



An Introductory Guide To Speech Recognition Solutions

Understanding the technology, the vendors and the market

A white paper prepared by Datamonitor

Sponsored by

Avaya, Cisco Systems, Convergys, Dimension Data, EDS, Envoy Worldwide, Fluency Voice Technology,
Genesys, HP, ICT Group, Intervice, Nortel, Nuance, RightNow, Tuvox, Voxeo, West Corporation

Publication Date: 8/06

www.datamonitor.com

Datamonitor Europe

Charles House
108-110 Finchley Road
London NW3 5JJ
United Kingdom

t: +44 20 7675 7000
f: +44 20 7675 7500
e: eurinfo@datamonitor.com

Datamonitor USA

245 Fifth Avenue
4th Floor
New York, NY 10016
USA

t: +1 212 686 7400
f: +1 212 686 2626
e: usinfo@datamonitor.com

Datamonitor Germany

Kastor & Pollux
Platz der Einheit 1
60327 Frankfurt
Deutschland

t: +49 69 9750 3119
f: +49 69 9750 3320
e: deinfo@datamonitor.com

Datamonitor Asia Pacific

Room 2413-18, 24/F
Shui On Centre
6-8 Harbour Road
Hong Kong

t: +852 2520 1177
f: +852 2520 1165
e: hkinfo@datamonitor.com

Datamonitor Japan

Aoyama Palacio Tower 11F
3-6-7 Kita Aoyama
Minato-ku
Tokyo 107 0061
Japan

t: +813 5778 7532
f: +813 5778 7537
e: jpinfo@datamonitor.com

ABOUT DATAMONITOR

Datamonitor plc is a premium business information company specializing in industry analysis.

We help our clients, 5000 of the world's leading companies, to address complex strategic issues.

Through our proprietary databases and wealth of expertise, we provide clients with unbiased expert analysis and in-depth forecasts for six industry sectors: Automotive, Consumer Markets, Energy, Financial Services, Healthcare, Technology.

Datamonitor maintains its headquarters in London and has regional offices in New York, Frankfurt, Hong Kong and Japan.

All Rights Reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher, Datamonitor plc.

The facts of this report are believed to be correct at the time of publication but cannot be guaranteed. Please note that the findings, conclusions and recommendations that Datamonitor delivers will be based on information gathered in good faith from both primary and secondary sources, whose accuracy we are not always in a position to guarantee. As such Datamonitor can accept no liability whatever for actions taken based on any information that may subsequently prove to be incorrect.

An Introductory Guide to Speech Recognition Solutions

© Datamonitor (Published 8/2006)

Page 2

This brief is a licensed product and is not to be photocopied

TABLE OF CONTENTS

ABOUT DATAMONITOR	2
INTRODUCTION	4
Speech recognition technology defined	4
Key milestones in speech recognition over the past 50 years	5
SPEECH RECOGNITION IN TODAY'S ENVIRONMENT	5
Companies that have deployed speech	6
Types of speech recognition solutions being deployed	7
WHY BUSINESSES ARE INVESTING IN SPEECH TODAY	8
The value proposition for speech	8
Greater flexibility and performance at lower prices	9
The market uptake for speech recognition solutions	10
THE COMPONENTS OF A SPEECH RECOGNITION SOLUTION	11
The vendors that provide these technologies and services	11
Snapshot of vendors in the voice business value chain	14
STRATEGIC RECOMMENDATIONS FOR BUYERS	17
CONCLUSIONS	19
APPENDIX	20

INTRODUCTION

Speech recognition was once viewed as a futuristic technology that would never leave the realm of science fiction. But over the past 50 years, key technology and commercial achievements in speech recognition along with increased CPU performance and lower hardware costs have helped make speech commercially viable for enterprises and service providers. Today, speech recognition is becoming increasingly prominent as a cost-cutting and value-enhancing solution for customer care and service enablement. Over 7 million calls are handled by speech-based self-service, routing and portal applications worldwide everyday and hundreds of mid-to-large sized businesses have implemented speech. Investments in speech are growing at rate of 22.6% per year. By 2009, businesses from across the globe will spend an estimated \$2.7 billion on speech recognition solutions and services.

As more businesses begin evaluating speech through the next few years many of them will find themselves in unfamiliar territory. Understanding the technology, the vendors and the market for speech recognition solutions can be a new and daunting task for these businesses as they get lost in the technical terms, market trends and vendor product differentiation. This white paper provides a simple introductory guide to speech recognition solutions and enables readers to:

- Learn about the technology, its history and its uses in today's environment;
- Understand the business benefits of speech solutions;
- Gain visibility into market adoption and trends for speech;
- Identify and learn of the various components that compose a speech solution;
- Distinguish between the products and services offered by the sponsors of this white paper.

