white paper

MultiModality in 2003: Making it Happen

A CLOSER LOOK AT CURRENT MOBILE DEVICES, NETWORKS AND THE APPLICATIONS THAT DRIVE MULTIMODAL ADOPTION FOR THE CARRIERS

FORWARD SECTION BY:

IGOR JABLOKOV (PRODUCT LINE MANAGER, IBM PERVASIVE COMPUTING DIVISION, CHAIRMAN, VOICEXML FORUM TECHNOLOGY COUNCIL)

WRITTEN BY:

ATUL SURI (DIRECTOR, STRATEGIC MARKETING, V-ENABLE)
SUNIL KUMAR (DIRECTOR, R&D, V-ENABLE)

V-ENABLE

Preface

This white paper is intended for wireless operators interested in launching MultiModal applications on their existing networks today, within the current and future device portfolio.

The purpose of this white paper is to educate the wireless carriers on launching revenue generating MultiModal applications on their existing networks. This paper will show case applications that can be developed using the V-Enable developer environment to deliver a richer user experience across all networks.

The white paper is based on V-Enable's current experience with operators' trials/ deployment processes and our expertise with all major wireless standards, including WAP, SMS, VoiceXML, IOTA, MMS etc…

About the Authors

Atul Suri, Director, Strategic Marketing
At V-Enable, Mr. Suri is responsible for driving strategic marketing initiatives and working with partners to develop complete end-to-end multimodal solutions. Suri holds a Masters Degree in Electrical Engineering from Clemson University and a B.S. from Delhi Institute of Technology, India.

Sunil Kumar, Director, Research & Development
At V-Enable, Mr. Kumar is responsible for driving standards efforts with both W3C and SALT Forum. Additionally, Kumar leads the innovative MultiModal technology group aimed towards strengthening the V-Enable patent portfolio. Kumar holds a Masters Degree in Computer Science from University of New Hampshire and a B.S. from University of Delhi.

V-Enable Inc.
4250 Executive Square, #200
La Jolla, California 92037
USA

Forward

Multimodality is a major focus area for IBM's Pervasive Computing Division, given recent advances in human-computer interaction research and the natural evolution of disparate user interfaces finally coming together in a cohesive fashion. With the industry facing slow growth due to a constrained investment environment, we must embrace technologies that add real value by simplifying the user experience (and that subsequently increase the number of given transactions performed). These interactions require extending existing infrastructure to support this and the recent wireless network revolution is a key factor in improving the state of mobile ebusiness applications. Ultimately, network availability, bandwidth, and reliability determine the nature and capabilities of the ebusiness applications. Historically, voice and data networks have been separated, with multimodal across network types (voice/data) and transports (wired/wireless) a complex and arduous task. Two incongruous factors are resolved by the drive for multimodality. The first being the large explosion of available data that a normal end user needs to navigate through on a daily basis. The second being the increased miniaturization of devices that need to access that large data store. We can clearly see the value proposition inherent in multimodality in solving these constraints beyond more traditional infrastructure improvements.

To succeed with multimodality we need your continued support in creating an ecosystem that is industry standards-based, allowing interoperability among the diverse players required for the most cost efficient and at the same time full featured end-to-end platform. Many of our respective customers and partners have a vested interest in VoiceXML through their existing technology and training investments, a fact that cannot be ignored in a cost sensitive environment. XHTML+Voice (X+V for short) builds on this foundation by allowing backwards compatibility with the large numbers of deployed VoiceXML-based applications and with other existing standards, such as XHTML and XML Events, while at the same time setting the stage for the growth opportunities multimodality will provide our customers. You can deploy a single application that then dynamically morphs based on a customer's preferences at the time of a transaction's execution. This allows end users to choose their method of interaction with these applications based on social factors, network availability, and device capabilities. To achieve these capabilities, X+V has been architected from the start to be highly adaptive, manageable, and scalable.

IBM Business Partners, such as V-Enable, are at the forefront of helping the market understand this area and how it can allow you and your partners to deploy revenue generating applications across multiple devices and networks. We look forward to your help in creating a standards-based whole product for end-to-end multimodality, which would add genuine value for our respective customers.

Igor Jablokov
Product Line Manager, IBM Pervasive Computing Division
Chairman, VoiceXML Forum Technology Council

EXECUTIVE OVERVIEW

Wireless users want to send text messages, browse information and services, carry out wireless commerce transactions, and access business applications, all from their mobile device and all with the click of a single button.

