Video Conferencing Technologies: Past, Present and Future

Jose Joskowicz [1]

[1]Universidad de la Republica

December 7, 2023

Abstract

This paper describes the historical trajectory of video conferencing systems, spanning from their earlier mechanical and analog origins in the 1920s to the sophisticated IP-based services delivered from the cloud in the 2020s. Each technological age is examined, highlighting the technical and functional aspects that characterized its evolution. Commercial landmarks of each age are presented providing a comprehensive overview of the most prominent offerings at pivotal moments in the timeline. By examining the past and speculating on the future, this paper aims to provide a holistic understanding of the development, current state, and forthcoming trends in video conferencing technology.

Video Conferencing Technologies: Past, Present and Future

Jose Joskowicz
Universidad de la República, Uruguay
josej@fing.edu.uy

Abstract— **This paper describes the historical trajectory of video conferencing systems, spanning from their earlier mechanical and analog origins in the 1920s to the sophisticated IP-based services delivered from the cloud in the 2020s. Each technological age is examined, highlighting the technical and functional aspects that characterized its evolution. Commercial landmarks of each age are presented providing a comprehensive overview of the most prominent offerings at pivotal moments in the timeline. By examining the past and speculating on the future, this paper aims to provide a holistic understanding of the development, current state, and forthcoming trends in video conferencing technology.**

I. INTRODUCTION

Video telephony and video conferencing systems were initially conceptualized and developed in the early 20th century. The first video call was made in 1927, even before the first television signal was broadcast on the air. However, it took almost a century, until the end of the second decade of the 21st century, for video calls and video conferencing to become part of everyday life.

This paper describes the development and evolution of video conferencing systems, from their mechanical and analog beginnings in the 1920s, to the advanced IP services offered from the cloud, in the 2020s. For each technological generation, the technical and functional aspects are presented and the most outstanding commercial offers at any given time are indicated. Finally, some technologies that will be incorporated and developed in the following years are ventured.

II. THE BEGINNINGS: ANALOG AND MECHANICAL SYSTEMS

The first public presentation of a video call was made on April 7, 1927, between the cities of Washington and New York, in the United States, through telephone circuits. This first technology demonstration consisted of a one-way video call. At one extreme, Walter S. Gifford, president of the American Telephone and Telegraph (AT&T) company, was in New York, using a telephone and image-receiving equipment, while Herbert Hoover, the U.S. Secretary of Commerce, was in Washington, D.C., in front of equipment capable of capturing and transmitting images. In the same presentation, immediately after the Washington-New York communication, a second video call was made, this time by radio transmission (i.e., wirelessly), between the 3XN experimental station in

Whippany, New Jersey and the same receiver in New York. In this demonstration, Engineer E. L. Nelson was seen speaking, explaining how the device worked.

The day after the demonstration, the New York Times headlined *"Far-off speakers seen as well as heard here in a test of television"*, *"Like a photo come to life"* the newspaper said prominently [1]. Figure 1 and Figure 2, published together with the article, show the protagonists of the first video call. Figure 3 shows Whippany's studies during the second wireless video call. It should be noted that the first television broadcast took place in 1928, a year after this demonstration.

Regarding the practical and commercial utility of these systems, Gifford mentioned [1]: *"What its practical use may' be I shall leave to your imagination. I am confident, however, that in many ways and in due time, it will be found to add substantially to human comfort and happiness"*.

The design and construction of this video telephony system was carried out by a team led by Engineer Herbert E. Ives [2]. While most of the basic concepts used by the system were already known, many technological and practical problems had to be solved in order to achieve proper operation. The June 1927 issue of "The Wireless World" gives a very good overview of the system [3]. The first aspect to consider is the transformation of an image into electrical signals. A solution to this problem had been proposed and patented in Germany several decades earlier, in 1884, by Paul Gottlieb Nipkow [4], using a spiral-shaped perforated disk, as shown in Figure 4. For the design of the videophone, engineers Frank Gray, J. W. Horton, and R. C. Mathes used such a disk [5][6]. Figure 5 shows a conceptual outline. A beam of light is focused to illuminate a limited area above the moving holes of the disk. A frame in front of the disk allows light to pass through only one hole at a time. A lens in front of the disk focuses an image from this moving aperture onto the object or person whose image is to be transmitted. As a result, the person or object is scanned completely in a series of successive parallel lines by a rapidly moving point of light. In the design of this video phone, it performed 18 full scans per second (the equivalent of 18 frames per second). As the point of light illuminates the person or object, the light is reflected and received by photoelectric cells, placed in front of the person or object being observed. The current output of the photoelectric cells is proportional to the light received, achieving the objective of obtaining an electrical signal that varies depending on the intensity of the light to be transmitted. Figure 6 shows a picture of the equipment used.

