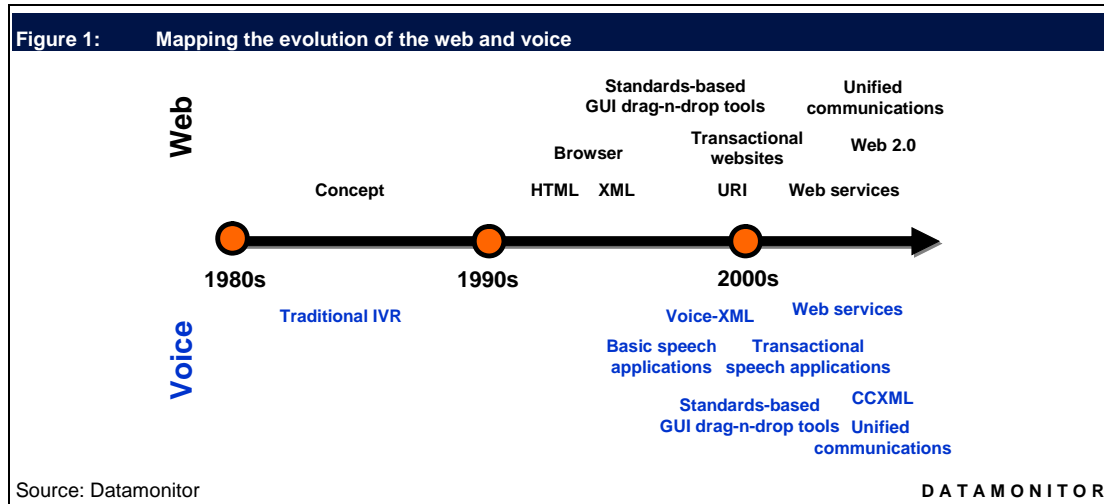


COMPARING THE EVOLUTION THE WEB AND VOICE CHANNELS

The web side story

Web design has changed dramatically over the past 10 years, driven by the constantly changing development and user environments. The HTML standard has evolved from a rigid structurally-based markup language to an extensible HTML and XML hybrid. The CSS standard helps separate presentation from content. Information, workflow and processes are assembled through connecting services together into applications that are presented to a browser. Browsers, like Internet Explorer and Firefox, are constantly evolving and introducing new features and functions. Websites which were initially no more than informational brochures have now become self-service terminals, allowing users to perform a wide variety of transactions from their desktops and mobile devices. No longer are websites slabs of text scattered with isolated GIFs and JPEGs, but now they support multimedia applications. From a back-end perspective, the barriers to entry for web application development have become much lower with the advent of GUI-based drag-and-drop tools, which greatly expedite development and deployment time.

As illustrated in Figure 1 the evolution of the web has played a pivotal role in the evolution of voice. Beginning in the 1990s standards such as HTML and XML and browsers set the stage for the web. This helped bring forth standards-based GUI drag-and-drop tools which provided developers with a quicker and simpler way to create websites. In the following years, developers expanded the use of these tools and created more sophisticated transactional websites (i.e. Amazon.com) which were more user-centric. And recently, web services have emerged in this space providing for greater interoperability and service oriented architecture (SOA) implementations.



The changing voice infrastructure

As shown in Figure 1, voice is building on innovations pioneered by the web. Voice-XML was created just a few years after XML and many Voice-XML IVR systems today are deployed in a web services framework. Proprietary languages and traditional IVR systems are losing sway in the market to standards-based languages and Voice-XML platforms. This coming year, 2008, will be the first in which the number of annual Voice-XML licenses shipped surpasses that of traditional,

legacy-based IVR. The growing adoption of Voice-XML is indicative of the fundamental shift in investment philosophy across many businesses. Whether in the web or voice environment, companies fear getting stuck with managing monolithic, expensive solutions that other elements in the network must conform with, be coded to and be designed around. As a result, businesses are graduating beyond siloed technologies and are looking to build on application paradigms to move towards a common standardized web architecture that provides interoperability across disparate systems. This is the driving force behind the SOA movement in the enterprise. SOA provides companies with the means to treat certain business processes and the underlying IT infrastructure as standardized, secure reusable components to address changing organizational needs. Looking forward, web and voice technologies in the enterprise will be deployed in an SOA environment.

