



T h e f u t u r e i s H E A R

## White Paper

SpeechBench  
*vox populi.*

Speech Desk, Inc.  
304 Weymouth Street  
Dix Hills  
NY 11746  
USA  
T +631-243-3410  
W [www.speechdesk.com](http://www.speechdesk.com)  
E [hasan@speechdesk.com](mailto:hasan@speechdesk.com)

May 2004.

## Contents

1	Information – The Lifeblood Of Business .....	3
<b>1.1</b>	<b>Voice Majority</b> .....	3
2	A Case For Voice.....	4
<b>2.1</b>	<b>Potential Market</b> .....	4
<b>2.2</b>	<b>Available Technologies</b> .....	5
3	SpeechBench .....	6
<b>3.1</b>	<b>Genesis</b> .....	6
<b>3.2</b>	<b>Inside SpeechBench</b> .....	6
<b>3.3</b>	<b>Architecture</b> .....	7
<b>3.4</b>	<b>Features &amp; Benefits</b> .....	9
<b>3.5</b>	<b>The Road Ahead</b> .....	11
<b>3.6</b>	<b>Target Market</b> .....	12
<b>3.7</b>	<b>SpeechBench In Action</b> .....	12

## 1 Information – The Lifeblood Of Business

Amidst the battle for the competitive edge, businesses today are striving to excel on many fronts. These efforts revolve around the ability to interface efficiently with clients, business partners and employees – as if in a race to provide stakeholders and consumers the desired information in time. Since time is money for both consumer and provider, business transactions need to be done faster and more accurately to insure satisfaction and profitability. It is the information age, and information drives business. As businesses mature, so does the quantity of information that is stored and processed. This very crutch of business today can become its burden. Information overload. How to sift through it rapidly to reach that needle in the haystack? That is the challenge before businesses around the globe.

In a world where you are not just what you know, but how fast you know it, speedy access to information is a vital aspect of communication, especially business communication. Today there are multiple channels between business stakeholders – Instant messaging, web, email, phone, and face-to-face meetings – all of which may be simultaneously deployed even in the smallest of establishments. But it is the most optimal use of these resources - to do it better, faster, cheaper, with innovation and even style and novelty - that is the competitive driver.

Even as businesses find their way through this maze, an option is emerging that has the potential to make an audible difference to the way business information is disseminated – voice. An option that is very obvious, that is most natural, and that which has always been freely available.

### 1.1 Voice Majority

Telephones are the most available communication devices. Speaking and listening comes most naturally to people, certainly more than hitting keys on the keyboard or pushing buttons on a phone instrument. In times when the Web and email has emerged as the most common or preferred means of business *communication*, it is estimated that over 90% of all business *transactions* still happen over the phone. Voice, as compared to other means, is obviously still the preferred medium. An interesting statistic. But these voice interactions come at a high cost, since voice implies the involvement of skilled people, which come at a cost.

Organizations are increasingly realizing that to meet their business communication needs cost-effectively, they must automate the channels. Since voice is the largest piece of the communication channels pie, it forms the bulk of their communication costs. Automation of voice communication suggests itself.

## **2 A Case For Voice**

Offshore outsourcing is a subject currently being discussed and debated in corporate and governments circles alike, both for its strategic benefits, and its ethics, or otherwise. Businesses are unlikely to relent in the face of increasing pressure from government, employee unions and society, simply given the undeniable benefits. But is also clear to them that as the practice gains wider acceptance, as the people costs in these low-cost geographies increase, its relative competitive edge would reduce. New models, paradigms would need to be invented. Those establishments first to deploy them effectively end up ahead of the competition.

Contact centers, which constitute a large percentage of work being contracted offshore, are no exception. Voice automation could be the next paradigm for them, leading to lower costs and higher transaction completion rates. Voice automation makes for a compelling business argument. According to Giga, the use of voice increases customer use of interactive voice response systems (IVR) from 20% to 60% over touch-tone. As per Nuance, a typical contact center can cut the cost per call from approximately \$15 down to \$0.20 with voice automation, while achieving the desired customer satisfaction. Other benefits include consistency, scalability, the ability to deliver efficient solutions for repetitive processes, and higher system utilization.