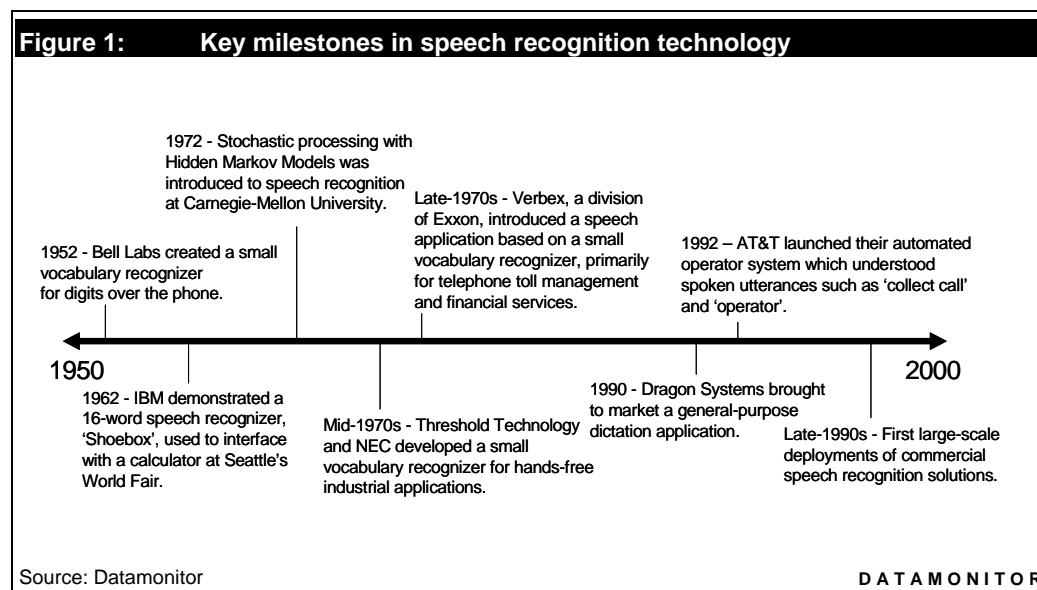
The white paper concludes with Datamonitor's strategic recommendations to organizations that are currently evaluating or deploying speech recognition solutions in their customer service and service enablement strategies.

Speech recognition technology defined

Speech recognition is a technology that uses natural speech as the interface between human and machine. It converts spoken utterances into sets of words through complex algorithms located on a computing device or system.

Key milestones in speech recognition over the past 50 years

The concept for speech recognition has been pursued by inventors, companies, and academic and government institution for well over a century. But it is the technical and commercial breakthroughs over the past five decades that have helped make speech recognition a commercially viable technology. A handful of these milestones are highlighted in Figure 1 below.



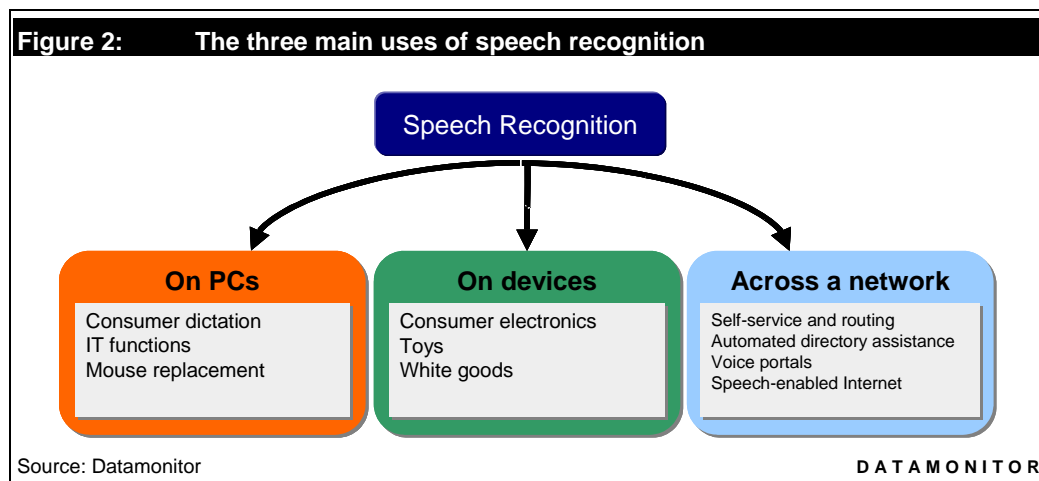
Significant R&D milestones were reached prior to the 1970s, such as limited vocabulary recognizers and the application of Hidden Markov Models to speech, but it was not until the 1990s that larger commercial speech recognition solutions emerged in the market. Vendors began to understand speech and language processes in greater detail as they related to human behavior. CPUs became powerful enough to substantially improve the performance of speech recognition solutions at a fraction of the hardware costs of previous decades. These factors combined helped vendors refine algorithms, introduce new languages for speech recognition and vastly increase speech vocabularies (the spoken utterances that can be understood by a speech recognition system) - setting the stage for speech in this next decade.

SPEECH RECOGNITION IN TODAY'S ENVIRONMENT

Today, speech recognition solutions are designed to engage the specific needs of the user/caller. Therefore, there are many different types of speech recognition solutions deployed in the market. These solutions range in function. Some service only niche markets, like field services, while others service

broader markets, like customer care. A basic speech recognition solution understands spoken utterances of the caller/user and executes a command, for example, voice-activated dialing on a mobile handset or basic call routing. Complex speech recognition solutions are more interactive in nature. They can both understand spoken utterances, and provide vocal responses. This enables a level of interaction needed to fully automate transactions. One example of this is speech enabled name and address capture/change applications.

Figure 2 shows the three primary environments for speech deployments: PCs, devices and across a network.



The largest market for speech recognition is 'across a network', as a result of the sharpened enterprise focus around call automation. This is where the majority of spending occurs for speech recognition solutions by businesses worldwide. Datamonitor calls this the voice business market, and defines voice business as *the application of speech recognition technology across networks and associated technologies*. The remainder of the white paper will focus primarily on the voice business aspect of speech recognition solutions.

Companies that have deployed speech

Chances are you have already interacted with a network-based speech recognition solution. Right now there are numerous speech implementations across major firms in communications, financial services, government, healthcare, retail, travel & tourism, technology and utilities markets. Just a snapshot of these markets will deliver an impressive list of global Fortune 1000 companies that have deployed speech, as seen in Table 1.

