IDC
Analyze the Future

WHITE PAPER

Can You Hear What I Mean? Polycom Delivers HD Voice

Sponsored by: Polycom

William Stofega

December 2006

EXECUTIVE SUMMARY

By all accounts, both consumers and enterprises are incorporating IP-based communications services in order to take advantage of potential cost savings, flexibility, and a rich feature set that only IP can deliver. IDC believes that the "sky is the limit" in terms of the potential for new applications and services. Although it is clear that the revolution implied by IP-based communications and services is clearly underway, both vendors and service providers have yet to create an end-user experience that can "amaze and delight." There is little doubt that the deployment of new multimedia applications will create a richer communications experience; however, IDC believes that voice communication is and will continue to be the most valuable part of the communications ecosystem — especially in the enterprise. Still, despite over a century of voice communications, little has been done to enhance the most important element of communications: voice quality. The adoption of new communications technology, especially IP, has created an opportunity to take the power of voice communications to the next level. Thus far, the emphasis has been on VoIP applications and cost reductions. IDC believes that the end-user experience in current implementations of VoIP can be improved in a manner that is simple yet clearly demonstrates the value and superiority of IP-based communications. This is especially important as businesses look to opportunities in emerging markets such as India and China. Doing business on a global basis demands new standards of clarity and consistency in communications. Foreign accents may obscure meaning and lead to costly misunderstandings. In addition, call centers located in non-native geographies must be able to communicate effectively with customers in order to resolve problems.

As voice communication evolves, new opportunities will present themselves. Lightweight-communications clients such as Skype are becoming more popular in both consumer and enterprise markets. Although these clients are relatively simple, since they traverse the open Internet the need for enhanced voice technology is especially acute. Voice is also beginning to be deployed in non-traditional domains such as gaming. Thus far the implementations are crude and the quality is spotty; however, IDC believes that an opportunity to offer an improved voice experience in gaming is especially appealing. Finally, in an ultra competitive environment full of new and traditional players, there are few opportunities to gain a competitive edge. IDC believes that enhancing the clarity of voice represents a Greenfield opportunity that could help service providers avoid the death spiral of commodity pricing.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

THE IMPORTANCE OF VOICE

With all the hype surrounding convergence, content, mobility, and multimodality, the single most powerful application that unifies and clarifies all forms of communications is the sound of the human voice. Although email, IM, SMS, and other forms of text-based messaging have assumed an increasingly important role in enterprise communications, text-based messaging at best is simply way to use voice more efficiently. In other words, messaging is used to schedule and arrange conversations or get answers to basic questions, such as "When is the project due?" "Can you make the meeting?" "How much will it cost?"

With its ability to add context through tonal subtleties plus the ability to glean feedback in real time, voice has an advantage that is practically unbeatable when compared to other forms of communication. When communicating with family and friends, voice is the first choice and provides an intimate sense of connection that cannot be conveyed in an email or IM session. For example, voice can relay concern from a grandparent or express the joy of a newfound relationship that goes beyond the meaning of the words used in the conversation.

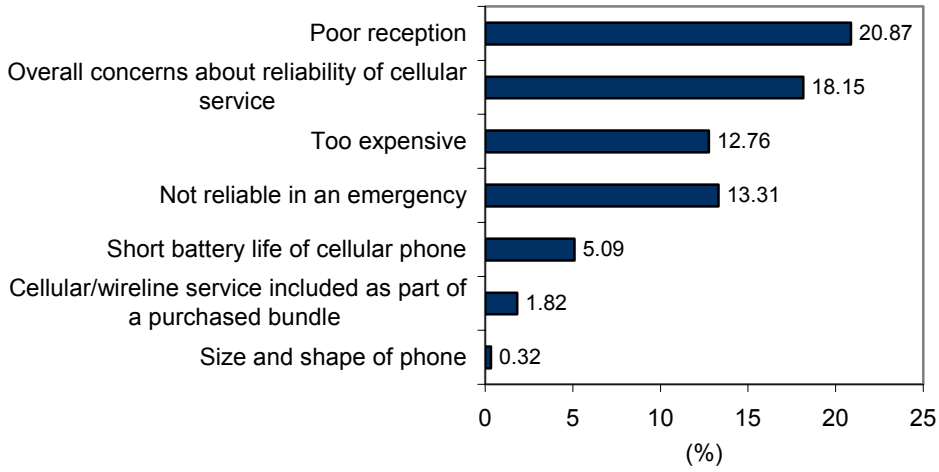
The demands on resources and time also amplify the value of a voice conversation. Still, given the reliability and availability of voice, most take it for granted. A short outage in voice service can be disruptive for businesses, while a prolonged lack of service can spell financial disaster. Even further, a missed phrase or nuance in a verbal financial transaction can threaten the stability of Wall Street.

