

Voice over IP (VoIP) Part 1

Dr. Andreas Steffen

©1999-2001 Zürcher Hochschule Winterthur

A. Steffen, 9.12.2001, KSy_VoIP_1.ppt 1

VoIP Scenarios

- Classical telecommunications networks
- Present: separate data and voice networks
- Future: unified networks
- Migration path to VoIP
- Least cost routing - today and tomorrow
- VoIP - pro and contra

H.323 Multimedia Communication Standard

- Overview

Video Compression

- H.261 / H.262 / H.263 DCT-based compression algorithms
- Macroblock based encoding of color information
- Motion vector estimation
- Group of pictures difference coding scheme
- Video conferencing picture formats

Audio Compression

- G.711 pulse code modulation @ 64 kbit/s
- G.722 adaptive differential pulse code modulation @ 64 / 32 kbit/s
- Linear prediction coder @ 2.5 kbit/s
- G.728 code excited linear prediction coder @ 16 kbit/s
- G.729 audio codec for frame relay @ 8 kbit/s
- G.723.1 low rate audio codec @ 6.3 / 5.3 kbit/s
- GSM 06.10 enhanced full rate coder @ 13 kbit/s
- Mean opinion score (MOS)

H.32x Family

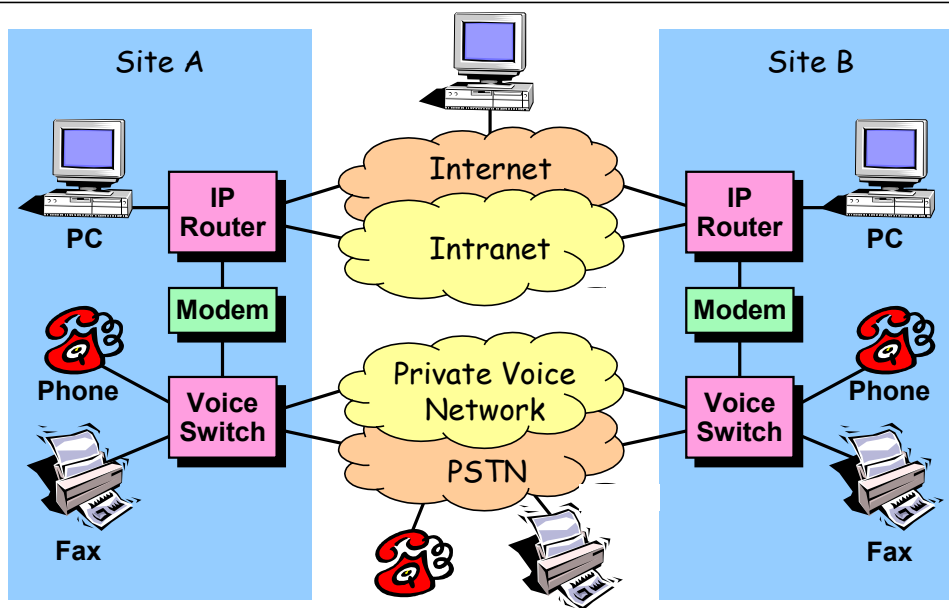
- H.320 Videoconferencing over ISDN / H.324 Videoconferencing over POTS

- **Theorie Teil 1 (heute)**
 - VoIP Anwendungen - Heute und Morgen
 - Übersicht H.323 Standard
 - Audio / Video Kompression
- **Theorie Teil 2 (nächste Woche)**
 - Call Setup / Capabilities Exchange / Channel Multiplexing
 - H.323 Gatekeepers / Gateways / Multipoint Conferences
 - IETF Session Initiation Protocol (SIP)
- **Praktikum (24./25.1.2002)**
 - H.323 Gateway und Gatekeeper Funktionalität
 - Siemens Hicom Xpress Call Center Applikation
 - H.323 SW Client „Microsoft Netmeeting“ / IP Phones



Voice over IP Scenarios

The Classical Approach Separate Voice and Data Networks



A. Steffen, 9.12.2001, KSy_VolIP_1.ppt 4

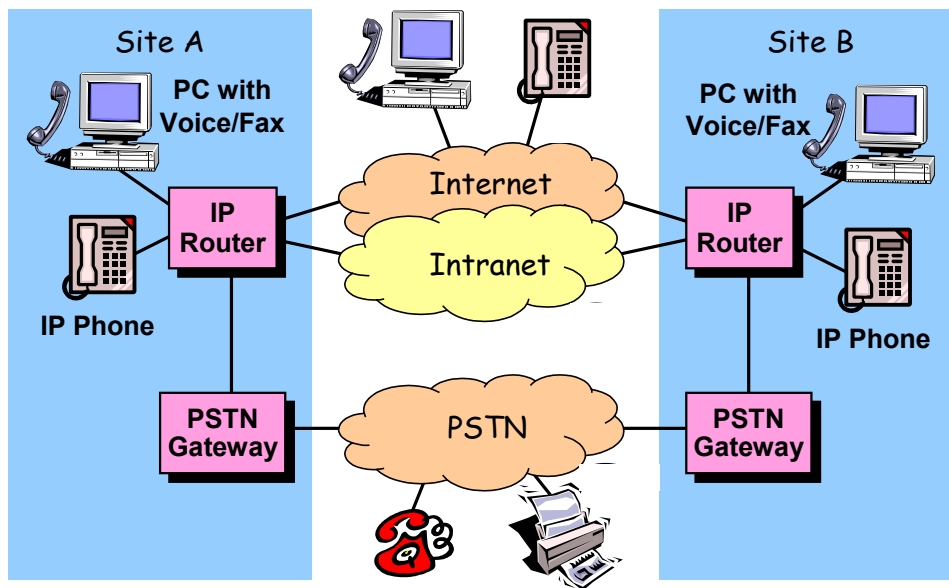
Past: Classical Telecommunications Networks

- The classical telecommunications network is voice-oriented. A company usually has a Private Branch eXchange (PBX) that switches internal voice calls and connects external calls to and from the **Public Switched Telephone Network (PSTN)**. Since Fax transmission is based on an analog modulation within the voice band, it is treated in exactly the same way as a voice connection.
- For connections between locations that have a high traffic volume, many companies use **leased lines** in order to minimize costs, thereby creating a **Private Voice Network**.
- When the first data applications turned up several decades ago, data was carried over voice lines using analog modems. Still today, remote login from teleworkers into the company data network is done mostly as direct dial-in over the PSTN.

Present: Separate Data and Voice Networks

- Today many companies have a separate data network carrying the backbone traffic between locations. This is often done over leased lines in order to minimize costs. When ATM is used as a transport medium then voice and data lines may share the same physical line but are otherwise treated as separate logical connections.
- Together with the LAN cabling at the various locations, these data connections form the **Intranet** of a company. External **Internet Access** is usually realized over a separate leased line to an Internet Service Provider (ISP). Using **Virtual Private Network** mechanisms like **IPsec** tunnels, Intranet connections can alternatively be carried over the Internet at lower costs and without compromising security.
- In the **Company Data Network** the **IP Routers** immediately attached to the end points of both Internet and Intranet data connections have the task of forwarding the IP datagrams within the network using packet-based IP routing mechanisms.
- In the **Company Voice Network** the **Voice Switch** has the task of setting up voice and fax connections using private/public switching mechanisms.

