

The Picturephone Is Here. Really.

THANKS TO THE POWER AND CONNECTIVITY OF TODAY'S MOBILE DEVICES, COMPUTERS, AND TELEVISIONS, VIDEO TELEPHONY WILL SOON BE EVERYWHERE

In the annals of technologies with long gestation periods, few can match video telephony. *Punch's Almanack* published a cartoon illustrating the concept way back in 1878. Then, throughout the next century, the idea resurfaced repeatedly in science-fiction comics, motion pictures, pulp stories, and novels. In the animated TV series "The Jetsons," starting in 1962, George's boss, Mr. Spacely, regularly appeared on a display screen to show George the latest sprocket design. In a memorable scene from the 1968 movie *2001: A Space Odyssey*, a weary space traveler videophones his daughter from a space station orbiting Earth.

Around the same time, videophones began showing up in the real world. AT&T announced its Picturephone service in 1964; the company even installed a Picturephone booth at New York City's Grand Central Terminal. But at US \$16 per 3 minutes of jerky images, the service never caught on.

Nevertheless, as with flying cars and jet packs, there is something about video telephony that people just can't let go of. And unlike flying cars and jet packs, a videophone is something you almost certainly have access to already, in the form of your computer, your smartphone, and almost every gizmo that communicates. The biggest computer firms have embraced the trend: Microsoft is now buying Skype for \$8.5 billion to further strengthen the video telephony capabilities already built into Windows, Office Live Meeting, and a number of other products. Apple's got FaceTime, and Google has begun rolling out multiuser video chat in its emerging Google+ social network.

With the exception of road warriors checking in with their kids at home, however, for most of us video telephony still isn't a part of our daily lives. But allow us to go out on a limb: It will be, and within just a couple of years.

The rap on video telephony is that people just didn't want it in their homes. They didn't want people seeing them bleary eyed

By Thomas Wiegand
& Gary J. Sullivan

and mused in the morning (or any other time, for that matter). Nor did they want their callers seeing that they were flipping through mail or making a grocery list while chatting on the phone.

So equipment manufacturers turned to the corporate world, introducing pricey videoconferencing systems designed to replace on-site meetings and reduce travel costs. While many companies invested in the technology in the 1990s, it typically gathered dust, unused. It looked as though people didn't want it anywhere.

Call us eternal optimists, but we believe that this conventional wisdom is wrong. For one thing, the vast majority of personal telephone calls occur between spouses or close relations: parents and children, siblings, and so on. These people have already seen each other bleary eyed and mused (and would probably overlook a little grocery-list making or other multitasking). We think the main reason people haven't embraced video telephony is that it has been clunky, owing to

PHOTOS: DAN SAEILINGER; STYLIST: WENDY SCHELAH/HALLEY RESOURCES

technical obstacles that prevented it from being done well. But most of all, videophone equipment was considered too expensive for most people for their private use.

One by one, those obstacles—hardware, networking, compression—have fallen away. And the final roadblock—standardization and interoperability—is teetering.

Let's start with hardware. Video telephony isn't all that complicated. It needs four basic things: a microphone and a camera to capture sounds and images, and a loudspeaker and a monitor screen to re-create them.

The call, of course, also needs a network to connect across. And there's one more basic requirement: a system to compress the data. The signals captured by the microphones and cameras contain more information than can be sent across the available communications networks, wired or wireless—and more than is necessary for an adequate video call. So the final piece of the videophone tech puzzle is a means for compression. On the sending end, a microprocessor and its software act as an encoder, compressing the signal—that is, reducing the number of bits that represent the video and audio data so they can be sent in real time over the available connection, be it wired or wireless. Of course, what gets encoded on one end must be decoded on the other; on the receiving end the microprocessor reassembles the audio and video from the bits. The compression system eliminates a vast amount of data, because today's communications networks, even broadband ones, don't have nearly enough throughput to send all the data created by the cameras and microphones.