It is important to note that although Voice-XML makes it simpler to develop speech applications by creating a standard application description language, platforms based on this open standard have been widely deployed for DTMF solutions as well. However, speech recognition technology is growing at a more rapid rate than that of DTMF, making inroads into the IVR landscape and sharpening focus on increasing automation rates, reducing costs and improving phone-based customer service. From a development perspective, proprietary tools are being phased out, as are legacy systems; new tools informed by the feel and presentation of GUI-based, drag and drop web development tools are gaining traction in the voice channel.

Leveraging the web paradigm in voice

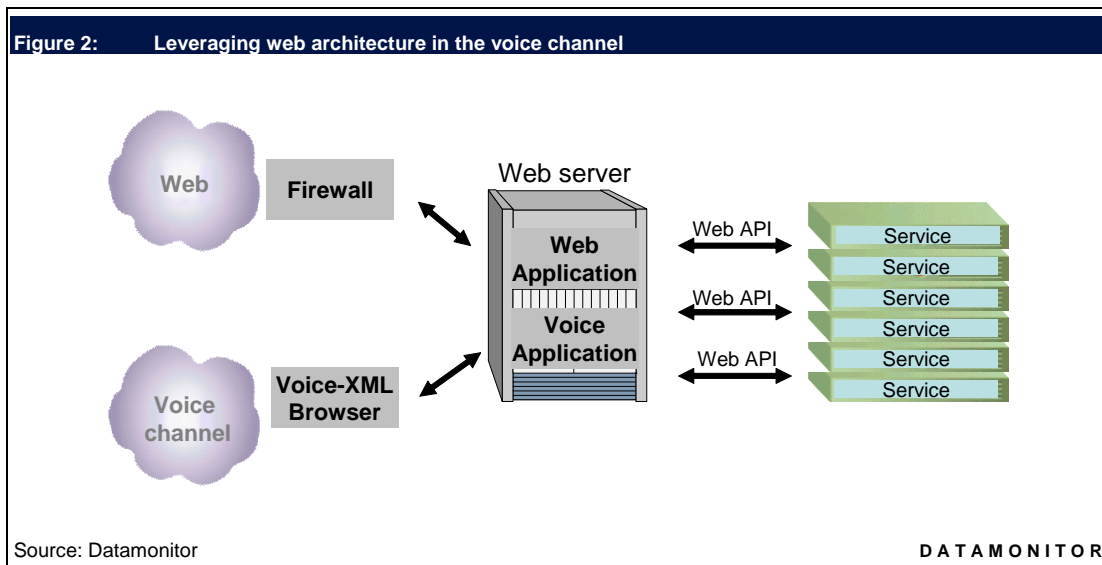
While the details of web and voice interface design are as different as music and painting, their core mission is identical. The success of both resides in the 'presentation layer,' which represents the harmonious relationship between input – whether through keystrokes or utterances – and output through browsers. The HTML browser displays images, sounds and text to users, just as the Voice-XML browser presents audio and text-to-speech (TTS) to callers over the phone. Both are key customer touchpoints, which must be optimized for cost as well as customer service. Therefore the web offers voice a valuable example, both architecturally and structurally.

In the earlier years of web design, corporate websites were used mainly as a broadcasting medium where the terms 'viewers' and 'audience' were frequently used to describe what we refer to today as 'users'. Since the mid 1990s, corporate websites have become more sophisticated, evolving from static to transactional, inflexible to extensible, textual to visual and audible. Today, corporate websites provide two-way interaction with customers and exist to provide users with rich self-service capabilities. User adoption is high and continues to grow year over year as more consumers turn to web self-service. Corporate websites could not have reached this level of sophistication without the introduction and subsequent evolution of standards, and of intuitive development environments.

Web and voice parallels are becoming instructive

The similarities of voice and web were not always so clear. Initially, presentation layer in the voice channel was extremely restricted; despite intense pressure to reduce live agent costs, DTMF simply did not offer a sufficiently flexible interface. Speech represented the most economical solution to reducing costs while increasing self-service sophistication in the voice channel. However, even with the advent of commercially viable speech recognition technology in the late 1990s, businesses were still constrained by inflexible proprietary environments of traditional IVR systems. Application development and call flow creation required extensive knowledge of vendor-specific proprietary programming and scripting languages.

