ABSTRACT

Today, Internet protocol (IP) based videophones are gaining faster market penetration because of the enhanced capabilities made possible through efficient integration of International Telecommunication Union’s H.264 video compression technology as well as network bandwidth utilization. While audio and video quality are the primary concerns for most users, design engineers must also keep in mind price, ease of use, latency, error resilience, interoperability and compatibility. This paper provides an overview of the market trends driving the demand for video telephony and the components and system considerations for building IP videophones.

Contents

1 Introduction .................................................................................................................. 1
2 IP Video Telephony: An Emerging Market ................................................................. 2
3 A Clarification of Terms ............................................................................................. 3
4 Efficient Use of the Bandwidth You Do Have ............................................................ 3
5 Basic Videophone Components ................................................................................ 4
6 Software Reuse ........................................................................................................... 6
7 Managing the Flow of Digital Media Content ............................................................. 7
8 Integrating the System Components .......................................................................... 9

List of Figures

1 Typical System Block Diagram of IP Videophone System ...................................... 4
2 Videophone Software Architecture ........................................................................... 7
3 the Near-End Clock to Track the Far-End Clock ....................................................... 9

Trademarks

TMS320C64x, C64x are trademarks of Texas Instruments.

1 Introduction

IP videophones refer to video client endpoints, which are terminals that can be used to make point-to-point video calls. It typically consists of a camera/LCD display and a base unit that executes video and voice compression algorithms to reduce network bandwidth, and transmits the stream via IP or ISDN networks. Video conferencing infrastructure systems include multipoint conferencing units (MCUs) which allow the connection of three or more video sessions in the same conference. The MCU manages the audio and video from each participant and enables live group communication sessions between all parties. This paper will focus specifically on IP videophones and the considerations a developer must take into account when creating such systems.
Early market drivers for videophones in business applications were the availability of broadband. Without sufficient bandwidth, the user experience is significantly hampered. Either the frame rate drops too low or the image quality is blocky and without detail. The expense and reliability of connectivity, however, has limited the adoption of video telephony primarily to the business and enterprise markets.

Typically, plain old telephone service (POTS) -based videophones are bandwidth-limited. Any opportunity to reduce the amount of data sent is essential to providing enough quality for the user. Even with sufficient processing resources – which raise the price of the videophone out of the consumer market – POTS connections are still limited in resolution. Alternatively, one of the key drivers for IP videophones in the consumer market is the almost ubiquitous availability of broadband connectivity to the WAN.

Broadband Internet access, either through DSL, cable modem, WiFi hot spots, or even emerging standards such as WiMax, is an essential ingredient for IP videophones. Quality is directly proportional to bandwidth and the quality of service (QoS) available within that bandwidth. Broadband data rates push the user experience of video telephony over the critical hurdle of acceptable quality and cost.

However, bandwidth is not the only technological consideration; if it were, videophones would have reached critical mass in parallel with DSL. Other key technologies that have to be place include QoS support within the WAN capable of supporting video, more efficient video codecs, reliable transport protocols, effective A/V synchronization and integrated silicon capable of bringing the cost of processing resources under control.

2 IP Video Telephony: An Emerging Market

Video-based communications is gaining increasing acceptance in the marketplace. According to Wainhouse Research, the group videoconferencing market is forecasted to grow from about $530 million in 2003 to just under $1.1 billion in 2008 – a compounded growth rate of about 15.5 percent.

Key applications are expected to be IP videophones running SIP or H.323 that work with IP PBXs/gateway. In this way, these IP videophones will be able to take advantage of the call processing capabilities that these servers already provide, keeping appliance cost down and simplifying design. These new appliances will quickly dominate the personal conferencing hardware market. The anticipated growth of this market segment is projected from $21 million from 2003 to almost $180 million in 2008, achieving a compounded growth rate of almost 53 percent.

One of the major barriers to IP videophone acceptance is the understanding of the value of video telephony. The capability to make "free" person-to-person calls face-to-face to anywhere in the world meets resistance for many behavioral and psychological reasons, such as familiarity with and comfort in using the technology. Fortunately, with a large display, IP videophones have a natural interface that lends itself to simplifying use of what would otherwise be complex functions. Additionally, with a direct connection to an IP PBX that provides all the key features users have come to expect from POTS-based voice phones, videophones become a superset of the standard telephone, offering the same functionality but more.

User interface design will be key in achieving a successful out-of-box experience for users. The additional functionality videophones offer cannot increase the complexity of using the appliance. As such, it stands as one of the more important differentiating features that designers bring to a design. Phone operation is as simple as plugging in the handset and dialing. Connecting a videophone to the IP PBX/gateway, whether it is collocated within the corporate network or connected to over a SOHO DSL line, needs to be as simple.

