Executive Summary

Speech is playing an increasingly important role in both the enterprise and service provider market segments, making it one of today’s fastest-growing technology areas. For example, Dataquest projects growth from $656 million in 2002 to $3.1 billion in 2008 in the global voice business market segment. The Gartner Group expects the speech market segment to grow from $127.5 million in 2002 to more than $258 million in 2007 [Fluss].

Improvements in advanced speech recognition technologies have brought new applications into mainstream use — particularly in the area of open, standards-based approaches using Voice Extensible Markup Language® (VoiceXML). The major goal of VoiceXML is to bring the advantages of Web-based development and content delivery to voice-enabled business applications.

VoiceXML enables an expanding range of solutions for telecommunications service providers looking for new revenue channels and enterprises looking for ways to improve customer service while reducing costs.

VoiceGenie, Intel, and ScanSoft have teamed up to produce a VoiceXML solution that provides telecommunications service providers and enterprises with a flexible framework for improving customer service and contact center operations while reducing the associated expenses. The solution uses a cost-effective and dependable speech-enabled IVR platform from VoiceGenie, based on Intel® architecture and well-suited to modular building block design. Telecom hardware from Intel accelerates the speech processing operations and interacts effectively with voice recognition and text-to-speech (TTS) components from ScanSoft.

(http://www.crmxchange.com/speech_technology/oct03.html)

June 2004
Business Opportunity

Challenged with managing rising operating costs and increased competitive pressure, service providers and enterprises are looking for new ways to increase efficiency and lower costs while delivering consistently outstanding customer service.

Providing customer self-service through the Web provides a partial solution, but this approach does not accommodate customers for whom network access is inconvenient or impractical. Interactive voice response (IVR) products have been available for many years to automate call processing, but until recently they were based exclusively on dual tone multi-frequency (DTMF), or touchtones. Applications built on these platforms provided a good return on investment (ROI), but often frustrated callers with lengthy menus and arcane command sequences.

Past-generation IVR platforms were based on proprietary application development environments and hardware platforms — limiting their integration flexibility and locking in the IVR vendor. Solutions that deliver the benefits of speech recognition and TTS in an open, standards-based environment offer the potential to increase flexibility and simplify deployment and integration tasks.

Today’s IVR products support sophisticated applications that can interact with callers in a simple way using natural speech. By incorporating TTS and speech recognition technologies, these IVR platforms continue to deliver a good ROI while raising the level of caller satisfaction. Greater automation rates can also improve savings, while broad call handling capacity reduces hold queue times for better service. Relieved of tedious and repetitive requests, contact center agents can function more efficiently, free to devote their time to higher-value tasks.

Speech Applications: Changing Business Models

Speech is an increasingly important component of total customer care strategies for both enterprises and service providers, offering numerous benefits including:

- **Increased automation of self-service transactions** — Organizations that have implemented IVR systems consistently strive to improve automation rates, or transaction completion rates (TCRs), to realize the full economic benefits of their IVR application. Unfortunately, traditional IVR systems can be frustrating to callers due to complicated calling trees or confusing menus. Many callers choose to bypass the automated system and opt out to a live agent. Speech-enabled applications can significantly increase customer usage of IVR systems, helping to reduce operational expenses and ensure the maximum benefit from IVR investments.

- **24/7 service** — Speech solutions enable organizations to provide a consistent level of service and give their customers the flexibility to access services anytime, anywhere.

- **Enhanced customer satisfaction** — Speech solutions provide mobile consumers with improved access to services and a significantly enhanced user experience over traditional touchtone IVR or Web-based services. With speech, self-service is available from any phone using only the caller’s voice, providing an easy-to-use, flexible interface that can significantly reduce calling times and dramatically increase the user acceptance of automation.

**CUSTOMER SELF-SERVICE FOR THE ENTERPRISE**

Contact centers are a necessity for any service-oriented business. Often a company’s first and most important point of contact with the public, the contact center represents the company’s public image and culture.

However, contact centers are expensive to operate and staff with trained and experienced agents. Moreover, for a typical contact center, most customer inquiries involve routine transactions — for example, account balance, status, or change of address information. Handling these basic tasks is an inefficient and costly use of an agent’s time.

Today, organizations across many industries — utilities, financial services, government, travel and entertainment, healthcare, and many others — are dramatically improving their customer service operations by adding speech technology.

By providing customers with the flexibility to access services 24/7 and conduct their own transactions using any telephone, speech solutions leave calling customers more satisfied, since their service issues handled quickly and efficiently.

