



What does it take to get
Good video telephony

Author:	Paul Sijben, EemValley Technology	Sijben@eemvalley.com
Date:	06-Jan-06	
Ref. Nr.	EV2005-99/EN	
For:	Cyburg, the Netherlands	

Executive Summary

Video telephony is one of the most appealing possibilities promised by broadband Internet. There are huge advantages to being in a telephone call and being able to see the other person. As has been known for quite some time, the nuances of face, body and arm gestures add a wealth of information to communication. Good video telephony adds quality to telecommunication that reduces the need for people to physically travel to meetings. Good video telephony can prevent the elderly and less mobile from becoming isolated. For video telephony to be widely accepted it needs to be easy to use, provide sufficient quality and be affordable.

Unfortunately good video telephony is not universally available. Today, video telephony usually implies that a group of people gather around a table and watch a TV showing a similar group of people around another table. Personal video telephony usually means watching postage stamp sized people in a PC screen, whose image is refreshed occasionally.

In this report we address the question; when a country like the Netherlands is 2nd in the world in broadband penetration and if video telephony is one of the most appealing broadband applications, why is good video telephony so hard to come by?

Many of today's videophones are difficult to use or privacy invasive. Adding a traditional receiver and some buttons familiar from regular telephones may address both of these issues. For a usability standpoint we would like to see the videophone as a standalone device with a telephone-like handset. A few buttons will allow the caller to "dial" the called and a simple button will allow calls to be accepted and terminated. A few buttons for starting and stopping the video streams and to mute the sound will enable an affordable videophone for anyone who can handle a regular telephone.

In this report we have shown that the state of the art is mostly determined by bandwidth limitations. However, broadband Internet is making huge strides forward. The introduction of *symmetrical* Internet with high *guaranteed speeds* offers opportunities for video telephony, finally there is the opportunity for good image quality at acceptable costs. The approach identified in this report will work with *symmetrical* bandwidth of 5-10Mb/s and will deliver an image quality comparable to that of VHS video with acceptable delays.

If there is the will and the bandwidth, nothing will stand in the way of general use of video telephony.

About the Author

Paul Sijben (36) is founder and chief technologist of EemValley Technology. In 1993 he graduated from the University of Twente, the Netherlands in computer science and continued there to do research on multimedia operating systems. In 1997 he joined Lucent Bell-Labs in the Netherlands and left in 2002 as a Distinguished Member of Technical Staff working on Voice over IP research and standards (accomplishments include substantial work on specification of the TIPHON architecture and the H.248/MEGACO protocol). In 2002 Paul became CTO of PicoPoint, a WiFi hotspot back office and roaming pioneer. In 2004 he left PicoPoint to found EemValley Technology to create the next generation of open and secure telecommunication infrastructure and services. He offers his knowledge and expertise to advanced networking and video telephony projects. Paul is currently also active in the ETSI TISPAN Specialist Task Forces on next generation services and security.

Paul's can be reached at Sijben@eemvalley.com.

Table of Contents

Executive Summary	3
About the Author	5
Table of Contents.....	7
1 Introduction	9
1.1 Existing video telephony applications	9
1.1.1 MSN Messenger.....	9
1.1.2 D-link DVC-1000.....	10
1.1.3 D-link DVC-2000.....	11
1.1.4 Polycom V500	11
1.2 Video telephony and bandwidth.....	11
1.3 Issues with video telephony today.....	12
2 Properties of good video telephony	13
2.1 Introduction.....	13
2.2 Image quality.....	13
2.3 Usability.....	14
2.3.1 Professional use	14
2.3.2 Private use.....	15
2.3.3 Domestic Privacy.....	15
2.3.4 User requirements	16
3 Anatomy of video telephony.....	17
3.1 Introduction.....	17
3.2 Bandwidth.....	17
3.3 Delay	19
3.3.1 Delay at capture	20
3.3.2 Network delay	20
3.3.3 Delays in processing the video.....	21
3.4 Streaming	21
4 Towards good video telephony	23
4.1 Introduction.....	23
4.2 Unlimited bandwidth.....	23
4.3 Conclusion	23
Annex A Camera Measurement	25
A.1 Introduction.....	25
A.2 Test setting	25
A.3 Base measurement.....	25
A.4 The secrets behind webcams.....	26
A.5 Improving webcam behavior	27
A.6 New technology; not faster.....	27

1 Introduction

Video telephony is one of the most appealing possibilities promised by broadband Internet. There are huge advantages to being in a telephone call and being able to see the other person. As has been known for quite some time, the nuances of face, body and arm gestures add a wealth of information to communication. Good video telephony adds quality to telecommunication that reduces the need for people to physically travel to meetings. Good video telephony can prevent the elderly and less mobile from becoming isolated.

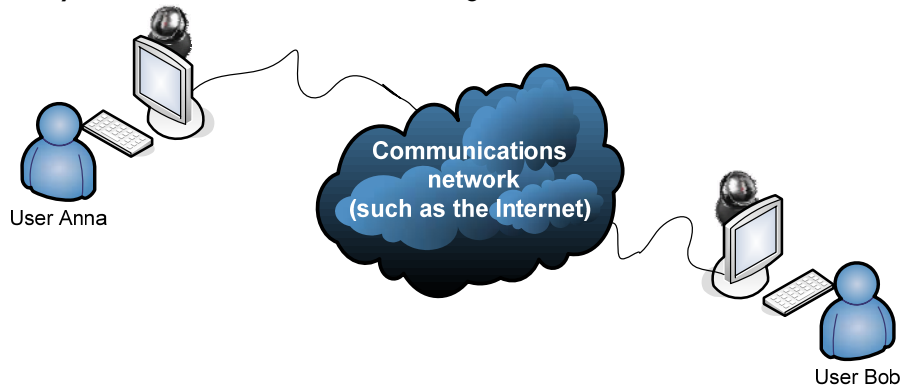


Figure 1. Video telephony

Unfortunately good video telephony is not universally available. Today, video telephony usually implies that a group of people gather around a table and watch a TV showing a similar group of people around another table. Personal video telephony usually means watching postage stamp sized people in a PC screen, whose image is refreshed occasionally, such as in Figure 2.

