Speaker Verification Extensions for WebSphere Voice Server

Enhancing Security in Speech Applications

Wendi Nusbickel

Jiri Navratil

Ricardo dos Santos

May 2006

Abstract

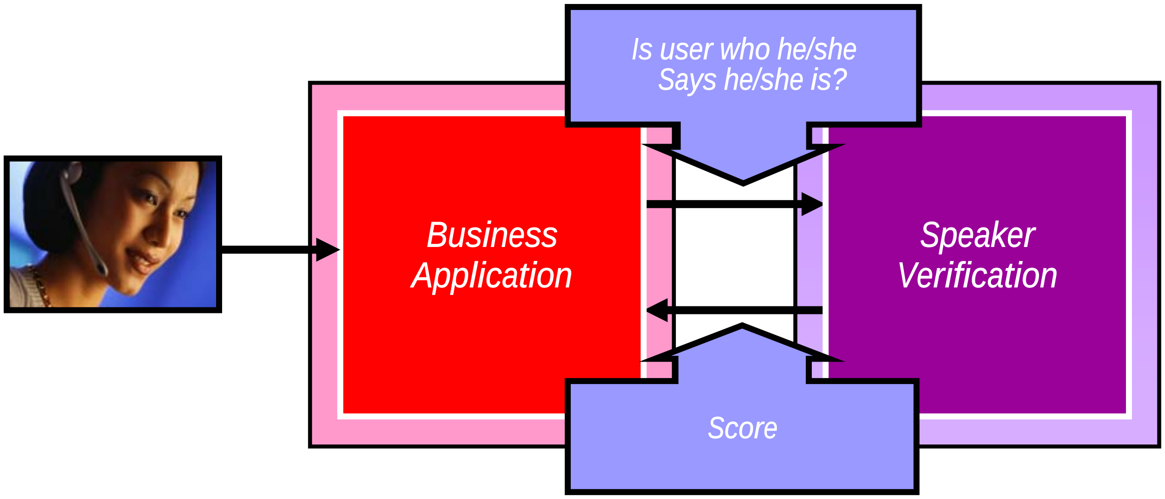
Secure access to sensitive data over the telephone has always been a challenge faced by companies providing speech solutions. From security questions asked by live agents to sophisticated automated voice authentication systems, a range of solutions is available in the market today. This document gives an overview of how IBM is incorporating speaker verification into its speech offerings and how easily customers can benefit from its award winning technology.

Overview

You can use voiceprints to verify a user with speaker verification.

When it comes to providing secure access for speech applications, a wide array of options are available in the market today, from Q&A sessions conducted by live agents to more automated solutions based on touch-tone PINs, or password matching based on traditional speech recognition technology. All these solutions are prone to fraud, as the authentication mechanism is based on information that can be provided by anyone. *How easily can someone's password be stolen these days?*

Conversational Biometrics (CB) technology enables a non-intrusive and highly accurate mechanism for authenticating users, based on the analysis of their voice. CB technology can be incorporated into speech solutions to provide much more accurate and secure applications. The ability to authenticate someone's identity based on his/her voice is referred to as *speaker verification*. It significantly reduces the risks of unauthorized access, since the authentication mechanism is based on the unique features of someone's voiceprint (much like a fingerprint is in the tactile sense).



A New Way to Authenticate WebSphere

The addition of speaker verification provides WAS with a new means for authentication – using voice.

IBM's speaker verification integrates into the WebSphere Application Server (WAS) product called WebSphere Voice Server (WVS). Speaker verification in WVS takes speech in, and matches against an enrolled voiceprint for authentication.

Not only is the WVS speaker verification engine implemented in 100% Java. It leverages the highly scalable and robust platform of Java 2 Enterprise Edition (J2EE). Because it runs inside of WAS, it brings all the WAS benefits to speaker verification, including: reduced deployment costs with integration into the IT infrastructure, central/common management, advanced system monitoring, increased reliability, and simplified problem determination.

