

Voice-enabled IT transformation: The new voice technologies

J. Christensen
B. Hughes

Voice technologies can enrich the ways in which people communicate and enable novel modes of collaboration, such as integrated voice- and text-based communication services, social and business networking facilities, and the evolving Internet technologies (collectively known as the *Web 2.0* technologies). These new technologies can alleviate many problems in interpersonal communications, including those affecting person-to-person, online meeting, contact center, and business-process scenarios. In this paper, we identify some of the new challenges enterprise employees face and discuss the potential of voice technologies to help with these challenges. We also examine the new business environment, the communication services it demands, and the challenges enterprises face in delivering these services.

INTRODUCTION

Voice-based communication and the way it is used are changing. While new voice technologies are gaining acceptance and are being implemented aggressively, users' experience of voice communication technologies has not yet changed in any significant way, nor have the technologies been exploited to provide users with a more productive communication experience. This paper examines the relationship of the new voice technologies with the changes under way in global enterprises, with the people who work in those enterprises, and with the information technology (IT) organizations in those enterprises. It is our goal to demonstrate that the new voice technologies can be of significant assistance to global enterprises and their employees in meeting some of the challenges of the twenty-first century workplace.

The voice communication infrastructure is going through a fundamental change that will have a profound effect on the communications industry and ultimately on how people communicate. Traditionally, the voice network and supporting enterprise voice systems have been highly proprietary systems, unique to each supplier and providing only minimal, rudimentary interoperability. For example, the traditional private branch exchange (PBX) solution involved vendor-proprietary processors, operating systems, application code, station and trunk cards, and end-user station equipment. From the enterprise

©Copyright 2007 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the Journal reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to republish any other portion of the paper must be obtained from the Editor. 0018-8670/07/\$5.00 © 2007 IBM

perspective, data networking services were managed on a completely separate basis, with various carrier services and classes of termination equipment. Any interoperability between the voice and data networks at the user level was the result of computer-telephony integration (CTI) capabilities, which also had a strong dependency on the equipment's vendor. CTI systems were expensive to build, and therefore mostly limited to contact center environments in application.

Whereas legacy voice networks were separate, dedicated networks used solely for voice signals, today's networks carry voice, images, video, and data. The convergence of data types started at the network layer¹ and was mostly transparent to end users and to voice applications and services. A significant turning point resulted when the Internet Engineering Task Force (IETF) introduced its Session Initiation Protocol (SIP), as defined in RFC3261 and subsequent RFCs (Requests for Comment).² SIP is a protocol that establishes session control (i.e., setup, change, and teardown) of any real-time communication (including voice, video, and instant messaging). Because SIP was established by the IETF, it has strong parallels to the Hypertext Transfer Protocol (HTTP). Hence, over time, the voice capabilities of converged networks have become usable by end-user applications and services using industry-standard platforms and techniques (for example, IBM WebSphere* and J2EE**). As a result, low-cost applications can offer users new communication options that include data and voice communication.

At the network level, separate and multiple wireless networks created for voice and low-speed data traffic are evolving, gaining bandwidth, and starting to interoperate with wired networks. At the user endpoints (the communication devices), handheld devices once used solely for voice (such as cell phones) or data applications (such as personal digital assistants, or PDAs) are becoming multipurpose devices, supporting voice, data, and even image and video services. Voice over Internet Protocol (VoIP) regulations are still evolving and will affect the adoption of new voice technologies. For a survey of topics related to VoIP, refer to Reference 3.

Convergence, or at least confluence, of datatypes is also appearing among the applications and services

that use the converged networks. The same networks that support voice communication now also carry entertainment channels and disseminate news, on-line publications, advertisements, and the distributed network-based data applications that have become the fabric of online enterprises and society at large.

For people involved in the communications sector, it is an extremely exciting time. The pace of change is as great as it has ever been. We hope to be able to give the reader a sense of the scope and depth of the changes that are already under way, and to describe the developments that may well fundamentally change the ways that people communicate.

In this paper, we explore these changes and describe how new voice opportunities support business transformation. We discuss some of the major business transformations taking place and the ways they affect people, their work, and their lives. We evaluate how well legacy voice systems measure up to these challenges. We examine the technology changes occurring in voice systems, including voice and application integration architectures, and how these transformations will alter the way people use voice in the business environment. Because the opportunities are extensive and an all-inclusive treatment would be quite lengthy, we have selected some representative areas to discuss.

We begin by placing the new voice technologies into context within the field of communications and collaboration and describe how they relate to other evolving technologies in that field. Next, we discuss how and why the new voice technologies impact global enterprises and their employees. We conclude with a brief look at the challenges that the IT organizations of these companies face as the new voice infrastructure expands.

VOICE TECHNOLOGY: CHANGES AND THEIR VALUE

Voice technology, technology in general, and the entire business world are changing. This paper advances the view that voice technology transformations will play a critical role in supporting business transformation.