Speech-based automation can also slash hold times for callers, who no longer get locked in long call queues. It can also help to shorten calls — dramatically reducing bottom-line costs.
A COMPELLING DIFFERENTIATOR FOR SERVICE PROVIDERS

Speech recognition systems are quickly becoming a fundamental way for telecommunications carriers to differentiate themselves in a crowded marketplace by delivering compelling content and services to their customers. Branded telecommunication services allow carriers to promote their own brands and build customer loyalty while they control costs.

By offering voice services to their enterprise and consumer customers, carriers also help their subscribers differentiate themselves in the marketplace and improve both their customer care and their bottom lines.

Telecom service providers are delivering and deploying a broad range of applications and services including:

- **Basic information portal** — This creates a branded information service where customers access information and content — for example, sports scores, traffic updates, weather reports, or news.
- **Voice portal** — This gives callers anywhere, anytime access to information using easy speech commands. Besides driving usage, it also acts as a competitive differentiator.
- **Enterprise portal** — This allows customers and suppliers to easily place and track orders and more.
- **Personalized services** — Offering compelling services helps to reduce customer churn.
- **Hands-free dialing** — This provides callers the ability to dial by voice, especially useful for cell phone users where small keypads make it hard to use touchtone services.
- **Workforce mobility applications** — This provides today’s mobile workforce with a single speech-enabled platform to access phone, fax, and email communications and enterprise applications such as customer relationship management (CRM).
- **Information manager** — This allows callers to manage personal contact information, calendars, etc.
- **Directory assistance** — This allows callers to easily retrieve phone numbers and addresses using any telephone.
- **Call management system** — This includes an auto attendant that welcomes callers, directs calls, and delivers company information.

The VoiceGenie and Intel Solution

VoiceGenie and Intel have teamed up with ScanSoft to produce a VoiceXML-based solution that gives businesses a highly adaptable framework for quickly and easily deploying speech-enabled applications. The solution provides an open-standards, speech-enabled IVR platform based on modular building blocks from Intel.

The solution delivers all the benefits of an IVR system and incorporates the latest TTS and speech recognition technologies. By using VoiceXML, the solution effectively taps into existing Web-based infrastructure for access to business rules, databases, and back-end integration. This distributes development and maintenance costs across both Web and telephone self-service options — further enhancing ROI.

The open-system architecture provides the framework to easily integrate and deploy the total solution, which consists of:

- **Rack-mount servers** based on Intel architecture running the open-source Red Hat* Linux* operating system.
- **Intel® telecom boards** for flexible telephony network connectivity.
- **High-performance OpenSpeech Recognizer® and Speechify® TTS technologies from ScanSoft.**
- **NeXusPoint® VoiceXML platform and operations and management software from VoiceGenie.**

Organizations across a broad span of industries are turning to servers based on Intel architecture and running the open-source Linux operating system. Solutions based on this platform can help reduce costs while improving performance and flexibility in all market segments.

Implemented as a horizontal technology, the VoiceGenie and Intel VoiceXML solution can provide benefits in nearly every situation where a telephone call is answered. Telecom service providers and enterprise contact centers can supplement their existing services by automating and expanding their reach with new services that add revenue. The solution can also be used for outbound calling, such as alerting employees to shift changes due to weather conditions or other factors.

Because of its reliance on open standards, this solution lets an internal information technology (IT) employee with solid Extensible Markup Language (XML) skills update corporate contact center and business applications using the VoiceXML stack instead of having to contract with expensive professional services from the legacy IVR vendors.
VoiceXML

VoiceXML is a World Wide Web Consortium (W3C) standard for speech-enabled applications, offering many technological benefits for applications requiring interactive TTS and voice recognition. VoiceXML uses “tags” to define a call flow, the dialog that occurs between a person and a computer over the telephone.

VoiceXML was designed for creating audio dialogs that feature:

- Synthesized speech
- Digitized audio
- Recognition of spoken and DTMF key input
- Recording of spoken input
- Telephony
- Mixed-initiative conversations

Many companies actively developing speech applications contributed to the design of VoiceXML working through the standards development process of the W3C. VoiceXML is a mature standard, first published in draft form in August 1999. The VoiceXML 2.0 Recommendation was approved and published in March 2004.

VoiceXML embodies many of the characteristics that HTML exhibited when the Web entered its rapid growth stage. As a markup language, it allows companies to reuse the business rules and back-end logic that already power their existing Web-based applications. VoiceXML also ushers in a new era of lower cost and vendor independence, since an application built with VoiceXML can run on any VoiceXML-based telephony platform. This feature translates into flexibility and investment protection for customers.

