



Criteria for Selecting a Speech Recognition Vendor

Introduction

Given that, you have already decided to use speech recognition technology within your call center as a means of cutting costs and increasing caller satisfaction, your next step is to determine which vendor or vendors you'll use to implement a solution. What are the criteria that you should use to make your selection? One factor to consider is, of course, cost. Another is the quality of the core technology – that is, how accurate is the speech recognition and how large a vocabulary can it support? But today, with many of the top speech vendors now boasting accuracy rates in the high ninety percent ranges for many tasks, and with the ability to recognize grammars containing hundreds of thousands and even millions of words, the core technology is no longer sufficient to differentiate vendors.. So what are the things you need to consider before choosing a speech vendor?

In this paper, we discuss some of the criteria that can be used to select a speech recognition vendor. We also provide a table that compares the four top vendor's call center speech recognition technology, Nuance Communications, Scansoft, IBM Pervasive Computing, and Microsoft, according to these criteria.

Core Technology Features

In recent years, speech recognition vendors have added a number of features to the core technology. Many of these features have greatly expanded the types of applications that can be built with speech applications, the caller populations that can benefit from the technology, and the environments (e.g., cellular) in which the technology operates. Depending upon the application needs of your business, some of the features you may want in a speech recognition engine include:

- Expanded language coverage - some speech recognition vendors provide extensive functionality in a single or very few languages. Other vendors provide components supporting numerous languages. When choosing a speech recognition engine, it is important to consider not only the language(s) offered, but the dialect of the languages (for example, U.S. English versus Australian English), since recognition accuracy will drop if the dialect does not match that of the caller population. Something else to consider is whether your application requires simultaneous recognition in multiple languages and if so, whether the engine supports that capability.
- Background noise elimination – with the continued push of hands free technology, especially in the mobile telecommunications area, background noise reduction is mandatory for the best recognition performance.
- Conversational natural language support – integration of statistical language model (SLM) technology into some vendor products allows callers to speak in unconstrained language. This has enabled the creation of applications such as call routers.



- Speaker verification – the ability to authenticate a caller’s claim of identity based upon biometric qualities of the voice. This adds a level of security to an application not possible with just a PIN. Depending upon the vendor, the languages supported for speaker verification, may or may not coincide with those supported for speech recognition. Some speaker verification engines can run independently of any speech recognition engine, while others required a speech recognition engine in order to operate.
- Hotword recognition - this allows the application to listen for specific control words ignoring other speech input
- Dynamic grammars – the ability to create and compile grammars in real-time enables applications with caller-specific data or those in which the caller options change frequently

Standards

Solutions that adhere to published standards or open specifications offer several advantages to the customer. These benefits include:

- Interoperability of components from different vendors. This gives customers the ability to mix components from different vendors thus obtaining the best overall solution.
- The ability to migrate easily from one platform to another
- The ability to leverage investments in existing web and IT infrastructure, thus reducing costs

Several standards and open specifications exist that are relevant to speech applications. The most important are:

- VoiceXML - a mark-up language used to create voice-user interfaces, especially for the telephone.. It is based on the Worldwide Web Consortium's (W3C's) Extensible Markup Language (XML), and leverages the web paradigm for application development and deployment. VoiceXML simplifies application development by taking advantage of existing web infrastructure. It facilitates code portability and reuse. It supports voice and DTMF inputs and text-to-speech and recorded prompt outputs.
- SALT – a speech-interface markup language that supports both speech-only and multi-modal applications.
- SRGS – The W3C standard for specifying the recognition grammar, that is, the set of the words, phrases, and/or sentences that the recognizer will be able to understand
- xHMI – a language used to define the high level dialog control



- SISR – for assigning meanings to recognized words, phrases, or sentences. (For instance, in many applications, the words “yes”, “yeah”, and “yep” would all be assigned the meaning “yes”)
- SSML– the W3C markup language used by application developers to control aspects of text-to-speech output such as pronunciation, volume, pitch, speed, and intonation.
- Media Resource Communication Protocol (MRCP) - a mechanism used to control speech recognition and text-to-speech servers in distributed environments. This allows distributed IVR platforms to be built.

Speech Development Tools

In recent years, speech technology vendors have begun to offer tools and methodologies that shorten deployment time and while increasing the cost-effectiveness of implementing speech technology applications. These tools simplify many of the tasks associated with implementing speech applications including: voice user interface design, grammar development, application coding, prompt creation, and tuning and analysis. Such tools may be essential to those businesses that decide to build and/or maintain their own speech recognition applications. Types of application development aids include:

Reusable Modules

Many vendors are now offering entire pre-built applications or application components. Often of these modules encapsulate the best practices of voice user interface, thus promoting well-designed applications. Additionally, use of these reusable pieces simplifies development and debugging. One thing to consider when evaluating re-usable modules is whether they are generic or were built for a particular application or industry. In general, the more generic the module, the more customization is required, and the more programming and speech expertise is required.