In this report we address the question; when a country like the Netherlands is the 2nd in the world in broadband penetration¹ and if video telephony is one of the most appealing broadband applications, why is good video telephony so hard to come by?

1.1 Existing video telephony applications

Video telephony applications have been available for years. The advent of broadband Internet and cheap webcams seems to make this application available to all. In this section we provide a number of examples. Others exist, such as video Skype.

1.1.1 MSN Messenger

MSN messenger is one of the most popular communications applications, in part this is because Microsoft offers the service for free. Figure 2 shows a screenshot of the application.

¹ In the Netherlands there are currently 3.8 million broadband connections in use out of a population of 16 million, a penetration second only to South Korea. "Basic" broadband connections are advertised with a bandwidth of 2-6 MB/s seconds downstream and 768 kb/s upstream.

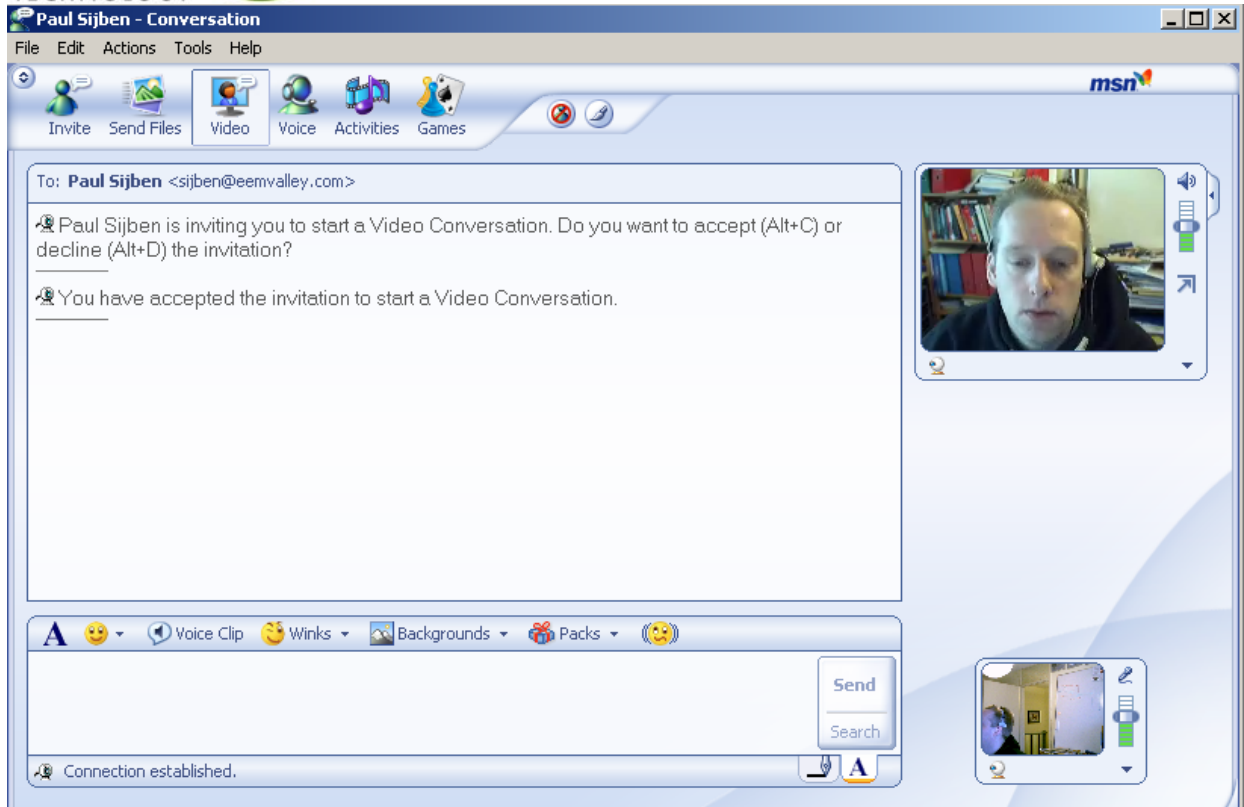


Figure 2. MSN messenger screen shot

We observe the following in using the application:

- **Small images.** The image of the other speaker is only a fraction of the screen area reserved for chatting and other actions. Setting the image to large screen (an option available with MSN messenger but not with Windows Messenger) stretches the image but this often makes the image look worse.
- **Cumbersome user interface.** It is easy to start a call with a known contact, starting a call with a stranger is difficult. It is not immediately apparent how to close the call.
- **Headsets.** Background noise and feedback loops reduce the sound quality. The MSN application has an echo suppression scheme but this is not sufficient. The use of a headset with microphone helps but does not look appealing.

1.1.2 D-link DVC-1000

D-link is a manufacturer of cheap consumer electronics. The DVC-1000 i2eye VideoPhone is D-link's first step in the world of video telephony. The system is advertised as easy to use but has problems with bad lighting conditions and no movable camera, this means that the speaker has to stay in one place in front of the screen. This device is easy to use, it offers a large image and is as easy to use as a regular telephone.



Figure 3. D-link i2eye

1.1.3 D-link DVC-2000

The D-Link DVC-2000 is a desktop video telephone. Variants of this device are offered by other manufactures. The device is very like a telephone in almost every respect but with the added bonus that one can see the other person. Unfortunately the small screen adds little as one needs to peer into the screen to see the other speaker.