Outside the contact center, a growing number of service providers, namely telcos and wireless carriers, are deploying speech recognition solutions in efforts to reduce costs, increase profit margins and average revenue per user (APRU). Solutions consist of directory assistance, prepaid account activation and value-add applications which include voice-activated dialing and voice portals.

WHY BUSINESSES ARE INVESTING IN SPEECH TODAY

We have now defined speech recognition and identified both the types of solutions available and the verticals in which they are being deployed. This brings us to our next question: what motivates businesses to invest in speech? The answer is clear; speech solutions are an economical and effective way to improve customer satisfaction, operations, workforce productivity and increase revenue.

The value proposition for speech

The average ROI for a speech recognition solution is typically achieved in the timeframe of 9-12 months and in some cases in as little as 3 months. The bulk of ROI benchmarks are calculated from the cost savings and increases in customer satisfaction in the enterprise, but speech also provides other distinct benefits to organizations. These include improved call completion rates and improved agent morale and utilization. Among service provider lines of business, speech provides companies with the ability to provide value-add services for new revenue generation and differentiation purposes.

The following provides a deeper dive into the value that speech recognition brings to organizations:

- **Cost savings** – Contact centers are able to cut toll and labor costs. Speech accelerates information access, improves routing accuracy and enables more levels of transactions than DTMF-IVR. In addition, as automation increases the number of agents needed drops. The savings are considerable. The average cost for a call handled by an agent in a US-based contact center is \$5.00. A call handled by speech self-service costs \$0.50;
- **Customer satisfaction** – Callers are often able to access information and complete transactions faster with speech than with DTMF-IVR or with a live agent. DTMF-IVR solutions have long menus, and agents typically have long queues. Accelerating transactions with speech has helped drive customer satisfaction higher. Speech technology can also quickly authenticate and route callers to the right departments, which facilitates faster call handling and resolution rates, further improving customer satisfaction;
- **Call completion rates** – Callers can conduct more types of transactions with speech than with DTMF-IVR. For example, a speech recognition system may ask 'where are you flying from?' The caller responds 'I'm leaving from San Francisco on Monday.' In that moment, the caller was able to input multiple variables without having to navigate through lengthy DTMF-IVR

menus to key in the correct response. Depending on the application, the customer can then go through a few more prompts and purchase an airline ticket or be routed to an agent that understands the nature and purpose of the call, as identified through the application. Overall, this increases both call completion and first call resolution rates;

- **Agent morale and utilization** – Deflecting low-level mundane calls to speech recognition solutions frees up agents to handle more complex, higher value calls. This boosts agent morale and leads to lower agent turnover rates, which translates to cost savings for the contact center. Moreover, staffing can be reduced and optimized, reducing agent idle time and increasing agent utilization rates;
- **Value-add services** – For service providers, speech provides a complementary interface that can improve transactions around the delivery of content to the user, e.g. mobile content, e-mail, streaming video, games and music. Given the limitations of the alpha-numeric keypad, consumers can obtain information and complete transactions more effectively and intuitively through a voice portal. This improves the breadth and depth of value-add services, in addition to introducing new differentiation points and revenue channels for companies.

Greater flexibility and performance at lower prices

In the last five years, the options available to businesses looking to invest in speech have dramatically increased. The advent of Voice-XML has freed them from vendor-lock in and given them a wider range of application designers to work with. New players have entered the larger speech market, lowering prices and increasing solution choice. And new deployment options have also proliferated; hosted and premise-based managed solutions have grown in popularity. Finally, best practices and customer behavior analysis has improved.

The move towards Voice-XML platforms

For over 20 years, proprietary IVR systems were the 'tried and true' phone-based automation, routing and service enablement application and accounted for the vast majority of IVR deployments in the market. These systems have since earned the name traditional IVR systems. But in recent years, as business needs have outgrown the constraints of rigid, legacy TDM-based telephony architectures - the native environments for traditional IVR systems - organizations have begun to move to more flexible Web services deployment models that favor Voice-XML, an open standard. This has helped change the direction of IVR investment. Today, more organizations are replacing traditional IVR systems with Voice-XML platforms to complement their speech strategies and migration to a Web-centric architecture, in addition to leveraging greater flexibility, interoperability and performance.

The entry of new players in the speech market

The emergence of open standards and the growth in speech deployments in the 1990s has attracted new players into the speech market. This has increased competitive practices and driven prices down and innovation up. The result is the availability of better and more affordable products in the market.

Growing demand for hosted and managed services

In the past five years, as networks matured and business practices changed, adoption of application service provider (ASP) and software-as-a-service (SaaS) models for enterprise and contact center applications grew tremendously. The demand for these deployment models transferred over to the speech market as more businesses looked to hosted and managed services to leverage the benefits of speech without having to heavily invest in the upfront costs for a premise-based speech solution.