Another approach for overcoming user resistance is to offer high-value functionality beyond standard telephony services. Features such as instant messaging and presence/availability management can themselves attract users, enticing them to experiment with the video aspect of telephony as a consequence. Access to e-mail and/or the Internet via browser can provide the same high-value appeal as well.
3 A Clarification of Terms

There are important differences between videophone and videoconferencing applications. Specifically, videoconferencing systems typically involve multiple locations or sessions instead of just two or three. Such a system must be able to broadcast to multiple sites but also accept multiple video feeds simultaneously. The challenge involved with handling multiple video and audio streams is that they involve a completely different set of design constraints. The processing power and bandwidth required to meet these challenges limits the market for videoconferencing equipment, making it a high-end technology.

Videophone systems, in contrast, are targeted for the mass consumer market. To be viable within this market, videophones support a lower resolution than videoconferencing systems. CIF is a commonly deployed resolution, offering a visual field of 352 x 288. Additionally, videophones must provide sufficient quality within a broadband connection. While ADSL promises data rates over 1 Mbps, most users are too far from the central office (CO) to achieve this level of bandwidth. To increase market penetration, a videophone should be able to operate within bandwidth on the order of 128 kbps. For either cable or ADSL, the upstream speed, which is typically in the range of 128-384 kbps, is the limitation factor, rather than the downstream speed.

Videophone systems are point-to-point devices, that is the videophone generates and receives a single video and audio stream. However, there is the consideration of how to handle call waiting. While the videophone may optionally process a second audio/video feed, this is not the same as a videoconferencing application. There is only one display, limited in resolution. Either the user must switch to the second stream – and stop processing the first stream – or the second feed is at a much lower resolution and is displayed in a corner of the display or in split screen mode.

4 Efficient Use of the Bandwidth You Do Have

Compression is a key technology for enabling videophones. For example, an uncompressed video stream at CIF resolution (352 x 288, 4:2:0 at 30 frames/seconds) requires more than 36.5 Mbps bandwidth. The H.263 standard was developed to offer higher quality than existing compression algorithms at even lower bit rates to enable SQCIF resolution over ordinary telephone lines or narrow band ISDN lines. However, it is the introduction of H.264 that is truly exciting for videophone applications. H.264 video codec technology enables VHS-quality video at about 600 Kbps, enabling high-quality video delivery over DSL lines. For CIF resolution video, H.264 typically achieves a 2X reduction in bandwidth over MPEG-2 and MPEG-4 in a majority of test cases, making it a serious contender to the entrenched MPEG incumbents. As a consequence, H.264 provides higher compression while maintaining low delay, critical factors for the videophone environment.

H.264 Baseline Profile, however, requires significantly more processing resources than the previous H.263 standard. One of the many innovations behind H.264 is improvement in the motion estimation algorithm. By working with a combination of 41 motion vectors, the algorithm finds a more ideal means of encapsulating movement within a frame. Computing these motion vectors and combing through the frame for their optimal placement, however, is not a trivial task and requires serious compute resources to achieve real-time encoding.
5 Basic Videophone Components

The basic hardware components of a videophone system are fairly straightforward (see Figure 1: Block Diagram of Basic IP Videophone System).

- An LCD displays both incoming video and serves as the user interface as well.
- A CMOS or CCD camera captures video and a video decoder converts the signal to a digital representation for further processing.
- A microphone captures audio and a headset or speaker plays back received audio.
- The appliance requires Ethernet connectivity and a healthy amount of SDRAM and Flash.
- The heart of the system is the digital media processor.

Figure 1. Typical System Block Diagram of IP Videophone System

The primary difference between low- and high-end videophones is the quality of the peripherals, implemented feature functionality, and flexibility of interoperability. There are significant cost variances across low- and high-end peripherals, including the display, camera, headset/speaker, and microphone. Low-end videophones will offer smaller screens with lower resolution than high-end appliances. High-end appliances will also offer optional feature-rich peripherals such as Compact Flash slots or USB (which enables connectivity features such as connecting to a PDA to transfer phone numbers or to connect to the Internet via the videophone network connection). User input can either be via keypad (low-end) or touchscreen (high-end). Available memory defines frame buffer capacity as well as how much room there is for additional applications and user data (i.e., email or browsing web pages).