Furthermore, even minor annoyances such as a garbled phrase during a cell phone conversation can be troublesome, especially if they reoccur. In fact, although mobile phones offer unsurpassed convenience, IDC survey data reveals that concerns about reliability and reception are the top inhibitors to cutting the cord and going completely wireless (see Figure 1).

FIGURE 1

Reason for Not Using Cellular Phone as Primary Home Phone

Q. What is the primary reason you would not use a cellular phone as your primary home telephone?



Source: IDC Consumer VOIP Survey, 2005

In order to understand the true value of voice communications, it is important to think about voice not just in terms of casual conversation but as a tool for real-time dissemination of critical information. In much the same manner that garbled text can create misunderstandings that have the potential to spoil relationships with key customers, the same scenario exists in voice communications. With this in mind, it is essential that voice conversations be transmitted and received with the best possible technology that creates the best possible quality to ensure every nuance is conveyed accurately and, most important, understood.

IP COMMUNICATIONS: FROM DISRUPTIVE TECHNOLOGY TO THE NEW STANDARD

Although commercially available, IP-based communications services and equipment began to appear in the late '90s, the real point of departure from "science project" to next-generation contender occurred in 2003. It was during this time period that commercially available service for both enterprise and consumers began to percolate upwards. Advances in technology that allowed for IP-based communications to replicate PSTN-like reliability and quality began to hit their mark. Furthermore, as "live" deployments began to proliferate, the advantages of IP telephony became much more tangible to potential end users. Also accelerating adoption was the declining cost of high-speed access facilities enabling businesses of all sizes to take advantage of new products and services, including IP telephony.

IDC believes that the flexibility and efficiencies inherent in IP-communications will ultimately lead to ubiquitous access to information — anytime, anyplace. No longer

will customers have to make allowances and adjust their lifestyle to the limitations of the communications network. Instead, the network will bend to the needs of the end user, both in terms of service coverage and the mode of the message itself.

The adoption of IP-based communications services by both consumers and enterprise customers is taking off, but there is still much unrealized potential and growth. Still, companies such as Skype have proven the popularity of an IP-based communications service that can straddle both enterprise and consumer markets. By 2010, IDC estimates that 40 million consumers in the United States will use some form of VoIP service to communicate. Furthermore, the efficiencies and capabilities of IP-based communications are hardly lost on enterprise customers, and by 2007 IDC estimates that in the U.S. revenue generated from the sale of VoIP equipment and services will reach $8.4 billion.

Voice Quality Over IP: Unrealized Potential

Still, despite a plethora of offerings, the true potential of IP-based communications has yet to be realized. Most of what is being delivered to both consumers and enterprise customers is simply a cheaper version of what is already available on the PSTN, especially when it comes to sound quality. A PSTN "toll quality" call is limited by the bandwidth (300–3,300Hz) utilized in the PSTN. This label of "toll quality" should by no means be accepted as an acknowledgement of the highest quality available. Instead, "toll quality" is simply a minimal base given the 3.3kHz bandwidth limitations built into TDM networks and is hardly reflective of the true potential of high-fidelity voice communications. IP-based voice communications can take advantage of technologies that may require additional bandwidth but can deliver voice quality that is superior by several orders of magnitude.

For example, the ITU's G.722 wideband speech codec for IP-based communications captures 7kHz of bandwidth. This is more than two times the amount of speech data captured in a TDM-based call. This additional bandwidth increases the flow of data enabling new levels of voice clarity. This additional clarity allows both parties involved in a conversation to better distinguish between vowels, consonants, and the many intonations used to communicate verbally. Furthermore, the higher level of quality can help reduce listener fatigue, improve concentration, and increase intelligibility.

THE NEED FOR BREAKTHROUGHS IN VOICE QUALITY

Globalization

The expansion of global business has increased the importance of accurate voice-based communication among end users. In order to grow revenue and increase profit margins, sales team members must travel abroad to look for new opportunities. As part of an effort to cut costs, businesses are outsourcing critical functions such as call centers and help desks to workers in foreign countries who are not native speakers.