Therefore, businesses were heavily dependent on vendor professional services making the TCO of a speech solution very high and unattractive. But with the advent and adoption of Voice-XML, the environment changed. Businesses now have an open standard and similar flexible underlying architecture as that of the web providing for application reuse and portability in some cases, long-term investment protection and intelligent transactions that did not require hordes of expensive professional services engineers. Because Voice-XML IVR solutions resemble distributed web applications it truly leverages existing web architecture. Voice and web applications can be parked on the same server. These applications can then access and assemble services through common web APIs, apply business logic to them and present the appropriate data to either an HTML or Voice-XML browser, as seen in Figure 2. Moreover business rules, data access and code assets are aggregated within the application server (i.e. Websphere). As a result, Voice-XML provides for IVR interactions to be determined from a consolidated application infrastructure which provides for more intelligent interactions with personalization capabilities over the phone at a lower cost.



The parallels between web and voice are now becoming instructive. Businesses can introduce new levels of sophistication in the voice channel by fully leveraging Voice-XML, web architecture and speech recognition and developing well-designed applications that help drive enthusiastic adoption by customers. Businesses that want to achieve this while keeping their costs under control should consider a long-term tooling strategy for the voice channel.

THE GROWING IMPORTANCE OF TOOLS IN THE VOICE CHANNEL

The need for tools is predicated on domain expertise, application type and complexity, resource availability and development costs. The majority of businesses today would not consider building their corporate websites from scratch through coding in a text editor as it is very labor intensive, expensive and time-consuming. In fact, labor accounts for 70% of IT budgets. In order for companies to achieve economies of scale and favorable ratios, businesses have come to rely heavily on tools. However, in the voice channel the use of tools for speech application development has not been widespread. Most companies have become heavily dependent on 3rd party professional services for speech application

development and design. To provide a frame of reference, a recent report by Datamonitor, *Defining Packaged, Pre-built and Custom Developed Speech Applications (BFTC1783)*, states that:

The vast majority of speech application development and management is handled by vendors because most clients lack the requisite skill sets in-house to design, develop, manage and optimize speech solutions. Therefore, clients are heavily dependent on vendor professional services spiking the TCO of a speech solution. In efforts to reduce the TCO of a speech solution a growing number of companies (namely early adopters) will look to become more self-sufficient in managing and maintaining a speech application over time...

Deploying a well-designed speech application requires commitment to the entire application life cycle; applications must be continuously fine-tuned and improved to drive customer adoption and satisfaction. Today, this is typically handled by a 3rd party vendor, however over time businesses should consider bringing more segments of the application life cycle management in-house as they become more adept with speech and look to reduce the TCO of a speech solution. Tools provide a long-term cost effective solution for businesses investing in voice.

Understanding tool options in the marketplace

Datamonitor defines a voice application development tool as the building blocks used to form the application framework needed to create, configure, test, debug and deploy a voice application, and defines a voice application management tool as an application that provides post deployment logging, monitoring and analysis capabilities. Sometimes referred to as service creation environments (SCE) or integrated development environments (IDEs), many of the voice tools available in the market stem from traditional IVR proprietary development environments. However, the industry's embrace of open-standards has resulted in the advent of new tools in the market that enable developers to create, test, debug and maintain Voice-XML-based applications. These tools typically are designed to work within a web framework and provide greater value than the tools used in proprietary development environments.

There are many choices in the market when it comes to speech application development and management tools. Most tools are bundled with IVR platforms and some are provided by application vendors and a select few by 3rd party tool vendors. The following comparative matrix of standard features and differentiators, Figure 3, is from a recent Datamonitor report, *A Deep Dive Into Speech Application Development, Management and Testing (BFTC1793)*. As the figure shows, there exist many standard and differentiating features of voice tools in the market. Over time as tools evolved many of the differentiators became standard features, including GUI drag-and-drop design and Voice-XML support.