Figure 4. D-link desktop videophone

1.1.4 Polycom V500

Polycom is well known for its hands free telephones and video high-quality telephony products for enterprises around the world. The V500 is a recent addition to its broadband portfolio. On the plus side this device offers a large screen, movable camera, Polycom's echo cancellation, so no headphones. On the downside this devices comes at a hefty price, \$1700 for the device without a screen.



Figure 5. Polycom V500

1.2 Video telephony and bandwidth

Video telephony requires much bandwidth. In the experiments shown in this report we used about 3 megabits for a one-way video stream. Although broadband Internet with megabit speeds is already advertised in some parts of the world does this not mean that those speeds are usable for video telephony.

Most broadband Internet offers today use ADSL². ADSL stands for Asymmetrical Digital Subscriber Line. The asymmetry of this technology is relevant for video telephony.

The providers offering this connection assume that the users of these lines will download more data than they will upload. This assumption is correct when the ADSL customers simply surf the web. In that case about 10 times as many bytes will be downloaded than uploaded. So the scarce bandwidth of the ADSL line is split in such a way that the download speed is much higher than the upload speed. In the Netherlands "basic" broadband connections are sold with a bandwidth of up to 6 MB/s seconds downstream and just 768 kb/s upstream.³

This assumption no longer holds with video telephony where both parties send roughly equal amounts of data. If both parties have a "basic" ADSL line their image quality is limited by the 768kb/s upload speed.⁴

² Broadband Internet via Cable companies uses a different technology from ADSL but has similar properties for our purposes.

³ ADSL2+ offers of 20MB/s and up generally do not lead to a significantly higher upstream bandwidth. 1MB/s is typically advertised. As the download/upload ratio is 10:1, this makes using 20Mb/s not feasible for web surfing.

⁴ And often less otherwise all bandwidth is take for other Internet applications.

The upload speed of 768 Kb/s is the theoretical bandwidth, the *effective* bandwidth may be much lower. ADSL lines are made cheap by selling the same bandwidth many times, this is called overbooking. Common overbooking factors in the Netherlands are 1:20 up to 1:40 where the same amount of Internet bandwidth is split over 20-40 households. This significantly lowers the effective bandwidth for video telephony when more than one user tries to communicate at the same time, thus reducing the quality that may be achieved.

1.3 Issues with video telephony today

The availability of broadband Internet has brought video telephony within the reach of households and smaller companies. Unfortunately the quality of the more affordable options is generally substandard. Besides, desktop applications are usually hard to use for people who do not like to fiddle with PC settings.

The effective bandwidth of broadband Internet is limited, so many of today's video telephony applications are built to this state of the art. This translates to small images with low refresh-rates. These settings are not changeable by the users, even if more bandwidth is available (such as on a high-speed company network) the quality is limited. Expensive enterprise video telephony devices use advanced encoding for the video and hence achieve better results, but these come at a significantly higher cost.

Bandwidth is the most common limiting factor for consumer video telephony.

New fiber optic networks remove this bandwidth limitation to video telephony quality. This report addresses the question what needs to happen to achieve cost-effective good quality video telephony assuming bandwidth is no longer a limiting factor.

We first address the minimum requirements for good video telephony. We next address the anatomy of a video telephony system and identify the current state of the art and which aspects of the system hamper the quality. We conclude this report with recommendations how good video telephony can be achieved with existing but perhaps not commonly used technology and networks.

2 Properties of good video telephony

2.1 Introduction

Good video telephony enables the participants to communicate naturally as if they were both in the same room. We subdivide this simple requirement into the following sub-requirements:

1. **Sound quality.** The quality of the sound must be as good as regular telephony or better (for example mp3 or even CD quality). The sound needs to be synchronized with the image.⁵ In practice the sound quality should be fine and only the synchronization should be a problem. We will not further discuss the sound in this report.
2. **Image quality.** The image quality must be comparable with real life, both in resolution and in refresh rate, otherwise the image will distract from the communication rather than enhance it. In this report we use the apparent resolution of VHS video as a lower margin. In Section 2.2 we will further address the video quality.
3. **Hardware and networks.** The endpoints (computers, videophones) and network between the sender and the receiver must transfer and process the video without too much delay and loss in quality. In Section 3 we will further address these technical issues.
4. **Usability.** The videophone must be usable by someone who can use today's television and (wired) phone. In Section 2.3 we will further address usability.

2.2 Image quality

The image quality of a videophone must be good enough that the participants will not be hampered by artifacts introduced by the technology. Current video telephony solutions too often deliver small images with low resolution or large images that come apart when the subject moves.

For the requirement on resolution we use a minimum value comparable with TV VHS quality image, however the larger the resolution the better. The effective resolution of a VHS recording is 352x240. This value has not been chosen arbitrarily as our lowest resolution. It is a quality familiar to the average user and which can be easily validated by anyone with a VCR. Lower resolutions quickly deliver a grainy image, especially when this is then enlarged to full size on a computer monitor or a TV screen.

For the dynamic behavior we again aim for the closest approximation of TV quality. Our eyes perceive separate images as a moving object when are refreshed at 24 times a second. Our eyes can become used to 12 frames a second for a while but this turns out to be tiring after a while.

Artifacts in the image (visible non-natural effects) when the subject moves⁶ quickly breaks the illusion that the person is in the same room. This shall be avoided. In practice this requirement will weigh heavier than the refresh rate.

⁵ Dolby Studios gives for movie theater films the requirement that the sound may not come ahead of the image by more than 5 milliseconds and not be behind the image by more than 15 milliseconds. Outside these margins the movie audience will be irritated by the lack of synchronization. We concur with these requirements.

⁶ Common in for instance the H.261 encoding in use in many of today's video telephony systems.

