# TMC SPEECH-WORLD™

### VoiceObjects Opens Its Voice Application Server Architecture To The Global Voice Industry

VoiceObjects, a provider of voice application management systems (VAMS) for voice-driven telephone services, has announced the launch of VoiceObjects XDK. The product was created to be the first open, XML-based dialog interface to a complete application lifecycle management framework. VoiceObjects XML Development Kit aims to allow service creators to more easily equip any voice or DTMF application with an interface to VoiceObjects' voice application management system (VAMS) that combines high-performance voice application server capabilities and comprehensive management tools. With VoiceObjects XDK, organizations using any service creation environment will be able to run their voice or DTMF application on a powerful and scalable voice application server. VoiceObjects XDK will help companies manage any existing or newly built application covering the whole lifecycle from deployment, monitoring, analyzing, maintenance and re-launch.
**www.voiceobjects.com**

### Definition Du Jour

Phoneme (fo' nem')

n. The smallest phonetic unit in a language that is capable of conveying a distinction in meaning, as the *m* of *mat* or the *b* of *bat* in the English language. (From the *American Heritage Dictionary, Third Edition*, Houghton Mifflin Company, 1992.)

> "Much speech is one thing, well-timed speech is another."
>
> — *Sophocles*

### Metaphor Solutions Announces Speech Application Development Tools

Metaphor Solutions, a player in the speech IVR (interactive voice response) arena, recently announced the first release of its Metaphor Conversation Manager (MCM). Metaphor Conversation Manager is an integrated development environment for developing automated speech applications that interact with callers, back-end data sources and Web services and with live service agents. MCM was designed to be easy to use and aid in speeding development. It works with any VXML gateway and the Microsoft Speech Server. It's an editor, linker, debugger and run-time interpreter that dynamically generates open-standard scripts in Voice XML and SALT from a high-level design-scripting language called Metaphor Script. It also includes Web-based end-user customization, as well as administration, logging and reporting tools. MCM aims to be five to ten times faster than developing speech applications in straight Voice XML or SALT, and three to four times faster than developing with graphical-based Voice XML form objects with property sheets. This is because Metaphor Script, the scripting language of MCM, leverages many inherited resources used in the best practices for dialog design and has more flexibility in controlling the conversation flow, as well as a simulation debugger. MCM includes a library of complete and fully tested package applications that allow the developers to customize and integrate speech IVR solutions for most common applications. Also included are an array of reusable and configurable dialog building blocks that help users build both speech and touch-tone voice applications that are customized to their business operations.
**www.metaphorsol.com**

# A Talk With Nuance's Lynda Kate Smith

The editors of *Customer Interaction Solutions*® recently spoke with Lynda Kate Smith, vice president and chief marketing officer of Nuance Communications (www.nuance.com), to hear her opinions on the present and future of speech technologies.

**CIS:** What needs do speech-based call center technologies serve that traditional touch-tone IVR systems do not?

**LKS:** More and more businesses, throughout a range of industries, are recognizing the business benefits of speech versus traditional touch-tone technology. Speech improves customer satisfaction and retention, reduces the overall cost of customer service and increases automation rates, as customers tend to stay within voice-enabled automated systems rather than opting out to a live agent. Analyst research indicates speech boosts customer use of automated systems from 20 to 60 percent, compared with touch-tone. On the business side, voice automation can dramatically reduce the cost of handling calls, delivering a return on investment in just a few months. Analysts at Cahners In-Stat and Giga report the average cost-per-call handled by agents ranges anywhere from $2 to more than $15. With voice automation, that cost is often cut to 20 cents or less per call.

From a caller experience standpoint, speech technology shortens the duration of calls by eliminating on-hold times and providing more streamlined, round-the-clock access to information. Customers get where they want to be with fewer steps and hassles — something very important since today's customers have neither the time nor the patience to wait on hold or navigate through a myriad of touch-tone prompts to get what they need. Speech is intuitive and user-friendly (there's no need to memorize choices) and is more responsive to customers' needs. In fact, a 2003 Harris Interactive study we commissioned revealed that nine out of 10 surveyed consumers preferred speech recognition to touch-tone.