Improved vendor best practices and customer behavior analysis

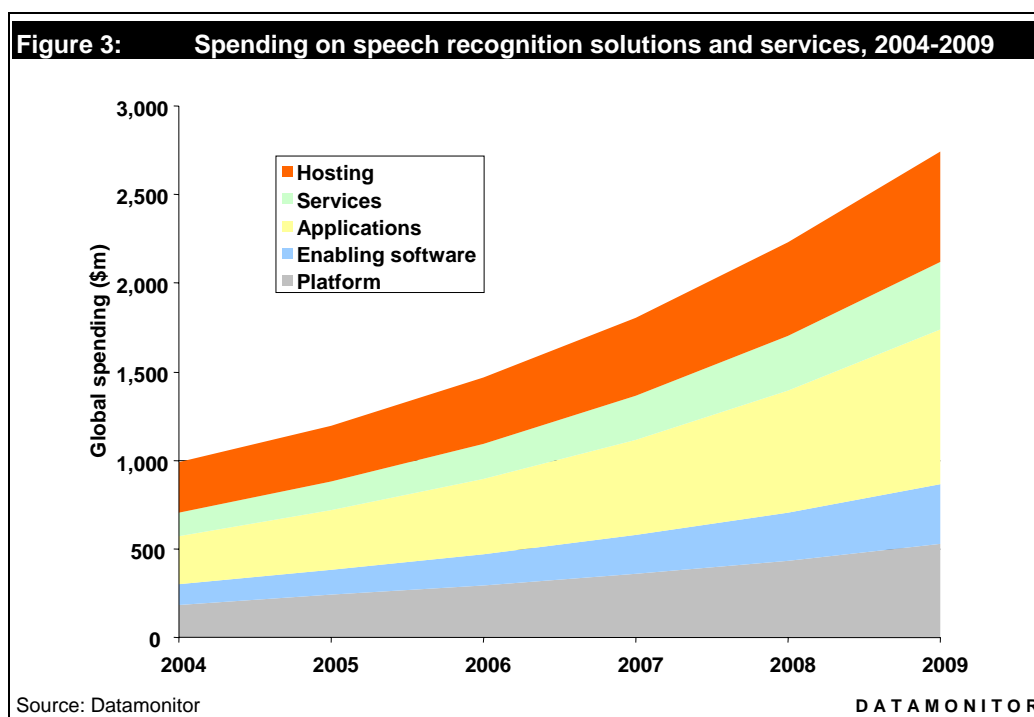
Recent improvements in processing power and speech recognition rates are enabling vendors to move past technological obstacles and towards tailoring their applications to business needs. They have been developing more structured best practices for speech application design, refining customer behavior analysis to provide solutions which zero in on concrete business objectives. Given that the application is the critical point at which the customer meets the speech solution—and the company—material progress in this area greatly improves the effectiveness of speech solutions.

Introduction of simplified application development environments and reusable components

Many vendors have introduced GUI-based, drag-and-drop speech application development tools and reusable application components that simplify and expedite speech application development efforts. These tools provide a higher abstraction layer for the development and rendering of speech applications so developers no longer have to engage in low-level coding unless needed. Many of these tools also support different development environments such as Java and C++ increasing the development options for speech application builds. Reusable application components are basic, pre-built speech application modules that can be assembled with other modules and customized to build a complete application.

The market uptake for speech recognition solutions

Greater flexibility and performance and lower prices have improved the value proposition for speech. As a result, spending on speech recognition solutions and services eclipsed the \$1 billion mark in 2005, providing validation that speech is a commercially viable technology. Businesses from across the globe spent roughly \$1.2 billion on speech recognition solutions and services that year. By 2009, spending is expected to more than double to reach \$2.7 billion at a CAGR of 22.6% as shown in Figure 3.



THE COMPONENTS OF A SPEECH RECOGNITION SOLUTION

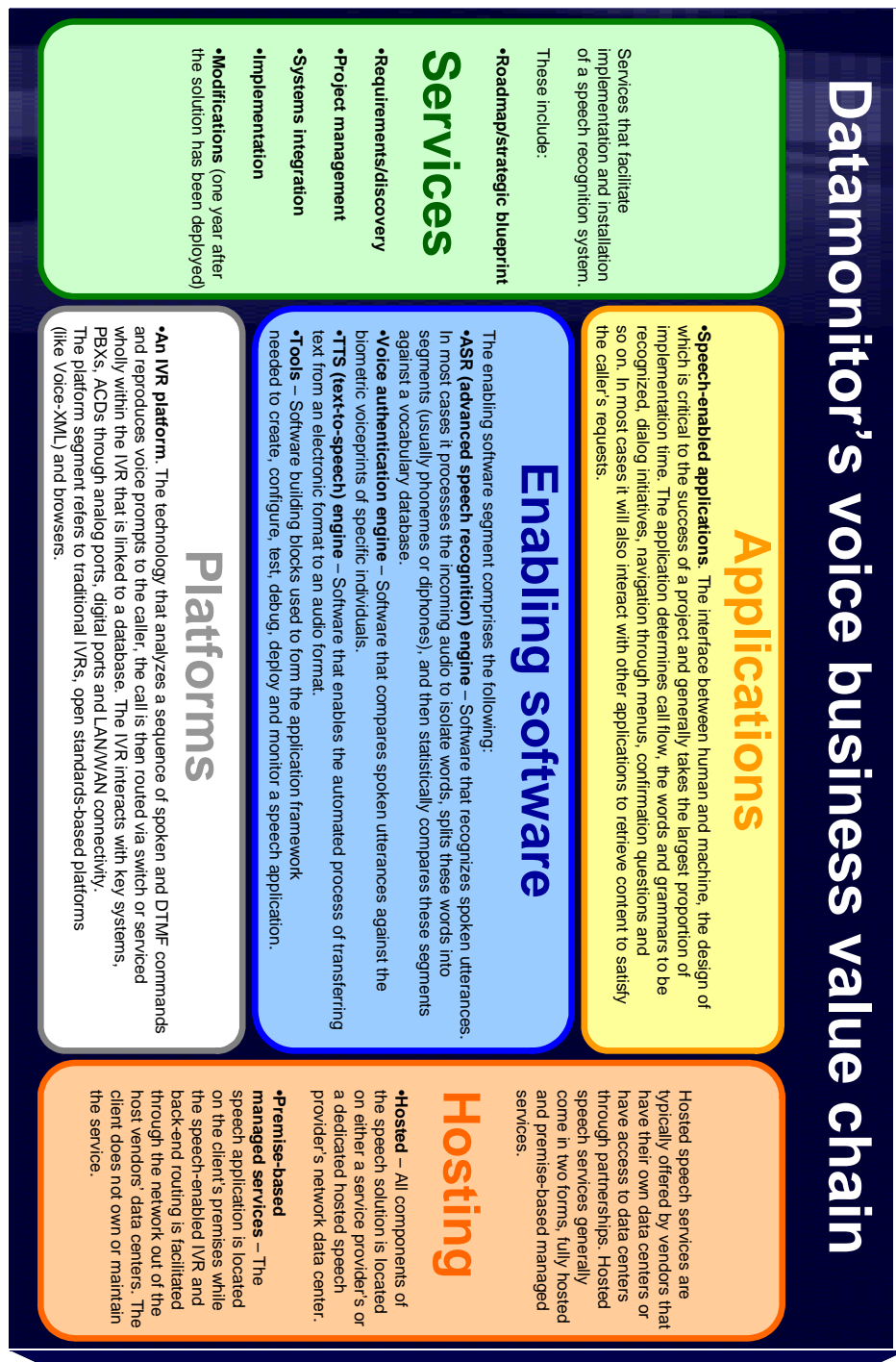
Network-based speech recognition solutions consists of a complete technology stack that includes applications that run on top of advanced speech recognition (ASR) engines and pre-recorded responses (by voice actors) or text-to-speech (TTS) engines layered on top of a speech-enabled IVR platform. Tracking spending on these specific products alone does not provide an accurate gauge for how much businesses are spending each year on speech recognition solutions. Services, hosting and tools account for a large percentage of speech recognition solution spend on an annual basis.