For example, one form of high-definition video has a resolution of 1280 by 720 pixels at a rate of 60 frames per second; uncompressed, that flows at about 660 megabits per second. Even if the resolution and frame rate were each cut by half, reducing the data flow to 165 Mb/s, that speed is still way beyond the capabilities of today's typical broadband networks, which operate at a tenth of that rate at best. So compression is essential. The algorithms used to encode the signals typically reduce the data by a factor of 250 to 1000.

Years ago, hardware that could run the compression algorithms fast enough to encode and decode good-quality video and audio signals in real time cost a lot. And in those days, the algorithms themselves weren't very efficient, imposing an even higher burden on the processors. To make it work at all required dedicated digital signal-processing hardware. The costs of that hardware and the need for high-speed network connections relegated video telephony to corporate conference rooms through the 1990s. There, groups of people could

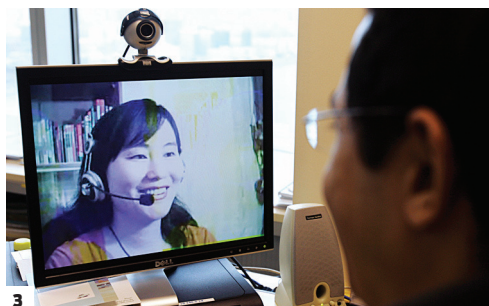
Video Telephony Through the Ages



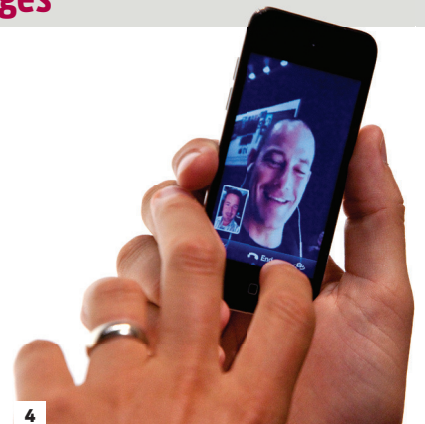
1



2



3



4

1 In 1965, the Bell System heavily promoted its new Picturephone service; this photo was published in the *New York Post*. **2** In the 1990s, business users were the target market for video telephony. Here, Intel marketing manager Jeff Abbate demonstrates a videophone at the company's Hillsboro, Ore., campus in 1997. **3** When Skype introduced a free and easy-to-use form of computer-to-computer video communications, it found a host of uses outside the business environment. Here, a user in Hong Kong takes a Mandarin lesson from a teacher on Hainan Island. **4** Today, Apple promotes FaceTime as a way to make an emotional connection with friends and family. **5** In 1991, users of business videoconferencing systems, like the US \$40 000 PictureTel Model 200 shown here, had to gather around the conferencing hardware to communicate. **6** Robotic telepresence systems like the VGo let the remote user navigate outside the conference room.

share a video camera and a screen connected to a similarly equipped remote conference room over expensive connections. And even so, the systems could send images with only limited resolution, about a quarter that of a standard-definition television picture, which itself is less than half the resolution of a high-definition television picture.

The cost and quality problems got even worse if the meeting called for participants at more than two locations to join. Then an expensive multipoint control unit (MCU) entered the mix, first to convert the variety of encoded data into a single package, and then to convert that package into all the formats needed for each participant's receiving hardware. This transcoding at the MCU not only degraded the image quality of the video but also created lags in the video stream, making communication unnatural and awkward.



5



6

The technology got a lot better, and a lot cheaper, around the turn of the millennium. Cellphones and laptops became enormously popular, with screens and cameras and vast processing power built into every unit. And when that occurred, video telephony became simply a software problem. Around the same time, Internet-style, packet-switched communications continued to replace traditional circuit-switched networks on the telephone system; this simplified the compression problem, because packet-switched networks typically make more throughput available to the average user, enough to pass along video images of at least a tolerable quality.

The rise of the portables flooded the market with relatively cheap flat-panel displays and cameras, lower-cost microphones, and chips fast enough to process audio and video. Today's smartphones, for example, have screen resolutions about as high as yesterday's standard-definition televisions.