Voice technology (or simply, *voice*) is of great significance in our daily business lives. Traditionally, voice has focused on live conversations, characterized by high levels of interaction, involving

two or more people. With new technologies, voice signals are converted to data packets and become recorded information files that can be “stored, searched, manipulated, copied, combined with other data, and distributed to virtually any device that connects to the Internet.”⁴ This adds a new dimension to the power of voice technologies. For example, a traditional voice-mail message is simply a digital recording of a voice audio segment. Currently, it can be listened to and forwarded to others. If it were an information file, new possibilities would emerge: multiple files could be sorted in different ways (e.g., by sender or topic), files could be searched for keywords, and files could be edited by deleting sections or adding a response directly into the message at relevant places. In a manner analogous to the exploding inventory of podcasts and MPEG-1 Audio Layer 3 (MP3) files, many opportunities exist to search, scan, and further process voice information.

Transformations are occurring in such areas as mobility, voice/application integration, PC-based voice, language translation, multimodal techniques, pricing and business models, and regulations. These changes are enabling new opportunities at an accelerating pace. However, this change also generates problems. For example, within an enterprise, people generally have a single e-mail ID, but often have multiple telephone numbers. How do sales executives, for example, make sure customers always know how they can be reached?

There are many methods available today for people to collaborate, such as instant messaging, text messaging, and Web conferences. Each method has strengths and weaknesses, and each plays a role in communication. However, effective communication and collaboration usually involve deeper interactions among people, such as discussing, informing, educating, negotiating, selling, interviewing, or buying. Voice conveys much more information than typical data formats (such as instant messaging, text chats, and Web conferences) because voice also communicates social context. For example, additional information is conveyed by the different ways people can speak the word “yes;” or a vocal inflection can change the context during a negotiation session; and a laugh in a conversation or meeting can alter the entire tone of a collaboration session.

Often, video and voice technologies are closely linked in many ways. However, in this paper, we have chosen to limit our discussion to voice for two reasons. First, the user experience of voice is different than the video experience. Second, video requires video-enabled endpoints and more network capacity than voice, and while this is not generally a problem in the United States, global enterprises are very concerned about the cost of bandwidth in other geographic areas. For these reasons, we acknowledge the video opportunities, but do not specifically address them here.

Voice and technology evolution

Voice is changing along with other technologies of various types; this coevolution presents significant opportunities for business transformation. Friedman has described 10 phenomena that have “flattened” the world (i.e., removed hierarchical structures).⁵ Referring to the first three phenomena (Windows**, Netscape**, and workflow software), he states that the unprecedented level of people-to-people communication enabled by these Web-based, application-to-application tools provides a new global platform for multiple forms of collaboration. The next six flatteners (“open sourcing,” outsourcing, “off-shoring,” “supply-chaining,” “in-sourcing,” and informing) represent new forms of collaboration that were enabled by these tools.

The final flattener includes digitized content; virtual, personal, multipurpose devices; and VoIP. With respect to VoIP, he predicts that it will certainly “amplify and further empower all the other forms of collaboration.”

The 2006 IBM Global CEO Study⁶ acknowledged these challenges and suggested ways to address them. If the enterprise is to survive in a flattened world, this study states, the enterprise must innovate in its business model, intensify its collaborative innovation, particularly beyond company walls, and orchestrate and foster innovation from the CEO level. To do this, the CEOs must “. . . create a more team-based environment, reward individual innovators, and better integrate business and technology.”

We now work in a global environment with people of many cultures speaking many languages, and this can greatly enrich our lives. When we work in global teams, the project workday is no longer nine

Table 1 Changing business models

Old model	New model
Video rental store	Netflix**
Bookstore	Amazon.com**
Music CDs	iTunes**
AAA road maps and guidebooks	Internet-based mash-ups; GPS systems
Fulfillment and repairs at corporate site	Fulfillment and repairs at UPS** depots

to five, and work on our projects often extends to the full 24 hours. These factors make life more challenging, rather than simpler. Business is more complex and can often appear to be disorganized. People are separated by distance and time and multiple markets, and regulatory environments play a role in business decisions. These factors make up the business transformation challenge.

TECHNOLOGY CHANGES AND THEIR IMPACT

The forces described previously are causing significant changes to corporate thinking, principally in the area of business models. According to one commentator, “What was once central to corporations—price, quality, and much of the left-brain, digitized analytical work associated with knowledge—is fast being shipped off . . . Increasingly, the new core competence is creativity . . . imagination, and, above all, innovation.”⁷ Thus, the environment calls for new and innovative business models. There are many examples of this theme appearing in practice today, as illustrated in *Table 1*.

There are two important characteristics of these new models. First, each of them requires a new ecosystem in order to work. For example, Netflix cannot function without the reliability of the postal service. Amazon.com acts as a channel between a huge customer base and a multitude of suppliers (including very small family businesses) to make a wide variety of items easily available. Travel is now easier through the use of Global Positioning System (GPS) devices that not only plot a route but also describe the restaurants, hotels, and interesting sites at each intersection (and will certainly, in the future, establish a communication channel with them to make reservations). Second, all of these business

model innovations result in totally new user experiences, not just a faster or cheaper version of the old experience. Not only are services and products offered in a different way, but the experience of other buyers is collected for all customers to use in making their buying decisions.

From these business model changes, several key challenges are emerging for the enterprise and for the user. An increasingly diverse set of enterprise data must be stored in a coherent way so that consumers can understand the semantics of the stored entities and the relationships among them. Enterprises place a high value on their very structured, application-focused data. But unstructured data, which includes voice, is growing in importance as well. As we have noted, voice can become data files that can be edited, stored, searched, copied, and manipulated. The emergence of voice analytics, language translation, and other technologies will enable new opportunities for leveraging podcasts, MP3 files, voice mail, and a variety of audio or video data files.