Figure 3 illustrates total spending across the five different components which Datamonitor defines as the voice business value chain. These are platforms, enabling software, applications, services and hosting which are defined and discussed in detail in Figure 4 on the next page.

The vendors that provide these technologies and services

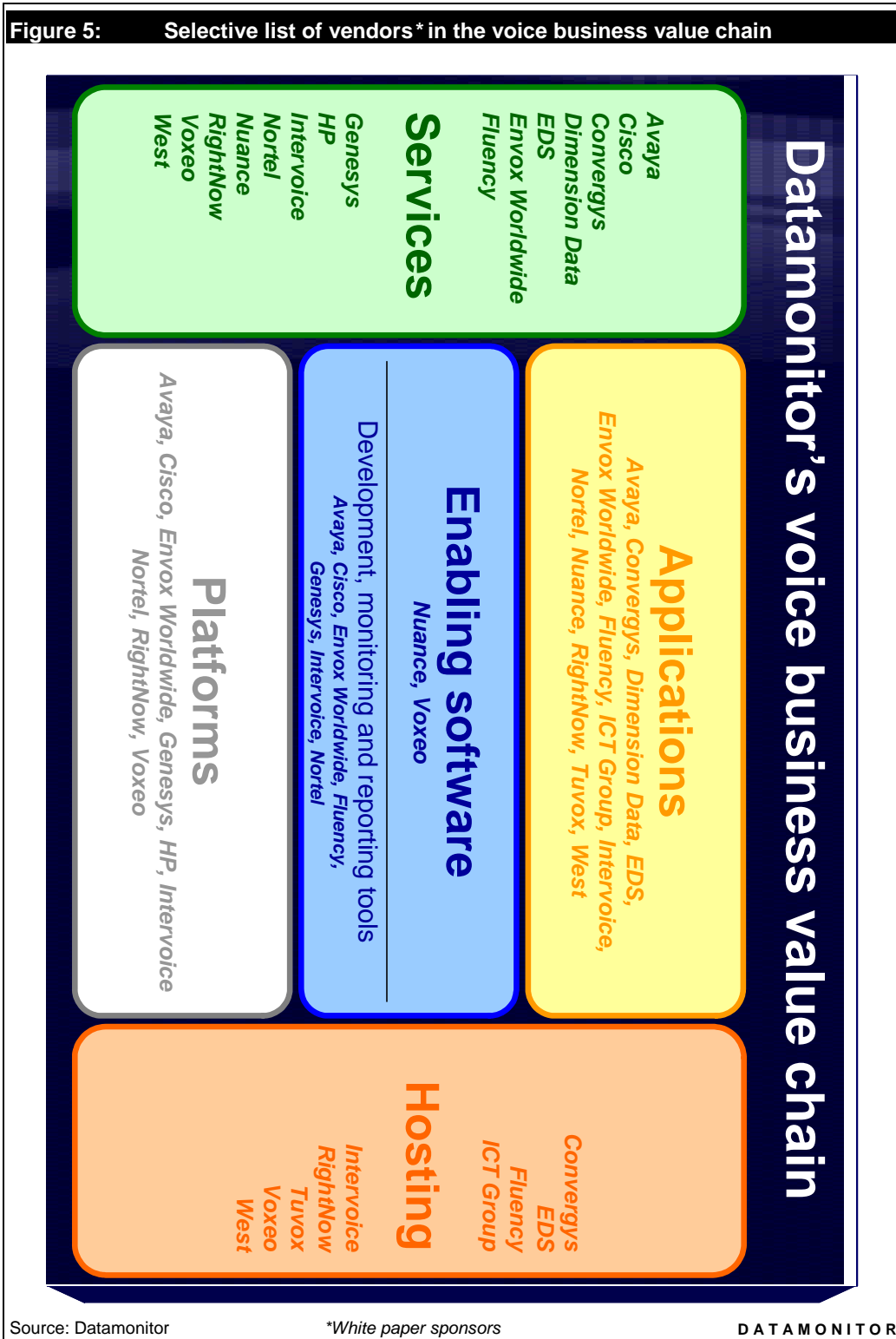
It is often difficult to understand what technologies certain vendors offer across the voice business value chain. Figure 5 provides a simple graphic that highlights some of the key vendors that compete within each component of the voice business value chain.