Delay in the communication may also detract. From the telephony world we know that a delay of 200ms per direction (one-way) is clearly noticeable and irritating because it complicates a natural conversation. In some cultures this threshold is much lower because of the speed at which people communicate. We use this 200ms as maximum acceptable delay with the remark that less delay is better.

We translate the requirements provided above to the following technical requirements:

- Minimal resolution of 320x240 pixels
- 16-24 bits color depth
- 20-30 image frames per second⁷
- 125-200ms delay one-way

2.3 Usability

The success of video telephony rides for a large part on usability. Usability is important in all applications, however in communications applications it is even more important as a difficult user interface may shame the user when they can not answer the call quickly or when the user has an audience when searching for the right way to perform a certain action in the call. Users should not be shamed by their use of the device! A good videophone is as easy to use as a desktop telephone today.

Video telephony is not common so we can not perform an empirical study to the uses of video telephony in daily practice. Firstly is it clear that in the stories where video telephony appear it is used just like a normal telephone. The current practice of someone wishing to make a video call, who goes up to a little room in the attic, after threats to the kids to stop their on-line games or they will be grounded⁸, after which some software updates need to be performed before the call can finally take place, is a clear hint of how not to do things.

We can find inspiration by how people who are being paid to look into the future, science fiction authors, describe video telephony. These people use future technology in their stories. A good author uses these elements as natural surroundings for the characters. This provides us with some insight how video telephony could take place in normal life.

2.3.1 Professional use

The popular TV-series Star Trek uses video telephony regularly. Usually this is used in the context of a functional conversation about work. When the admiral calls the captain during his sleep, the latter is first warned by a communications officer. This provides the recipient to dress and presentable before taking the call or, failing that, to take the call on audio only. While the conversation is going the parties can send each other information without breaking eye contact for long. They may conference in other people on-demand or put one person on hold while speaking to another.

⁷ Taking a higher refresh rate than 24 frames per second may reduce overall delays.

⁸ As it uses bandwidth that is needed for the video call.

Bruce Sterling uses in his book "Islands in the Net" video telephony in the professional context as well, mostly in the form of conference calls among the principals in geographically distributed company. Any person participating in such a video conference dresses up and applies make-up to appear their best. One exception is made where an personal important message needs to be delivered.

These two examples provide some insight how video telephony might be placed in the professional context. Users of this technology may use it everywhere, for business or private calls. This is completely unlike how video telephony is used today in businesses, people going to a special meeting room to watch a similar group on a large screen television.

The requirements for professional video telephony usability seems to lean towards it being an integral part of the regular workspace so people make a video call as easily as a telephone call or sending an email.

2.3.2 Private use

Sci-Fi authors usually place video telephony in a business setting. Fantasy apparently has less use for video calls in a domestic setting.

Experiments with video telephony in home-care, such as the CamCare project from Sensire, show how home-care customers are very happy to have this good means of communication with care-workers. Experiments at Cyburg have shown that people will be creative and come up with all kinds of uses of video telephony once it is available (keeping in contact while they do not venture out into the streets at night, playing games together, contacting their GP, kids etc.)

2.3.3 Domestic Privacy

It is clear that it is undesirable to have a direct video link into one's life when one is not ready to be seen by others (say, half-dressed). This would mean an unacceptable invasion of privacy. So a domestic videophone will even more than a traditional fixed telephone have to provide a clear indication of who is calling and based on that provide the opportunity to refuse the call or accept it as a voice-only call until one is ready to take the video as well.

A domestic videophone should, even more than an enterprise one, be separate from a desktop computer. Also privacy is even more important in a domestic setting than at work. The videophone places the participants virtually inside their home and they may not wish to share their home with a caller. A videophone shall therefore require some shielding from sounds and images from the rest of the home and not just put a camera and microphone into one's living room. This may be achieved by placing the videophone in a kind of cell like the telephones of old, or to provide them with a receiver like today's telephones, see for instance Figure 6.



Figure 6. Traditional receiver for a videophone?

Video calls are not limited to known contacts (such as in MSN messenger) but one should be able to call everyone. However one should take the example from current messaging-based systems to temporarily block certain groups of people from calling.

2.3.4 Usage requirements

We foresee a standalone device what may take the same place at today's telephone in the workplace and at home. Users will have the ability to call everyone and to be called by everyone they wish. The caller needs to be explicitly identified before the call is accepted or rejected. Accepted calls may be set to audio-only at first and video may be added in later.

The user interface is comparable with that of today's telephone offering a simple way to call known contacts from an address book and to dial anyone else by number or name. Calls may be started and stopped at the touch of a button.

3 Anatomy of video telephony

3.1 Introduction

In this section we make an inventory of the components of video telephony to understand why today we do not have good video telephony.