Compression technology has also improved significantly in recent years. The main players here—the Visual Coding Experts Group (VCEG) of the International Telecommunication Union and the Moving Picture Experts Group (MPEG) of the International Organization for Standardization—cooperatively brought out a new compression standard under the name H.264/MPEG-4 Advanced Video Coding (AVC) to supplant the H.263, MPEG-4 Part 2, and MPEG-2 video-compression standards of the 1990s. (The alphabet soup of acronyms comes from the fact that there are two different standards organizations involved and that the video standards are subsets of larger sets of audio and video standards.)

The new standard reduced the bit rate for the same video quality to at least half that of its predecessors. For example, it takes HD video with a resolution of 1280 by 720 from a raw data rate of 660 Mb/s down to 2 Mb/s or less. That means clearer, smoother video images, video calls across standard Internet connections, and the ability to connect with multiple people simultaneously. Already, about a billion devices, including iPods, mobile phones, and other consumer devices, use the new standard to display broadcast TV, Blu-ray movies, Windows Media or QuickTime files, and YouTube videos.

Problem solved? Not quite. H.264/MPEG-4 AVC, even though it was originally intended for video telephony as well as for consumer devices, isn't quite good enough. Problem No. 1: It is simply a video-encoding format. It does not cover any of the other aspects of telephony, such as audio coding and all the protocols that define, for example, how the system tells the receiving equipment that it's getting a video call. Those protocols involve a vast array of different specifications.

Problem No. 2: Even in the area of video coding, this format is limited. For example, if even one or a few packets are dropped during a transmission—pretty common in today's networks—the result may be catastrophic. The image the user sees is completely garbled. It might look something like a wet painting that has been wiped by a hand. Or if software masks the smearing, the video instead freezes in not just one but multiple frames. These distortions can last a half second or more, and a half second seems quite long to a viewer.

Removing redundancies from video images is good for compression but bad for robustness. To understand the problem, consider Short Message Service texting: People use abbreviations and other conventions to remove superfluous characters. However, even one or two typos in one of these cryptic SMSs can alter its meaning or render the entire message incomprehensible. The upshot is that encoding and decoding technology needs to be more forgiving when data drops out of the transmission.

Problem No. 3: lack of interoperability. Skype, for example, uses a proprietary video-compression system; it was already out in beta when the earliest draft of the new standard was published. Apple, Google, and Microsoft all start with one established coding scheme or another—Apple's FaceTime uses H.264/MPEG-4 AVC, Google+ uses another alteration called scalable video coding (more on that later), and some Microsoft products are currently based on a third standard called VC-1—and then add their own nonstandard technologies for the system negotiation and transport protocols. Today, these systems can't talk to each other, although you can get a Skype app that runs on the iPhone. So at least for now, if you have any hopes of using video telephony regularly, you'll have to plan your calls carefully. If you're using an iPhone, you'll likely use FaceTime to call other iPhone users. But if you want to make a video call to an Android user, you'd likely text the person and suggest your friend open up a Skype app, switch yourself from FaceTime to Skype, and then make the call.

Your phone will not figure all this out automatically. But help is on its way. The companies that manufacture communications gear are trying to hash out their differences under the auspices of a burgeoning

alliance called the Unified Communications Interoperability Forum, which was founded by HP, LifeSize, Microsoft, and Polycom. Their plan is to work with standards organizations, companies, and government regulators until the currently disparate technologies evolve into products that can seamlessly communicate.

Along with better interoperability, making video calls as common as text messages will require something else: the ability to talk to at least a small group of people at the same time. People take this for granted in the voice realm—consider the success of three-way calling and conference-calling services. To date, though, having more than two people participate in one video call strains existing video-telephony systems, creating unacceptable delays. Mixing people communicating using high quality Ethernet connections in an office, say, with participants using a hotel's strained wireless network means that everyone on the call must suffer the low resolution or jerkiness of the hotel user's connection; the lowest common denominator typically prevails. The same thing happens when some participants are using devices that are "smarter" than other devices in the same call: a high-powered laptop versus a low-end phone, for example.

