be efficient, consistent and intuitive. Use confidence scores to avoid confirming every entry; make sure navigational commands are consistent throughout all applications in the portal; conversational flow should "make sense" to most users.*
- *Prompts must be carefully crafted, short but not ambiguous. It should be clear to most users exactly what to say to the system.*
- *Help prompts must be short. Users can simply not remember much more than one piece of information per prompt.*
- *The system should "reveal itself" to users at appropriate times. Context-dependent help can be used when the conversation bogs down and shortcuts can be offered when things are going well, to help users learn the system incrementally.*

Another difference between visual and voice interfaces is portal navigation. Visual portal design is based on the concept of a page, but most users don't care

---

Highlights

---

*Applications may be put on a page because they fit well together visually, but a different organization makes more sense in a voice menu*

*Many voice targets may be implemented as shortcuts rather than normal menu choices in order to keep prompts short*

*Voice application access provides a superior application design environment for directed dialogue applications*

*Tools facilitate iterative implementation and debugging, and teach best practices for conversational design*

that components on a page are made from different portlets. The navigational problem for visual portlets is finding the page that contains the right application. In order to support a large number of applications, the portal can group similar pages into a *page group*.

These visual concepts are not useful to a voice interface. While there will be some overlap -- for example, major categories may be the same between visual and voice -- the navigational menu structure for voice is likely to be quite different for several reasons:

- *There will be some visual-only portlets and some voice-only portlets*
- *Applications may be put on a page because they fit well together visually, but a different organization will make more sense in a voice menu.*
- *Many voice targets may be implemented as shortcuts rather than normal menu choices in order to keep prompts short.*

In other words, a voice interface is much more than simply a voice enablement of a visual interface.

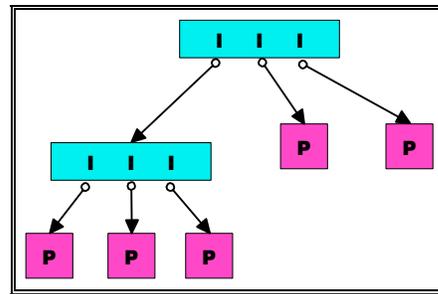


Figure 4 -- Directed Dialogue Menu Structure. Vertical bars represent menu choices managed by the aggregator, and P's individual voice portlets.

## The Voice Application Interaction Model

The majority of voice applications will be directed dialogue designs, as these are the simplest to create and in many cases the easiest to use. Directed dialogue designs are controlled by the automated system, offering a specific set of choices. This paradigm, also known as system initiative, is the easiest for users to learn, but for complex applications it can be inefficient and tedious. Mixed initiative dialogue designs allow both the system and the user to take control of the dialogue as appropriate.

Because the majority of voice access applications will be directed dialogues, IBM WebSphere Voice Application Access provides a superior application design environment for directed dialogue applications. The emphasis is on tools that facilitate iterative implementation, debugging and enhanced designs, using best practices for conversational dialogues.

---

## Highlights

---

*As the VoiceXML 2.0 specification nears final approval, IBM WebSphere Voice Application Access will quickly move to support it*

*IBM WebSphere Voice Application Access combines with other IBM portal offerings to create a platform for multimodal applications using server based voice technology*

*Mixed initiative applications allow more efficient voice access*

*Businesses can combine current technologies in new and exciting ways*

*This product provides the development tools and technologies needed to facilitate fast design and deployment*

*Now is the time to consider enabling enterprise applications with an open standard voice interface*

## Roadmap for Future Functions

### Support of 3rd Party Speech Technologies

Support of popular vendor speech technologies is a critical product objective. Building on the VoiceXML standards allows IBM WebSphere Voice Application Access to work with third party browsers and their associated underlying speech recognition and text-to-speech technologies.

### VoiceXML 2.0

As the VoiceXML 2.0 specification nears final approval, IBM WebSphere Voice Application Access will move quickly to support it.

### Multimodal Applications

New wireless networks and devices capable of supporting both voice and data channels require multimodal applications that combine the power of voice and visual interfaces. IBM WebSphere Voice Application Access is designed to combine with other IBM portal offerings to offer a platform for multimodal applications using server based voice technology.

### Mixed Initiative Application Design

Mixed initiative applications allow more efficient voice access. These applications are characterized by multiple transactions, which are often related to each other, and transactions requiring input to multiple fields. Future versions of IBM WebSphere Voice Application Access will support such application designs.

## Summary

IBM WebSphere Voice Application Access integrates voice channel access into the WebSphere application platform. The architecture is based on well-accepted open standards for both Web and voice application design. This allows businesses to choose their implementation from a large variety of computer hardware and operating system software, combining many current technologies in new and exciting ways. Furthermore, this product provides the development tools and technologies needed to facilitate fast design and deployment.

IBM WebSphere Voice Application Access has enabled the addition of voice access to information by enterprise employees, business partners and customers. Now is the time to consider enabling existing applications with an open standard voice interface.

**Extending an Enterprise with IBM WebSphere Voice Application Access**  
Voice Access as a Competitive Advantage

For More Information

For more information about IBM WebSphere Voice Application Access, in North America please contact 1-800-Talk-2Me, otherwise e-mail [pervasive@us.ibm.com](mailto:pervasive@us.ibm.com)

WebSphere Portal Server

<http://www.ibm.com/software/webservers/portal/>

IBM WebSphere Portal Server with VoiceXML

<http://www.w3.org/Voice/>

# Extending an Enterprise with IBM WebSphere Voice Application Access

## Voice Access as a Competitive Advantage

© Copyright IBM Corporation 2002

IBM Corporation  
Route 100  
Somers, NY 10589  
U.S.A.



Produced in the United States of America  
11-02  
All Rights Reserved

IBM, the IBM logo, and WebSphere are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Linux is a registered trademark of Linus Torvalds.

Other company, product and service names may be trademarks or service marks of others.

This paper discusses strategy and plans, which are subject to change because of IBM business and technical judgments.

All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only.

References in this publication to IBM products or services do not imply that IBM intends to make them available in any other countries.

Performance results obtained in other operating systems may vary significantly from your results. There is no guarantee that these measurements will be the same on your systems.