The boundaries between the enterprise and its customers and partners will become more important as more and more collaboration occurs outside the enterprise. This can be observed in production systems, but boundaries must be crossed even earlier in the product life cycle. The IBM CEO study found that most CEOs feel that the best ideas come from customers and partners. Thus, collaboration tools that restrict nonemployees from fully participating or require them to load special software will not succeed.

Speed of execution is critical for enterprises to succeed. The United Parcel Service (UPS) model (with, for example, service and inventory facilities at shipping hubs) shortens the transit time and cost needed for a repair cycle. The Amazon.com model allows resellers to start business transactions almost overnight.

Employees’ increasing mobility makes the provisioning of appropriate devices and software more challenging. With mobility and globalization, the number of face-to-face conversations has been reduced, and the development of close working relationships has been greatly limited. Additionally, the business day can become longer for people involved with global teams.

Given these changes, the pressure on enterprises and their employees increases dramatically. People must adapt to rapid change, increased partnership with others inside and outside the enterprise, and work efforts that span time zones. In the next section we examine the challenges people face, with an emphasis on the challenges prompted by the new voice technologies.

Challenges for employees

As companies adapt to the demands of the global economy, employees find themselves with new challenges. While the new voice technologies cannot begin to solve all of the challenges that enterprise employees face, they can help with the interpersonal communication aspects of these problems. Five interrelated challenges are described in the following. In subsequent sections, we examine how people deal with these problems using legacy and new voice communication technologies.

Among the challenges faced by enterprise employees are: (1) The need to complete more work objectives each day; (2) The need to make good use of their time by prioritizing tasks and selectively communicating with others to complete the most important tasks first; (3) The need to keep track of communications in progress with others; (4) The need to communicate with people without knowing the best way to do this effectively for each person (for example, when to interrupt, what medium or device to use, the relative importance of the communication, and how to best respect personal preferences, privacy, and cultural protocols); and (5) Finding ways to build trust, achieve consensus, hold effective discussions, learn, and achieve goals that were traditionally best met with face-to-face meetings, even while the new communication technologies allow more business dealings to be done without face-to-face interaction.

These challenges not only change the way that enterprise employees do their jobs, but also the way they conduct their personal lives and coordinate personal and professional tasks each day. The increased pace and scope of business forces employees to multitask; as a result, active tasks must be prioritized and tracked. This leads to the shifting of work and communication with business colleagues during and outside of the normal workday. Thus, employees find they need to fit work and personal tasks together in new ways. Often, person-

to-person communication (including voice communication) is needed to accomplish these tasks.

These employees face similar challenges in their personal lives. As family members are multitasking and becoming more mobile, opportunities for regularly scheduled face-to-face communication become less common. For example, families in which both parents are enterprise employees have to integrate the demands of two jobs into their family life. The care and supervision of children and elderly dependents also becomes more difficult, as more demands are placed on parents' time. These challenges call for new interpersonal communication strategies, and they motivate the adoption and accommodation of the new communication technologies.

Legacy voice technology

The voice user experience has been well established as a key business capability for decades. Unfortunately, typical voice services are limited in functionality and have not grown with business needs. The legacy voice environment can be characterized by a high degree of supplier control, stability and inflexibility of technology, limited interoperability, and government regulation. While stability has often been viewed as an advantage with respect to voice, because it implies a level of reliability, stability can characterize a system that is slow to change and difficult to adapt to new business demands.

This inflexibility has a significant impact on employees. In enterprises with a high degree of employee mobility, people may confront multiple phone devices, multiple phone numbers, multiple voice mailboxes, and multiple voice pass codes. Furthermore, it is unreasonable to expect the end user to determine the best way to place outbound calls (e.g., by using a cell phone, a PC-based software client, a desk phone, a hotel phone, or by charging the call to a credit card).

In light of these conditions, levels of complexity, and the challenges identified herein, how well can the legacy environment meet the new challenges?

Table 2 summarizes some key points. In this table, it is clear that the legacy environment has difficulty keeping up with the requirements of businesses in transition. Enterprises today need to be able to add, delete, and change capabilities very quickly—they

Table 2 Challenges of legacy voice technology

Challenge	Legacy voice technology	Gaps/Risks	Impact
Get more work done each day	Requires use of multiple services that are not integrated	Multiple phone numbers, callers, and customers Management of multiple services is time consuming No support for social or business networks	Important calls may be missed or delayed Lost opportunities due to difficulty in communicating
Use time well; get the most important work done first	Encourages “interrupt-driven” behavior	More interrupts, due to limited communication choices No method to manage interrupts based on priority	Attention is focused on handling interrupts rather than work priorities
Keep track of communications in progress	Causes challenges in coordination and management of communications from multiple devices and services	Voice mail is sequential regardless of caller, task, or priority No comprehensive way to organize communication	Time is wasted in attempts to conclude dialogues in progress
Communicate effectively; bid for people’s time	Makes it difficult to determine the availability of the person one wishes to contact	Selecting best time and method of communication requires knowledge of contact’s state	Working relationships suffer from frustration in remembering and managing tasks
Achieve the benefits typical of face-to-face communication	Difficult to read communication cues	Difficult to determine when to spend “quality time” with colleagues	Overuse of face-to-face meetings limits the total work accomplished

need an environment in which they can control features and functionality. Individuals need richer functionality presented in simpler user interfaces. For example, a user attempting to communicate with another user should be able to use a simple interface and rely on the system (behind the scenes) to find that user’s phone number, availability, and other parameters, and even make conversions between spoken voice and text as specified in each user’s preferences.