The Future Approach Voice/Fax over IP - A Unified Network

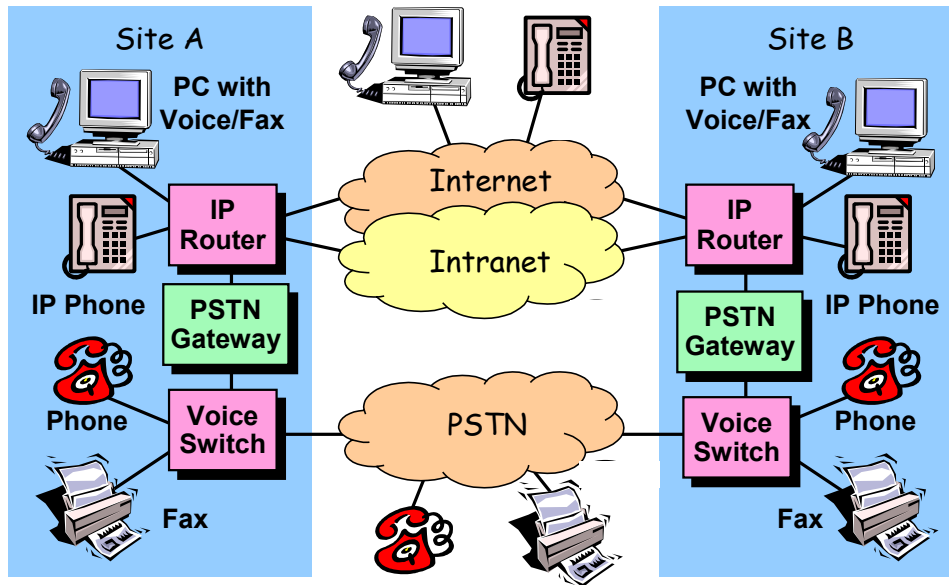


A. Steffen, 9.12.2001, KSy_VoIP_1.ppt 5

Future: Unified Networks

- Already today, the fast-growing world-wide data traffic volume has reached or probably even surpassed the total traffic generated by traditional voice connections. In the not too far future, voice calls will therefore take up only a small percentage of the total traffic. The largest part of a telco carrier's investments go nowadays into the expansion of their data networks. The global Public Switched Telephone Network (PSTN) represents a huge installed base and will still be maintained over the next couple of years or even decades. But no new major investments will go into the traditional voice network.
- **Voice over IP (VoIP)** technology allows the transportation of voice traffic over pure data networks, thereby eliminating the need for a separate voice network.
- As a VoIP terminal, most users will prefer a comfort telephone set equipped with an IP interface, a so called „**IP phone**“. Other users will use software-based multimedia applications like e.g. Microsoft Netmeeting running on their PCs that additionally allow the real-time exchange of user-data.
- Since the existing Public Switched Telephone Network will be in use for many years to come and many subscribers will never change to VoIP, a **PSTN Gateway** is required as an interface between the packet-based VoIP network and the connection-oriented PSTN.

The Intermediate Approach Migration Path to Voice/Fax over IP

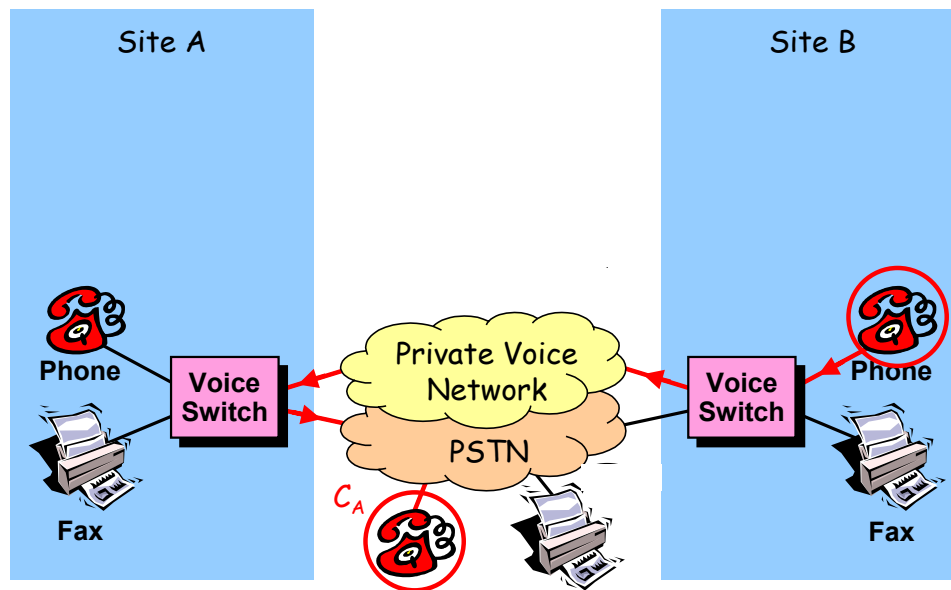


A. Steffen, 9.12.2001, KSy_VoIP_1.ppt 6

Typical Migration Path to VoIP

- Most companies have PBXes that still have a life time between 5 and 10 years. Therefore at the beginning only selected groups will be equipped with VoIP. First-choice VoIP candidates are e.g. customer support, organized in distributed call centers, or software developers who work at their PCs most of the time anyway. The majority of a company's staff will keep on to their traditional phone sets and fax equipment and will be migrated at a relative slow pace.
- Therefore a PSTN Gateway is usually inserted between the IP-based VoIP network and the company's voice switch, making it possible to establish both incoming and outgoing voice connections to and from the internal and external telephone networks.