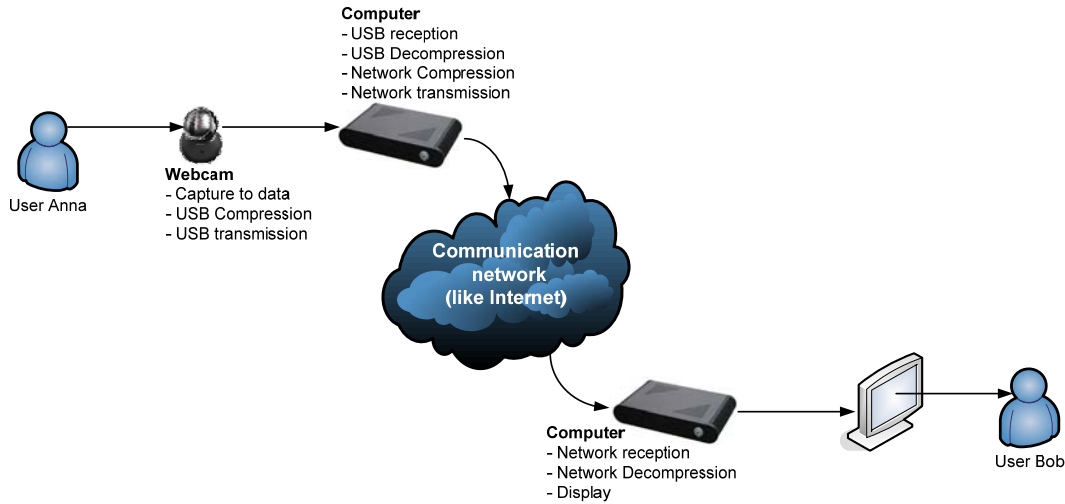


Figure 7. Anatomy of today's video telephony

Figure 7 shows the relevant components of today's video telephony. We pass through the complete flow from Anna's image to the Bob's eyes. Anna's image is recorded by her webcam. This camera will convert the optical image to a data stream which is passed via its USB connection. The bandwidth of the USB cable may be too limited to carry all the data so the images may be compressed to fit the available bandwidth before sending them. Anna's computer will decompress the images after reception of the data off the USB bus. Then the data will (again⁹) be compressed before the data is sent over the network. The network between the two computers will transport the video stream just like any other data. Bob's computer will receive the video data, will decompress the images and display them on Bob's screen.

Next we focus on aspects of video telephony we have just identified:

- The bandwidth necessary to transmit the images
- The delay in each of the steps.

3.2 Bandwidth

Bandwidth use in video telephony, just like with regular telephony, is symmetrical. Both parties send audio and video to the other. This symmetrical behavior differentiates video telephony from

⁹ Unfortunately the compression of the data is specific to Anna's type of webcam so this can not be used to send the video compressed to Bob.

surfing the web, listening to internet radio or watching web-TV. The latter applications are asymmetrical, a small request is followed by a lot of data from the server.

Video telephony generates enormous amounts of data. When we take the requirements for good quality video telephony from Section 2.2 we come to the amounts of data provided in Table 1, a good quality video stream takes from 23-53 Mb/s. These quantities are too much for today's networks. The data must be compressed to fit into the available bandwidth. Table 2 gives an overview of common video coding methods and the corresponding data rates.

		min	-	max
Resolution in pixels	320x240			76800 pixels
Bits per pixel	16-24	1228800	-	1843200
Kilobytes per frame		150		225
Frames per second	20-30	24576000	-	55296000
Kilobytes per second		3000		6750
Megabits per second		23		53

Table 1. Raw data quantities

Coding	Kilobits per sec.	Image frames per sec.	Resolution
ISO MPEG-4	72	15	176 x 144
	560	30	352 x 288
Microsoft MPEG-4	28	10	176 x 144 or less
	56	15	176 x 144 or less
	256	15	320 x 240
	512	30	320 x 240 (16 bits color)
	768	30	320 x 240 (24 bits color)
	128	10	176 x 144 or less
	256	15	176 x 144 or less
H.261	512	30	176 x 144 or less
	768	30	320 x 240 (16 bits color)
	1000	30	320 x 240 (24 bits color)
MPEG-1	512	24	352 x 240 (16 bits color)
	750	24	352 x 240 (24 bits color)
	1000	30	352 x 240
MPEG-2 Half D1	2000, 2500, 3000, of 3500	30	352 x 480
MPEG-2 Full D1	3000, 3500, 4000, 4500, 5000, 5500, of 6000	30	704 x 480

Table 2. Life video coding and their bandwidth (source: cisco.com)

H.261 and H.263 are popular coding methods for today's video telephony systems. Unfortunately their quality is limited. Due to their low resolution and image frame rate. Also rapid movements give artifacts in the image for a long time after the moment has happened. Their successor, H.264, is one of the MPEG-4 modes.

The table shows us that today we have a number of ways in which we can bring the flood of image data down to manageable quantities that can be transmitted over a network. A problem with coding video data is that it takes time and hence adds to delay, which we wanted to keep as low as possible.

3.3 Delay

We have identified that the delay from source to target should be less than 200ms and that lower delay is better. There are a number of factors that contribute to delay in video telephony. Figure 8 shows the operations on video images in video telephony solutions. This figure provides us with a framework in discussing the delays. The figure shows Anna as a video source through her webcam and Cary via a video camera and a video capture card in his computer.

The sources of delay may be summarized as:

- Video Capture
- Transmission in the network
- Video Coding/decoding
- Other manipulations of the video