From the enterprise perspective, the picture is also complex. Traditionally, the enterprise must subscribe to a variety of services from multiple providers. From the network transport perspective, public service telephone network (PSTN) carrier networks provide transport and interconnectivity using traditional technologies, while cellular carrier networks provide similar services for mobile voice users using a variety of incompatible technology standards. Interconnectivity among these users is achieved by routing traffic through the PSTN. Enterprises also subscribe to data networks using a multitude of technologies. For each of these types of networks, the enterprise engages with multiple carriers within multiple geographic areas, each with

differing offerings, rate structures, and regulatory environments.

Similar situations exist in the equipment market. Proprietary voice PBXs are available from multiple vendors with little interoperability (except through the PSTN) and often with associated voice mail systems that increase the linkage to specific vendors even more. Enterprises often focus on PBX systems with a small number of vendors to minimize the number of proprietary system types in their environment, but at the cost of restricting their ability to negotiate pricing and upgrade to better products as they become available.

Some specialized areas, such as contact centers, have specific services and equipment. In the contact center, the enterprise typically uses automatic call distributors (ACDs) with restrictive characteristics similar to the PBX industry. Larger ACD vendors also offer interfaces with computer systems as well, but again, often in a proprietary manner.

Thus, the legacy environment lacks the capability to meet many requirements for adjusting to new ways of doing business.

NEW VOICE TECHNOLOGIES

In this section, we examine each of the five challenges cited earlier in this paper and discuss ways in which new voice technologies can address them.

Challenge 1: Enterprise employees need to get more work done each day

It is already common practice in IBM to replace face-to-face meetings with teleconferences, saving considerable travel time and expense. The new voice technologies can make teleconferences more effective by enabling the integration of multiple communication channels (e.g., voice, video, file and screen sharing, or instant messaging), and by providing some of the communication cues typical of face-to-face meetings that have been lacking in teleconferences. For example, these cues may indicate if a participant wants to ask a question or interject a comment or if a participant has just joined or left the meeting. In addition, the system can provide a side channel (in text or voice) to a few close colleagues during a meeting.

For many employees, the best way to get more work done each day is to accomplish more during regular business hours. Increasingly in IBM, this means doing more multitasking. The following are examples of voice communication in parallel with other tasks: reading or writing e-mail during teleconferences; taking teleconferences on a mobile phone while in transit; or having one or more instant message sessions active during phone calls.

The new voice technology can help time-shift voice communication as well. Advanced recording systems for a meeting could allow a user to pause and quickly catch up, replay a question that was missed while multitasking, join a meeting late and catch up, or review a meeting the user missed entirely, but in a fraction of the time it took to hold the original meeting.

When fully synchronous voice communication is not needed, the new VoIP technology can help time-shift communication over longer intervals. Simple one-way communication can be recorded, compressed, indexed, and then consumed when convenient. While time-shifting is commonly done today with broadcast streams (e.g., through podcasting), VoIP technology can provide time-shifting for every teleconference and online meeting in the enterprise.

Another common practice is to combine instant messaging and voice communication. For example, one of the most common uses of instant messages in IBM is to set up a voice conversation (for example, a phone call or a “voice chat”). With better integration between voice conversations and instant messaging, one could imagine shifting between fully synchronous voice conversation and semi-synchronous voice messages as needed.

A more advanced form of time-shifting could be applied to the voice communications between two people who would have had difficulty connecting using legacy voice solutions. One might have a time-shifted voice conversation with someone using the same style of interaction in use today with instant messaging, but using voice instead of text. Extensions of this paradigm include real-time transcription between voice and text, a persistent transcript of the voice conversation that could be reviewed and reused, and extensions to multi-person conversations. This last example is effectively a time-shifted, online meeting among a group of people who are not only physically dispersed, but also not all available at the same time. This may be the future of online meetings in a global company like IBM.

Challenge 2: Enterprise employees need to prioritize tasks and selectively communicate with others

This challenge exposes a few general and important communication problems—namely, bidding for someone’s time and selecting a time and mode for a communication. By “bidding for someone’s time,” we refer to the problem we all face when we need to talk with someone about something. We intend to differentiate this task from simply “informing someone of something,” which only requires one-way communication. We examine how the bidding process presents the initiator and the recipient of the bid with different challenges.

When people bid for our time using a voice channel, it is difficult to prioritize and schedule that bid. For example, an employee may walk over to a colleague’s office to talk, only to have that conversation interrupted by a phone call from a third party. In the same way that e-mail and instant messaging allow one to prioritize an inbound text message, the new voice technologies provide opportunities to manage inbound voice communications.

An inbound voice communication may not need a response in real time. While the new voice technologies enable voice communication that is not in real time, the person initiating the communication needs enough feedback to know how and when to expect a response. While some scenarios such as this were discussed earlier in this section, socially acceptable protocols have not yet emerged for this style of voice interaction. The person initiating the bid need not choose the recipient's device (for example, the recipient's office phone, cell phone, or home phone). The new technologies would allow the initiator simply to select the target person and allow that person to map the bid to the device of his or her choice, or alternatively to queue the bid until it can be handled.