Solution Architecture

The complete VoiceGenie and Intel VoiceXML solution consists of:

- A carrier-grade server based on Intel architecture running the Red Hat Linux operating system
- Communications hardware: Intel® NetStructure™ DM/V960A-4T1 Combined Media Board with Intel® Dialogic® Continuous Speech Processing Technology
- VoiceXML Interpreter* and software from VoiceGenie — the NeXusPoint* framework, available in Enterprise and Service Provider editions.
- OpenSpeech Recognizer* speech recognition engine from ScanSoft
- ScanSoft’s Speechify* TTS VoiceXML scripting and recorded audio prompts
- A VoiceXML service creation environment
- Operations, administration, and maintenance (OA&M) software
- Enterprise back-end database integration
- VoiceXML applications

The VoiceXML Gateway* interprets VoiceXML documents, following commands to play prompts, synthesize speech, recognize utterances, and transfer calls. An application server delivers the VoiceXML documents through a Hypertext Transfer Protocol (HTTP) or Hypertext Transfer Protocol Secure (HTTPS) connection, much as an application server provides

Benefits of the VoiceGenie and Intel VoiceXML Solution

- **Exceptional ROI** — Many companies realize a substantial ROI within three to six months based on increased automation and savings from greater operational efficiency.
- **Improved customer satisfaction** — Customer satisfaction improves because of shorter waits on the telephone, reduced transaction times, and a user-friendly and natural speech user interface.
- **Increase revenue opportunities, competitive advantage, and customer churn** through enhanced services.
- **Stronger job satisfaction** — Contact center employees experience greater job satisfaction when freed from tedious tasks to focus on more challenging issues.
- **Rapid integration** — The open-standards, Web-based architecture provides rapid integration to existing systems that have been developed for information access over the Internet, lowering application development costs.
- **Simplified platform administration** — The solution provides easier platform administration over the long term through the shared Web infrastructure.
- **Enhanced investment protection** — The open-standards framework contributes to a long product lifecycle, easy future upgrade path, and lower IVR platform costs by relying on equipment based on a standard PC architecture.
- **Best-of-breed components** — The solution provides the assurance that comes from working with leading solution vendors such as VoiceGenie, ScanSoft, and Intel.
HTML pages to control the display of a Web browser. The application server typically executes Java® scripts, acting as a link between the VoiceXML Gateway and back-end database servers. Figure 1 shows a functional business diagram.

The NeXusPoint framework® interprets VoiceXML documents, following commands to play prompts, synthesize speech, recognize utterances, and transfer calls. An application server delivers the VoiceXML documents through a Hypertext Transfer Protocol (HTTP) connection, much as an application server provides HTML pages to control the display of a Web browser. The application server typically executes Java® applications, acting as a link between the VoiceXML platform and back-end infrastructure. Figure 1 shows a functional business diagram.

The VoiceXML-based NeXusPoint platform from VoiceGenie connects to a service provider’s telephony switch or an enterprise’s corporate PBX system. The platform is based on a rack-mount server based on Intel architecture with Intel NetStructure telecom boards handling the calls and VoiceGenie’s NeXusPoint VoiceXML Interpreter and ScanSoft speech engines managing the dialog.

This combination provides a bridge between the traditional phone system and the speech-driven application that accesses Web infrastructure, business logic, and database servers. Figure 2 shows the software architecture.
The VoiceXML application controls all interaction with the telephone caller. It includes grammars that specify all the words and phrases that can be recognized and the associated actions for each spoken command. VoiceXML uses XML tags to identify the grammars, prompts, and logic that describe the call flow — i.e., how the dialog proceeds between a caller and the automated system.

This solution stack is based on the VoiceXML-based NeXusPoint platform from VoiceGenie, a comprehensive solution based on the Linux operating system that integrates a 100% VoiceXML 2.0-compliant Interpreter, innovative and highly functional extensions, powerful call management capabilities, and the carrier-grade OA&M software.

Engineered using an open-standards model, the solution delivers exceptional scalability, robustness, and reliability — critical factors as enterprises and service providers grow their telephony services to match their customer bases. The solution also incorporates ScanSoft speech recognition and TTS software optimized for the VoiceXML standard. OpenSpeech Recognizer® provides speech recognition for multiple languages and delivers high accuracy with landline, wireless, and hands-free input. Speechify TTS provides natural-sounding synthetic speech in a variety of voices. Together, these technologies support construction of powerful, easy-to-use speech applications. Figure 3 shows the systems architecture.
The telecom boards from Intel incorporate Intel Dialogic Continuous Speech Processing Technology, a set of software and firmware features that support high-performance, speech-enabled applications on open computing platforms. Continuous Speech Processing Technology enhances existing speech technologies by processing real-time voice signals, which identify human speech input, and presenting the input to the host platform for speech recognition. The real-time functions include echo cancellation and voice activity detection (VAD). This approach frees host platform resources for more complex speech recognition tasks such as analyzing and recognizing the speech input for an application.