Another benefit of speech is enhanced security. It's been estimated by the Federal Trade Commission that identity theft and fraud affect nearly 10 million people and cost businesses upwards of $60 billion per year. Companies are searching for ways to secure their call centers while reducing costs and promoting customer convenience. Voice authentication (voiceprint technology) allows businesses to offer secure, personalized over-the-phone customer service. Many financial institutions have implemented voiceprint technology to verify customers and to reduce I.D. theft and fraud. Before transactions (i.e., checking a bank account balance) are authorized, users' spoken voiceprints are compared to those previously enrolled and on file. The transactions are performed only after an exact match is made, based on the unique characteristics of the caller's voice. With automated voice authentication, more callers remain in the automated system, and agents can focus on addressing high priority customer issues.

**CIS:** At this point, have we seen all of the enterprise and consumer applications for speech, or do you think what we've got thus far is the tip of the iceberg?

**LKS:** What we've seen so far is only the tip of a very large iceberg. In the past year alone, we've seen a tremendous increase in demand for voice automation solutions. Speech is no longer an add-on for big companies with ample discretionary budgets; it's a necessity for any company, large or small, that uses the telephone as a means of communicating with customers. We anticipate that speech applications for enterprises and consumers will evolve, continuing the current momentum.

Voice automation is following a similar path that we've seen enterprise software markets travel over the past few years: it's moving from custom applications to packaged applications. Packaged applications speed speech application deployment and time-to-market while increasing customer satisfaction. Nuance's packaged applications leverage the company's best practices in voice automation yet also have the flexibility to meet company-specific business and technology requirements. Packaged applications empower companies to configure, manage and maintain their own applications, serving as an easy-to-deploy alternative to built-from-scratch speech applications. This trend will further drive adoption in all markets.

**CIS:** I've heard it said that the actual code behind speech recognition is some of the most complicated code ever to be written. Can you expound on why the creation of the actual engine is so complex?

**LKS:** The combination of speech recognition (the ability to interpret words and commands) coupled with natural language understanding (the capability to conduct conversations between humans and machines) is very complex. Nuance alone has been refining this technology for more than 10 years as an independent entity and for many years prior as a part of the Stanford Research Institute.

There are two reasons for the complexity. The main reason, of course, is the complexity of human speech. It is widely varied, subtle, contextual, accented, and the engine must overcome the variations in spoken language to reach a correct interpretation of the request with a high degree of accuracy. The second challenge is deployment environment. Telephones are used everywhere, from the relative quiet of a home or office, to airports, trains and automobiles. They're both wired and wireless. Ultimately, it's any device with a microphone and a network connection. The speech engine must be able to overcome all of the challenges associated with the deployment environment and, to a further degree of complexity, in the product design.


## Statement of Ownership, Management and Circulation

(Required by 39 U.S.C. 3685)
Publication Title: Customer Interaction Solutions
Publication Number: 68-2830
Filing date: 10/1/04 Frequency: monthly
Number of issues published annually: 12
Annual subscription price: requester (none to qualified US, $49 nonqualified US, $49 Canada, $85 Foreign)
Mailing address of Known Office of Publication and Headquarters or General Business Office:
Technology Marketing Corp., 1 Technology Plaza, Norwalk, Ct. 06854
Publisher: Nadji Tehrani, Technology Marketing Corp., 1 Technology Plaza, Norwalk, Ct. 06854
Editor: Tracey E. Schelmetic, Technology Marketing Corp., 1 Technology Plaza, Norwalk, Ct. 06854
Owner: Nadji Tehrani, Technology Marketing Corp., 1 Technology Plaza, Norwalk, Ct. 06854
Issue Date for Circulation Data Below: September 2004

| Extent and Nature of Circulation | Average # of Copies each issue during Preceding 12 months | No. copies of single issue published nearest to filing date |
|---|---|---|
| Total# of copies: | 72,815 | 72,800 |
| Total Paid and/or Requested Circulation through mail: | 70,222 | 70,240 |
| Free Distribution by Mail: | 1,560 | 1,714 |
| Free Distribution Outside the Mail: | 139 | 500 |
| Total Free Distribution: | 1,699 | 2,214 |
| Total Distribution: | 71,921 | 72,454 |
| Copies Not Distributed: | 894 | 346 |
| Total: | 72,815 | 72,800 |
| Percent Paid and/or Requested Circulation: | 97.64% | 96.94% |

The good news is that both challenges listed above have been overcome by Nuance. Other companies have also attempted to solve these technical challenges, and many have abandoned their efforts over time because they were not successful. Instead, they've chosen to partner with Nuance, take our core technologies and build applications on top of them.