- **Low-Level Components.** These are relatively low-level building blocks that can be used to construct applications. For instance, a development toolkit might include components for capturing dates, times, money amounts, etc. Each component could include a call-flow, pre-recorded prompts, tuned grammars, application code, and/or tuned text-to-speech for prompting. These are generic in nature.
- **Shrink-Wrapped Applications.** These are already built applications that address a specific need (such as call-routing or directory assistance), or a targeted industry (e.g., banking). Pre-packaged applications generally require only a small degree of customization. Depending upon their complexity they can be deployed in anywhere from a few days to several months.
- **Application Kits.** These fall somewhere between low-level components and shrink-wrapped applications. Application kits consist of sets of components that have already been customized for a particular application or industry. They require less work to use the low-level



components but still may require programming and/or customization to meet the application needs. Application kits can decrease the time required to develop a speech application by as much as fifty percent.

Development Tools

If you decide to build or maintain your speech applications in-house rather than have them built for you by the speech vendor, then you should consider also purchasing a toolset designed to aid in building speech applications. Some of the things to consider in choosing a toolset include:

- Whether the tool supports all the tasks necessary to deploy a speech application including: dialog design, grammar design and building, usability testing, application coding, recognition testing, and prompt creation.
- How well the tool supports integration into back-end, web infrastructure and legacy systems.
- What type of development environment(s) does the tool provide? Depending upon their background, some programmers may prefer a graphical interface, whereas, others may prefer the ability to program in the native programming language (e.g., VXML).

Logging and Monitoring Tools

Unlike most software development, speech recognition applications require one or more post-deployment tuning cycles. Logging and tuning tools facilitate this task by providing event logging, viewing, and report as well as analysis tools.

Other Considerations

Some of the other factors to consider when deciding upon a speech vendor include:

- The Operating Systems which the software runs on
- Compatibility with IP telephony – this is important if you are migrating towards IP.
- The scalability of the solution – how many ports can the software support?

Conclusion

Speech recognition technology can increase your call center effectiveness by cutting costs and enhancing caller satisfaction. However, in today's market, with so many speech recognition vendors available, the decision about which technology to choose is a difficult one. Depending upon your business and application needs, there are a number of factors which should be considered when selecting a vendor. We hope that this paper serves to help you in that selection process.



Speech Recognition Engines Comparison Matrix

Nuance 8.5, Scansoft OSR 3.0, Microsoft Speech Server, IBM WebSphere Voice Server

Vendor:	Nuance	ScanSoft	Microsoft	IBM	
Core Speech Recognizer	Product	Nuance 8.5	OSR 3.0	Microsoft Speech Server 2004 R2	WebSphere Voice Server
	Languages/Dialects Supported	Arabic (Jordan) Cantonese Chinese Mandarin Chinese (China, Taiwan) Czech, Danish, Dutch English (Australia, France) French (Canada, France) German (Austria, Germany, Switzerland) Greek Hebrew Italian Japanese Korean Norwegian Portuguese (Brazil) Spanish (Castilian, North-Latin America) Catalan Swedish Turkish	Cantonese (Hong Kong), Catalan (Spain) Czech (Czech Republic), Danish (Denmark), Dutch (Netherlands) English (Australia, U.K., India, Singapore, USA), Euskera (Spain), Finnish (Finland) French (Canada, France, Luxembourg, Switzerland) German (Austria, Germany, Luxembourg, Switzerland) Greek (Greece), Flemish (Belgium) Hebrew (Israel), Hungarian (Hungary) Italian (Italy, Switzerland) Japanese (Japan) Korean (Korea), Mandarin (PRC, Taiwan) Norwegian (Norway), Polish (Poland) Portuguese (Brazil, Portugal) Russian (Russia), Slovak (Slovakia) Slovene (Slovenia) Spanish (Argentina, Colombia, Spain, USA) Swedish (Finland, Sweden) Turkish (Turkey), Walloon (Belgium) Welsh (UK)	English (USA) Spanish (USA) French (Canada)	<i>Note: Not all languages are available on all operating systems</i> Cantonese Chinese Dutch English (Australian, US, UK) French (Canada, France) German, Italian, Japanese Korean Portuguese (Brazil) Spanish (Castilian, Mexico) Traditional Chinese
	Simultaneous Multi-lingual recognition?	Yes	Yes	No	No
	Ability to create new language for customer	Yes	Yes	No	No
	Background Noise Detection/Elimination	Patented process for noise detection	Yes	Yes	Yes
	Conversational Natural Language Support?	Yes - Accuroute	Yes - Speak freely	Yes	Yes
Maximum grammar size (words)	100's of millions	over 1 million	Unlimited	1 million	