While the new voice technology can simplify the selection of the recipient's voice end device (for example, a cell phone), it can also provide more alternatives for voice communication and hence complicate the initiator's decision process. For example, a business traveler wishing to start a voice communication from a hotel room might use a PC-based VoIP service, a cell phone, or the hotel phone, and choose among various ways to pay for the call. In addition, as technology develops, telephony service providers add features to try to win market share. While this accelerates the pace at which new technology is adopted, it also keeps users guessing with respect to how to use, configure, and troubleshoot successive releases of their business and personal voice devices and services.

Opportunities to help with the bid for someone's time as the initiator of the communication will be discussed in challenge 4. The emergence of so-called "presence and awareness" technology provides another way to inform the bidding process. When someone (the bidder) wants to communicate with another person, this technology makes the bidder aware of information about the other person's ability and willingness to communicate. While the technology promises to help both parties in the communication process, the technology is immature and socially acceptable practices have yet to emerge. Presence data provides information about the state (or status) of an entity. The IETF standards for presence allow for and encourage flexibility in the definition of entities which can include (but are not limited to) people, groups of people, phones, data terminals, rooms, documents, sensors, or anything for which a state or status can be reported. Likewise,

state can take many types of value, such as: available, busy, in transit, or in a meeting.

When others bid for our time in a face-to-face situation, we expect that they will take our situation into account. For example, if I am talking to another person, I do not expect a third party to start conversing with me. Current presence and awareness technologies afford others with some social cues in non-face-to-face situations. For example, if others could see that I was already talking on the phone, in a meeting, driving, or sleeping, I could expect them not to attempt a real-time bid for voice communication at that moment. See Reference 8 for a case study that was performed to analyze how IBM managers deal with and manage interruptions and to examine the implications of these behaviors for the use of presence and awareness information.

People are sensitive about being "tracked." As technology supplies more information about people's whereabouts and activities, it is important that users feel in control of that information. The IBM Grapevine project⁹ explored the value of presence and awareness for making communications decisions, and the importance of respect for personal privacy.

Challenge 3: Enterprise employees need to keep track of communications in progress with others

The e-mail in-box is one of the simple, common tools that enterprise employees use to track work in progress. Unfortunately, the e-mail in-box is not integrated with instant message transcripts or with a user's multiple voice message queues. Even the so-called *unified messaging* solutions available today do not do much to integrate, organize, and track all of the text and voice communications users deal with daily, much less relate them to the documents and processes that are referenced by those messages.

Because VoIP puts voice communication, text communication and documents on the same networks, servers, and end devices, a truly unified portal could be designed for inbound communications, presenting them in the context of the tasks that relate to those messages. In addition, improvements in voice processing technologies (for example, speech recognition and speaker identification) would provide opportunities for personalized management of communication events and switching between text and voice modes.

Challenge 4: Enterprise employees need to communicate with people without knowing the best way to do this effectively for each person

As examples of this challenge, employees may not know when to interrupt, what medium or device to use, the relative importance of the communication, and how to best respect personal preferences, privacy, and cultural protocols.

The magnitude of this challenge is hard to describe, but it is clearly a problem that is getting worse as the need to communicate more broadly and in a more timely and efficient manner grows. Factors such as outsourcing of work, using offshore labor, and projects that span more than one enterprise all require employees to communicate with people who are unfamiliar to them. As the demographics of our business contacts become more diverse, so do the challenges for effective communication. In many cases, one has to locate the appropriate contact person, and this requires added communications within our business and social networks.

In bidding for someone's time, one may be the initiator of the bid and have the challenge of selecting a time and a method to present the bid. The initiator needs to consider the urgency of the request and how to convey that sense of urgency; the medium (e.g., voice, chat, e-mail, Short Message Service [SMS], or pager); how to make sure the recipient understands the bid (e.g., how much context to include with the bid); and how to present the bid in a way that the recipient will find appropriate.

This is clearly a significant problem, and one that must be solved in many cases without complete information. Yet this problem is one that people solve every day and adapt to the information and technology at hand. One interesting example of this, which involves voice communication beyond simple telephone calls, is the de facto bidding procedure adopted by users of Sprint Nextel** push-to-talk phones. The push-to-talk function of Nextel phones enables a user's voice to be heard on the speaker of another Nextel user's phone at the discretion of the person initiating the message. It is uncommon, however, for Nextel users simply to begin talking on the speaker of another user's phone without permission. Instead, users often use a protocol to bid for another user's attention by selecting the target person and then pushing and immediately releasing the push-to-talk button without saying anything.

This causes the target user's phone to chirp and displays the identity of the person initiating the bid. The recipient can then choose whether to accept the bid and start talking to the bid's initiator.

While the new voice technologies make voice communication easier to manage and provide some time-shifting options previously unavailable, effective solutions to challenge 4 will likely require not only new voice and text communication techniques,

■ A truly unified portal could be designed for inbound communications, presenting them in the context of the tasks that relate to those messages. ■

but also increased use of presence and awareness information, new social and business networking tools and skills, and perhaps some degree of integration of communications with business processes and transactions within and outside of the enterprise. Reference 10 examines the use of presence and awareness to help people communicate using the voice and text communication services deployed in IBM.