Least Cost Routing of Voice/Fax Today: Over the Private Voice Network

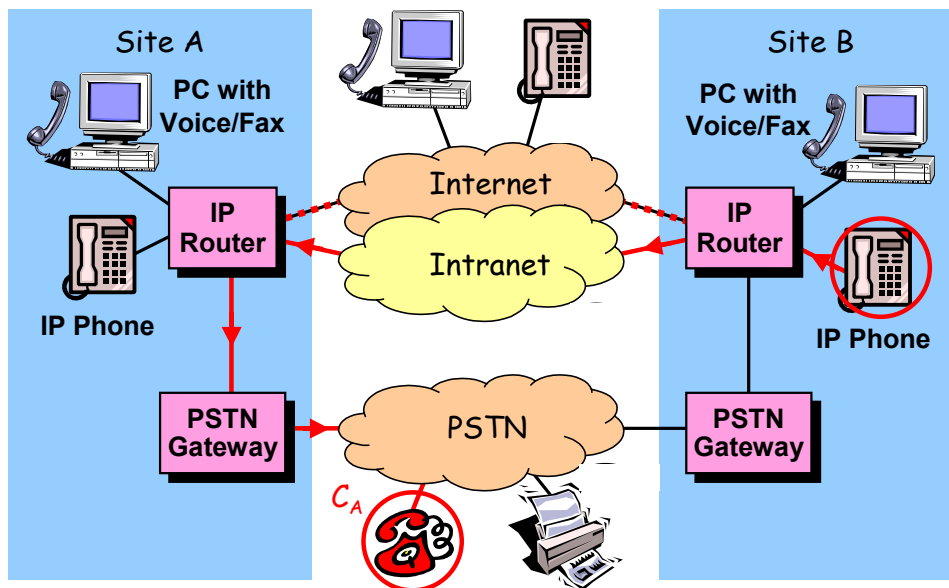


A. Steffen, 9.12.2001, KSy_VoIP_1.ppt 7

Least Cost Routing over a Private Voice Network

- The scenario is the following: An employee at site B wants to make an external call to a customer C_A . The least cost routing algorithm running in the voice switch at site B determines the cheapest route to C_A at the current time of day.
- As a result of this search, instead of setting up a direct long distance call over the PSTN from site B, the call is first switched over the private voice network to the PBX at site A where a local call is set up over the PSTN to the customer C_A .

Least Cost Routing of Voice/Fax Tomorrow: Over the Intranet / Internet



A. Steffen, 9.12.2001, KSy_VoIP_1.ppt 8

Least Cost Routing over the Intranet or Internet

- We now assume a company voice network based completely on VoIP. The PBXs have been replaced by PSTN gateways.
- When an employee at site B wants to call an external customer C_A , the least cost routing algorithm running at the local H.323 gatekeeper determines the PSTN gateway at site A as the optimal exit point to customer C_A . All VoIP packets for this call are now routed either over the private Intranet or over the public Internet to the PSTN gateway at site A.

■ Advantages

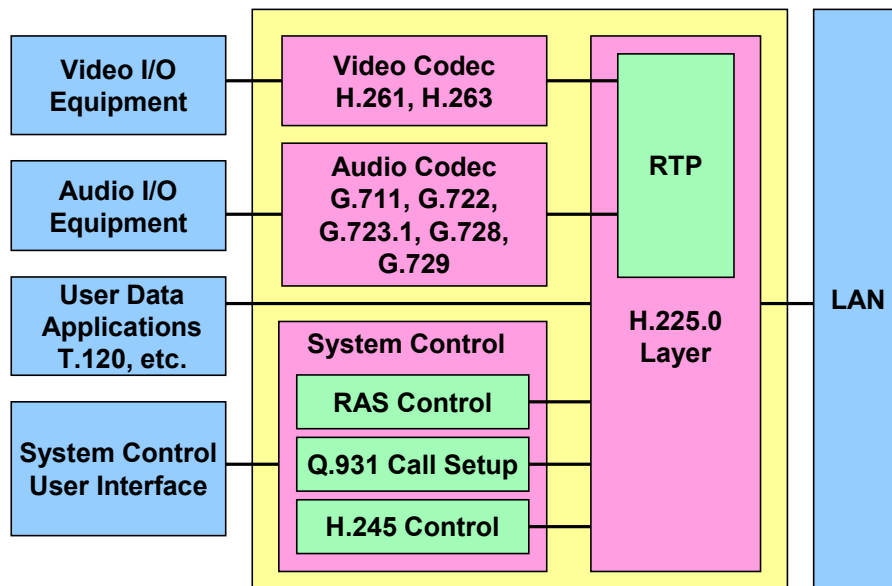
- Cost savings on long distance calls
- Fewer leased lines for private networks
- Single RJ-45 connector at the workplace for all services
- Elimination of expensive voice switches
- Enables new multimedia features, e.g. human operator assisted e-commerce

■ Problems / Open Questions

- Control of delay, jitter and packet loss over IP-based networks
- QoS guarantees (RSVP, ATM traffic contracts)
- Universal directory services (X.500, LDAP)
- Interoperability (compliance with ITU-T and IETF standards)

H.323 Standard

ITU-T H.323 : Packet-Based Multimedia Communications Systems



A. Steffen, 9.12.2001, KSy_VoIP_1.ppt 11

The ITU-T Standard H.323

- The H.323 recommendation for „Packet-Based Multimedia Communications Systems“ is an **umbrella standard** comprising the numerous ITU-T recommendations listed below:

Video Streams (Video Codecs H.261, H.263)

- These recommendations define various digital video compression algorithms.

Audio Streams (Audio Codecs - G.711, G.722, G.723.1, G.729, G.729)

- These recommendations define various digital audio compression algorithms.

User Data Applications (T.120, etc.)

- These recommendations define how user data applications are encoded. Examples are shared whiteboard applications and user file transfer.

System Control

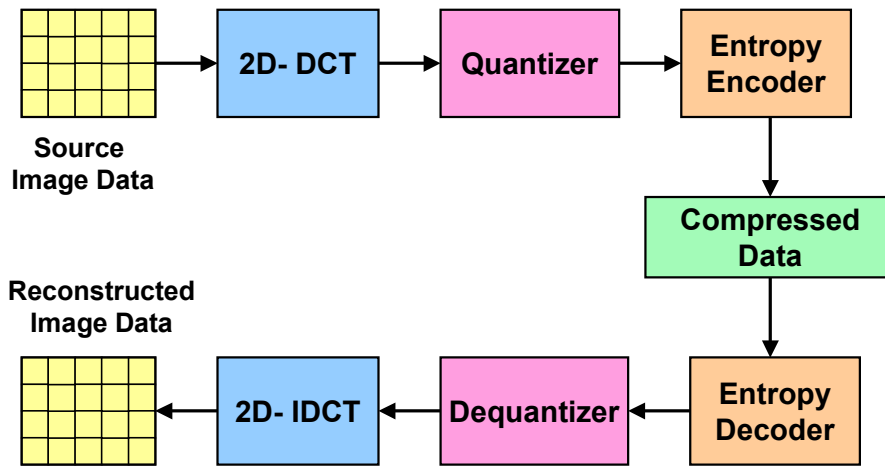
- **RAS Control:** Registration with a H.323 gatekeeper. Name resolution services.
- **Q.931 Call Setup** with an VoIP peer using the ISDN basic call control protocol.
- **H.245 Control:** Negotiation of common multimedia properties between VoIP peers.

Call Signalling Protocols and Media Stream Packetization (H.225.0)

- Defines the packet-based multiplexing of the various video and audio streams using the unreliable **Real-time Transfer Protocol (RTP)**.
- Defines the transmission of user data streams using the reliable TCP protocol.
- Defines the format and structure of the various system control messages.