In spite of all its advantages, voice has its own limitations – repetitiveness, artificialness in cases of synthesized speech, and hence a lack of personal touch. These need to be addressed through innovative techniques & technologies.

### **2.1 Potential Market**

The Kelsey Group estimates that by 2005, around 45 million wireless phone users in North America will access voice-enabled portals and websites on a daily basis. This is expected to create a \$12 billion voice-enabled market. Allied Business Intelligence estimates a market size of 250,000 voice websites and portals by 2005. These forecasts do not; however, seem to be materializing due to reasons such as high cost & unproven technology, evolving standards and dearth of voice programming skills. The factors are linked – voice is an emerging technology, the adoption is not widespread. Tools are expensive, skills are costlier still. Increased adoption will drive down these costs. But affordable costs drive adoption. However, cost is not the only inhibitor – there are perceived gaps in the available technology, and issues such as quality of synthesized voice are very real.

## 2.2 Available Technologies

IBM & Microsoft are leading the way in the race for voice supremacy. These players have set the pace by developing standards - VoiceXML (VXML) and Speech Application Language Tags (SALT) respectively, for the development of voice technologies, on which several products have been based. VXML requires telephony-heavy programming skills, as against SALT, which can use which can use the Visual Basic platform to assist easier creation of speech applications. Some offerings in the market include

[InVision Studio](#) - from a long-established player, Intervoice Brite - is a mature and effective voice development environment for the creation of VXML applications. Highly effective for developers creating high volume voice enabled applications. The drawback - it is a standalone development environment.

Positioned as a Rapid Development Tool, [VBSALT](#) is Microsoft.NET centric voice development tool. A reasonably good tool that will find greater adoption as SALT gains traction.

[BeVocal](#) provides VXML development libraries and tools free of cost. The developer hosts applications on BeVocal servers and pays for the service. The tools are good but not novice-friendly.

Microsoft has recently launched its [SpeechServer](#) products. Microsoft's vision is to create a flood of speech-capable programmers via this suite. Also, the development environment would necessarily have to be upgraded to .NET; specialized tools such as VB.NET would have to be bought, apart from the investments in SpeechServer. Making the Total Cost of Ownership (TCO) that much higher.

[SpeechBench](#), SpeechDesk's voice technology, aims to fill in certain crucial gaps not addressed by other players. We believe that this approach would truly take voice to the masses. Adoption of these (above) technologies demands extensive investments - in the acquisition and maintenance of development tools, server technologies, and the ongoing high costs of specialized programmers. With SpeechDesk technologies, existing ecosystems suffice and most any programmer can create voice applications. A closer look in the following section.

## 3 SpeechBench

### 3.1 Genesis

For the purpose of enabling voice in generic websites, SpeechDesk needed to create tools that would permit handling of volumes, and also fit in seamlessly with existing software development platforms familiar to developers. To achieve this, we created an Integrated Development Environment (IDE) plug-in, used for designing websites and applications – SpeechBench. SpeechBench has been designed as a developer-friendly voice-enabling tool where the actual process of tagging is automated and handled behind the scenes – the very philosophy that made “visual” development tools (VB, VC, Forte for Java, etc.) widely acceptable.

What IBM and Microsoft offer with their VXML & SALT programming standards are tags that are interpreted by speech engines. The voice application designer must still create an effective Voice User Interface. The SALT/VXML developer will also need to understand telephony. Over years of R&D to create and stabilize, SpeechDesk has developed [SpeechFirst](#), possibly the most affordable, intuitive and effective voice user interface. The combination of SpeechBench & SpeechFirst lets developers create a voice enabled website or application at design time – without having to understand the intricacies of voice technologies. Just one click, literally, and the selected portion will be enabled for voice or telephony, or both. [SpeechBench alleviates the pain involved in creating voice-enabled websites, by virtually eliminating the learning curve and the drudgery of developing voice and telephony applications.](#)

### 3.2 Inside SpeechBench

Simplicity, speed, low cost of ownership, of development, of people – these are the various factors that make up SpeechBench’s philosophy.