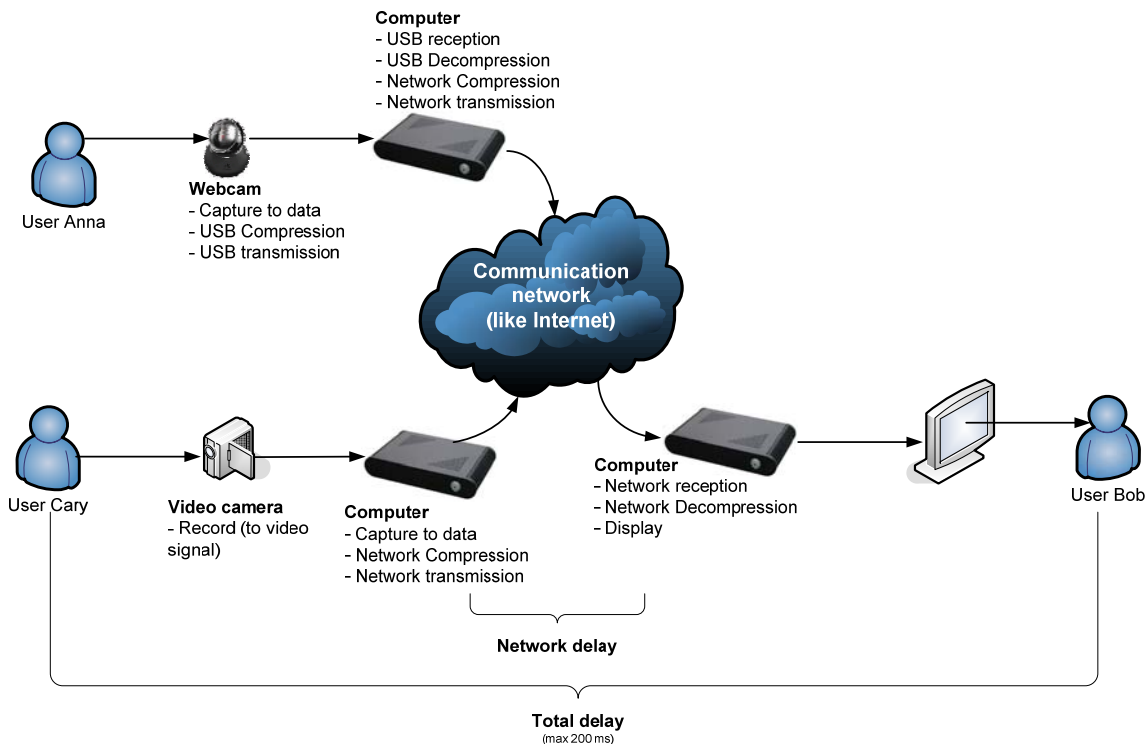


Figure 8. Delays in video telephony

3.3.1 Delay at capture

Recording moving images may be done via either a webcam or via a video camera with a hardware capture card. The last method is the most effective, the video camera converts the images into a video signal, which is turned into data by the capture card.

Figure 9 shows in the right hand side of the figure a millisecond clock which is recorded by a camera, on the left hand side of the figure the recorded image is displayed after it has passed through the computer. The difference between the two times as shown by the clock is the delay caused by capture, processing and displaying of the image.



Figure 9 Video camera delay, 70ms



Figure 10. Webcam delay, 200ms

Figure 10 shows the same setup for the webcam. These two figures show that the delay with a webcam is significantly higher. This higher delay with the webcam is caused by the automatic white balance and gain control settings, as well as the coding for the USB link as mentioned earlier, this is further elaborated in Annex A. We can reduce this delay to 100ms by better configuration. However video telephony applications that do not address these issues may struggle with unacceptable delays because the entire delay budget is taken even before the data is sent to the remote side.

3.3.2 Network delay

Network delay is caused by a number of issues:

1. Speed of the connection
2. The number of routers and switches through which the connection passes.

Figure 11 shows the delays in today's Internet from an ADSL line connected to Dutch quality ISP XS4all to www.amazon.com. The figure shows per line the time needed to reach the destination of each of three test packets and to reach each of the routers in between.

The figure shows a number of things:

1. The delay to the other side of the Internet may easily reach 115ms, more than half our delay budget of 200ms.
2. Delay is not always constant, sometimes the delay does not increase but decrease between two routers measured.
3. It takes 73ms to cross the Atlantic (UK to US, line 7-8)
4. 1/5th of the delay, about 23 ms, is caused by the first mile, the ADSL link. If the same test is performed not from a computer behind ADSL but from a computer at the ISP's datacen-

ter the time to reach 0.so-6-0-0.xr1.sara.xs4all.net (the fourth line in the figure) the delay is just 0.5 ms and all subsequent times are correspondingly lower.

traceroute to www.amazon.com, 30 hops max, 38 byte packets			
1	10.100.0.1	0.785 ms	1.059 ms
2	195.190.249.25	25.799 ms	22.992 ms
3	42.10ge-4-0-0.xr2.d12.xs4all.net	24.363 ms	29.126 ms
4	0.so-6-0-0.xr1.sara.xs4all.net	23.698 ms	23.655 ms
5	mpr2.ams1.ge8-0.pni.xs4all.above.net	25.643 ms	24.189 ms
6	so-0-0-0.mpr1.ams5.nl.above.net	26.488 ms	24.882 ms
7	so-5-0-0.cr2.lhr3.uk.above.net	38.605 ms	37.850 ms
8	so-7-0-0.cr2.dca2.us.above.net	110.547 ms	110.432 ms
9	so-5-0-0.mpr2.iad1.us.above.net	116.885 ms	113.746 ms
10	so-3-0-0.mpr2.iad5.us.above.net	114.003 ms	117.127 ms
11	so-0-0-0.mpr1.iad5.us.above.net	110.614 ms	110.688 ms
12	amazon-above.mpr1.iad5.us.mfnx.net.	118.708 ms	114.536 ms
13	72.21.201.27	119.799 ms	111.123 ms
14	72.21.205.24	111.739 ms	114.515 ms

Figure 11. Internet delays

3.3.3 Delays in processing the video

Every processing step on video data causes delays. If the image has to be processed as a whole, either because of the nature of the process, the coding chosen or clumsy implementation, this will cause an additional delay as the entire frame needs to be collected. This collection process takes 33-50ms depending if 30 or 20 frames are captured per second. This explains why a higher frame-rate may result in a lower overall delay, as the time between the two images is lower.

The delays incurred through compression and decompression depend on the type of coding used. Some coding types have been especially created with low delays for live interactive sessions in mind. H.261 was once created with that aim in mind, however this coding was used for ISDN connections using 64kb/s or 128 kb/s bandwidth. The more modern H.264 (part of MPEG-4) achieves a low delay by processing small tiles of 8x8 pixels¹⁰. Coding and decoding of the media data is limited to just these tiles, which reduces the delays. However each of these processing steps take processor time to execute, modern coding types such as MPEG-4 require either a fast processor or special hardware coding/decoding.

The ultimate display step will require a full frame, if parts of the image are displayed before the rest, the image will appear to flicker. Collection of this full frame will take 1/30 to 1/20 second, 33-50ms.