Challenge 5: As more business dealings are done remotely, enterprise employees must find ways to achieve goals that were traditionally achieved with face-to-face meetings

Face-to-face interaction has been and is likely to continue to be a superior way to create and develop business and personal relationships. Remote meetings will nevertheless grow in number, and business success will go to those who can adapt. In the following, we offer some observations on this phenomenon.

The integration of voice with other collaboration technologies can help people communicate when they are not collocated. This will make remote meetings more effective. Presence and awareness can provide some of the cues that are needed for effective communication.

While it is clear that voice provides a richer channel for communication than text (because voice carries the speaker's tone of voice and emphasis), videoconferencing should provide more advantages than

voice alone. Experience with videoconferencing has been mixed, however. Many people derive very little added value from a videoconference when compared to a high-quality audio teleconference. Others have found videoconferencing useful in specific contexts. For example, some enterprise employees have frequent videoconferences with their core colleagues, all of whom are equipped with an advanced videoconferencing solution.

Within IBM, instant messaging has displaced phone calls for many interactions. Voice messaging (or voice chat) has not experienced widespread adoption, however, in the United States. Conferencing with established colleagues appears to be more effective (and feel more natural to many) than with new and unknown colleagues—i.e., it is useful in maintaining a relationship remotely, but not for creating a new one.

IMPACT ON IT ORGANIZATIONS

Throughout this paper, we have discussed the challenges and opportunities presented to individuals within enterprises and to enterprises themselves in the context of the transformation of business and the workplace. We feel that the transformation of technology and the new architectures this transformation yields will become the enablers for this change. However, several issues will need to be addressed by IT organizations to best leverage these technology transformations. There are no industry-wide solutions for these issues, as a great deal depends on the individual enterprise and the opportunities it pursues. In the following, we identify and clarify some of these issues.

When they are not colocated, users need better tools and technologies to share, discuss, and debate information in a secure fashion. Customers and partners must be fully enabled participants in collaborative sessions, with access to all meeting tools and appropriate access to databases.

Users' worlds are quite complex, and they need a communications service that is not difficult to learn and use. The user experience must be reliable, simple, and intuitive. It should not matter to our users whether they are using a cell network, public Wi-Fi network, or enterprise wireless network, nor should users be penalized by the types of devices they use and their richness of function or access to

data. User interfaces should always look and feel the same (on phones, PCs, and other devices). Users should be easily able to change the state of a session (e.g., to add people or to add tools) at any time during the session. There may be a great deal of behind-the-scenes automation (such as user-interface normalization, network optimization and neutrality, bid/response mechanisms, and state-change mechanisms) that is required to accomplish this, in order to shield the user from complexities.

Consumer-style services must be brought into the enterprise. Employees are also consumers of communications services in their off-hours. More than ever, they are exposed to highly functional and inexpensive services on publicly available infrastructure. Often, employees (especially younger employees) expect parity between consumer and enterprise services.

As presence systems increase the probability of first-attempt connections, we would expect the number of voice mail messages to decline. Similarly, when voice mail can be routed to e-mail, the number of places one needs to check for messages goes down as well. Instead of accessing voice mail after being inaccessible only to hear "you have 115 new messages, all of them marked urgent," we will have fewer messages and be able to manage them more easily.

What is the role of voice mail in the future? Similarly, when people are working remotely and are increasingly mobile, is the plethora of PBX and IP-PBX features still needed? How many users know how to transfer calls or set up a three-way conversation using a traditional PBX system? Speech-recognition engines might allow an inbound caller to speak a message that is then translated into text for instant delivery to an instant messaging or SMS service. Communications applications might allow the user to respond to these instant messages or SMS messages by selecting a preestablished response message such as "I will call you back in 30 minutes" or "the system will check your calendar and schedule a call at your earliest availability." In all of these cases, users are served in a more real-time fashion and the need for leaving a message in a voice mailbox (for action at a later time) is diminished.

The enterprise IT staff needs to change to support the new voice technologies. As converged voice and

data networks and solutions replace legacy voice-only networks and appliances, the demands on the enterprise IT staff will shift. Experience with the new VoIP protocols and components will be needed, and problem determination and resolution skills and tools will need to be developed. IT organizations will continue to find open standards valuable to expand deployment options.

Voice has traditionally been managed as a separate resource, with separate networks, separate suppliers and vendors, and different endpoints. The traditional parameters (media type and cost) are yielding to a new set of parameters: mobility, flexibility, simplicity, collaboration, and cost. Our user endpoints are converging with smart phones and PC-based “soft phones.” There is little difference between a two-person voice call and a collaboration session with two people. Stored voice content should be leveraged similarly to other nonstructured data for the underlying intelligence it contains. Voice endpoints are increasingly capable of accessing and providing information from or to virtually any application.

Many enterprises have relied on suppliers to provide well-managed, integrated services. This model will continue to be chosen by many enterprises in the future. However, this leaves the enterprise at the mercy of the supplier in determining features and functionality and which downstream systems are supported (i.e., integrated). The evolution of these services is subject to risks associated with the suppliers’ analysis of market demands for change and the suppliers’ development cycle and financial picture. We believe that many enterprises will be reluctant to take this risk, as it impacts their ability to transform their operations. Some industry segments are progressing quite well, for example, with the implementation of Internet multimedia subsystems (IMS) that, if implemented properly, present selected supplier capability to the customer, often in service-oriented architecture (SOA) formats.