So one item on the wish list is the ability to see different callers in a group at different resolutions, rather than just at the worst one. Another is more flexibility in these multiperson calls—the ability to make the video image of one participant larger on your screen than others, for example, without requiring each device to open up a separate connection to every other device, which dramatically multiplies the demands on throughput and processing.

Toward that end, VCEG, MPEG, and the Unified Communications Interoperability Forum have been working on the interoperability issue. VCEG and MPEG have jointly developed a new standard technology, scalable video coding, or SVC, publishing it in November 2007. The SVC design is an extension of the H.264/MPEG-4 AVC standard, not an entirely new scheme, so it is relatively easy for people who use the base standard to enhance their products to also support SVC.

SVC specifies what a bit stream has to look like to be read by all devices following the standard, and how the decoder of those devices translates that bit stream into images. The SVC technology gives the devices that use it all sorts of options in terms of video quality, including different resolutions and frame rates, by allocating one section of the bit stream to the lowest-quality options—sort of like a short text message summarizing a longer message. Devices having minimal processing power or communicating over low bandwidth can grab this small set of bits and ignore the rest. More sophisticated devices with faster network connections can also pay attention to bits that enhance quality, using these to display smoother, higher-resolution images. As in the text message analogy, the bits are adding details to the basic information that's in the short text message; however, the recipient can get the gist of the message without those details.

There is a small price to pay for breaking up the video signal in this way. Were the system to select a level of video quality and encode it separately, it might use about 10 percent fewer bits than it takes to use SVC—that is, including all the possible levels of video quality in the transmission and letting the receiving hardware do the selection. However, the 10 percent overhead is worth it, because SVC can make multiple video quality options available at once. This enables true device interoperability—people at big-screen computers using blazingly fast connections can participate in a video call involving someone using a smartphone in a hotel room without giving up the large HD images of the rest of the callers.

With SVC, users can also selectively size the video images on their receiving devices, choosing to see coworkers as smaller images and saving bandwidth to make the boss bigger and clearer—or the other way around.

The SVC approach also makes it easier to guard against those transmission errors that cause the annoying video glitches, because it doesn't take a lot of extra data to protect the small subset of data that encodes the lowest-quality images. Using the SMS example, the system could easily make sure that the SMS summarizing the larger text arrives error free, by transmitting another copy of the original summary or by supplementing it with a mathematical check of its accuracy. This would be harder to do for the larger message. In video telephony, protecting the "summary" means that receivers will almost always be able to display at least this most basic video image, no matter how bad the Internet connection; instead of smears and freezes, network glitches will only mean the image resolution will drop briefly, a far less vexing effect.

Companies that make corporate videoconferencing equipment have quickly embraced SVC. Vidyo pioneered some of the first SVC-based systems before it was fully clear to many in the industry that SVC could be especially useful in this application; that company and its partners, including Google, Hitachi, Ricoh, and others, have since adopted SVC technology in a number of products. Vidyo's implementation of SVC is also behind Hangouts, the multipoint video-chat system in Google+. Other companies, including Microsoft, Polycom, and Radvision, have also introduced videoconferencing systems based on SVC or announced plans to do so.

If companies that make consumer devices follow these business-equipment manufacturers, soon every device will indeed be able to talk to every other device. That could possibly happen in two or three years.

The use of video telephony on mobile gadgets does face other obstacles besides standardization, like ambient noise, poor lighting, battery drain, and strained cellular network capabilities. And, to be realistic, when you're in a crowded and noisy place, you're unlikely to find a phone's video feature useful. However, when you're in a coffee shop or a hotel room, video capabilities can significantly add to a conversation. This is even more the case with the larger tablets, which with scalable technology will be able to take advantage of more features than smartphones can—such as higher resolutions or making several people visible—even on the same call.

And the initial awkwardness felt by video callers in a public place will quickly fade. When the telephone came into the home more than a century ago, people feared everyone would be able to listen in on their private conversations; today, people chatter on cellphones in public with abandon. Indeed, people may soon forget that phones were once meant to be listened to, not watched.

By the way, "The Jetsons" was set in 2062. We're way ahead of them. □

 **POST YOUR COMMENTS**
online at <http://spectrum.itee.org/videophone0911>