In the receiving station, a disk similar to that of the transmitting station, also with small holes arranged in the shape of a spiral, rotates in synchronism with that of the transmitting station. The observer looks at a small rectangular hole or frame in front of the disk, as seen in Figure 5. This frame only allows to see one hole of the disk in the field of view. As the disk rotates, the holes pass through the frame one after the other in a series of parallel lines, each displaced a little from the last, until in one full revolution the disk covers the entire field. Behind the disk is a special neon lamp. In this lamp, the cathode is a flat metal plate with enough area to completely fill the field defined by the frame of the front of the disk. In this way, the image is projected onto a small area, 5 cm × 6.3 cm, that can only be observed by one person. Figure 7 shows a receiving station and the detail of the neon lamp used.

Figure 1. Walter S. Gifford, chairman of AT&T at the first public demonstration of television. Mr. Gifford talks to Secretary Hoover in New York and is able to see Mr. Hoover on the screen immediately in front of him. (Mr. Hoover is on the right.) [1].

Figure 2. Herbert Hoover, Secretary of Commerce, has become the first televised personality. He is seen in front of the device located in Washington, broadcasting his voice and face to the audience located at Bell Labs in New York [1].
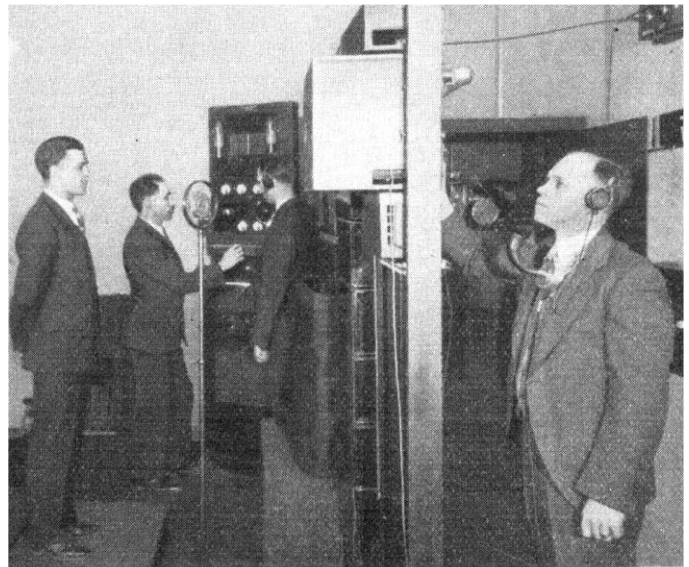

Figure 3. Broadcast equipment at studios in Whippany, New Jersey [9].

For presentation to a large audience, the system includes another device, 61 cm × 76 cm, consisting of a neon tube, bent into 50 parallel sections, as shown in Figure 8. The tube has an inner electrode and 2500 outer electrodes, 50 for each section. A mechanical switching system sequentially energizes each of the external electrodes, generating illumination proportional to the current received, which, in turn, is proportional to the light intensity of the corresponding point on the transmitted person or object. This screen, in today's terms, had 2500 pixels.

The New York Times reporter mentioned: *"When the television picture thrown on a screen two by three inches (5 cm × 6.3 cm), the likeness was excellent. It was as if a photograph had suddenly come to life and begun to talk, smile, nod its head and look this way and that. When the screen was enlarged to two by three feet (61 cm × 76 cm), the results were not so good"* [1].