Infrastructures that are capable of rapid change must be built. Each enterprise must determine how to do this, but generally, the answer lies in a new style of architecture for voice (one which is currently being adopted by the enterprise for its other applications)—namely SOA.

Figure 1 shows a high-level logical view of such an architecture for voice. As the figure shows, voice is

no longer an isolated tower or silo. Voice services are presented, through Web services, to any application or business process that needs them. This is not limited to the traditional collaboration-style services. Certain business processes that require human intervention can be included, such as a sales process that might require management approval of a bid, a procurement process that might require price negotiation, or a human-resource process that requires employee or manager review of annual performance.

We no longer look solely to typical vendors for transport and equipment. In an enhanced relationship, an enterprise can also expect them to provide Web services that we would then present to our users in a manner that is consistent with internal presentation standards. Services such as GPS and road traffic alerts are extremely important in a mobile environment.

Component design and reuse of components is significant in this context and can reduce development time and the speed at which new features and functions are offered. Voice will interact with a broad selection of corporate authentication and authorization systems, billing systems, directories, presence systems, databases, and collaboration applications. We will no longer be able to rely simply on a single supplier to provide everything, because our applications will exercise real-time control over sessions and decisions will be made based on calendars, user locations, user preferences, and privacy instructions. Enterprises will need to act increasingly as integrators, and must demand that suppliers comply with key standards (such as SIP) and practices (such as SOA).

CONCLUSION

In this paper, we have identified some of the new challenges faced by enterprise employees. We have noted that using voice technologies simply to make and receive calls is shortsighted. The new voice technologies should be used to help people with the new challenges they face, and can expand the ways people can talk to one another. These technologies can enable new and enhanced ways of integrating voice services with text-based communication services, presence and awareness services, social and business networking facilities, and evolving Internet technologies. We examined the new business

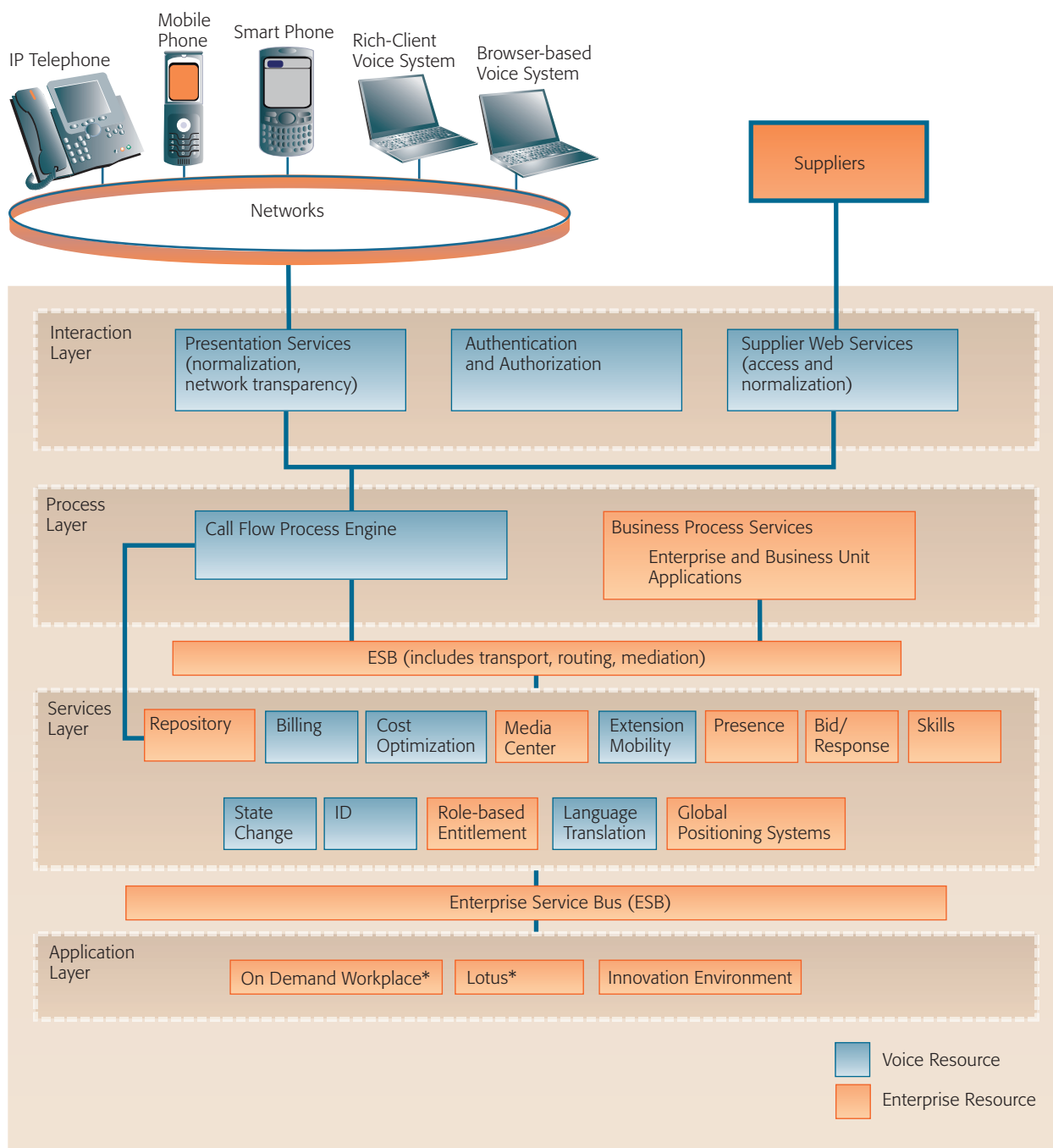


Figure 1
High-level voice architecture

environment that needs better communication services, and the challenges enterprises will have in delivering these services.