Figure 4: The voice business value chain defined



Source: Datamonitor

DATAMONITOR



Snapshot of vendors in the voice business value chain

There are many vendors, outside of those listed in Figure 5, that also provide products and services across multiple areas of the voice business value chain. Some of these vendors are highlighted in the Appendix. This section provides an overview of each vendor, as provided by the vendors, represented in Figure 5. The vendors are listed in alphabetical order.

Avaya – Avaya designs, builds, and manages communications solutions for 1m+ businesses worldwide, including 90%+ of the FORTUNE 500®. Avaya offers end to end lifecycle services and flexible support from managed services to hosted solutions via Avaya on Demand. In 2005, Avaya released Voice Portal and Dialog Designer. Voice Portal is a Linux based platform that features license pooling and automatic failover, delivering superior reliability and business continuity. Dialog Designer is a free Eclipse-based IDE that includes powerful simulation tools that reduce deployment time and risk. Avaya preserves today's software investments for Conversant users via common tools and technologies, open standards support, and transferable software platform licensing. *Currently serving all global markets.*

Cisco Systems – Cisco Customer Voice Portal provides organizations with a world-class automated voice self-service solution complete with call-management and call-treatment capabilities. Through the use of VoiceXML, the solution can use the same information available to customers on the corporate Web site to enhance the customer service experience. With support for automated speech recognition (ASR) and text-to-speech (TTS) capabilities, callers can obtain personalized answers to their questions and conduct business in innovative ways, without the costs of interacting with a live agent. Cisco Customer Voice Portal protects existing investments in contact center technology assets since it can be deployed in TDM-based, IP-based, and hybrid contact center environments. *Currently serving all global markets.*

Convergys – Convergys Self-Service provides hosted speech recognition solutions to organizations seeking to automate high call volume. Complete application design and deployment services for the Convergys VoiceXML-based SpeechPort redundant platform are available. Key differentiators include an Open Hosting model that provides more external control to a hosted application for customers and partners, an exceptionally secure infrastructure, and web-based reporting for business and technical analysis of call activity. Experience as the world's largest outsourcer of contact center agents facilitates integration between Convergys speech and multi-vendor agent systems. *Currently serving all global markets.*

Dimension Data – Dimension Data is a specialist IT services and solutions provider with expertise in networking, security, operating environments, storage and customer interaction technologies. Through our customer interactive solutions, we've deployed some of the most innovative, usable speech solutions in the world, demonstrating cutting-edge expertise in Voice User Interface and Persona Design. Our excellence in speech applications is acknowledged through multiple international awards including: Nuance 2004 Global "Best Practice" and "Best Vision"; SpeechTek 2004 International

“Innovation” and “Best Personality”. For more information, visit www.dimensiondata.com. *Currently serving North America, EMEA and APAC.*

EDS – EDS globally delivers a portfolio of information technology, applications and business process outsourcing services to clients worldwide in the manufacturing, financial services, healthcare, communications, energy, transportation, consumer & retail industries and to governments. Customer Self-Services is a managed hosted IVR/Speech recognition service that enables enterprises to deliver customer support, complete transactions, and provide content-rich information. Enterprises can transition to a reliable, fully managed solution without business disruption while tapping EDS’ proven methodologies, deep industry and technology expertise. *Currently serving all global markets.*

Envox Worldwide – Envox Worldwide is a leading global provider of voice solutions. With over a million ports deployed and one million contact center agents served worldwide, the Company’s products dramatically reduce the time, cost and complexity of automating and personalizing customer interactions through voice self-service and contact center solutions. Envox enables customers to leverage legacy infrastructure investments while ensuring a smooth migration to new standards and technologies. Available through a global partner network, Envox’s award-winning products include: Envox Communications Development Platform; Envox VoiceXML Studio; Envox CT Connect; and, Envox CT ADE. *Currently serving all global markets.*

Fluency Voice Technology – Fluency Voice Technology is a leading provider of packaged speech recognition applications for use in call centers. Fluency’s solutions enable enterprises to significantly reduce costs and enhance customer service by automating much of the high volume call center activities. Fluency most recently launched the Virtual Speech Agent (VSA) Suite, a packaged application developed to automate call center workflows and solve common call center tasks across industries. Fluency continues to perfect its expertise as a developer and provider of speech solutions offering customers both packaged and custom deployments in a hosted or customer premise environment. *Currently serving North America and EMEA.*

Genesys – Genesys, an Alcatel company, is one hundred percent solely focused on open and standards-based (VoiceXML, HTTP, SIP, MRCP) self-service platforms since 1999. Genesys allows enterprises to achieve key business objectives by integrating the self-service platform with other call center assets in both traditional telephony and IP environments, enabling sophisticated routing and reporting capabilities. Genesys is the global leader in VoiceXML platforms with solutions spanning OEMs, enterprises and service providers. Genesys is also the market leader in migrating enterprise customers from legacy IVR systems with over 500 customer migrations in the last four years. *Currently serving all global markets.*