Another problem to be solved for the system was the synchronism between the transmitter disks and the receiver's two screens. Engineers H. M. Stoller and E. R. Morton worked on this aspect, designing a control mechanism that allowed the phase shift between the emitter and receiver disks to be adequate (less than 4.3 arcminutes, according to the theoretical design). It was decided to build synchronous motors with 120 pairs of poles, at a frequency of 2125 cycles per second, using a reduction factor of 120, thus achieving a rotational frequency of 17.7 revolutions per second [7]. The synchronism reference was taken from the receiving equipment and distributed to the neon screen (in the same place as the receiver) and to the remote transmitter system.

A diagram of the telephone circuit transmission system is shown in Figure 9 [8]. One line was used to send the video (with a backup line), one line for synchronism signals, one line to send the audio, and another line to receive the audio from the remote station. The video was sent in baseband, with frequencies between 10 Hz and 20 000 Hz. The synchronism was transmitted in a modulated form. A primary carrier frequency of 1575 kHz was selected for the wireless

transmission for the video, and a secondary frequency of 1545 kHz for the audio channel [9].

On May 24 of the same year, 1927, Engineer John L. Baird made a demonstration similar to that of Ives, in this case through telephone lines between the cities of Glasgow and London, in England. Professor E. Taylor Jones, of the University of Glasgow, was present during the demonstration, and published the following notes: *"The receiving apparatus was set up in a semi-darkened room, the lamp and shutter being enclosed in a case provided with an aperture. The observer looking into the aperture saw at first a vertical band of light in which the luminosity appeared to travel rapidly sideways, disappearing at one side and then reappearing at the other. When any object having 'contrast' was placed in the light at the sending end, the band broke up into light and dark portions forming a number of 'images' of the object. The impression of side way movement of the light was then almost entirely lost, and the whole of the image appeared to be formed simultaneously. The image was perfectly steady in position, was remarkably free from distortion, and showed no sign of the 'streakiness'.... The size of the image was small, not more than about two inches (5 cm)... The amount of light and shade shown in the image was amply sufficient to secure recognizability of the person being 'televised', and movements of the face or features were clearly seen"* [10].

The ideas and technologies used by Baird and Ives were similar, although the implementation was different. In his early systems, Baird used a system with multiple disks [11], as shown and explained in Figure 10. In later systems designed by Baird, line sweeps were done vertically, from top to bottom, and from right to left. Figure 11, taken from the video [12], shows the point of light on the person's face, the vertical line it produced when the disk rotated at its nominal speed, and the image obtained at the receiver.

It is worth mentioning that, in the following year, in 1928, Baird performed the first intercontinental television broadcast, between London and New York [13].

Three years after the first one-way video call, in 1930, Ives' team introduced a two-way video telephony system [14]. Its principle of operation is the same as that used in the 1927 demonstration, with the following improvements: 72-holes disks were used, instead of the 50 of the previous system; A blue light lamp was selected at the sending station, as photoelectric cells were much more sensitive for that wavelength; the power of the lamps at the receiving station was increased and the video transmission bandwidth was increased to 40 kHz (twice that used in the previous system). Ives patented the system in April 1930 [15].

Transmission between sites required five telephone circuits: two for video signals, two for audio signals and one for synchronism signals. In addition, a backup circuit and two other circuits were available to turn on a remote monitoring light that indicated that the cabin was occupied. Figure 12 shows a conceptual schematic of the entire system. Figure 13 shows a detail of the booth specially designed for interlocutors.

Regarding the new system, Walter S. Gifford said: *"Although the research and refinement of the last three years has led to a great improvement and simplification of the equipment required for television, it is necessarily complicated and still expensive. It must be attended by a technician and requires a large number of devices. This is due to the scientific requirements that are essential for a satisfactory television transmission. Thus, although considerable progress has been made on the technical side of the issue, the commercial potential of television is still uncertain"* [16].

On May 19, 1932, a two-way video system, based on Baird's previous work, was presented in Paris. In this system, the disks spun at 750 revolutions per minute, producing 12.5 frames per second. According to a report at the time, *"The features are easily recognizable and the play of expression on the face is remarkably clear. The movement of the lips can easily be followe*d" [17].