Further, wireless technologies are decreasing in price[1] and becoming more standards-based, making them easier to develop and deploy. Consumers and business users are demanding additional productivity and flexibility in accessing their mobile applications.

MultiModality is the next step in the evolution of data and speech services for wireless carriers. The promise of this new enabling technology is to accelerate the adoption of current data and speech services in the network, and to also leverage the operators' investment in the current infrastructure by increasing the utilization.

This new technology is available for integration in 2003 with revenue generating applications that support the business case for deployment. The end-user devices to support such services are broadly deployed today. The speed and capacity offered by existing networks are quite sufficient for offering exciting MultiModal services. And, the standards that will assure protection of carriers' investments in rolling out such services are now in place.

The Revolution has Begun!

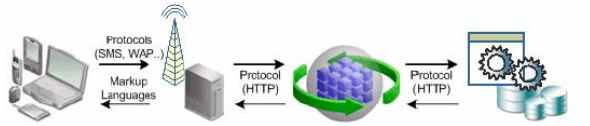
The wireless Internet is about to start its second phase, where business needs drive enabling mobile technology. Though the current wireless Internet experience pales in comparison to the desktop, due to the "walled garden" approach to content adopted by most carriers, the technology is quickly allowing for the convergence of user experience within these environments. With significant advancements in speech technology, the limitations of small keypads (e.g. the difficulty of entering search strings or Web addresses, especially for ideographic languages with several thousands of characters) can be made a challenge of the past.

Effective *mobile* computing is about delivering the information users want; on the device they need it on and in the *way (mode)* they want to access that information. Different situations and job functions require different *modes* to access the same information. Also, depending on the type of mobile application, the mode changes. With more available modes of access, the value of the mobile application increases as it becomes accessible under different circumstances. A mobile worker, for example, needs access his email in real-time via wireless; visually when in a noisy environment, and hands-free when driving in an automobile. With multiple modes of interaction, users have the choice of using their voice, or an input device such as a key pad, keyboard, stylus or other input device. For output, users would like to be able to listen to spoken prompts and audio, and to view information on graphical displays.

Though the ever-increasing availability of new device types, form factors, wireless content and network types create numerous challenges in delivering a consistent multimodal interface, such complex integrated transactions can be offered today with the introduction of a wireless software platform that can seamlessly integrate network elements for wireless messaging, content browsing, voice access, and mobile commerce.

The Technology to "Make It Happen"

The wireless Internet is a medium where content is accessed by a variety of devices: laptop, phone, PDA, etc., and modes: text, Web, voice, etc. There are many infrastructure components that interoperate in the wireless value chain as described below.



Wireless infrastructure Components

[1] Voice minutes are already a commodity

Mobile Devices and Browsers[2]

- Wireless devices and browsers are used to access the mobile Internet. Each wireless device generally runs a browser to display received information. Based on the wireless services available today, there can be many types of browsers on a phone: Data (WAP, xHMTL, other data format), VoiceXML (server side or distributed), SMS, MMS and Video browser.
- Wireless markup language is the language that a browser speaks – the markup language specifies how information should be presented on the device. Common Markup Languages include VoiceXML, WML, and xHTML.

Mobile Devices and Browsers – "Make It Happen" Takeaway!

The only requirement for a carrier's offering a MultiModal experience is a server side VoiceXML browser and an SMS browser that is available on practically all devices in the installed base. With additional browsers, the MultiModal experience will be further enhanced and appeal to a wider target of content applications.

Wireless Network

- Networks are the underlying infrastructure that is used by the wireless carriers. An important characteristic of networks is the bandwidth and connection type. Today, the networks have evolved from 2G, where only a single channel (voice or data) is possible, to 2.5G networks, where both data and voice channels can co-exist[3]. Further, with the evolution to 3G networks, all handsets will be capable of simultaneous data and voice channels.
- Wireless protocols[4] are used to deliver content to devices. Examples of protocols are WAP, VoiceXML, SMS, MMS (Multimedia Messaging Service) and Video (MPEG4).
- Wireless Gateways translate the protocol request to the standard HTTP protocol. Gateways speak a variety of protocols such as WAP, SMS, MMS, Video, VoiceXML and others.