HP – 40 of the top 50 operators rely on HP OpenCall Media Platform which has a proven track record of more than 500,000 voice ports deployed successfully worldwide. HP’s OpenCall Media Platform is an open standards-based, broadly featured platform that supports a wide variety of voice and video applications and allows operators to consolidate their IVRs onto a single, easy-to-manage platform

enabling cost control and increased business agility. As video capabilities and IP Multimedia Subsystem (IMS)-based services are supported, an investment in HP OpenCall will return strong benefits long into the future. *Currently serving all global markets.*

ICT Group – ICT Group is a leading global provider of customer management and related marketing, technology and BPO solutions with operations in 9 countries. Through ICT Global Interactive, our CRM Technology Solutions organization, we provide speech-enabled IVR, outbound alerting (e-mail, SMS, voice), web self-service (e-mail management, contact management, knowledge base, co-browsing), hosted multi-channel support (ACD, CTI, digital recording/monitoring), and application design/development, transcription, interactive personality creation, scripting, voice recording, and more. Key differentiators for ICT Group include an expansive global footprint, over 20 years' experience and solid financial performance. *Currently serving all global markets.*

Intervoice – Intervoice provides leading enterprises and network operators with the Intervoice Solutions Framework, including the platform, **tools**, applications and services necessary to optimize the customer experience through the delivery of voice automation solutions. Intervoice Media Exchange, the open, standards-based multi-media platform, offers unparalleled flexibility for advanced multi-media messaging, portal, IVR and payment applications. Intervoice is focused on the enterprise and network markets, providing solutions that improve operational efficiencies, drive revenue and increase customer satisfaction and loyalty. *Currently serving all global markets.*

Nortel – Nortel is recognized as a leading provider of self-service solutions for more than thirty years; deploying commercial speech applications more than fifteen years ago. Nortel offers advanced speech technology, multimedia platforms, GUI and web based tools, and network services. Nortel Applications Center provides a framework for improved customer service via advanced speech, IP-enabled video communications, real-time multimedia, collaboration, unified messaging, computer-telephony integration, and virtual contact center. This enables a cohesive application environment with a common media server, tools, and reporting and administration. *Currently serving all global markets.*

Nuance – Nuance is the leading provider of speech and imaging solutions for business and consumers around the world. Its technologies, applications, and services make the user experience more compelling by transforming the way people interact with information and how they create, share, and use documents. The combination of Nuance's unique HumanTouch™ technology and leading speech professional services delivers the best possible experience for your callers and the best return on your speech investment. Every day, millions of users and thousands of businesses experience Nuance's proven applications. For more information, please visit www.nuance.com. *Currently serving all global markets.*

RightNow – RightNow Voice™ helps organizations deliver a superior customer experience by offering voice-driven applications that dynamically present a wide variety of options and content to callers. Pre-built voice applications act as "live" agents and seamlessly co-exist with customer service agents and other resources to either fully or partially automate calls. RightNow Voice™ supports complete multi-

channel self-service strategies, ensuring that customers will receive consistent information across channels. It can be fully hosted or deployed on-premise where speech infrastructure is already in place and delivered as a managed service to assure ongoing success. *Currently serving all global markets.*

Tuvox – TuVox provides a wide range of On Demand speech applications including natural language call routing, over 50 self-service applications and knowledge based diagnostics. By delivering a superior caller experience, TuVox improves the adoption of voice self-service and delivers higher automation and cost savings. TuVox speech applications are delivered in both hosted and on-premise managed services and can quickly be updated in response to business changes. Customers only pay for automated calls and there is no need for expensive infrastructure. TuVox applications have a total cost of ownership of 50-60% lower than custom coded applications. *Currently serving North America, EMEA and APAC.*

Voxeo – Voxeo Corporation makes telephony applications as easy to deploy as web applications. Voxeo offers hosted VoIP and IVR services, premise VoIP and IVR systems, and maintains one of the most active voice application customer portals in the world. With over \$40 million invested in IP and XML technologies and infrastructure, Voxeo serves customers ranging from small startups to the Fortune 500 who value speed and simplicity where complexity once reigned. For more information visit <http://www.voxeo.com> or call +1 (407) 418-1800. *Currently serving all global markets.*

West Corporation – West Corporation is based in Omaha, Nebraska, and had 2005 sales of \$1.5B. The company has provided interactive services since 1989 and today is a leader in the deployment of Natural Language ASR solutions. West is the largest IVR hosting service in North America with over 120,000 ports including 26,000 ASR and 46,000 VoIP-enabled ports. West's platform is a distributed system based on open industry standards such as VXML, MRCP and SIP. West also offers network-based, on-demand ACD services already supporting more than 11,000 CSRs, a real-time CSAT survey tool and an automated notification service capable of reaching more than 400,000 contacts per hour. *Currently serving North America.*

STRATEGIC RECOMMENDATIONS FOR BUYERS

Choosing and deploying a successful speech recognition solution can be difficult without the right strategy. The endeavor typically involves a team of individuals working towards one common goal – successful creation, deployment and ongoing management of a speech recognition solution. While application developers, systems integrators, consultants and VUI specialists focus on design, development and interoperability, business managers must focus on ROI, customer service and operational aspects of the speech solution. This will help ensure a successful speech deployment that will yield measurable ROI results, improve customer service and profitability. The following are Datamonitor's recommendations for companies that are investing in speech recognition solutions:

1. **Do your research** – This is the first step. It is important that businesses conduct extensive research to understand the technology, the vendors and the market. Businesses should review case/ROI studies across relevant vertical markets. This will provide an understanding of what similar businesses have accomplished with speech. They should also track the market movement for speech adoption. This will provide perspective into the market adoption rates for speech recognition in addition to the different types of application deployments in the market.
2. **Create a plan** – The success of a speech recognition solution hinges on careful planning and the right execution. Businesses should take into account the findings from their research and create a tactical and strategic plan that clearly defines short- and long-term goals.
3. **Pick the right vendors** – Having specified goals, businesses must now choose the right vendor or vendors that will help them achieve their goals. Towards this end, businesses must evaluate vendors against the following: best-in-class technology, financial viability and investment protection, level of expertise, proven success and the strength of their partner ecosystem.
4. **Select the right deployment model** – Speech recognition comes in three different deployment models: premise-based traditional, premise-based managed services and hosted services. Each deployment model has its distinct advantages.

Premise-based traditional – All components of the speech solution (platform, engine, application) are located and operated on the client's premises. This is the most common deployment model for speech and provides businesses with autonomy over the solution.

Premise-based managed services – The speech application is located on the client's premises while the speech-enabled platform and back-end routing is facilitated through the network out of the vendor's data centers. The client does not own or maintain the IVR platform and routing. Ports or platforms are typically shared in a multi-tenant environment, clients take ownership of the application and security can be provided by either the client or the provider. With this model clients are able to focus resources on managing the speech application while connectivity and routing are delegated to the provider. The client relinquishes some control to the managed service provider.

Hosted services – All components of the speech solution are located on a dedicated hosted speech provider's network data center and all speech application development and management and routing is outsourced. Ports or platforms are typically shared in a multi-tenant environment and application management and security is provided by the hosting vendor. Hosted services provides several cost advantages over other deployment models as it enables clients to evaluate and leverage the benefits of speech without having to absorb the heavy upfront costs. Clients do not have to hire additional personnel or pay for heavy systems integration services. Most of the costs for the entire solution stack and professional services

are amortized across minutes, calls and transactions in a hosted services model. However, the client relinquishes much of the control to the hosted speech provider.

Businesses are not confined to one specific deployment model. In fact, many organizations today have deployed speech across premise-based traditional, premise-based managed services and hosted models to improve business continuity and cut costs.

5. **Start small and scale up** – Businesses should work closely with their vendor to deploy a small pilot solution to test the waters. They should always start small and scale up as applications are refined and callers/users become more familiar with interacting with a speech application. Businesses should deploy simple applications and then progress to more complex applications that fully automate transactions.
6. **Focus on lifecycle management** – Simply building, deploying and tuning a speech application is not enough to achieve long-term success. Businesses must dedicate resources to continuously improve and optimize speech recognition on an ongoing basis. This requires significant post deployment monitoring and tuning but will produce the best user experience for callers/users and help drive automation and customer satisfaction rates higher. The focus should be on the customer, and businesses should choose a vendor that tailors its solutions to customer preferences and is constantly reaching for higher levels of customer satisfaction.

CONCLUSIONS

Speech recognition has matured over the past 50 years from a concept to a reality. It is a complex technology but offers practical business benefits to organizations. As a solution it provides an economical and effective way to reduce overhead costs while improving customer satisfaction, operations and workforce productivity. Looking forward, as people become more accustomed to interacting with a voice user interface, speech recognition will be integrated into a wider range of services and products outside of the contact center and service provider environments. The technology will continue to improve and proliferate in the market helping drive greater cost savings, top-line revenue and workforce productivity.

APPENDIX

Other vendors competing in the voice business value chain (not an exhaustive list)

Accenture, Aiptera, Aspect, AT&T, Bearing Point, Bell Canada, BeVocal, BT, Comverse, Datria, Eckoh, Empirix, Gold Systems, Holly Connects, IBM, Interactive Intelligence, Loquendo, Lumenvox, Metaphor Solutions, Microsoft, MTI, Oracle, Prarie Systems, Pronexus, Qwest, Roger's Business Solutions, Sprint, Syntellect, Telephonetics, Telisma, Tellme, Telstra, Telus, Unisys, VeCommerce, Verizon, Vicorp, Viacore, Voice Vault, Voice Objects, Voxify.

Definitions

Dual tone multi-frequency (DTMF) - The signal to the phone company that a caller generates when he/she presses keys on a telephone's keypad. In North America and it is commonly known as touchtone phone (formerly a registered trademark of AT&T). DTMF has generally replaced loop disconnect ('pulse') dialing.

Interactive voice response (IVR) - A technology that analyzes a sequence of spoken and/or DTMF commands and reproduces voice prompts to the caller, the call is then routed via switch or serviced wholly within the IVR that is linked to a database. The IVR interacts with key systems, PBXs, ACDs through analog ports, digital ports and LAN/WAN connectivity.

Natural language understanding (NLU) – A statistical language- and semantic-based approach that interprets the meaning of a string of words. Rather than using the traditional 'finite state grammar' approach, which lists possible utterances, NLU enables callers to speak in infinitely varied ways (using disfluencies, such as 'um') and get understood. (This definition is provided by Nuance Communications)

Author

Daniel Hong, Senior Voice Business Analyst

Daniel heads Datamonitor's voice business research and is considered a thought leader in the speech recognition industry. He is the author of several reports and studies that examine the current opportunities, relevant issues and future direction of the global voice business market. Daniel has written many industry white papers on speech recognition solutions and engaged in numerous consultancy projects for various Fortune 1000 companies. He can be contacted at dhong@datamonitor.com