Video Compression

H.261/H.263 Video Compression based on the Discrete Cosine Transform (DCT)

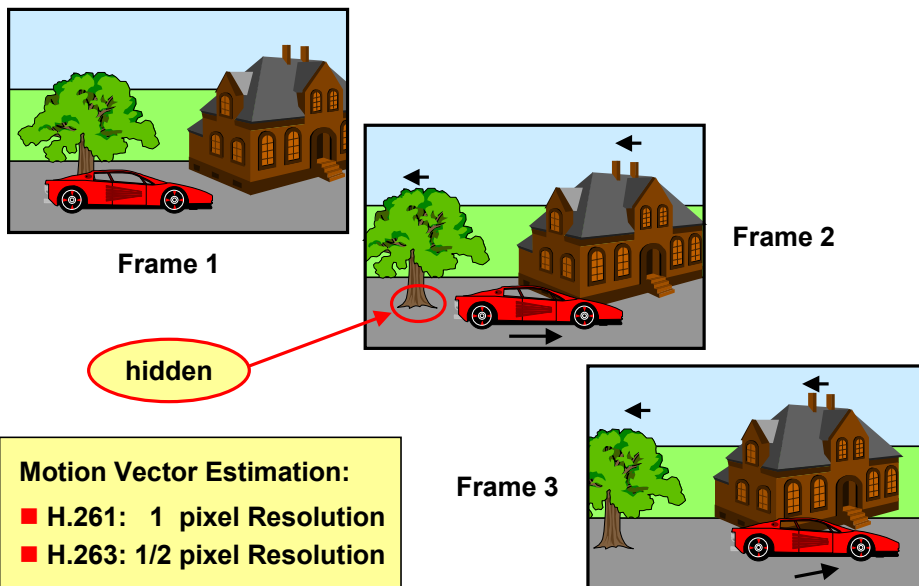


Video Compression Algorithms based on the DCT

- In H.323 packet-based multimedia applications one of the two ITU-T standards H.261 or H.263 (which were optimized for small transmission bandwidths) are commonly used.
- When a sufficiently large transmission bandwidth (2-6 Mbit/s) is available, also the MPEG-2 standard H.262 could be used for H.323 based video communication.

Building Blocks

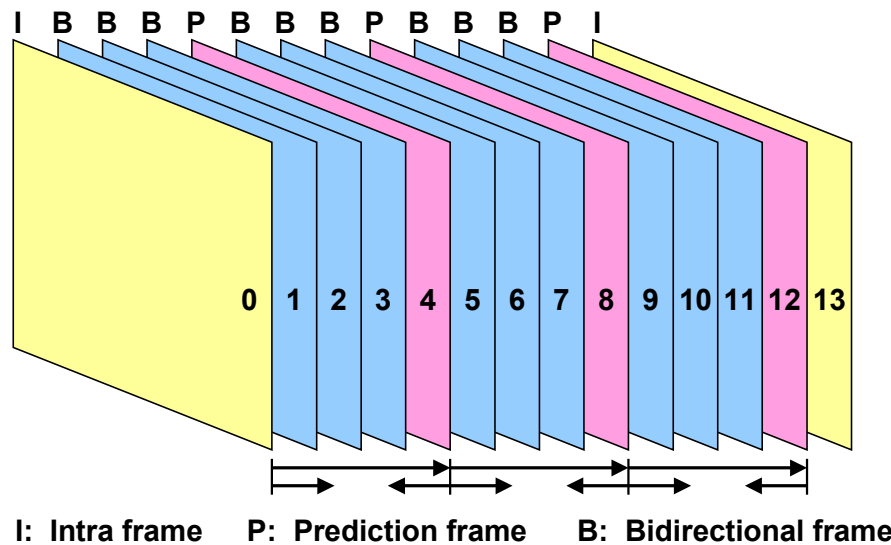
- **2D-DCT / IDCT:** The two-dimensional discrete cosine transform (DCT) transforms image sub-blocks of 8x8 color pixels into a frequency domain representation of 8x8 frequency components. The luminance and chrominance components are transformed separately. The DCT is a **lossless operation** since the inverse discrete cosine transform (IDCT) restores the original image pixels given the frequency distribution information.
- **Quantizer / Dequantizer:** Using quantization tables, each luminance and chrominance frequency component is quantized to a variable number of bits. Usually the number of allocated bits decrease with increasing image frequencies since the human eye is less sensitive to errors in high image frequencies.
Due to the finite number of 2^b quantisation steps corresponding to the number of b allocated bits, quantization in the sender is a **lossy process** that cannot be reversed by the dequantization process in the receiver. This translates into a tradeoff between the achieved compression factor and the resulting image quality.
- **Entropy Coder / Decoder:** The resulting bit values of the quantized luminance and chrominance frequency components are optimally compressed using a Huffman run-length encoding. This is a **lossless compression operation** that can be reversed by the entropy decoder at the receiver.



Motion Vector Estimation

- High video compression factors are achieved by coding the changes between subsequent video frames only. Since both the camera and the frame objects can move, a motion vector must be estimated for each moving object. This is done on the basis of the 16x16 macroblocks. H.261 does motion vector estimation with 1 pixel resolution, whereas the new H.263 standard achieves better results with an improved resolution of half pixels.
- For each macroblock an optimal horizontal and vertical offset relative to the previous frame can be determined, thus minimizing the difference information between the two multiblock areas. This is called **forward prediction**.
- Often some part of a scene that has remained hidden in a previous frame, is suddenly exposed at a later time. In this case it would be more advantageous to code a motion-compensated image difference relative to the later frame. This is called **backward prediction**.
- By using the **Group of Pictures (GOP)** scheme described on the next slide, coding of both forward and backward differences is made possible.

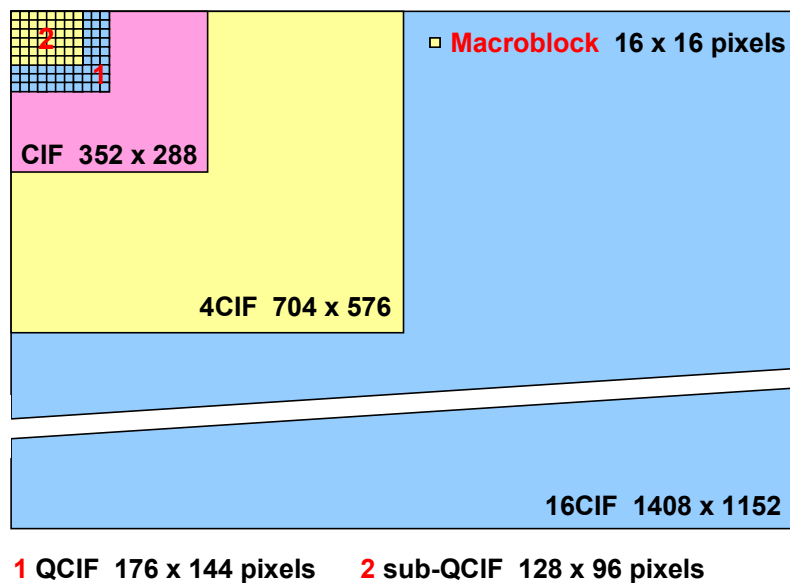
Group of Pictures (GOP)