Feature for feature, each voice-development tool offers varying degree of advantage over the other – most related to the capability and knowledge of the developer. SpeechBench has been planned as a novice-friendly tool. It is the only tool that is designed to plug into a developer-friendly IDE such as MS FrontPage, Macromedia Dreamweaver, or Eclipse. SpeechBench is probably the only tool that [enables building of a multi-modal and telephony access application in one process.](#) SpeechBench as a standalone is the only product that will support the existing IT ecosystem without upgrades. We expect it to gain wide adoption.

Neither VXML nor SALT seem to address the sea of existing HTML content; content that would have to be converted in either technology for voice access. [SpeechBench helps enable existing HTML, at an affordable Total Cost of Ownership \(TCO\).](#)

SpeechBench [uses openly published standards](#) such as VXML. It uses Microsoft’s royalty-free SAPI 5.1 speech engine, to minimize licensing costs. It is simple, requires no knowledge of voice or telephony programming, and empowers Web developers to create robust and scalable voice-enabled applications rapidly, for a fractional TCO.

SpeechDesk aims to provide a superior suite of tools for wide adoption by developers and end users. Our technologies support the [design-once-use-anywhere approach](#). With a typical voice and 4-port telephony application starting at less than \$ 10,000 – including, hardware and integration costs – we are positioned to garner a large share of the target market. Using SpeechBench, we can guarantee that a ‘speech-novice’ developer can create a deployable voice application within 48 hours of starting.

### 3.3 Architecture

The following figure illustrates the components of SpeechDesk’s speech technology architecture:

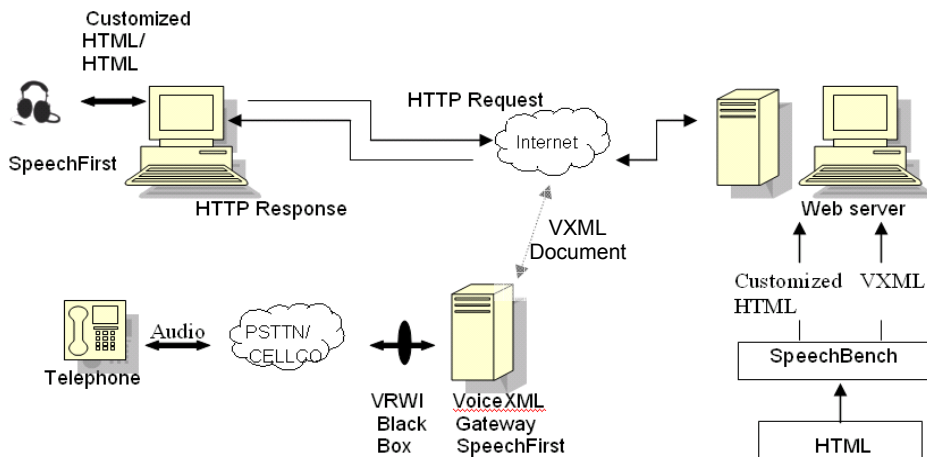


Figure 1

#### VRWI BlackBox

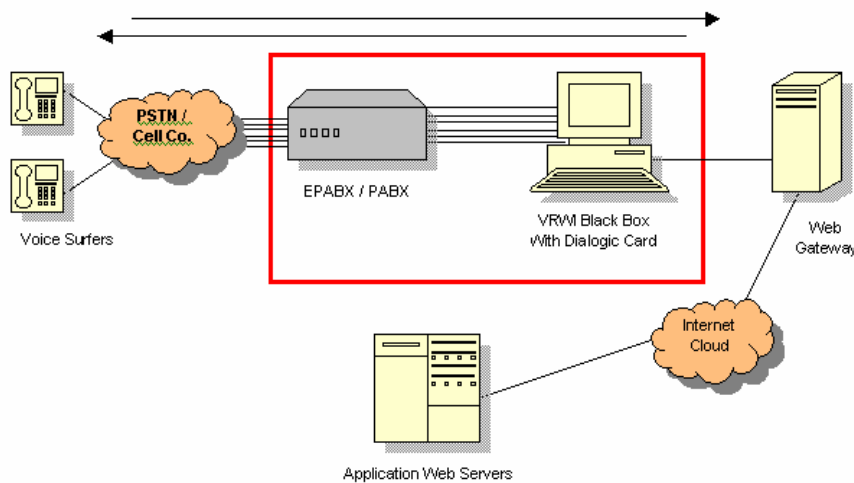


Figure 2

There are five key components to speech-enable a Web application using SpeechDesk's speech technology:

1. **SpeechBench.** SpeechBench enables speech-enabling of normal HTML pages by adding SpeechFirst's custom speech attributes to the HTML page and generating VXML output for the same content for telephony services.
2. **Web Server.** The Web server will serve speech-enabled Web pages and VXML files to corresponding devices such as Web browser and telephony gateway with SpeechFirst.
3. **Telephony Gateway.** The telephony server connects to the phone network. The server incorporates a voice browser interpreting the VXML and then serves the output to the user on the phone.
4. **SpeechFirst.** SpeechFirst is a desktop-based application, which will help the user to interact with websites using voice. It interprets the voice-enabled HTML page and decides the action to be taken for a particular command.
5. **VRWI - VRWI (Voice Response Web Interface)** is a "Black Box" – a software component that will be running on a server interfacing with a Voice Telephony card like Dialogic. This "Black Box" is resilient, fail-proof and capable of interpreting Voice Commands from the Surfer and converts Text responses received from the answering Web Server into speech to be relayed back to the caller.

### 3.4 Features & Benefits

A quick run through SpeechBench's features and resulting benefits:

- ✓ **Lower cost of development and deployment** through the use of open standards like VXML & SALT. Wherever used, SpeechDesk's own standards would be published, making the architecture open.
- ✓ **Ease of development** via the Web programming model. Speech interfaces can be developed using familiar markup-style development language.
- ✓ **Lower investment in training and maintenance** since existing Web developers can be used for speech application development.
- ✓ **Lower TCO** – through the use of royalty-free components like SAPI. Open, standards-based voice automation enables businesses to reduce their costs by leveraging investments in existing Internet and IT infrastructure and operate on inexpensive, off-the-shelf hardware and computing resources.
- ✓ **Server-easy** – SpeechBench's architecture is **designed to minimize the traffic between the Web client and the server**. Only the voice tags reside on the website / server. When the page is opened, these, along with the other HTML material, are sent to the browser. The SpeechFirst client recognizes the tag and "speaks" accordingly. Since all the processing takes place on the client machine, there is practically no resource burden on the server. If at all, perhaps 1% to 2% involved in sending a few extra tags. In the case of telephony, the processing takes place on the telephony server.
- ✓ **Barge-in** – **allows the user to interrupt** while the system is still speaking.
- ✓ **Error Control** – Reducing chances of errors is achieved by restricting number of active words (commands) at any given point in time, to address background noise, mispronunciation, etc.

However, a possible limitation of SpeechBench, due to factors beyond its scope. If the HTML application being voice-enabled is not optimally planned, it is quite possible that the best programming effort would not yield a very voice-friendly application.

As a case in point, a broad comparison between SpeechBench and Microsoft .Net Speech SDK for multi-modal and telephony:

<b>Factor</b>	<b>SpeechBench</b>	<b>Microsoft .Net Speech SDK</b>
<b>Ease Of development</b>	Any person having basic HTML knowledge will be able to develop speech application.	Developers will require knowledge of SALT, Scripting, ASP.net and SAPI.
<b>Speech Engine</b>	Microsoft SAPI 5.1. Can support other engines.	Microsoft SAPI 5.1.
<b>Supported clients</b>	PC, Telephony, PDA support in future.	PC, Telephony.
<b>Supporting Web Browsers</b>	Internet Explorer currently (in future, all browsers).	SALT compliant browser (probably a future version of Internet Explorer).
<b>Learning curve</b>	Practically none –plug-in for leading development environments.	Steep – requires understanding of server and other technologies.
<b>Ability to enable existing Web pages</b>	Any web page can be easily voice enabled.	The Server side code will have to be rewritten.
<b>Developer training</b>	None other than getting accustomed to drag-n-drop. Use existing development tools.	Developer will have to be trained in ASP.Net And Microsoft Speech SDK. Need to license Speech .NET and other .NET software.
<b>Client software</b>	Client side download of SpeechFirst required.	No download for client side required.
<b>Usage</b>	Can be used by any developer, lowers the barrier, “democratizing” speech.	Will need knowledge and training in .NET Framework and investment in the Visual Studio, tool for specialized developers.