On March 1, 1936, Germany's Minister of Postal Services, Eltz von Rubenach, inaugurated the two-way television and telephone service of the German Post Office, connecting Berlin and Leipzig, at a distance of about 190 km, by cable. The service was promoted and marketed at 3 Marks 50 Pfennigs for three minutes of communication. According to the journal Nature, *"The quality of the pictures produced appears to be good, 18O-line definition and 25 frames a second being used. The head and shoulder image of a person is clearly produced. The effect is comparable to a small size projection of a substandard cinema film.... Details like the bands of a wrist-watch or a ring on the hand holding the telephone are said to be clearly visible. The apparatus used in Berlin was constructed by the German P.O. laboratory and that used in Leipzig by the Fernseh-Aktiengesellschaft of which Baird Television Ltd. holds a quarter of the shares"* [18]. Figure 14 shows the cabin used.
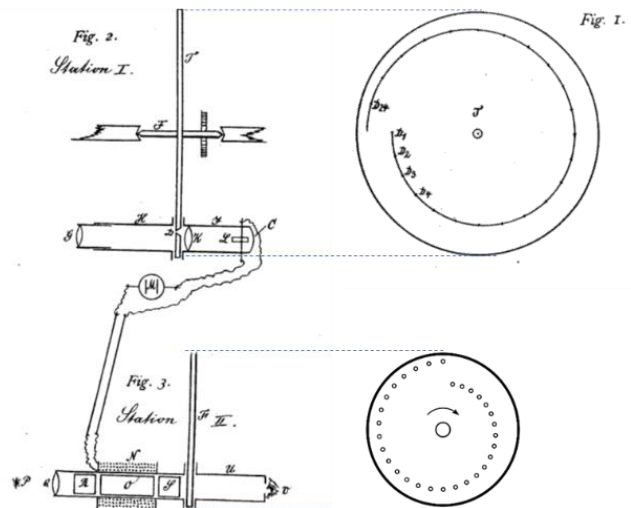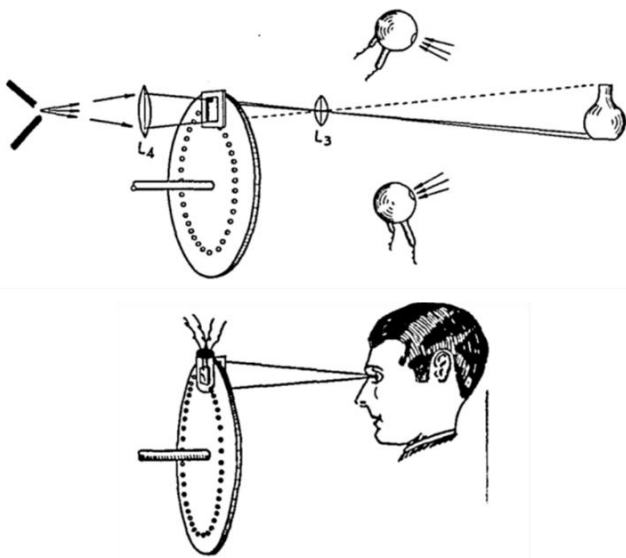

Figure 4. Schematic of the Nipkow disk, based on [4].
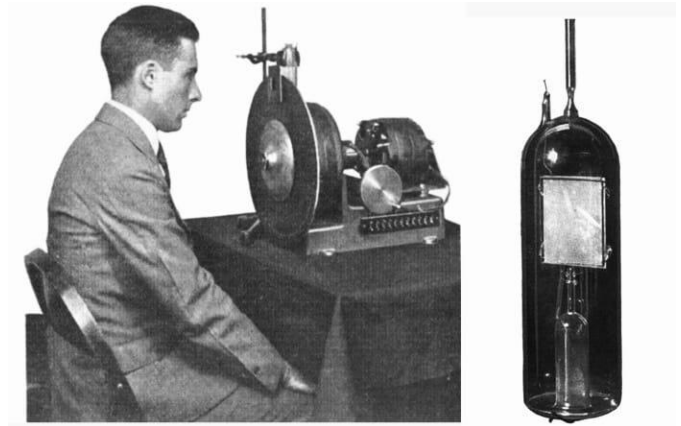
Figure 7. A neon lamp (shown in detail on the right of the figure) works with a current proportional to the light intensity of the original image and illuminates the disk, which has a series of small holes. As the holes traverse the field of view, the intensity of the lamp varies, and the observer receives the original image [5].