The majority of voice tools come with a dialog / call flow editor and have design features that include: grammar, prompt and offline development. Standard testing and monitoring capabilities include dialog, advanced speech recognition (ASR) / grammar, logging and runtime analysis. The differentiators, as seen in Figure 3, are representative of the differentiating tool features in the market today. As the dividing lines between IT and telecom environments blur, the need for separate web and voice tools will disappear and these IDEs will begin to merge. Today, a growing number of voice tools today can be used as plug-ins that can be added to Eclipse offerings of web application servers.

Figure 3: Standard features vs differentiators of speech application development tools

<u>Capabilities</u>	<u>Standard features</u>	<u>Differentiators</u>
Design	<ul style="list-style-type: none"> • Dialog / call flow editor • Grammar development • Prompt development • Offline development 	<ul style="list-style-type: none"> • Call recording • Availability of component libraries • Application portability
Testing and monitoring	<ul style="list-style-type: none"> • Dialog testing • ASR / grammar testing • Logging tools • Runtime analysis 	<ul style="list-style-type: none"> • Real-time monitoring • Real-time analytics • Multi-channel analytics • Predictive analytics
Interface and support	<ul style="list-style-type: none"> • GUI drag-n-drop • Voice-XML compliant 	<ul style="list-style-type: none"> • Support of multiple Voice-XML platforms • Open, web-based IDEs • Eclipse support

Source: Datamonitor DATAMONITOR

In the past couple of years there has been strong market demand for improved analytics and investment protection. As a result, vendors have introduced these new features into their tool offerings to help them separate themselves from the rest of the pack. While the standard features are table stakes, the differentiators are unique in the market and vendors have different approaches to these features. The following provides a snapshot of the types of tool offerings in the market by vendor category:

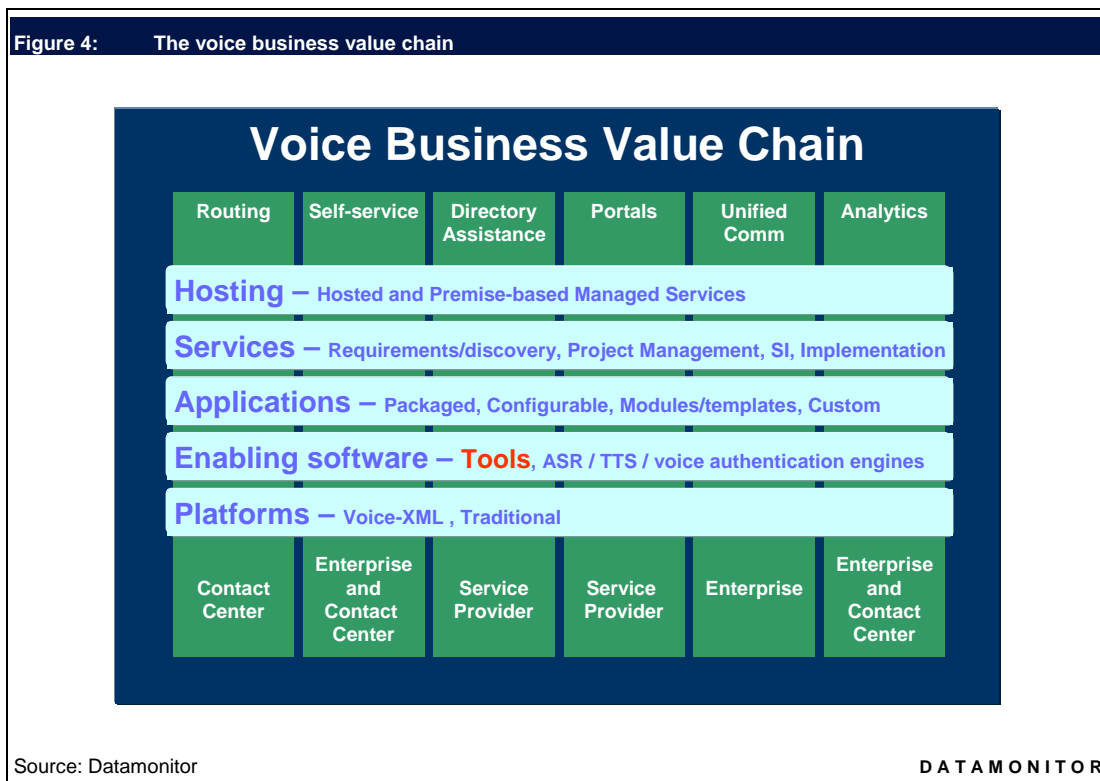
- **Platform vendors** – This category includes IVR platform vendors that provide speech application development tools that are tightly integrated with their respective platforms. Most of the tools, which support Voice-XML, stem from previous traditional IVR proprietary development environments. Although proprietary, tools that serviced speech application development for traditional IVR has a proven track record of success and platform vendors hope to support both proprietary and open environments for speech application development with their tools. Typically these tools cannot work with other platforms. Platform vendors normally absorb any run-time licensing fees into the overall solution cost for their tools as the core business revolves around selling IVR platform licenses, services and sometimes applications;
- **Application vendors** – This category includes speech application vendors that empower their customers with more control over the application. Most application vendors design and deploy the speech application for their clients and offer a tool to enable them to make minor changes, typically these tools are as bespoke as the