Higher integration of peripherals, or at least their drivers/controllers, lowers SC system Bill of Materials (BOM). Vendors offering digital media processors typically offer a family of pin- and software-compatible devices addressing a wide range of applications. Thus the same base hardware design can serve across the low- to high-end, from a processing perspective, by upgrading the digital media processor. Memory capacity for high-end applications can be increased in a similar fashion by populating slots left empty for low-end applications.
Often, the cost economies of maintaining a single modular design outweigh the cost optimizations of a series of individual designs. Software device drivers; therefore, need to be able to support the range of peripherals you might support; for example, a higher quality microphone may require a higher quality A/D codec to capture the higher quality. Chip manufacturers track what OEMs require in higher end devices, and adjust the peripherals with these needs in mind. Systems integration is essential for reducing appliance cost: fewer individual components reduces packaging, power consumption, glue logic, board space, and the need for time consuming design integration, all factors that reduce overall system cost. The videophone market has already reached the integration phase, and integration will help drive the emerging market forward.

Texas Instruments, for example, offers the DSP-based DM64x family of digital media processors, which is based on the architecture of TI’s TMS320C64x™ generation of digital signal processors (DSPs). This full suite of digital media processors is code-compatible with TI's other C64x™DSPs and can interface with up to six simultaneous 8-bit BT 656-compliant video streams. The DM643, DM642, DM641 and DM640 processors offer on-chip integrated video ports, glueless Ethernet and multichannel audio. The processors are capable of performing up to four simultaneous MPEG-2 main-profile-at-main-level (MP@ML) video decodes at full D1 (720 x 480) resolution in real time. The suite of devices is also capable of full video encoding in real time and offers code support for broadcast-quality Windows Media 9 encode and decode technology. In addition, these devices support the latest industry standard algorithms including MPEG-4 AVC (H.264) decode. The suite of processors enables the developer to select the appropriate product for the level of system performance required by the targeted end equipment, and the integration of video-specific peripherals allows the developers to further reduce the cost of the system. The DM643 digital media processor is well suited for IP-based videophone designs with the right mix of peripherals and system integration.
6 Software Reuse

There is also much differentiation possible at the software level. However, with a common, base hardware platform design, software has less impact on overall system cost than the quality of peripherals selected. This is because high-end software features are code-compatible with low-end processors. What this means is that designers can migrate high-end functionality down to low-end devices to differentiate their products over time as the market matures without incurring the significant cost increase that comes with changing hardware.

Primary software components include networking protocol and communication stacks, audio and video codecs, data processing and user applications (see Figure 2: Videophone Software Architecture). Currently there are two protocols vying for dominance in the video telephony space: H.323, which is the primary protocol used in existing videoconferencing applications, and SIP, which is the primary protocol used for VoIP.

To some degree, the technical merits of H.323 and SIP are irrelevant since this is a battle that will decided by the carriers. OEMs building videophones will need to support both if they want to compete. Interoperability will be a key enabler for the videophone market; if the users can not talk to each other using equipment from different vendors, this will stall the market to everyone's detriment.

The same issue arises from a video codec standpoint. H.263 is the current dominant video codec, but H.264 video codec offers too many advantages to ignore. Given the relative low load of the audio stream, another key differentiating feature will be the ability to support high quality voice codecs that are not required within the H.323 or SIP specs. MPEG codecs also retain a certain marketshare, so maintaining interoperability again becomes important. Additionally, videophones can be made more attractive if they offer enhanced functionality such as the ability to view photos, live streaming of digital media content or play audio files.

Pre- and post-processing of video and audio streams is another primary software requirement. Some of these features are optional, such as enhancing video before compressing them. Others are optional in how you implement them; consider the length of an AEC (acoustic echo cancellation) tail. Optional processing consumes limited processing resources but can yield a significant increase in perceived quality. Other types of processing may be absolute necessary, such as converting a camera sensor output from Bayer pattern RGB to YCbCr 4:2:0.

User applications are also an important part of a videophone system. For example, users need to be able to manage phone numbers in a relatively easy fashion. Synchronizing numbers across videophones, PCs and PDAs would present high-value to business users with large contact databases. Designers can choose to create new applications that run native on the videophone, use off-the-shelf software from third parties that specialize in application software, or even manage data remotely using a web server architecture.

Ideally, software components need to be able to be dynamically implemented, for a number of reasons. For example, if the network connection suddenly becomes constrained, the videophone needs to be able to adjust processing to reduce the bit rate, either by changing the codecs, frame rate, or compression ratio. Alternatively, these components can be optionally implemented depending upon whether the user has purchased a device with low- or high-end peripherals (the digital media processor may need to scale the output of different cameras/CMOS sensors to match the required resolution), or remotely upgrade the videophone to unlock the use of advanced features, introducing an aftermarket revenue stream.