Group of Pictures

- A group of pictures (GOP) starts with an **intra frame** or **I-frame** that is coded as a stand-alone picture without reference to any other frame.
- Depending on the configured parameters of the GOP, every **nth** frame is coded as a forward **prediction frame** or **P-frame**. The first P-frame is coded relative to the I-frame, the second P-frame relative to the first P-frame, the third P-frame relative to the second P-frame, etc.
- The macroblocks in the $n-1$ **bidirectional frames** or **B-frames** between two subsequent P-frames are coded either in forward direction or backward direction whichever gives the better result.
- Every **k** frames a new I-frame is transmitted, thereby starting the next group of pictures. This limits error propagation to a single GOP and allows film cuts at the GOP boundaries.

Video Conferencing Picture Formats Common Interchange Format (CIF)



Video Conferencing Picture Formats

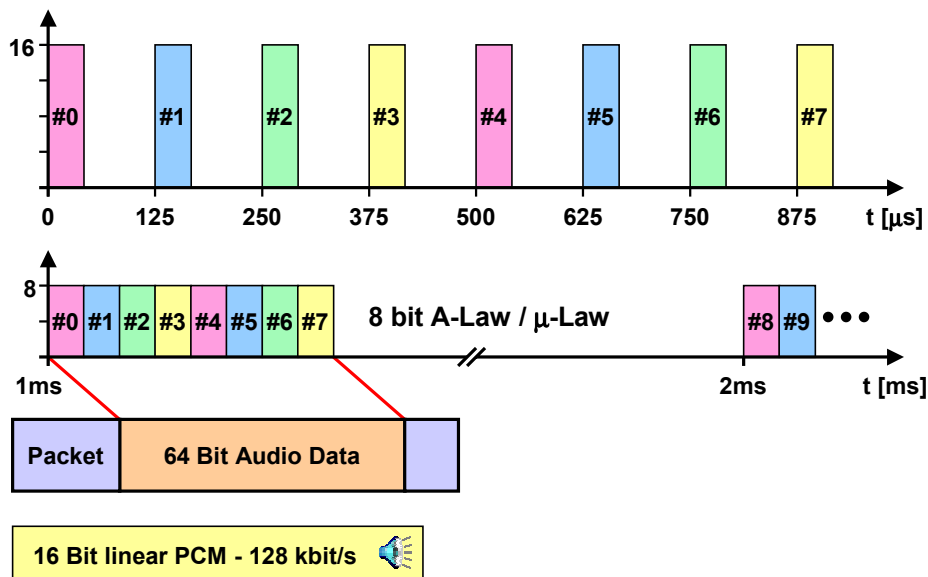
- For video conferencing applications a **Common Interchange Format (CIF)** with a standardized picture size of 325 x 288 pixels has been defined.
- For applications over narrow-band connections (e.g. 33.6 kbit/s POTS or 64-128 kbit/s ISDN channels) and/or low-end multimedia terminals (little computing power, usually with SW-based encoders/decoders), the standard has been extended to include the smaller Quarter-CIF and the even smaller sub-Quarter-CIF formats.
- For applications using broad-band connections (> 2 Mbit/s) and powerful multimedia terminals (usually with HW-based encoders/decoders), the standard has been extended to include the TV and high-resolution formats 4CIF and 16CIF, respectively.
- All standardized picture sizes are multiples of the 16 x 16 pixel macroblocks used by the DCT-based video compression algorithms to encode the chrominance information. The required processing power for real-time encoding and decoding is directly proportional to the total number of macroblocks in the picture.

Picture Sizes

• sub-QCIF	128 x 96 pixels	8 x 6 =	48 macroblocks
• QCIF	176 x 144 pixels	11 x 9 =	99 macroblocks
• CIF	352 x 288 pixels	22 x 18 =	396 macroblocks
• 4CIF	704 x 576 pixels	44 x 36 =	1584 macroblocks
• 16CIF	1408 x 1152 pixels	88 x 72 =	6336 macroblocks

Audio Compression

G.711 : PCM - 64 kbit/s A-Law / μ -Law Pulse Code Modulation



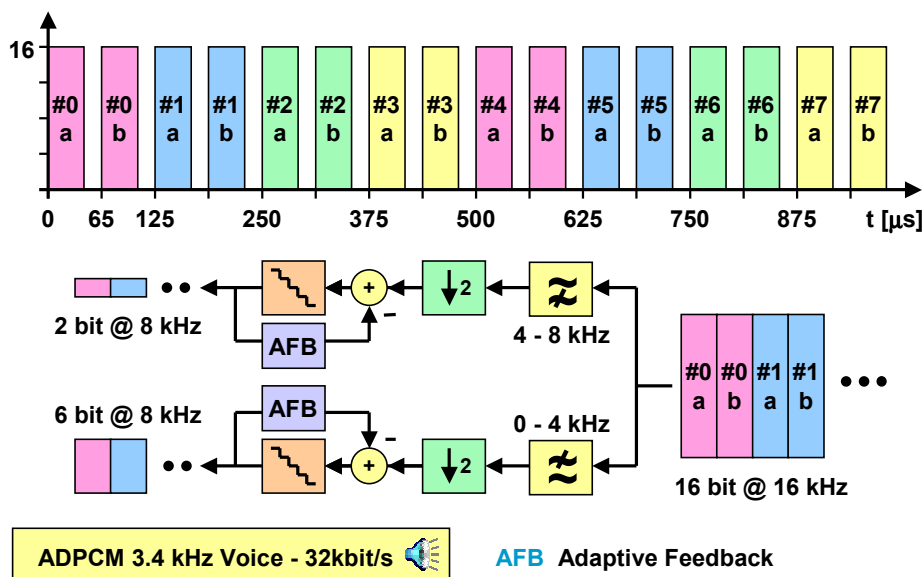
ITU-T G.711 - Pulse Code Modulation

- In telephony systems the voice band is usually limited to 300 Hz .. 3.4 kHz. According to the Nyquist theorem this band-limited speech can be sampled with an aliasfree sampling rate of 8 kHz. This means that a signal sample is taken every 125 μ s and converted into a 16 bit linear PCM value.
- As a next step the dynamic range of the 16 bit linear PCM value is compressed into an 8 bit amplitude value by using a logarithmic mapping function. In North America the μ -Law function is used, whereas in Europe and the rest of the world the A-Law function is applied.