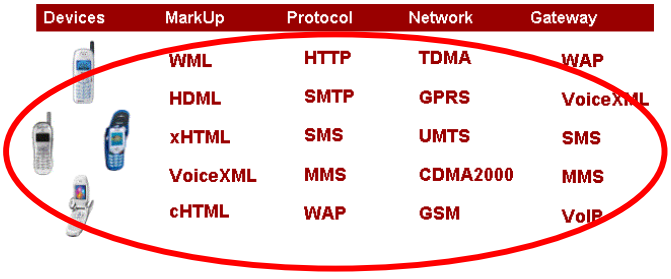
Wireless Networks – "Make It Happen" Takeaway!

The only requirement for a carrier's offering a MultiModal experience is a network architecture that supports SMS and WAP protocols and supports a single channel at any time. VoiceXML can be integrated into the solution either within the network or as a hosted solution. With the inclusion of packetized voice in the network (VoIP), a single data channel can deliver a simultaneous MultiModal experience seamlessly.

Application Servers, Applications and Content

- Application servers have come into play to increase the efficiency of application development, deployment and management. A wireless application server connects the wireless content source over the wireless network to the wireless gateway.
- Applications/Content have a wide variety of forms including database information, personalization content, alerts, e-mail, location services etc. The large amount of content sources increases the complexity of having a manageable way to deliver each application to every type of device in the most optimized fashion. There are over 50,000 existing VoiceXML applications developed today, and over a million WAP enabled sites that provide mobile content.

Summing it all up, the MultiModal challenge is in delivering a consistent content experience across all devices, networks, and protocols.



MultiModal Sphere of Influence

[2] For the purposes of this paper, we will limit our discussion to mass market mobile devices

[3] Depending on the type of mobile handset

[4] Wireless protocols are more efficient over the wireless networks than the standard HTTP protocol.

MULTIMODAL CATEGORIES

There are currently four categories of MultiModality based on the experience to the end user on their wireless device.

- **SMS-Based:** Use of SMS as the visual interface, and speech as an input or output. Works on all wireless devices, limited to a plain textual output interface.
- **Sequential:** Use of data and speech for both input/output, depending on the application and the user environment and choice. Works on all browser-ONLY wireless devices and provides a richer textual experience. The experience introduces latency while switching between the modalities.
- **Synchronous:** Similar to sequential, except it provides an enhanced experience on intelligent clients such as JAVA/ BREW/SYMBIAN devices. The enhanced experience is due to the reduction in latency, half-duplex voice channel and the control of the application over the switching state.
- **Simultaneous:** The promise of a pure MultiModal interaction is to get both speech input/output and data input/output seamlessly in a single session using multiple channels. This experience is dependent on networks and handsets that support multiple channels simultaneously.

MultiModal Categories

	SMS-Based	Sequential	Synchronous	Simultaneous
Networks	2G/2.5G/3G	2.5G/3G	2.5G/3G	2.5G/3G
Solution	Server-Side	Server-Side	Distributed	Distributed
Handset Type	2G/2.5G/3G	Data Browser (WAP, xHTML)	Intelligent Client (BREW/JAVA)	Smart Phones (SIP/SYMBIAN)
# of Handsets	over 1 Billion	400-600MM	3-5MM	Less than 1MM

WHERE ARE THE STANDARDS?

The wireless revolution brought a variety of wireless devices and as many different web programming models. HTML, WML, cHTML, and VoiceXML, are only a small fraction of the markup languages used to create wireless and voice applications. The W3C standards body is currently reviewing an X+V (xHTML+VoiceXML) specification submitted by IBM, Motorola and Opera. X+V neatly separates all elements of typical data application, making proper multimodal adaptation of content possible. These elements include presentation, style, instance data, user interface events and business logic. HTML and XML savvy Web developers will find X+V easy to learn since xHTML is a clean XML version of HTML. Application developers will find that their investments in current applications are protected with these new standards along with room for growth. Browser based, messaging or voice applications can be implemented using the device and channel independent X+V programming model.

Another MultiModal standard that is gathering momentum is SALT. Speech Application Language Tags is a speech interface markup language. It consists of a small set of XML elements, with associated attributes and DOM object properties, events and methods, which apply a speech interface to web pages. SALT can be used with HTML, xHTML and other standards to write speech interfaces for both voice-only (e.g. telephony) and multimodal applications. SALT does not extend any individual markup language directly, rather it applies the speech interface as a separate layer which is extensible across different markup languages. The dialog framework, which drives the SALT speech interface, can be as loosely or as tightly coupled as necessary to the underlying data structure (e.g. an HTML form), so that speech and dialog components can be reused across pages and across applications.