### 3.5 The Road Ahead

The power of SpeechBench lies in its simple approach and ease of use, and SpeechDesk continues to work at enhancements that will help developers build more complex applications, faster, better and with greater functionality than most. In line with this thought, a range of enhancements have been planned for SpeechBench. As under:

SpeechDesk will introduce the Emotional Quotient in Synthesized Speech (EQSS) feature in future versions of SpeechBench. EQSS adds a [measure of expressiveness to machine-like sound](#), providing a quick and affordable means of adding variation and richness to monotonous voice. (A note - AT&T has spent multi-million dollars to create Natural Voice. With this, AT&T can clone almost any voice, but is expensive to deploy. In line with SpeechBench's philosophy, EQSS will be a cost-effective way of achieving quality results.)

VXML 2.0 is now the accepted W3C standard, but SALT is fast catching up. Its version 1.0 is under consideration at the W3C, and developer-adoption is growing. We would look to leverage the awareness created by these giants by ensuring SpeechBench's compatibility to these two dominant standards. SpeechBench could be purposed for VXML- or SALT-compliant output. Another feather in its TCO cap - no need to invest in new technologies.

To broad base its appeal, an [OpenSource version](#) of SpeechBench is also being considered. The open source community has experienced a boon in Linux gaining traction as the server platform and the Linux OS for the desktop is beginning to see higher adoption. SpeechDesk intends to address this section of the user community in the near future.

SpeechBench is currently available for FrontPage and Dreamweaver. [A version for Eclipse](#) is planned. There are plans to make SpeechBench more versatile - [support for other speech engines](#) is in the offing.

The next iteration of SpeechBench will allow developers to build an increasing number of ['what-if' scenarios](#). For example, the developer can define that after three failed recognition attempts,

- A - System shifts into IVR mode
- B - Transfer call to a live agent
- C - Hang up

*An example: If a travel agent offers 24x7 information on cruises via an automated voice system, developer can effortlessly program the appropriate 'what-if' scenario. The caller surfs the tours on offer, checks out the pricing, then checks the availability. At this point there is the potential of converting the interested caller to a sale and the call could be transferred (a) to a live agent at the office if the call has been received between regular work hours, or (b) to a call center, etc.*

### 3.6 Target Market

Simplicity and low cost of ownership drive the philosophy behind SpeechBench. Lower costs mean that businesses do not need to stake high investments into voice automation, an area where results have not been uniform. However, the premise (of voice) is attractive to businesses and the low investments required by SpeechBench make the decision easier. Simplicity, ease of use, speed of result – ensure that organizations start seeing results quickly, and at low entry costs; compelling factors that drive the adoption of any new, untried technology.

SpeechBench is hence [an attractive proposition for the mid-market](#) – the segment that makes its investments, and justifiably so, based on cost and quick returns. Given that the mid-market constitutes a large slice of the business pie in any market, SpeechBench's aims at popularizing speech and making it ubiquitous in the Small-to-Medium Enterprise (SME) segment.

Adoption in large volumes will have an effect that will build on itself – higher volumes will translate into lower costs, further accelerating adoption. As with any technology, this will drive the market towards maturity, in turn assisting the adoption of higher-end voice technologies, on the back of increased consumer confidence. Leading to the creation of a larger pie, and hence larger slices of the pie, for the players.

It is clear today that pure play speech companies have not succeeded in reaching their original forecasts in market adoptions. These companies have failed to create the basic building blocks that would have ensured increased voice adoption. SpeechDesk aims to fill that gap.

### 3.7 SpeechBench In Action

SpeechBench has been selected as a tool for voice enabling a [20,000-page web encyclopedia \(www.rajashivaji.com\)](#) on an Indian warrior king – Shivaji. This is purportedly the largest encyclopedia on the Web and is due to be launched in April 2005 by the President of India. SpeechBench will be integrated into a content generation tool used by the websites' developers.

SpeechBench has also been used to create a [2,000+ page multi-modal website](#) for Disabled Peoples Association of Singapore ([www.dpa.org.sg](#)).

All trademarks belong to their respective owners.