Vendor:		Nuance	ScanSoft	Microsoft	IBM
	Hot word recognition	Yes	Yes	No	No
		Yes	Yes	Yes	Yes
Speaker Verification	Product	Nuance Verifier 3.5	SpeechSecure	None	None
	Languages	English (Australia/New Zealand, South Africa, U.S./Canadian, UK) Brazilian Portuguese Cantonese Chinese Castilian Spanish Dutch French (Canada, France) German (Germany, Switzerland) Italian, Japanese Korean, Swedish Spanish (Latin America) Mandarin Chinese	Language independent – any language		
	Requires recognizer engine?	Yes	No		
	Configurable security level	Yes	Yes		
Standards	VoiceXML 2.0	Yes	Yes	No	Yes
	SALT	No	Yes	Yes	No
	SRGS	Yes	Yes	Yes	Yes
	xHMI	No	Yes	No	No
	SISR	No	No	Yes	Yes
	MRCP	Yes	Yes	No	Yes
	SSML	Yes	Yes	No	Yes



Vendor:		Nuance	ScanSoft	Microsoft	IBM
	supported:	SIP, SNMP, GrXML, RTP	Aurora, VoIP	T1 ISDN PRI T1 CAS (FX) ECMA 323 HTML, HTTP, XML, SOAP, WSDL, SSL	CCXML H.323 Java TM
Application Components	Product(s)	There are a number of built-in grammars that are supplied with the product	Open Speech Dialog Modules (OSDM)	a grammar library reusable prompts	Websphere Voice Toolkit
	Domains covered		yes/no alphanumeric, dates, times social security #s, credit card #s credit card dates international postal codes, money digits natural numbers, phone numbers names, address, email collection voice enrollment, unconstrained alphanumeric	For prompts: Number, date Time, navigation	alphanumeric list selection confirmation credit card #s credit card expiration dates, currency dates, directions, durations, email addresses, numbers social security numbers, street types, telephone numbers time URLs, major US cities, US states US time zones.
Application Kits	Product(s)	Nuance Flexible Application Suites (FAS)	SpeechPak Application Kits SpeechWorks Automated Directory Assistance (ADA) SpeechWorks Call Navigator	Microsoft Speech-Enabled ASP.NET Commerce Starter Kit	Wizard-based interface for combining components into applications
	Industries/Applications Supported Industries	Credit card services Retail banking, Insurance Utilities, Telecommunications (wireless, wire line) U.S. Address Capture	Healthcare Utilities Automotive (for embedded technology) Directory Assistance Call Routing	Commerce	
Packaged Applications	Product(s)	Nuance Call Steering 1.0 Nuance Caller Authentication	Open-Speech AutoAttendant Speech Attendant Speech Attendant Large Enterprise Edition	Available from independent software vendors	No
	Industries/Applications Supported Industries	Call Routing Caller Authentication	Auto-Attendant		



Vendor:		Nuance	ScanSoft	Microsoft	IBM
	Estimated Time to Deployment	50% normal deployment time (probably about 4 to 5 months versus 9 months)	a few days		
			Available in these languages: US English, UK English, Australian English, Canadian French, European French, German, Dutch, and American Spanish		
Development Tools		Nuance Application Environment (NAE)	Scansoft doesn't develop own tools, but has a wide variety developed by their platform partners		Websphere Voice Toolkit
	Tasks supported	Design Development Tuning Analysis Grammar development Pronunciation		Design Development Debugging Analysis Grammar development Prompt creation	Call-flow design Application coding Debugging Grammar building & testing Prompt creation Pronunciations Call Control Natural Language Understanding model
	Development environment	GUI		Visual Studio .NET	Eclipse IDE
Logging and Monitoring Support		Provides tools	Everything written to log files – no tools	Yes – supports call flow analysis and statistical analysis of call data integrated with Windows Media Player and SQL Server™ Reporting Services (SSRS).	Administration panels to facilitate configuration, monitoring, and troubleshooting.
Other	Operating Systems	Windows Server 2002 Windows 2000, IBM AIX Sun SPARC Solaris 2.8 x86 Solaris 2.8	Windows Server 2003, Windows 2000 Windows NT, Linux, UNIX, Solaris	Microsoft Windows Server™ 2003 Enterprise Edition	Windows 2003 Linux AIX



Vendor:		Nuance	ScanSoft	Microsoft	IBM
	Maximum Port Capacity	Unlimited (some deployments accept 1 million calls/day)	Limited only by system capacity; typical installation is 120 ports	<ul style="list-style-type: none"> • Handles up to 96 telephony ports of continuous, fully loaded speech recognition per TAS node • Unlimited scale-out by adding more nodes 	Scales by adding additional servers
	Compatible with VoIP	Yes	Yes	Yes - optional Vail SIP Telephony Interface Manager	Yes
	Multi-modal capabilities?	No	No (but other Scansoft products provide)		