Figure 5. Up: Use of the Nipkow disk to project a beam of light as a small dot onto the object to be transmitted. The reflected light is received in several photoelectric cells.
Bottom: Apparatus to reconstruct the image. The neon lamp behind the disk is powered by a current proportional to that generated by the photoelectric cells. Based on [6].
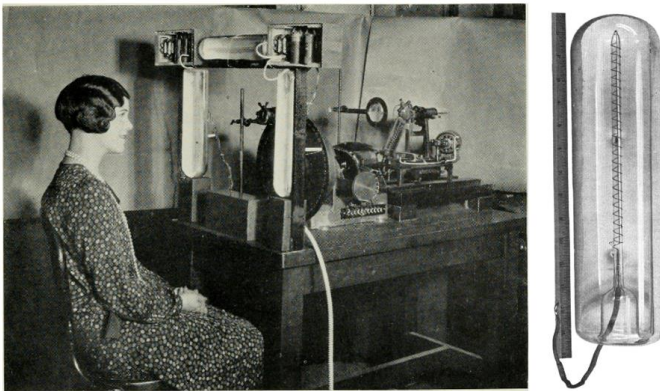
Figure 6. The disk rotates 18 times per second, using a synchronous motor. The disk has a spiral of holes, each of which allows the light generated by an arc lamp to pass through, and projects it onto a moving point of light on the person. The light reflected by the person is collected by three large photoelectric cells (see detail in the right of the figure), located at the top and sides of the frame [5].
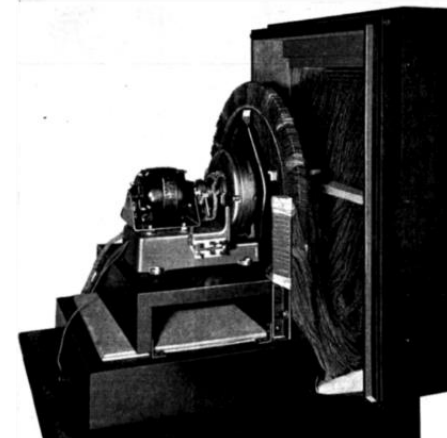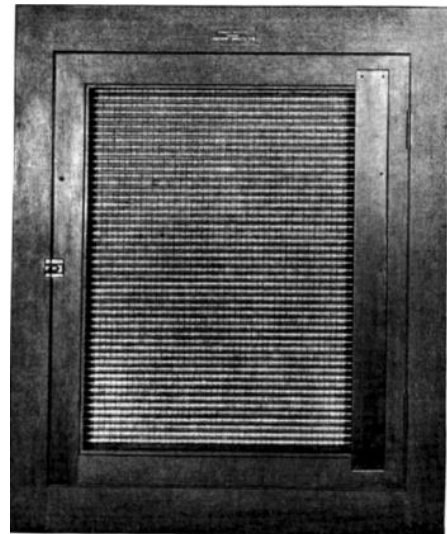
Figure 8. Up: Screen formed by a neon tube, folded into 50 rows and with 2500 electrodes. Bottom: Switching and wiring system. The current is distributed through 2500 wires to successive electrodes, synchronized with the rotation of the emitter's disk [5].