applications themselves. Application vendors do not typically charge for any run-time licensing fees and rarely focus on selling tools, rather they focus on selling professional services and application licenses;

- **3rd party tool vendors** – This category includes vendors whose primary products are speech application development and management tools. Tools from these vendors have an advantage in that they support a variety of platforms providing options for their customers. In addition, 'the touch and feel' of these tools are usually 'niftier' than those provided by platform or application vendors, as tool vendors invest heavily in tool design. Tool vendors typically charge fees for both run-time licenses and developer seats.

TOOLS IN THE CONTEXT OF THE CHANGING VOICE LANDSCAPE

Tools are part of a larger solution that consists of a complete technology stack which includes applications that run on top of ASR engines (for speech solutions) and pre-recorded responses (by voice actors) or TTS engines layered on top of an IVR platform. These are implemented through a bevy of services and deployed on premises-based solution or in a hosted or a premise-based managed services model. All these components create the voice business value chain. Figure 4 shows the voice business value chain aligned with application types and environments.



Within the voice business value chain, tools (in red font) are classified as enabling software alongside ASR, TTS and voice authentication engines. Tools are viewed as an enabling component of speech as they provide the means to design, create, deploy and manage an entire speech solution from the platform to the application levels. Over the last few years

platforms and enabling software have become more commoditized shifting value towards the higher valued areas in the value chain, namely applications, services and hosting. With commoditization came consolidation. In fact, over the last five years, there has been heavy consolidation across the entire voice business value chain indicating a maturing market. As a result of the heavy M&A activity, the number of vendors providing tools (whether they are platform, application or 3rd party tool vendors) has shrunk. At this point, Vicorp is the only major 3rd party tool vendor focused exclusively on the voice channel.

Q&A WITH VICORP

Vicorp provides a unique approach to creating, configuring, testing, debugging, deploying and managing voice applications. The vendor's tool enables rapid development and deployment cycles as well as application portability across Voice-XML platforms. To better understand Vicorp's tools and services offerings Datamonitor interviewed the company's executive team.

Why Vicorp?

- Vicorp is the largest independent Voice Service Creation vendor;
- Specialist in Voice-XML tools, platforms and voice applications for over 15 years;
- Vicorp has first class reference clients;
- Vicorp gives service independence, allowing clients the freedom to choose their mix of ASR / TTS and media server infrastructure;
- Vicorp enables huge cost savings by enabling deployments in a mixed media infrastructure, so allowing clients to better sweat their assets;
- The Company is regarded as being best-in-class;
- Vicorp's platform product, xMP, allows lower cost support over whole service life cycle;
- Time to market can be greatly improved. Vicorp's RapidApp methodology cuts the voice application development cycle time significantly by leveraging the use of xMP as a build environment;
- Comprehensive application reporting capabilities;
- Allows 100% application portability across media server platforms and application servers.

What is the key ingredient above all else that will make Vicorp successful?

When clients see how quickly Vicorp creates advanced voice services they are usually amazed to see how the whole engineering paradigm has changed. Vicorp usually shows this capability to its clients by building sample applications 'on the spot' so that the power of the tools can speak for itself.

The vision of xMP is that the future creation of voice services will become something that almost anyone can do for themselves at a desktop.

What does Vicorp offer in terms of product and services?