Just as WAS brings benefits to speaker verification, the reverse is also true. WVS speaker verification brings many benefits to WAS. Identity theft is the number one crime in America today. Speaker verification lets the customer feel they can sign up for a service with confidence that someone else is not going to get into their account. The ability to use voice for authentication brings numerous advantages. Essentially, it adds an extra layer of protection to sensitive information. For example, if a user's account ID and password are stolen, the imposter would be detected by the system when he tries to get in pretending to be someone else. The use of voiceprints increases the reliability of identity verification and makes it much more difficult for someone to break into a user's account.

IBM's Speaker Verification Technology

The technology behind the speaker verification feature of WVS provides customers with a competitive edge.

IBM's Speaker Verification technology offers significant advantages. It provides a grammar, language, and text independent authentication mechanism. You can enroll saying anything, in any language, and have it verify you, saying anything, in any language! Some of the benefits of the speaker verification feature of WVS include:

- Language Independence
    - This means that one speaker verification engine can handle all languages.
    - Speaker can enroll in one language and be verified in another
- Text Independence
    - User can say anything, not bound by a grammar.
- Speaker Tracking
    - Continuously monitor entire call for assurance that verified speaker answered all prompts.
- Speaker Change Detection
    - Can alert when different speaker is present in call (For example, a person calls in but then boss takes over the conversation.)

Your speech application can take advantage of this flexibility and provide a truly integrated and non-intrusive verification process. Since anything you say as part of a transaction dialog can be used to verify your identity, there is no need to remember pass-phrases or go through a separate verification process. For instance, you can prompt for an account number, have it recognized and the caller verified within the same dialog.

IBM Speaker Verification technology was ranked top among numerous worldwide participants in recent evaluations conducted annually by the US National Institute of Standards and Technology (NIST).

IBM Speaker Verification technology was also ranked top across all test conditions in a recent evaluation conducted by the *Bundesanstallt fuer Fernmeldestatistik* (BFST), Germany comparing numerous participants from the industry. The test conditions included landline, and cellular telephony as well as radio channels.

IBM has over 60 patents and 30 papers associated to its Speaker Verification technology, including a patent selected as one of Five Killer Patents by the MIT Technology Review Magazine (May/2004 issue).

Additionally, the Speaker Verification feature of WVS can be complemented by a sophisticated policy manager. Many customers provide their own policy management. IBM also provides a policy manager component which compliments the speaker verification feature of WVS.

Use of Standards

With WVS, IBM continues its commitment to standards, using J2EE, MRCP, and the W3C Speech Interface Framework.

In today's industry, the use of open standards has proven to be a driving force towards lowering solutions costs. This is particularly true in the speech business, where applications are more and more based on a vast collection of industry standards.

Standards can be used both inside (internal) and outside (external) of a product. IBM's speech server of WVS with WAS provides both.

Internal standards reflect how a product was architected. A product based on Java 2 Enterprise Edition (J2EE) has a plethora of standard J2EE facilities at its disposal. Typically less code needs be developed when using such a framework, thus saving development expense. A J2EE product uses a common means to integrate to a J2EE provider, like WAS. When a product is architected to a standard, it is very flexible. WVS with speaker verification is architected on J2EE and WAS.

External standards directly affect the possible users of it. When a product supports a standard interface, it reduces the customer's risk of using the product. The customer isn't locked into a specific product or proprietary technology.   Media Resource Control Protocol (MRCP) is the standard interface for accessing speech resources in a speech server. MRCP is the standard interface to use WVS. MRCP defines speech resources including speech synthesizers, recognizers, and verifiers. An example of actions one can perform for speaker verification via MRCP include: start/end of a verification session, query/delete/create a voiceprint, verify, and get an intermediate result.

Along with MRCP WVS supports the W3C Speech Interface Framework of standards. This includes standards such as voice grammars (SRGS), and speech markup (SSML).