3.4 Streaming

Existing video telephony applications deliver a low quality. There are applications created to stream over the Internet that do deliver a good quality video. Unfortunately these applications have not been designed for interactive use. These applications incur a delay both on the sending

¹⁰ Better compression can be achieved if there is no requirement that the media is processed from a live data source with the least delay, however this is obviously not suitable for video telephony.

and the receiving side of 300-500ms to buffer for variations in available bandwidth. This makes these applications unsuitable for video telephony.

There are other ways to achieve these goals. The Java Media Framework makes video streaming experiments easy¹¹. Figure 12 shows the streaming delay from a webcam through one computer to another on the same LAN. We used MJPEG coding¹² and the RTP streaming protocol and achieved 140ms delay between recording and display. The bandwidth used in this experiment was significant, 2,5Mb/s.

3.5 Delays conclusion

The delays in video telephony add up. We can expect a delay between 70-110 ms for capture and display, 30 ms for streaming and up to 90ms for Internet transmission (assuming fast access to the Internet). Our budget of 200ms is stretched but with careful implementation and deployment we will be able to stay within the budget.



Figure 12. Streaming using the Java media framework, 320x240 30fps, MJPEG coding, delay 150ms, bandwidth use: 2.5 Mb/s

¹¹ But the time to set up a stream makes it impractical for a real application.

¹² MJPEG encodes each image using JPEG image compressing. MPEG uses JPEG frames once in a while to send a complete image and subsequently sends only changes to that image.

4 Towards good video telephony

4.1 Introduction

Dissecting existing video telephony solutions has clarified the bottlenecks in existing video telephony applications. These applications were created with the idea that bandwidth is the limiting factor and that no more can be done to increase the quality. We reverse this assumption.

4.2 Unlimited bandwidth

We specify here an *affordable* videophone assuming much symmetrical bandwidth, say, 25Mb/s for one household. Part of this bandwidth may be reserved for the duration of the video call, say, 5-10Mb/s. With this amount of bandwidth in mind we will (re)design a videophone.

Although 5-10Mb/s is a lot of bandwidth, this is not enough to carry all the image data of a video call uncompressed. However our measurements show that we can manage even with old-fashioned MJPEG compression. We can even use a higher resolution than the bare minimum defined at 320x240. Scaling up to 640x480 will take four times as much bandwidth.

An alternative is coding using the more modern H.264. This available amount of bandwidth can be used to send video with a very good quality. TV quality will require about 2Mb/s. H.264 will require either a more expensive processor or specialized H.264 (de)coding hardware. This kind of chips are getting cheaper all the time, but we will need to consider high quantities of chips before they will be cheap enough for this videophone.

We have shown that affordable webcams, if well configured, will be able to reach the required resolution. We have also shown that video cameras will achieve better results, if our videophone can handle a composite video input. Such chips are available too. They too will add costs to the videophone and simple video cameras that may be used for this application costs as much as a complete webcam. The pricing probably is based on the quantities involved, webcams are available everywhere while small video cameras are mostly used in small quantities for surveillance purposes.

For a usability standpoint we would like to see the videophone as a standalone device with a telephone-like handset. A few buttons will allow the caller to "dial" the called and a simple button will allow calls to be accepted and terminated. A few buttons for starting and stopping the video streams and to mute the sound will enable an affordable videophone for anyone who can handle a regular telephone.

4.3 Conclusion

Video telephony offers huge advantages in communication and may help people to better stay in touch and to reduce business travel and hence traffic jams.



For video telephony to be widely accepted it needs to be **easy to use**, provide **sufficient quality** and be **affordable**.

Many of today's videophones are difficult to use or privacy invasive. Adding a traditional receiver and some buttons familiar from regular telephones may address both of these issues.

In this report we have shown that the state of the art is mostly determined by bandwidth limitations. However, broadband Internet is making huge strides forward. The introduction of **symmetrical** Internet with high **guaranteed speeds** offers opportunities for video telephony, finally there is the opportunity for good image quality at acceptable costs. The approach identified in this report will work with **symmetrical** bandwidth of 5-10Mb/s and will deliver an image quality comparable to that of VHS video with acceptable delays.

If there is the will and the bandwidth, nothing will stand in the way of general use of video telephony.

Annex A Camera Measurement

A.1 Introduction

We showed in Figures 9 and 10 that there are huge differences in the capture delay of video cameras and webcams. In this annex we investigate these differences.

A.2 Test setting

Figure 13 shows the test setting for the base measurements. A computer displays a millisecond clock on its screen, the camera captures this image and displays this on the other screen. The camera also records parts of its own screen so the clock is shown twice in the captured video. This allows us to easily compute the delays incurred.