G.711 embedded in RTP-Packets

- When VoIP applications are operated over a 10 Mbit/s or 100 Mbit/s LAN, as it is usually the case in call centers, then the G.711 speech format is used. Since no speech compression algorithm must be applied, the PCM speech samples can be transmitted with little delay and with a high speech quality. In order to keep the overhead due to the UDP and RTP headers at a reasonably low level, H.225.0 recommends to group 8 PCM samples together and to transmit them in a UDP datagram every 1 ms.

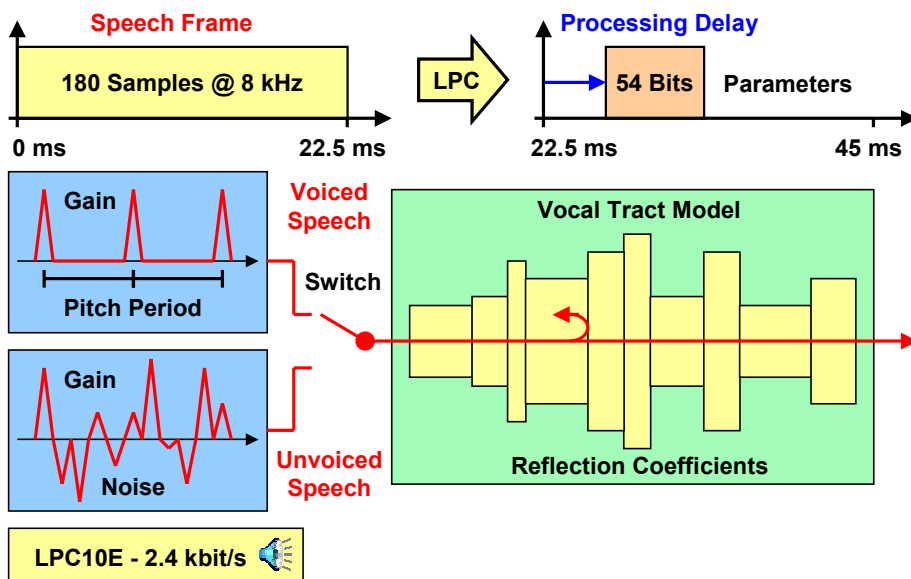
G.722 : ADPCM 7 kHz Voice - 64 kbit/s Adaptive Differential Pulse Code Modulation



ITU-T G.722 - Adaptive Differential Pulse Code Modulation

- Extensive studies have found out that the perceived quality of a real-time video conference depends to a large extent on voice quality. If speech quality is good, people are ready to accept imperfections in the transmitted video stream.
- Therefore in high-end video conferencing systems the voice bandwidth is often increased to 7 kHz, requiring a doubling of the sampling rate to 16 kHz.
- The goal is to have an increased voice bandwidth but without increasing the bit rate of 64 kbit/s. This can be achieved with Adaptive Differential Pulse Code Modulation (ADPCM) that works as follows:
 - The 16 bit linear PCM speech samples acquired with a sampling rate of 16 kHz are applied to a filter bank consisting of a low pass filter with a passband of 0 - 4 kHz and a high pass filter with a passband of 4 - 8 kHz. Through filtering the effective bandwidths have been reduced to 4 kHz, so without incurring aliasing effects, the sampling rate can be halved to 8 kHz by throwing away every second sample coming out of the filters. This process is called „down-sampling“.
 - The amplitudes of the filtered high-pass and low-pass signals are now quantized to a finite number of bits. Since the human ear is less sensitive to high frequencies, only 2 bits are assigned to the high-pass signal, whereas the low-pass signal gets 6 bits, resulting in a total rate of 8 bits @ 8 kHz = 64 kbit/s.
 - Since G.711 assigns 8 bits to a 4 kHz signal, the quality of the low-pass signal in G.722 would be worse with only 6 bits. The trick behind ADPCM consists of an adaptive feedback loop (AFB) that subtracts the quantized output signal from the original input signal, so that only the changes of the speech signal get encoded. Since the difference signal has much smaller amplitudes, 6 bits and 2 bits for the low-pass and high-pass signals, respectively are sufficient.
- A variant of the G.722 ADPCM method differentially encodes a 3.4 kHz signal at 8 kHz with only 4 bits per sample, resulting in a bit rate of 32 kbit/s. This algorithm is used e.g. by DECT.

Linear Prediction Coder (LPC) - 2.4 kbit/s Vocal Tract Model / Parameter Estimation



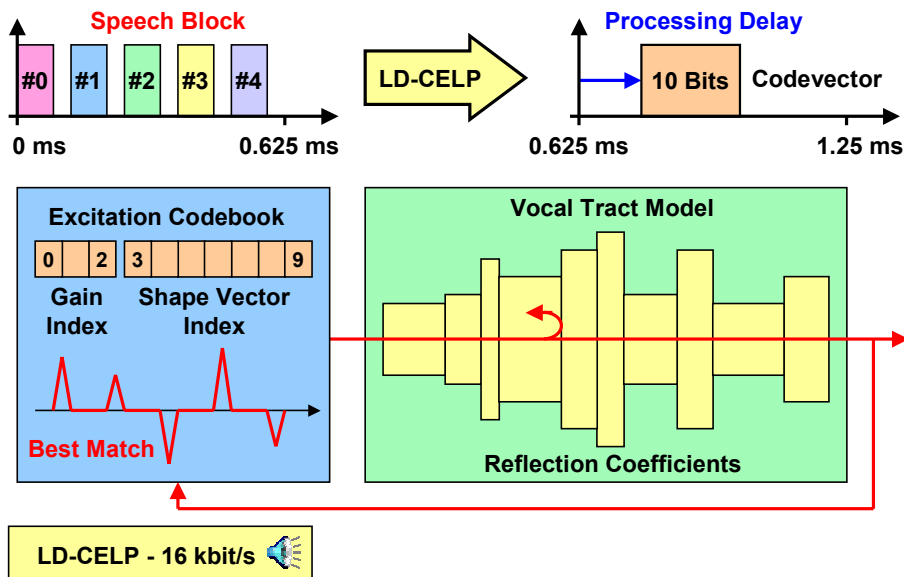
Speech Compression Algorithms based on Vocal Tract Models

- If speech rates below 16 kbit/s are to be achieved, then the way how human speech is generated in the vocal tract has to be modelled:
- The cavities of the mouth, nose, throat and larynx, as well as the influence of the tongue can be modelled by a series of cylindrical rings with variable lengths and diameters. These cavities are traversed by sound waves generated by the vocal chords in the larynx, with the effect that the spectral components of the sound wave are filtered according to the current shape of the vocal tract.
- A diameter change from one cylindrical segment to the next is equivalent to a corresponding change in the wave impedance. This causes part of the travelling sound wave to be reflected at the segment boundaries, the exact amount depending on the set of **reflection coefficients** that can be derived from the diameters.