- Best-in-class Voice-XML tools - gradually covering the entire voice application life cycle;
- Run-time independence;
- Ease of clustering and creating scalable highly fault-tolerant voice architectures;
- Ease of use (usability);
- Multi-tenancy (especially for Managed Service Providers) across clustered media or application servers;
- Speed and ease of maintenance;
- Turnkey application development;
- Independent advice on any aspect of service services, including architecture and migration options;
- A complete professional services offering from design to deployment, the approach allows inter-working at every level with other partners in the solution offering.

How does your offering stack up against others in the voice channel?

xMP is best-in-class for dynamic feature rich service creation. Virtually every other tool on the market is either:

- Proprietary (e.g. only of use in conjunction with a proprietary media server or similar); or
- Non open standards-based (e.g. providing non standard Voice-XML outputs that are not transportable across platforms); or
- Low function open source frameworks or limited use open source tools that require considerable custom engineering and no support.

What advantages can a business expect to gain in the short- and long-term as a Vicorp client?

- Break away from proprietary IVR stacks - saving large amounts of OpEx and CapEx;
- Service independence - no wasted investment;
- Direct control in managing critical voice interface to customers;
- Multiple platform choice - ability to 'overflow' peak traffic to your network providers Voice-XML platform;
- Rapid service building for lower cost;
- Better customer service from voice self-service and personalization;

- Very quick ROI.

IMPLEMENTING A SUCCESSFUL TOOLING STRATEGY IN A SHIFTING MARKET

Applications in both web and voice environments can be developed using a text editor; however, tools that provide compiling, debugging and testing capabilities are critical to expediting development and supporting application life cycles. The fundamental value of tools across the web and voice worlds is essentially reduced development time which translates to cost savings. Most voice tools today provide this on some level.

To make a name for themselves in the competitive voice market, some 3rd party tools vendors, like Vicorp, have had to introduce unique features and functions that differentiate themselves from the tools provided by platform and application vendors. These are highlighted in Figure 3, but the key advantage that 3rd party tools vendors, like Vicorp, provide is flexibility. Businesses are not limited to the specific platform IDE because 3rd party tools typically support multiple Voice-XML platforms. This becomes especially important as application portability becomes crucial for businesses over the next several years with increased SOA implementations in the enterprise.

Growing importance of application portability in deployment models

Hosted and premise-based managed services are on the rise for speech solutions in the voice channel. They are economically more palatable than premise-based traditional speech solutions as CapEx is significantly lower. In addition, due to limited expertise available in the market, businesses are investing in these models to tap the expertise and best practices of the hosted and managed services vendors in application design, tuning and everyday operations.

As more businesses and hosted / premise-based managed services providers deploy a SOA and web services framework, commonalities in infrastructure and integration points will become more prevalent and application portability will become more widespread. At this time, application assets can and will be transferred between hosted / premise-based managed services provider and clients with relative ease if created through a 3rd party development tool like Vicorp. This will provide a new level of investment protection, deployment flexibility and vendor agnosticism, and in doing so, break down barriers to investments in speech technology and further divorce the platform from the application layer. Application portability is further enhanced if both hosted / premise-based managed services provider and client are using the same development tool. As a result, businesses are able to closely examine business requirements to engage in deployment models that are a better fit to their organizational needs and long-term IT strategy.

Utilizing tools as the abstraction layer

By utilizing a 3rd party tool as the abstraction layer, businesses are no longer confined to vendor-specific tools that come bundled with platforms or applications but have the flexibility to create and deploy applications across multiple platforms. IVR technology is rapidly commoditizing with the emergence of Voice-XML. Platforms can be swapped out or upgraded while the application assets can be maintained and protected. As a result, economies of scale can be improved significantly through standardizing on a 3rd party development tool for application development and management. Application assets are protected in the event a business changes Voice-XML platforms or engages in a different deployment models such as hosted or premise-based managed services.

Towards this end, businesses should consider standardizing on a 3rd party tool to complement their existing investments and strategies in the voice channel. Tools provided by platform and application vendors typically do not provide flexibility

beyond the vendors' specific products while 3rd party tools like Vicorp are flexible across multiple platforms. As businesses begin taking ownership of proven and simpler voice applications they will need to develop the requisite skill set to effectively tune and improve the performance of the application. The ability to own application assets provide organizations with greater flexibility and vendor agnosticism which translates to lower TCO over time. Without an effective tooling strategy this will be extremely difficult.