**CIS:** Not very long ago, text-to-speech technologies had a reputation for being highly unreliable. How has this changed and improved in the last several years?

**LKS:** Text-to-speech (TTS) is often confused with speech to text (in other words, dictation) software. And it's true that many people's experiences with these primarily desktop products were poor. These products are "speaker dependent," requiring significant amounts of training to be able to recognize words and phrases reliably. Telephony-based speech recognition, like that which Nuance provides, is "speaker independent," requiring no training by the user for accurate recognition.

As for TTS, the principal issue has been vocal quality, not accuracy or reliability. This software has primarily been used as part of speech applications that deliver frequently changing information that is difficult and expensive to handle with recorded prompts (i.e., train or plane schedules). Speech recognition is used to understand a caller's request, recorded prompts are used to communicate information that rarely changes, and TTS software is used to speak dynamic information, such as the name and address, for a listing — information that is stored as text in a database.

Nuance is continually improving the vocal quality of its TTS product to make it much more human and natural sounding and less mechanical. Recently, several TTS-related developments have contributed to this, including things like SSML (speech synthesis markup language) that enable developers to add emphasis and tune pronunciations.

**CIS:** Why has Nuance chosen to focus only on enterprise, and not consumer, products?

**LKS:** Nuance software powers a range of consumer-focused products, including the Personal Calling feature in GM Onstar, voice dialing from Sprint and other wireless carriers, and much more. All these products use Nuance's core expertise in telephony speech recognition, also known as network-based speech. We believe the network approach to offering consumer services is far superior to the "embedded" approach some companies are taking primarily because of lack of processing power in today's devices. This limitation forces companies to go with a speaker-dependent approach to speech recognition. Just as with the desktop dictation example, speaker dependency has many drawbacks.

**CIS:** I find the concept of Nuance's "persona creation services" quite fascinating. Do you have a team of psychologists testing and refining to find out what kind of "artificial voice" customers prefer to hear? Does the preference vary from industry to industry?

**LKS:** Personas are a fun and interesting aspect of Nuance's speech applications, as for callers they are the "face" of a company/brand. Nuance research indicates that callers associate the system voice with specific characteristics — even if the system was not designed to exemplify any particular character — and are attracted to the personality that they deem most similar to their own. Callers are also drawn to specific persona traits (i.e., confident, warm, easy-going). For example, a bank may want its persona to convey authority, security and clarity, so perhaps an ideal persona would be a 60-year-old retired investment banker who worked for years on Wall Street. We have a team of Nuance persona consultants who work with partners and customers to develop personas.

Nuance persona design and production consultants include experts in creative writing, conversational linguistics, sound design and audio production. The persona consultants use their deep working knowledge of natural language speech recognition applications to define the persona early in the design process, based on the identity of the target user group, user feedback about the task they are completing and the imagery associated with the company's brand. Once the persona is created, we run through a series of steps prior to going live. These include customer usability tests, focus sessions, customer surveys and recognition testing. We then conduct a performance analysis post-launch and make any necessary adjustments.

We take this process very seriously, as a company's persona helps establish a consistent, positive experience for users that underscores its branding goals and enhances the specific image conveyed by its speech system. It is our goal to create the best (virtual) ambassador possible to represent a company's brand.

**CIS:** Can you name a few of Nuance's OEM partners that integrate Nuance technologies into their own products?

**LKS:** Nuance works with a variety of partners to bring speech solutions to market. We do have OEM partners, who bundle our core speech technology into theirs as part of an overall solution they are delivering. These include companies like Avaya, Cisco, Edify, Intervoice, Nortel and Syntellect, among others.

Nuance also has relationships with value-added resellers, who deliver speech solutions to their customers based on Nuance's platform product, NVP. Representative partners in this category include Bell Canada, Digital DataVoice, Telus Business Solutions, Versay and WorkForce Technologies, among others.

Our ecosystem is further rounded out by a wide range of technology partners who offer complementary products and solutions.