Will all of these factors change the way that people talk to one another? The degree to which personal use of telephone calls to business colleagues has changed

over the past 10 years, as e-mail and instant messaging have taken hold, suggests that change is likely. In this paper we have advocated the position that a great deal of change is imminent. The emerging voice technologies are positioned to change not only the way that people talk to one another but also the way people work with each other.

ACKNOWLEDGMENTS

The authors have been immersed in the evolution of voice technologies and their intersection with the enterprise data network and the wider Internet for quite some time. The lessons and insights in this paper come from our collaboration with many talented colleagues inside and outside IBM. While listing some of them here will necessarily omit some very important colleagues, we accept the risk, apologize in advance for the omissions, and would very much like to thank Paula Stewart, Juerg von Kaenel, Brent Hailpern, John Turek, John T. Richards, Wendy Kellogg, Jeremy Sussman, Steve Levy, William Bennett, Tracee Wolf, Amy Travis, Bill Rippon, Konrad Lagarde, Tom Erickson, Harry Reichlen, Fred Spulecki, Dave Newbold, Jon Peluso, Bill Bodin, Flor Brouard, Mike Pryzgoda, Ed Bonkowski, Frederic Surand, Steve Wootton, David Petch, Terry Shaw, Will Morrison, Keith Fancher, Richard Teevan, and Craig Hayman.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

**Trademark, service mark, or registered trademark of Sun Microsystems, Inc., Microsoft Corp., AOL LLC, Netflix, Inc., Amazon.com, Inc., Apple, Inc., or United Parcel Service, Inc., or Sprint Nextel Corp., in the United States, other countries, or both.

CITED REFERENCES

1. *Specifications for the Network Voice Protocol*, Internet Engineering Task Force RFC 741 (November 1977), <http://www.ietf.org/rfc/rfc741.txt>.
2. J. Rosenberg et al., *SIP: Session Initiation Protocol*, Internet Engineering Task Force, RFC 3261 (June 2002), <http://tools.ietf.org/html/rfc3261>.
3. Special Issue on VoIP, *ACM Queue* **2**, No. 6 (September 2004).
4. K. Werbach, "Using VoIP to Compete," *Harvard Business Review* **83**, No. 9 (September 2005).
5. T. Friedman, *The World is Flat*, Farrar, Straus and Giroux, New York (2005).
6. *Expanding the Innovation Horizon: Global CEO Study 2006*, IBM Corporation (2006), http://www.ibm.com/ibm/ideasfromibm/us/enterprise/mar27/ceo_study.html.
7. B. Nussbaum, "Get Creative!" *Business Week Online* (August 2005), http://www.businessweek.com/magazine/content/05_31/b3945401.htm.
8. J. M. Hudson, J. Christensen, W. A. Kellogg, and T. Erickson, "I'd Be Overwhelmed, but It's Just One More Thing to Do: Availability and Interruption in Research Management," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, Changing Ourselves*, ACM Press, New York (2002), pp. 97–104.
9. J. Richards and J. Christensen, "People in Our Software," *ACM Queue* **1**, No. 10, 80–86 (2005).
10. J. Christensen, J. Sussman, S. Levy, W. E. Bennett, T. V. Wolf, and W. A. Kellogg, "Too Much Information," *ACM Queue* **4**, No. 6, 50–57 (2006).

Accepted for publication June 6, 2007.

Published online November 7, 2007.

Jim Christensen

IBM Research Division, 19 Skyline Dr., Hawthorne, NY 10532 (ibmjim@us.ibm.com). Mr. Christensen has spent several years working in the IBM Research Division at the Thomas J. Watson Research Center, most recently on a joint project with the IBM CIO and CoE (Center of Excellence) on the IBM internal audio communication service known as Rendezvous or IBM Enhanced Audio Conferencing. His background is in software engineering and he has focused his career on applying and extending new technologies to help people get things done. He has B.S. and M.S. degrees in computer science from the University of Illinois at Urbana-Champaign.

Bob Hughes

IBM CHQ, Enterprise on Demand, 294 Route 100, Somers, NY 10589 (rbhughes@us.ibm.com). Mr. Hughes is currently in the IBM CIO office in the Infrastructure Strategy and Enterprise Architecture group with specific responsibility for voice strategy and SOA infrastructure. Having been involved in new voice technologies for many years, he has provided direction for the internal adoption of the new voice at IBM from the initial IP phone and IP transport phase through the current efforts toward integrating voice with applications. Prior to joining IBM, he managed voice and data networking globally for Avon Products and Emery Air Freight. He has a B.S. degree in mechanical engineering from the University of Pennsylvania and an M.B.A. degree in finance and organizational behavior from the University of Michigan. ■