Datamonitor's strategic recommendations

As voice tools further evolve and more features are introduced and refined, tools will become more of a strategic asset. The need for specialized resources will diminish as will the number of developers required for application development. As a result, significant cost and strategic advantages will stem from effective use of tools. Datamonitor recommends that businesses consider the following when forming a tools strategy for the voice channel:

- **Consider speech recognition** – 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;
- **Determine how much control of the application you want in the short-term and long-term** – Businesses have unique needs. Some will benefit from delegating application management to a vendor while others will find it more economical and practical to bring application management in-house. Determining if and when to take off the training wheels for speech application management is something that should be considered and mapped out;
- **Map out your customer service strategy across channels** – Many businesses moving to a unified multi-channel CRM strategy are looking to improve consistency in customer service levels across each channel. Given that over 80% of customer interaction in contact centers occurs over the phone, IVR interaction should be treated as a part of an enterprise workflow. Therefore, businesses should work with a vendor that provides the capability to integrate IVR call data and analytics with the corporate data warehouse. Data from other channels like web, e-mail, fax and mail transactions should also be fed into the same repository. This will enable businesses to analyze CRM data across multiple channels to form a cohesive corporate CRM strategy;
- **Invest in a tool that provides application portability** – Business needs are constantly changing in the face of regulatory, competitive, industry and customer-driven events. As a result, organizations need the ability to quickly introduce new products and services and terminate existing products and services according to shifting patterns in the market. In the context of IVR, an organization's ability to engage in and shift between different deployment models is increasingly becoming more attractive and can be easily facilitated with application portability. Therefore businesses should invest in a tool that provides application portability unless the businesses will be staying with the same platform or application vendor over the long-term;

- **Choose a strategic partner and not just a vendor** – Choosing a tool ultimately means choosing a strategic partner rather than a vendor. A strategic partner's product roadmap should be in alignment with the businesses' vision. Who wants to choose a vendor that will not invest further resources in the advancement of its tool?;
- **Deploy a well-designed speech application** – 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 self-service solutions on an ongoing basis. This requires significant post deployment monitoring and tuning but will produce the best user experience for customers and help drive the automation and customer satisfaction rates higher. The focus should be on the customer, and businesses should choose a tool that can provide detailed levels of monitoring, analytics and intelligence.

CONCLUSIONS

Whether in the web or voice environment, companies fear getting stuck with managing monolithic, expensive solutions that other elements in the network must conform with, be coded to and be designed around. Standards in the web have freed businesses from these monolithic software prisons and Voice-XML is having the same effect in the voice world. As voice continues building on innovations pioneered by the web there will be more parallels between the two. Tools, like in the web environment, will play a crucial role in the ongoing evolution of voice applications. While tools won't get rid of the need for VUI designers and grammar specialists (for complex applications) they do provide businesses with the means to create, analyze, optimize and manage voice applications in a cost effective and strategic manner.

APPENDIX

Definitions

ASR (automatic speech recognition)

Software engine that listens to and recognizes spoken words. In most cases it processes the incoming audio to isolate words, splits these words into segments (usually phonemes or diphones), and then statistically compares these segments with a linguistic database. Depending on the word spoken, a value is returned, normally with a degree of confidence.

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.

Platform

An IVR (interactive voice response) platform. 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. The platform segment refers to traditional IVR, standards-based platforms and browsers.

Voice-XML

Voice-XML is the World Wide Consortium's (W3C) standard markup language based on XML used for creating voice user interfaces that use advanced speech recognition (ASR) and text-to-speech (TTS) technologies. Since its commercial release in 2000, Voice-XML has emerged as the dominant open standard in IVR technology and is today's most widely deployed open standard for IVR implementations, with a growing sphere of deployments and developers surrounding and supporting it. Although it was initially created for speech recognition solutions, platforms based on this open standard have been widely deployed for DTMF solutions as well.

Author

Daniel Hong

Lead Analyst, Customer Interaction Technologies

dhong@datamonitor.com

Disclaimer

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.