X+V: A CLOSER LOOK[5]

xHTML+Voice brings spoken interaction to standard WWW content by integrating a set of mature technologies such as xHTML and XML Events with XML vocabularies developed as part of the W3C Speech Interface Framework. The profile includes voice modules that support speech synthesis, speech dialogs, command and control, speech grammars, and the ability to attach Voice handlers for responding to specific events, thereby re-using the event model familiar to web developers. X+V first re-formulates VoiceXML 2.0 as a collection of modules. These modules, along with Speech Synthesis Markup Language and Speech Recognition Grammar Format are then integrated with xHTML using xHTML modularization to create the X+V profile.

Delivering Applications via Voice

Accessing applications via voice entails an end user calling a server on a traditional phone line and interacting with an audio interface. The end user has two possible input methods: either through speech or the number pad. Prior to speaker-independent speech recognition, users had to train the software to recognize their utterances. Today, speaker-independent speech recognition and VoiceXML are sufficiently mature for large-scale deployment enabling carrier grade voice applications to be developed and deployed for telephone access. There are three components in such a deployment, the voice gateway, the application server, and the content source.

The main advantage of a MultiModal architecture (such as X+V) lies in its ability to deliver the same content to multiple modes at no additional costs. Web content access is no longer restricted to a particular mode (i.e. voice) or device (i.e. WAP). In fact different modes can be used within the same application, utilizing the particular advantages of each mode to the benefit of better usability. These multi-modal applications use two or more modes in the same application or user context. For example, order information could be pushed to a messaging device using SMS (mode #1), and the user could respond by voice (mode #2), or browser (mode #3). The need for applications developed specifically for wireless becomes diminished because every standard-compliant Web application becomes automatically 'multimodal' enabled. The results are reduced development costs for wireless and voice applications and higher return on investment (ROI).

[5] X+V definition at www.w3c.org

THREE STEPS TO BEING MULTIMODAL

Well-designed, effectively implemented mobile MultiModal solutions can be delivered today that provide real consumer value by offering a richer user interface to wireless applications. MultiModal applications can be categorized into three functional areas that lend themselves to effective customer segmentation: Information, Communications and Entertainment (ICE). Adoption considerations for these applications mandate a minimal behavioral change and thus a very short learning curve (if any), at the same time enhancing the user experience significantly to

create demand and uptake. V-Enable provides application developers with a consolidated environment for developing multimodal applications across all access channels – voice, SMS, WAP Push, MMS, LBS and Video. The scalability of the MultiModal platform to incorporate new wireless technologies such as LBS, video, others, makes it ubiquitous for developing new applications.

APPLICATION BUSINESS CASE FOR GOING MULTIMODAL

When creating the business case for going multimodal, carriers should begin with a focus

on simple business applications they can streamline to cut costs or drive new revenue. Initially, carriers should focus on existing applications that have a proven ROI and adoption rate, and look at ways to Multimodalize them! The advantage of making the existing top-sellers (such as SMS based applications, Email, Games) more useable will clearly drive new revenue opportunities from existing content/ applications with a quick time-to-market. Additionally, carriers could focus on SMS-based MultiModal applications such as yellow pages access, information services, and others that can be enhanced with voice and other data interfaces (MMS, LBS, etc…). Other horizontal applications such as mobile E-mail and Personal Information Management (PIM) can be multimodal-enabled to make the business mobile professionals more productive. Sales and field force automation, customer relationship management, personnel-intensive tasks such as order entry, expense-reimbursement and process control are other key areas that can be multimodal-enabled.

APPLICATION SCOPE/USAGE

The most important part of deciding on the scope of the application is to understand the detailed situation of the targeted end user, and map out the relevant usage environment. Only when the user context and expectations – as well as the new application flows – have been determined, can the actual design of the application go forward. The ability to embed multiple user interfaces in the application will directly determine its usability and adoption.

APPLICATION DEVELOPMENT

As the number of access channels increase, developers have traditionally been forced to re-write the same application once for each channel or device. This "single-silo" approach to application development creates increased complexity, higher maintenance costs, and longer development cycles. V-Enable's veSTUDIO™ is a developer portal for quickly building**,** testing and deploying multimodal applications. It lets any developer, systems integrator or independent software vendor quickly develop multimodal applications, immediately making them accessible from all devices. The veSTUDIO™ enables developers to focus on their business logic, while we focus on the delivering a consistent multimodal experience across all devices, networks and protocols.

MULTIMODAL APPLICATION SNAPSHOTS