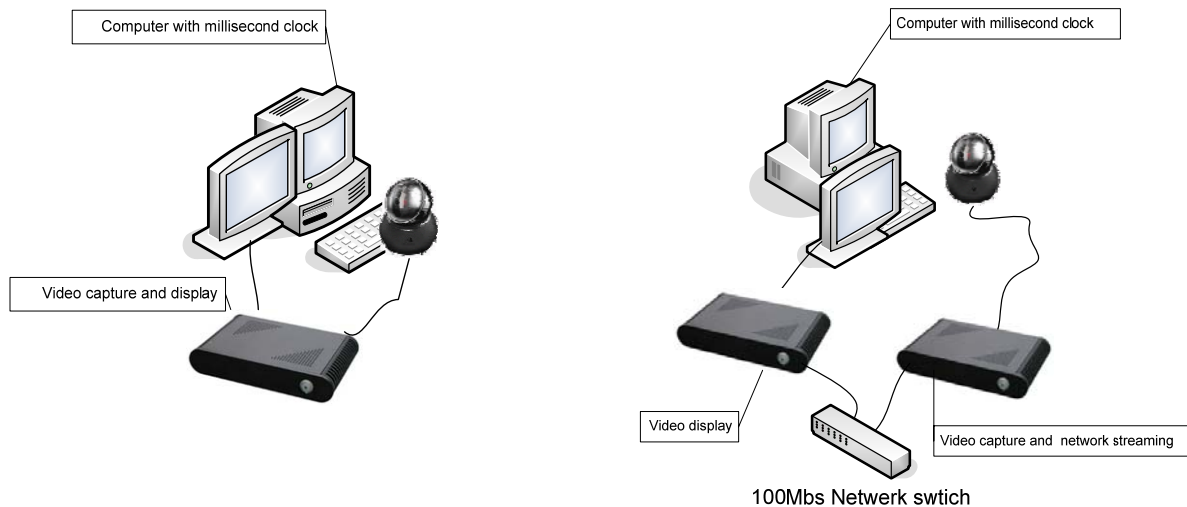


Figure 13. Test setting for the base measurements

Figure 14. Test setting for the streaming measurements

The follow-up measurements have been performed with the test setup as shown in Figure 14. Here a second computer is inserted between the camera and the screen. The two computers are connected via a 100Mb/s network switch.

A.3 Base measurement

The base measurement of capture delays was done using a Sony Handycam, connected to a Hauppauge PCI video capture card. The recorded image was displayed on the computer's monitor via TV viewing program¹³ this showed a promising 70ms delay.

The base measurement of the webcam was performed with a Logitech® QuickCam® Sphere webcam and displayed with the same program. This resulted in a disappointing 200ms delay.

¹³ Xawtv on linux.



Figure 15. Video camera delay; 70ms



Figure 16. Webcam delay; 200ms

A.4 The secrets behind webcams

Both measurements showed that the test setup worked but also showed a huge issue with webcams. As it turns out webcams come in various flavors. For this test we already used the best ones we could find. The more expensive Logitech models use light sensitive sensors (CCDs) from Philips. These CCDs provide a far superior image over the CMOS chips in use in the cheaper webcams.

As it turns out, the specs on the box do not always match the device's capabilities. Especially megapixel claims turn out to be exaggerations in practice and are usually interpolations by the driver software¹⁴.

Philips has two series of CCD chips that are sold to Logitech (amongst others); the PCVC645/646 with a maximum resolution of 352x288 pixels and the PCA675/680/690/730/740 series with a maximum resolution of 640x480 pixels.

Below we have printed three tables (source: PWC FAQ¹⁵) with these chip's available resolutions and hence the real resolutions available from the cameras that they are used in. On the vertical you will find the image resolution and on the horizontal the frame rate. A ✓ in a field shows that this resolution is supported at the frame rate given. A ★ indicates that this resolution and frame rate are only supported in the so-called "compressed mode".

These tables show that (1) not all resolutions are available at all frame-rates and (2) the so-called "compressed mode" has to be used for the qualities we need. This compressed mode is a Philips invention to overcome the bandwidth limitations on the USB line. The webcam compresses the images with a Philips-proprietary compression algorithm so they will fit on the USB cable. The computer then has to uncompress them before further processing.

¹⁴ Logitech makes the subtle difference in its product descriptions, "1.3 Megapixel sensor" as opposed to "4 Megapixel software enhanced".

¹⁵ <http://www.lavrsen.dk/twiki/bin/view/PWC/FrequentlyAskedQuestionsPWC>

Frame rate	PCA 645/646& VC010							
	3.75	5	7.5	10	12	15	20	24
sQCIF 128x96		✓	✓	✓	✓	✓	✓	✓
QSIF 160x120								
QCIF 176x144		✓	✓	✓	✓	✓	✓	✓
SIF 320x240								
CIF 352x288	✓	★	★	★	★	★		
VGA 640x480								

Frame rate	PCVC 675/680/690					
	5	10	15	20	25	30
sQCIF 128x96	✓	✓	✓	✓	✓	✓
QSIF 160x120	✓	✓	✓	✓	✓	✓
QCIF 176x144	✓	✓	✓	✓	✓	★
SIF 320x240	✓	★	★	★	★	★
CIF 352x288	✓	★	★	★	★	★
VGA 640x480	★	★	★			

Frame rate	PCVC 730/740/750					
	5	10	15	20	25	30
sQCIF 128x96						
QSIF 160x120	✓	✓	✓	✓	✓	✓
QCIF 176x144						
SIF 320x240	✓	★	★	★	★	★
CIF 352x288						
VGA 640x480	★	★	★			

Table 3. Resolution of Philips CCD chips in webcams

A.5 Improving webcam behavior

Despite these webcam limitations we can achieve better results than the base measurement showed. Looking into the driver settings we found that the white balance can be set to manual and the light gain may be manually adjusted as well. This dramatically improves the webcam response times. Figure 17 shows the delay measurement after manually changing the camera settings and setting the camera to 30 frames per second. The delay has been almost halved!



Figure 17. Improved Webcam delay

A.6 New technology; not faster

Logitech introduced a new product line in 2005 using different CCD chips. These chips utilize the full rate of a USB 2.0 link and hence have more bandwidth to the computer at their disposal. These cameras support higher resolutions without software enhancement and offer a much

clearer picture. Unfortunately the delay is comparable with that of the webcams investigated earlier. Again turning off automatic light controls show the same delay improvements. Figure 18 shows a measured delay of 120ms.

This delay is so close to that of the earlier measurements which suggest similar causes. We infer that there are a fixed number of steps that need to be taken into before the image may be displayed and that these images need to be processed whole. (110-120 ms delay and 33ms between two images suggests 4 processing steps with three image transitions in between.) These steps are probably (0) capture, (1) coding+transmission on USB, (2) receiving from USB and decoding, (3) display.



Figure 18. Delays using the Quickcam Fusion