Linear Prediction Coders

- In order to determine the current settings of the vocal tract model, a speech frame of 22.5 ms duration must first be recorded. This introduces a significant delay to which the processing delay required to encode the speech frame must be added.
- Using sophisticated algorithms the optimum parameter settings for the reflection coefficients are computed out of the 180 collected speech samples. The synthesized output from the model should be as close to the original speech frame as possible.
- It is a crucial task to determine if currently a vowel or a consonant is spoken. In the first case a periodic pulse train producing harmonics is applied to the vocal tract model, in the second case a noise generator is connected. The estimation of the current **pitch frequency** of the pulse train generator is even trickier.
- For each speech frame only the parameter settings of the speech model are transmitted. An exact copy of the speech model at the receiver resynthesizes the speech frame. This leads to an intelligible but rather impersonal type of speech. Therefore pure LPC coders are only used by the military and in cheap toys.

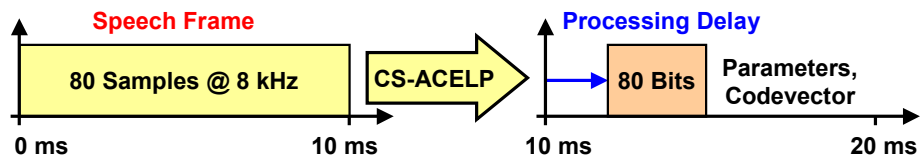
G.728 : LD-CELP - 16 kbit/s Low-Delay Code-Excited Linear Prediction



Code Excited Linear Prediction Coders

- The use of either a noise or periodic train generator exciting the vocal tract leads to a rather impersonal type of speech. With a linear prediction coder it is further nearly impossible to encode non-voice signals like e.g. music. Therefore in order to give synthesized speech the timbre and characteristics of the original speaker the excitation source has to be improved.
- This can be done by using a code book where in the case of the G.728 LD-CELP algorithm 128 different excitation vectors and 8 different gain factors are stored, resulting in a 10 bit long code vector characterizing the excitation. An optimization algorithm determines for each speech frame the optimal code vector that when applied to the vocal tract gives the smallest difference to the original speech sequence.
- Only the 10 bit code vector must be transmitted. The parameters of the vocal tract model is updated relatively slowly, making it possible that the corresponding model at the receiver side can automatically track and rebuild the state of the sender model. Thus there is no need to transmit any model parameters.
- The biggest advantage of the G.728 algorithm is its small speech frame size of 5 voice samples, resulting in an extremely small delay of only 0.624 ms.
- The G.728 codec at 16 bit/s produces a speech quality comparable to G.711 at 64 kbit/s. Even music is reproduced surprisingly well.

G.729 : CS-ACELP - 8 kbit/s Conjugate-Structure Algebraic CELP

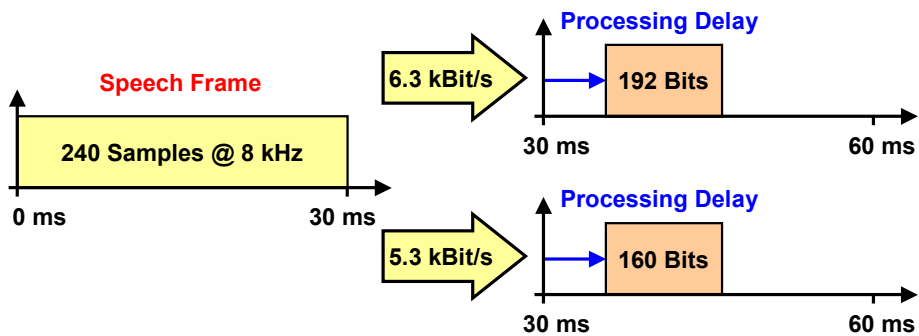


- Variant of a Code-Excited Linear Prediction Coder
- Default Voice Coding Algorithm for Frame Relay


CS-ACELP - 8 kbit/s



G.723.1 : MP-MLQ / ACELP - 6.3 / 5.3 kbit/s Multi-Pulse Maximum Likelihood Quantization



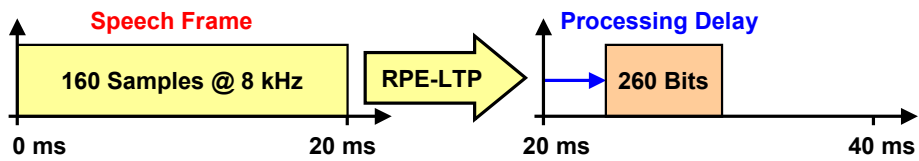
- Variants of a Code-Excited Linear Prediction Coder
- Default Voice Coding Algorithm for H.324 over POTS

CELP - 4.8 kbit/s 

The G.723.1 Codec - an Engineering Miracle !

- The G.723.1 audio codec is one of the most ingenious algorithms presently available, producing a surprisingly good speech quality at the low rates of 6.3 kbit/s and 5.3 kbit/s, respectively. The only drawback is the rather high speech frame duration of 30 ms, leading to significant signal delays.

GSM 06.10 : RPE-LTP - 13 kbit/s GSM Enhanced Full Rate Coder (EFR)

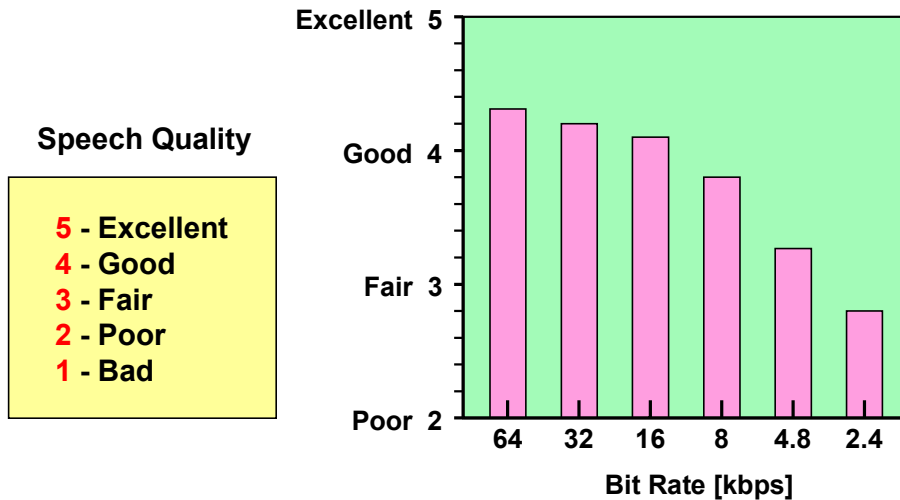


- Regular Pulse Excitation Long-Term Predictor (RPE-LTP)
- Variant of a Linear Prediction Coder

EFR - 13 kbit/s 

GSM 06.10 Enhanced Full Rate Codec



























- The new EFR codec is much superior to the original FR codec. Its modern architecture delivers at 13 kbit/s a speech quality comparable to G.711 at 64 kbit/s.