Of course, one of the advantages of IP-based communications is the ability to eliminate the costs associated with distance, thus enabling the use of outsourced

foreign workers engaged to service the needs of customers in the United States as well as other countries. Furthermore, the drive to utilize workers in foreign countries is not just about reducing cost in non-professional occupations, rather it addresses a growing shortage of qualified professionals in a variety of different fields. This shortage is especially acute for those with an expertise in IT. A 1997 report by the U.S. Department of Commerce entitled *America's New Deficit: The Shortage of Information Technology Workers* predicted the explosive growth in demand for IT workers. Although this demand has slowed somewhat, IDC predicts that the worldwide offshore IT service market will grow from $13.1 billion in 2005 to $29.4 billion in 2010, a five-year compound annual growth rate (CAGR) of 17.6% (see *Worldwide and U.S. Offshore IT Services 2006-2010 Forecast*).

Given the increasing proliferation of international workers, the need for improvements in voice fidelity is more apparent than ever because grammar, pronunciation, and even word selection are much different than the listener expects. A Korean speaker of English, for example, will commonly substitute "p" for "f" ("faint" becomes "paint," "coffee" becomes "copy"). A Turkish speaker may insert extra syllables ("stone" becomes "istone" or "sitone").

When presented with a continual string of such verbal puzzles as the meeting progresses, the listener is distracted. Pieces of the conversation are lost on these mental detours, trying to deduce what words were used. As this occurs over and over in a meeting, mental fatigue increases, while comprehension and interaction drop. The listener has to frequently divert attention to figuring out what words were spoken, instead of staying with the flow of the conversation. Too much time is spent in unraveling the intended meaning, instead of understanding it.

This problem can be troublesome for outsourced professionals, including those with expertise in software development, engineering, and banking, but is especially acute in outsourced call centers. Competitive pressures have pushed corporations to pursue a relentless drive to cut costs, which has in turn forced companies to pursue highly skilled but less costly pools of workers, many of whom reside in foreign countries. India and China have emerged as global powerhouses for outsourced call centers and help desk expertise. Providing responsive and accurate customer service is the most important task of any corporation and is roughly akin to "whispering in the customer's ear."

Even more important, the outsourcing phenomenon is worldwide. Companies in Europe and Asia are taking advantage of the efficiencies and access to new pools of professional workers that outsourcing can bring to the table.

Whether utilizing offshore workers to staff a call center, beef up a software development team, or simply increase interactions with foreign customers and prospects, voice fidelity plays an important role in accurate delivery and reception of communications. Given the wide range of dialects and speaking capabilities plus factoring in the variety and quality of communications devices and end points, the potential for misunderstanding looms large.

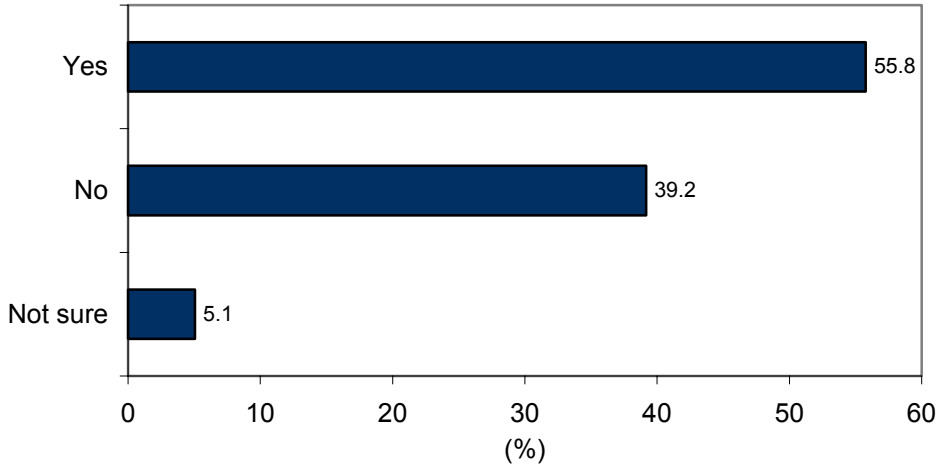
Remote Workforce

In order to attract and keep a talented workforce, many companies allow geographically dispersed employees to telecommute from home. Depending on the distance from the office, this arrangement can be on a permanent basis or for one or more days per week. In these remote environments, VoIP telephony services are often used for both personal and work-related communications. In fact, IDC survey research reveals that more than half of consumer IP telephony usage is work related (see Figure 2).