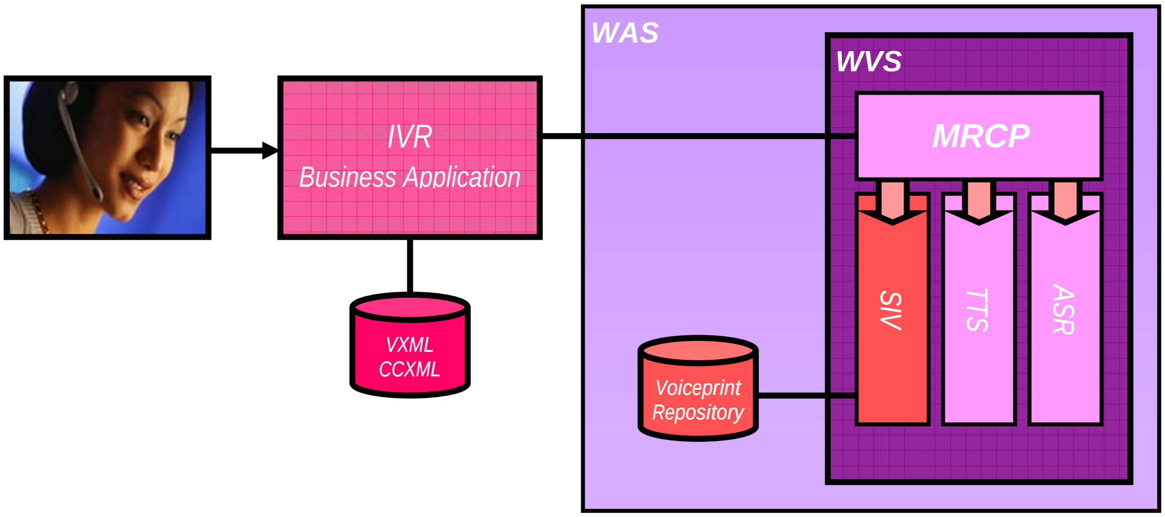
Speaker Verification for WebSphere Voice Server

IBM offers a new feature for its WVS product, called ‘speaker verification’.

IBM WebSphere Voice Server confirms IBM's commitment to support, adopt, and drive open standards. It ties together more than 30 years of worldwide speech research and technology expertise with the infrastructure provided by the IBM WebSphere platform.

WebSphere Voice Server builds on the base of other IBM WebSphere products to provide scaling, load balancing, failover, recovery, systems management, logging, tracing, and problem determination consistent with the IBM WebSphere family.

The WebSphere Voice Server (WVS) plays a major role as the foundation for IBM speech solutions. It provides a robust and scalable platform, leveraged from a premier J2EE based architecture provided by Websphere Application Server. The following picture shows how the major WVS components interact with your business application.



Essentially, WVS is a speech server that provides access to speech resources (ASR/TTS engines) through a standard MRCP interface. This functionality can now be extended by enabling the Speaker Verification Extensions, which adds an extra speech engine (SIV) for speaker verification.

The Application Model

The speaker verification feature of WVS is easy to use via standard VoiceXML.

The programming model for voice applications is Voice eXtended Markup Language (VoiceXML). VoiceXML is part of the W3C Speech Interface Framework. Voice servers such as WVS are accessed from a VoiceXML browser, executing a Voice XML application. The latest approved version of VoiceXML (2.1) does not include support for speaker verification. As a stopgap measure, IBM provides a simple web interface to the speaker verification feature of WVS.

Summary

This paper documents IBM's commitment to speech and standards. With conversational biometrics, IBM offers a new way to authenticate with WebSphere – using your voice. It highlights the *speaker verification* feature of WebSphere Voice Server. The technology behind this feature is a leader in the industry. And it can be easily integrated into your web-based voice applications using VoiceXML. IBM continues to show its leadership and strength in the WebSphere line, through this new feature in WVS.