Mean Opinion Score (MOS)

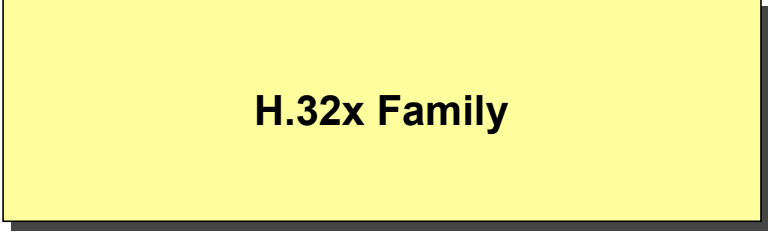
- Today the most reliable method to determine speech quality on an absolute scale is through a subjective opinion test where speech sequences are played to several people who are asked about their opinion immediately afterwards.
- The opinion **scale** is standardized and goes from **bad** (1) over **poor** (2), **fair** (3) and **good** (4) to a maximum rating of **excellent** (5). The MOS value is then computed as the average or mean over all answers.
- The ITU-T recommendation P.800 meticulously defines the laboratory setup for the MOS listening tests, including speech and noise levels, the speakers (at least two males and two females), ... up to detailed instructions for the subjects.
- The MOS graph as a function of the speech transmission rate clearly shows that speech rates down to about 16 kbit/s are generally perceived as having good quality, whereas the military-grade LPC algorithms with rates of 2.5 kbit/s and less are judged to be significantly below fair. Depending on the particular speech compression algorithm, the rates in the range from 5 .. 13 kbit/s come to lie somewhere between good and fair.

Speech Samples in Various Environments

	G. 711 64 kbps	ADPCM 32 kbps	G. 728 16 kbps	G. 729 8 kbps	CELP 4.8 kbps	LPC 2.4 kbps	GSM 13 kbps
Space Shuttle							
Shuttle Crew							
Music							
Bit Errors 0.1%							
Bit Errors 1%							

Various Speech Environments

- **Space Shuttle:** A single speaker plus the typical background noise occurring in the cockpit of a space shuttle. Speech compression algorithms should not only work well under studio conditions but should also behave robustly in noisy environments.
- **Shuttle Crew:** Sentence spoken by a group of people. Shows the ability of the algorithms to model several voices simultaneously. LPC as a single vocal tract model fails miserably.
- **Music:** Most speech compression algorithms model music quite acceptably, the exception being again the LPC that tracks the lead singer only.
- **Bit Errors:** Bit errors in G.711 PCM produce clicking noises, in differential PCM also sudden volume changes can occur. Speech models are quite robust as long as the error rate remains small (0.1%) but they start to get unstable with higher error rates (1%). The GSM enhanced full rate coder is a typical example for this behaviour.



H.32x Family

H.320 Videoconferencing over ISDN



- $n \times 64 \text{ kBit/s}$ (1 .. 6 ISDN B-Channels)
- H.261 / H.263 Video Codecs: QCIF / CIF
 - 15 CIF Frames/s @ 128 kBit/s
 - 30 CIF Frames/s @ 384 kBit/s
- G.711 / G.728 / G.722 Audio Codecs

Professional Video Conferences

- Large-scale conferences with several participants usually require a transmission rate between 256 - 384 kbit/s (i.e. 4-6 ISDN B-Channels). This is a large improvement over the 2 Mbit/s that were required just a few years ago. With 384 kbit/s about 30 CIF frames per second can be transmitted, which gives a large picture with smooth, film-quality movements. For high-quality 7 kHz speech the G.722 @ 64 kbit/s can be used, taking up one ISDN B-Channel.

Private Video Conferences

- For video conferences between two people using a videophone or a PC-based video terminal an ISDN BRI connection is sufficient. For a QCIF sized picture a transmission rate of 128 kbit/s with the older H.261 video codec and 64 kbit/s with the improved H.263 codec are sufficient. The frame rate is usually below 15 frames/s, but which is enough for the quasi-static picture of a videophone. For speech compression the 16 kbit/s G.728 LD-CELP algorithm is generally selected.

H.324 Videophone over POTS I



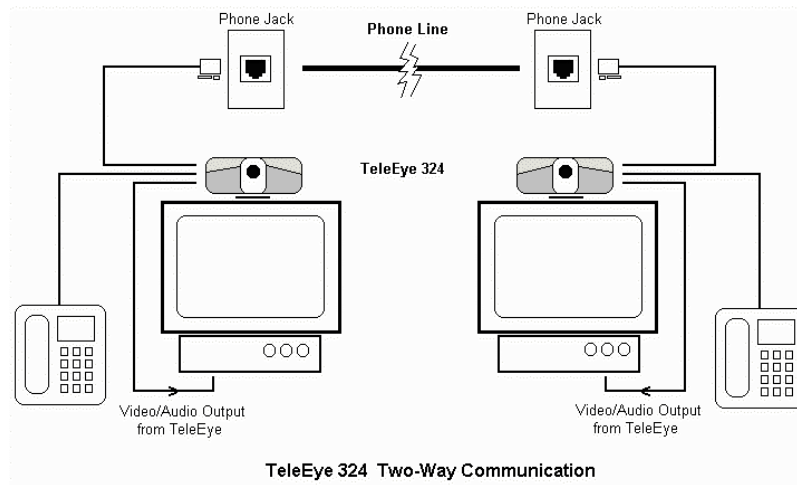
- V.34 Analog Modem
- H.263 Video Codec
- G.723.1 Audio Codec



Video Conferences over POTS

- The H.324 standard is an attempt at videoconferencing over the **Plain Old Telephone System (POTS)** using analog modem technology. Since for a bidirectional transmission between two modems (V.34 or V.90) the maximum achievable speed is only 33.6 kbit/s, both video and audio have to be heavily compressed.
- For audio transmission the extremely efficient G.723.1 codec with compressed speech rates of 5.6 and 6.3 kbit/s was developed. The remaining bandwidth allows the H.263 codec to transmit sub-QCIF pictures at a not very exciting rate of 10-15 frames/s.
- H.324 based telephony has rather remained a gadget for the consumer market that up to now hasn't found a large acceptance.

H.324 Videophone over POTS II



TeleEye 324 Two-Way Communication

Modem Speed: 33.6 kBit/s {
- Audio: 5.3 / 6.4 kBit/s
- Video: sub-QCIF / 15 Frames/s

Overview of ITU-T Recommendations for Multimedia Communication

	H.320	H.323	H.324
Approval Date	1990	1996 / 1998	1996
Transport Medium	ISDN	IP	POTS
Video	H.261 / H.263	H.261 / H.263	H.261 / H.263
Audio	G.711 / G.722 G.728	G.711 / G.722 G.728 / G.729 G.723.1	G.723.1
Control	H.230 / H.242	H.245	H.245
Multiplexing	H.221	H.225.0	H.223
User Data	T.120	T.120	T.120