FIGURE 2

Use of VoIP Telephony Service for Work-Related Calls

Q. Do you use your VoIP telephony service for work-related calls?



Source: IDC Consumer VOIP Survey, 2005

Enhanced voice quality is an imperative in these environments due to the detrimental impacts of slower home Internet connections and higher levels of background noise.

Differentiation

Whether delivered via a premise-based product or an off-site hosted solution, the look and feel of IP-based communications solutions from different vendors is becoming less differentiated. Customers looking to utilize IP-based communications as a competitive differentiator are looking for something beyond feature bloat and cost savings to help further extend the utility of their technology investment. IDC believes that in order for next-generation communications equipment and services to evolve beyond their current state, IP vendors and service providers must create value for customers allowing them easier and more productive work environments. Technology that is not focused upon creating value by enhancing the end-user experience in a simple and straightforward fashion will find limited usage in any market segment. As

much as vendors and services providers look to applications such as presence to enhance the value of IP-communications, they often overlook the most critical application that is provided over an IP communications system: voice.

EMERGING USE CASE

The Skype Effect and Beyond

IDC also believes that IP-based voice will begin to change the very notion of communications. In many ways applications such as Skype are already beginning to realign communications. Perhaps one of the most important realizations about Skype is that its voice quality is incredibly rich despite the fact that the service is a best efforts solution that rides over the open Internet. This high-fidelity sound is certainly one of the drivers of Skype's exceptional growth, and it is a perfect example of using the additional bandwidth available in IP-based networks to allow for wideband codecs to enhance the sound of an IP call. The rich quality of a Skype call is especially evident on communications that take place entirely on an IP network with no conversion to TDM. Of course, once a call touches the PSTN, the conversion process requires that a Skype call adhere to the bandwidth limitations implicit in TDM networks, and the call quality can suffer. Thus, although many pundits point to voice-quality problems in IP communications, most of the blame resides in the limitations inherent in the TDM-based networks.

It is also important to point out that Skype is a communications platform that seamlessly coexists as a business application as well as a communications tool for personal usage. Casual observations by IDC confirm the pervasive presence of Skype on corporate laptops. At a recent technical conference, a large percentage of those attending were either using the Skype client to make calls or communicate via an IM session while listening to presentations. Thus, without spending a dime, Skype has been able to appear on some of the most coveted real estate in the world: the corporate desktop, while effortlessly transitioning to the mass market. The success of Skype is often attributed to viral marketing, and the company certainly has validated this low-budget marketing technique.

Although Skype enables its customers to experience the power and potential of IP communications, IDC believes that the future of voice telephony is less about self-standing segregated services and more about integrating voice into other applications and services. Instead of voice being relegated to traditional roles, through the use of IP, voice will become embedded in business process and search technologies. For instance, voice capabilities added to RFID tags could allow a lost suitcase or personal item to "call" its owner to come and retrieve it. Furthermore, as technologies become less proprietary and cheaper, the whole notion of a single-purpose centralized service provider begins to fade away. Business could create smart and disposable voice applications to enable communications with their customers. Finally, voice-driven Web search could realize a world of smart agents that communicate with end users to efficiently organize and assist in both business and personal matters. Although interfacing with smart agents will involve the use of a keyboard to generate text-based commands and instructions, IDC believes that IP-based voice will also serve as an interface. This will free end users from computer terminals and allow access to

information and services from any location. Of course these and other futuristic services will require clear hi-fidelity sound to convey accurate instructions.

Further advancing the revolution in IP resources, universal availability will mean that the network is no longer the limiting factor in terms of voice quality. Furthermore, although IP-based voice services have just started to impact traditional wireline service, it is clear that it will also begin to seep into and disrupt traditional notions of mobility. Dual-mode handsets that can seamlessly roam over cellular and WiFi networks are fast becoming a reality. Some cities are already building municipal WiFi networks that will be used to transport voice and multimedia applications. These networks will not only be used by consumers but also by municipal workers who will need crystal clear voice communications to help ensure optimal productivity.