Programmability of the digital media processor is key for maintaining device flexibility and interoperability. The more codecs a videophone can support, the more value it has to its consumer because there will be fewer times that a connection cannot be resolved because of lack of interoperability. Also, existing designs can age more gracefully if they can easily interface with different peripherals over time without a redesign. As important is the ability to update equipment already deployed. As with any emerging industry, sticky interoperability issues can arise over time. Even more likely is the introduction of new and improved codecs; consider the rapid evolution of Microsoft's Windows Media technology codecs. Devices that implement codecs in software can evolve with these changes. Full software programmability gives designers the option to avoid having to guess what the future holds by giving them the ability to meet the changing needs of the market however they manifest.
Managing the Flow of Digital Media Content

The key programming challenge is managing the flow of digital media content through the IP videophone. With the exception of user interface and application-level functions, which do not require a real-time response, all other processing within a videophone involves time-sensitive data. Managing these flows so that real-time deadlines are met is essential. Latency is the problem here, and must be balanced between sending and receiving streams; take too long to capture, process, and transfer the audio and video streams and you increase the chance that the data will arrive too late to be useful.

There are five primary data streams in a typical videophone application. The transmit video stream (1) is captured by the camera and distributed to both the display and network. The incoming receive video stream (2) arrives over the network and is sent to the display as well. The processor must resolve how both streams appear on the screen (the transmit stream is typically picture-in-picture within the receive stream). The transmit audio path (3) has two distribution paths as well (speaker and network), and must be resolved with the incoming receive audio stream (4). Finally, a control stream (5) based on H.323 or SIP manages the connection. The application must also be able to simultaneously manage the user interface and execute any user applications.

Video and audio are time-sensitive streams. The human ear is capable of noticing a delay of 200 ms; in other words, if it takes longer than 200 ms for a voice packet to reach its destination, the listener will be able to discern the delay. Latency requirements for video are much more relaxed. In any case, latencies in the audio stream tend to be more noticeable to users than latency in the video stream. One of the reasons for this is that unless the scene has changed drastically, the pixels from the last frame are close in proximity to their position in the new frame. Recovering from a lost video packet is fairly straightforward: use the pixels from the last good image and the user probably won’t be able to tell the difference unless he or she is actively focused on that part of the display. If consecutive packets are lost or corrupted, they represent a different part of the display, so even a long run of lost packets is not overly troublesome.
Audio packets, on the other hand, are much more fragile than video packets. Each packet is completely different so a lost packet can be masked by repeating the last packet but not ideally. The loss of even a few consecutive audio packets can be devastating in that there is no way to prevent the user from hearing the loss.

For these reasons, audio streams typically need to be treated with a higher priority than video streams, in addition to the fact that more relevant information is transferred over audio than video. It is important that the IP network support multiple priorities for real-time digital media content to prevent video and audio packets from contending with each other.

It is also important to note that audio quality is almost a secondary consideration. The bit rate of a voice codec represents only a small percentage of the total bandwidth requirements of a videophone. Depending upon the application, it may be more cost effective to use a higher bandwidth codec like G.711 than a low bit rate codec like G.728; G.728 is a more processor-intensive codec than G.711 so the bandwidth savings of using G.728 are worth much less than the cost of additional processing resources required to support G.728.

One of the biggest challenges for early video telephony appliances has been audio/video synchronization (see Figure 3: Adjusting the Near-End Clock to Track the Far-End Clock). Since audio has a higher priority than video, it must be transferred in its own stream. Separating the audio and video creates the problem of putting them back together. A/V sync is one of the most noticeable components of the user experience: when lip movement is not synchronized with the audio sound, the effect is very disconcerting. There are three components to A/V sync. The first involves synchronization of local audio and video clock at both near and far end to prevent the A/V clocks drifting from each other over time. This aspect of A/V can be mostly resolved by splitting a single source clock with PLLs (Phased Lock Loop) to clock the two streams.

The second aspect of A/V sync concerns the dynamic nature of streaming data. For example, if the network is constrained, the far-end (transmit-side) videophone may reduce its frame rate. In order to place the frame in the right time slot, the H.323 or SIP protocols employ timestamping. With timestamps, the digital media processor marks the time when a frame or voice packet is captured and places the timestamp in the packet header. The near-end (receive-side) videophone parses the header to retrieve the timestamp, which tells the videophone when to use the data.

The third element of A/V sync involves synchronization between two videophones and is primarily a clocking issue. The near-end and far-end clocks may operate at slightly different rates. If the far-end clock is faster, the A/V buffers will begin to overflow since the near-end appliance plays back data at a slower rate than it receives the data; likewise, if the far-end clock is slower, the A/V buffers will underflow. Both cases are a problem.