Continuous Speech Processing Technology supports important application-friendly features such as barge-in, the ability to interrupt speech prompts by speaking over them. Barge-in allows callers to control the pace of a conversation and complete the interaction more quickly, resulting in a more pleasant user experience and a more efficient use of platform resources. Barge-in saves host-system resources, improves system utilization, and can reduce phone charges by shortening the duration of the call. The VAD functionality in Continuous Speech Processing Technology provides a pre-speech buffer that produces better voice recognition using less host processing, and enhances the accuracy of speech detection. In addition, Continuous Speech Processing Technology was designed to be flexible and permits VAD to be used or disabled in conjunction with the speech detection algorithms created by speech technology developers.

The Intel NetStructure DM/V960A-4T1 Combined Media Board provides, in a single PC slot, a robust media feature set including voice processing and speech recognition (and on select media loads, fax and/or conferencing capabilities), combined with an extensive suite of network protocols. The board is available in both H.100 (PCI) and H.110 (CompactPCI*) compliant universal form factors. This platform is ideal for service providers and large enterprise applications, with flexibility that lets developers build single applications to be deployed on either industry-standard form factor.

Carrier-grade servers based on Intel architecture provide the high performance required to support computationally-intensive speech recognition and TTS operations.

**Summary**

The VoiceGenie and Intel VoiceXML solution incorporates components from Intel to automate customer service operations over the telephone using advanced voice recognition and TTS technologies.

Telecom hardware from Intel offloads much of the processing workload and ensures reliable, consistent voice recognition through Continuous Speech Processing Technology.
This solution is ideal for both telecommunication service providers and business enterprises. Service providers gain a cost-effective, easily deployable package constructed to carrier-grade specifications and designed for maximum scalability as service volumes increase. Support for VoiceXML and the open-standards infrastructure provide a flexible, highly adaptable framework for rapidly implementing applications to support client requirements and open up new revenue channels. Medium-to-large-scale enterprises can substantially reduce their operating costs and provide a consistently higher level of service through the automated capabilities of this solution. Freed from the strictures of proprietary IVR systems, staff members in IT departments with XML expertise can design and implement speech-enabled applications and more easily maintain the system.

The VoiceGenie and Intel VoiceXML solution uses best-of-breed components to produce a capable, low-maintenance, cost-effective framework for delivering speech-enabled applications that are suitable for both the telecommunications industry and individual business enterprises.

### Definitions and Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRM</td>
<td>Customer relationship management</td>
</tr>
<tr>
<td>DTMF</td>
<td>Dual tone multi-frequency</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Markup Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>HTTPS</td>
<td>Hypertext Transfer Protocol Secure</td>
</tr>
<tr>
<td>IT</td>
<td>Information technology</td>
</tr>
<tr>
<td>IVR</td>
<td>Interactive voice response</td>
</tr>
<tr>
<td>OA&amp;M</td>
<td>Operations, administration, and maintenance</td>
</tr>
<tr>
<td>ROI</td>
<td>Return on investment</td>
</tr>
<tr>
<td>TCR</td>
<td>Transaction completion rate</td>
</tr>
<tr>
<td>TTS</td>
<td>Text-to-speech</td>
</tr>
<tr>
<td>VAD</td>
<td>Voice activity detection</td>
</tr>
<tr>
<td>VoiceXML</td>
<td>Voice Extensible Markup Language</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>

### References

- ScanSoft — [http://www.scansoft.com/speech](http://www.scansoft.com/speech)
- VoiceGenie — [http://www.voicegenie.com](http://www.voicegenie.com)
LEARN MORE ABOUT THIS INNOVATIVE SOLUTION

For general information, proof points and case studies about the products described in this white paper, visit:


1515 Route Ten
Parsippany, NJ 07054
Phone: 1-973-993-3000

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT, EXCEPT AS PROVIDED IN INTEL’S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS. INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

The Intel products referred to in this document are intended for standard commercial use only. Customer is solely responsible for assessing the suitability of the product for use in particular applications.

Intel, VoiceGenie, and ScanSoft may make changes to specifications, product descriptions, and plans at any time, without notice.

*Other names and brands may be claimed as the property of others.

Intel, Intel Dialogic, Intel NetStructure, Intel Centrino, and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Printed in the USA Copyright © 2004 Intel Corporation All rights reserved. Printed on recycled paper. 06/04 00-9127-001