CHALLENGES AND OPPORTUNITIES

Challenges

The conversion from a traditional business telephone system to IP telephony raises fundamental quality of service (QoS) considerations. The path of a voice phone call can be broken down into several elements, each of which affects quality. For the WAN, operational performance and end-to-end QoS must meet basic user expectations for voice reliability and toll-quality sound. Data network services are very familiar with SLAs specifying latency, throughput, availability, packet loss, mean time to repair, and mean time to report. A converged network running voice, video, and data over a packet infrastructure has to measure and guarantee the traditional data SLAs as well as a set of voice/video-specific SLAs. Three key parameters are applied to ensure that user expectations of voice quality are met. These are:

- **Clarity.** Clarity is a measure of the quality of the voice sound to the listener. Two techniques are used to measure clarity of a voice signal after it has been transmitted from a phone: perceptual speech quality measurement (PSQM) and perceptual analysis measurement system (PAMS).

- **Delay.** Delay is the time it takes the voice signal to travel from the caller to the person called. In the data communications world, this is referred to as latency. Delay is introduced through the time required to code and decode voice into and out of packet. In addition, the switches and routers add small amounts of time.

- **Echo.** Echo is the sound of the speaker's voice returning via the same phone. This is usually caused by electronic misalignment between the trunk line and the phone line.

Opportunities

Although much can be done to create a stable network environment for IP voice, proper and ongoing network assessment must be made to assure that end users are getting the quality they require. Enabling the best voice quality without paying attention to both infrastructure and low-quality end points is tantamount to listening to a MP3 file over a monophonic earpiece.

Still, the opportunities for high-fidelity sound loom large in IP-based telephony. These reasons are not only related to helping workers communicate effectively and avoid costly misunderstandings but also include the continual cycle of new and younger employees entering the workforce. IDC has continually stressed the importance of understanding trends and expectations that exist in the youth market as they relate to the enterprise. Simply stated, the youth of today will become the employees of the rapidly approaching future. They have grown up with a different set of expectations that includes quality sound from all of their devices both at home and work. Those employers that understand these needs will be able to bring together the best and brightest, while those that do not understand these needs will be at a competitive disadvantage and will eventually fade away.

Voice is also beginning to play an increasingly important role in online gaming. Players are creating virtual worlds in which IP-based voice communications are used to bring the virtual experience closer to reality by enabling the creation of communities populated by animated personalities that speak. Other gaming scenarios include virtual teams that compete against other teams. Each side utilizes voice communications to plot and direct their efforts against the other side. Voice is also becoming part of the Web 2.0 and social networking phenomena. Interactive dating as well as popular youth-oriented sites such as MySpace are implementing voice capabilities. Thus far the sound quality of online voice communications is poor and certainly does not match the visual experience, but this is only temporary given the increasing expectations by gamers and those involved in social networking for high-quality voice communications.

Perhaps a bit closer to realization and widespread adoption is video conferencing and video telephony. Video as a communications tool has always held great potential, but the underlying technology to support the application has conspired to prevent widespread adoption. However, with the advent of cheaper high-speed facilities and plug-and-play technology, video communications is beginning to climb the adoption curve. Given the digital home experience in which consumers regularly experience theater-like video and sound, video communications must also move to high definition.

POLYCOM HD VOICE

Although given its almost universal availability, the importance of voice communications is often forgotten. Furthermore, the potential of voice has been stymied by both economics and complacency. However, similar to advances in technology that enhance the ubiquitous experience of television, the use of wideband codecs, in conjunction with the appropriate hardware design, will enrich and enhance voice communications. Leveraging over 15 years of research in communications technology, Polycom has developed a technology it calls High Definition (HD) Voice that is designed to electrify voice communications well beyond the traditional experience.

Polycom HD Voice delivers enhanced clarity over ordinary phone lines and over IP for life-like, vibrant conversations. This increased clarity enables much more natural conversations, which significantly boosts recognition and can help enhance

productivity. You can hear every word without repeating, which saves time and cuts down on misunderstandings. As global business and remote collaboration continue to increase, the need for clear communications has become more critical than ever. Accents and cultural differences can lead to costly misunderstandings. The clarity delivered by Polycom HD Voice makes remote collaboration easier, enabling geographically dispersed teams to communicate as effectively over the phone as they can in person.

By utilizing the technology developed by Polycom, HD Voice can deliver a new level of voice quality over the phone using the following:

- Wideband technology for over twice the clarity of ordinary phone calls
- Patented Acoustic Clarity Technology 2, which generates full duplex, echo cancellation, and noise reduction with advanced voice processing
- System design that integrates the Polycom software components together with its hardware design that can maximize the overall sound quality of a phone conversation

Experience Polycom's HD Voice technology at www.polycom.com/hdvoice.