One solution is to increase the size of the A/V buffers. Unfortunately, given the sensitivity of human perception, these buffers cannot be very large or users will notice the delay. The preferred solution is to adjust the A/V clock source and requires the ability to adjust the clock in hardware. In this way, the videophone can play back data at a different rate than it captures it. The near-end videophone does this by monitoring the level of data it has buffered. As the buffers fill, the videophone adjusts its clock to run faster to run the buffers down. When the buffers are low, it adjusts the clock to run slower to allow the buffers to fill. In this way, the near-end videophone is able to estimate the clock rate of the far-end videophone, as well as accommodate, to some degree, network latencies.
8 Integrating the System Components

When you consider all the functionality and applications an IP videophone must support under the hood, there’s quite a bit going on, especially with all the real-time deadlines that must be managed. A robust embedded operating system can provide a great deal of the underlying framework required to manage the multiple tasks running concurrently such as managing the various A/V and control streams, minimizing latency, handling interrupts, and so on. Integrated features such as a TCP/IP networking stack provide a reliable operating environment upon which to build application software and reduce time-to-market considerably.

The design of IP videophones used to be a matter of deciding what balance of resolution, frame rate, image quality, and bit rate yielded the best user experience. With the wide availability of broadband connectivity, highly integrated SoC digital media processors, and off-the-shelf audio and video codecs optimized for performance, achieving sufficient video and audio quality is no longer a concern; IP videophones are viable and cost-effective to produce. Instead, designers are challenged to further reduce cost or decide where they want to exceed quality expectations first to differentiate their products from their competitors.

With the feasibility of IP videophones no longer in doubt, the video telephony market is on the exciting edge of expanding in ways no one has imagined. Initial deployments will necessarily focus on the basics to keep costs down and encourage mass adoption. However, it will not be long before the IP videophone becomes another appliance as common as the POTS voice-phones.
IMPORTANT NOTICE

Texas Instruments Incorporated and its subsidiaries (TI) reserve the right to make corrections, modifications, enhancements, improvements, and other changes to its products and services at any time and to discontinue any product or service without notice. Customers should obtain the latest relevant information before placing orders and should verify that such information is current and complete. All products are sold subject to TI's terms and conditions of sale supplied at the time of order acknowledgment.

TI warrants performance of its hardware products to the specifications applicable at the time of sale in accordance with TI's standard warranty. Testing and other quality control techniques are used to the extent TI deems necessary to support this warranty. Except where mandated by government requirements, testing of all parameters of each product is not necessarily performed.

TI assumes no liability for applications assistance or customer product design. Customers are responsible for their products and applications using TI components. To minimize the risks associated with customer products and applications, customers should provide adequate design and operating safeguards.

TI does not warrant or represent that any license, either express or implied, is granted under any TI patent right, copyright, mask work right, or other TI intellectual property right relating to any combination, machine, or process in which TI products or services are used. Information published by TI regarding third-party products or services does not constitute a license from TI to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property of the third party, or a license from TI under the patents or other intellectual property of TI.

Reproduction of information in TI data books or data sheets is permissible only if reproduction is without alteration and is accompanied by all associated warranties, conditions, limitations, and notices. Reproduction of this information with alteration is an unfair and deceptive business practice. TI is not responsible or liable for such altered documentation.

Resale of TI products or services with statements different from or beyond the parameters stated by TI for that product or service voids all express and any implied warranties for the associated TI product or service and is an unfair and deceptive business practice. TI is not responsible or liable for any such statements.

Following are URLs where you can obtain information on other Texas Instruments products and application solutions:

<table>
<thead>
<tr>
<th>Products</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplifiers</td>
<td>Audio</td>
</tr>
<tr>
<td>Data Converters</td>
<td>Automotive</td>
</tr>
<tr>
<td>DSP</td>
<td>Broadband</td>
</tr>
<tr>
<td>Interface</td>
<td>Digital Control</td>
</tr>
<tr>
<td>Logic</td>
<td>Military</td>
</tr>
<tr>
<td>Power Mgmt</td>
<td>Optical Networking</td>
</tr>
<tr>
<td>Microcontrollers</td>
<td>Security</td>
</tr>
<tr>
<td></td>
<td>Telephony</td>
</tr>
<tr>
<td></td>
<td>Video &amp; Imaging</td>
</tr>
<tr>
<td></td>
<td>Wireless</td>
</tr>
</tbody>
</table>

Mailing Address: Texas Instruments
Post Office Box 655303 Dallas, Texas 75265

Copyright © 2005, Texas Instruments Incorporated