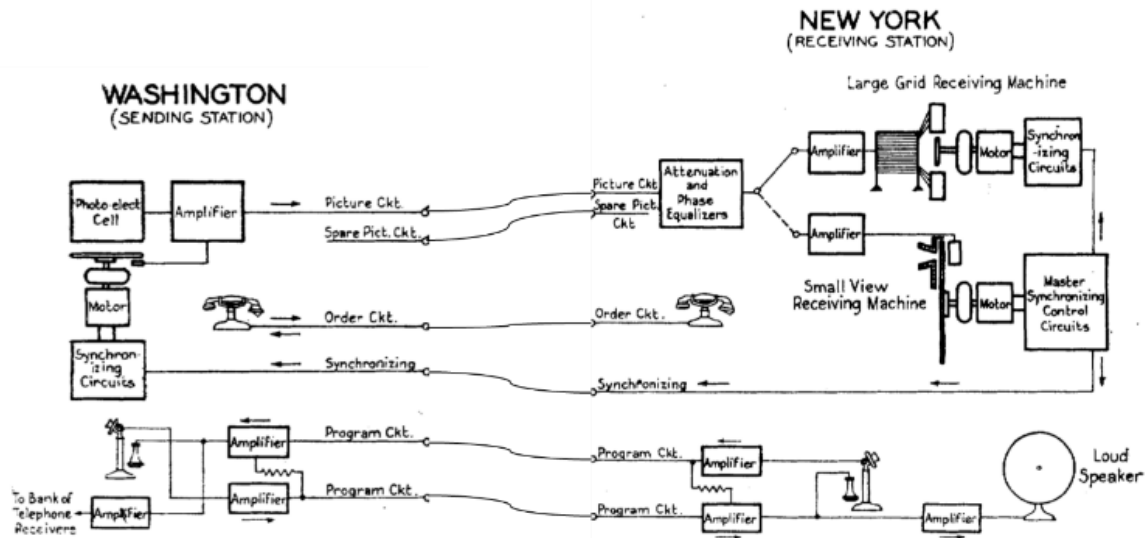
Figure 9. Schematic of the complete system, based on [8], including the transmitting station in Washington and the receiving station in New York. Communication between the two sites was carried out by telephone lines: one line to send the video (with a backup line), one line for synchronism signals, one line to send the audio and another line to receive the audio.
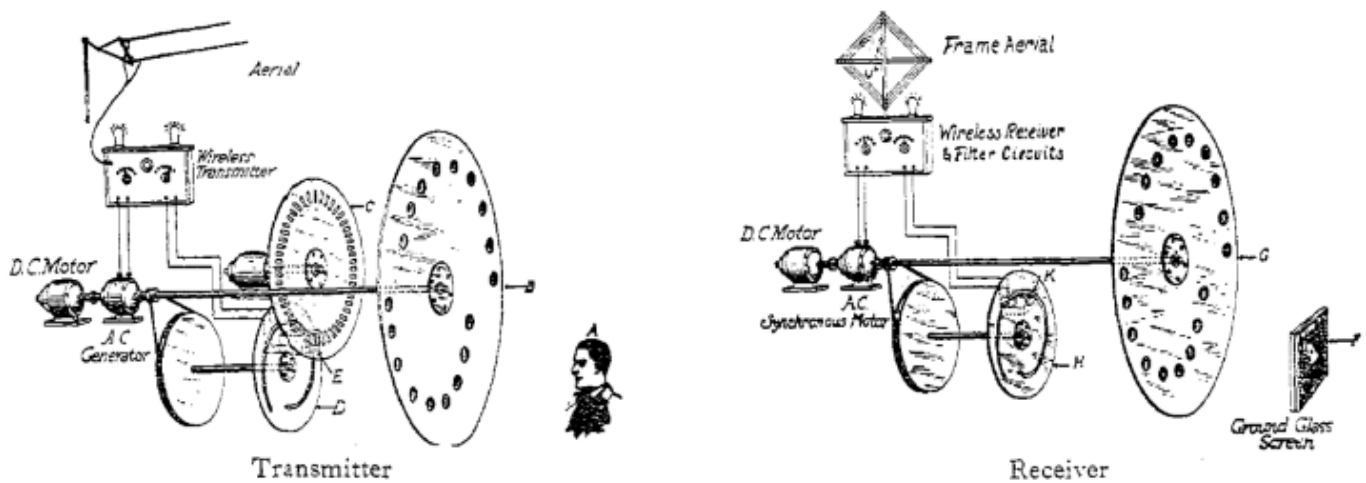

Figure 10. Bird's transmitter (left) and receiver (right) system. By means of a disk B containing lenses in staggered formation, a series of strips of the image is passed through E, the opening of the light-sensitive cell, after being interrupted by the slotted disk C. The spiral disk D makes a finer grain to the image by subdividing the strips. Based on [11].

Figure 11. Scanning by Baird's system. On the left, a point of light with the disk fixed. In the center, a vertical sweep line with the disk spinning. On the right, image obtained from the receiver [12].
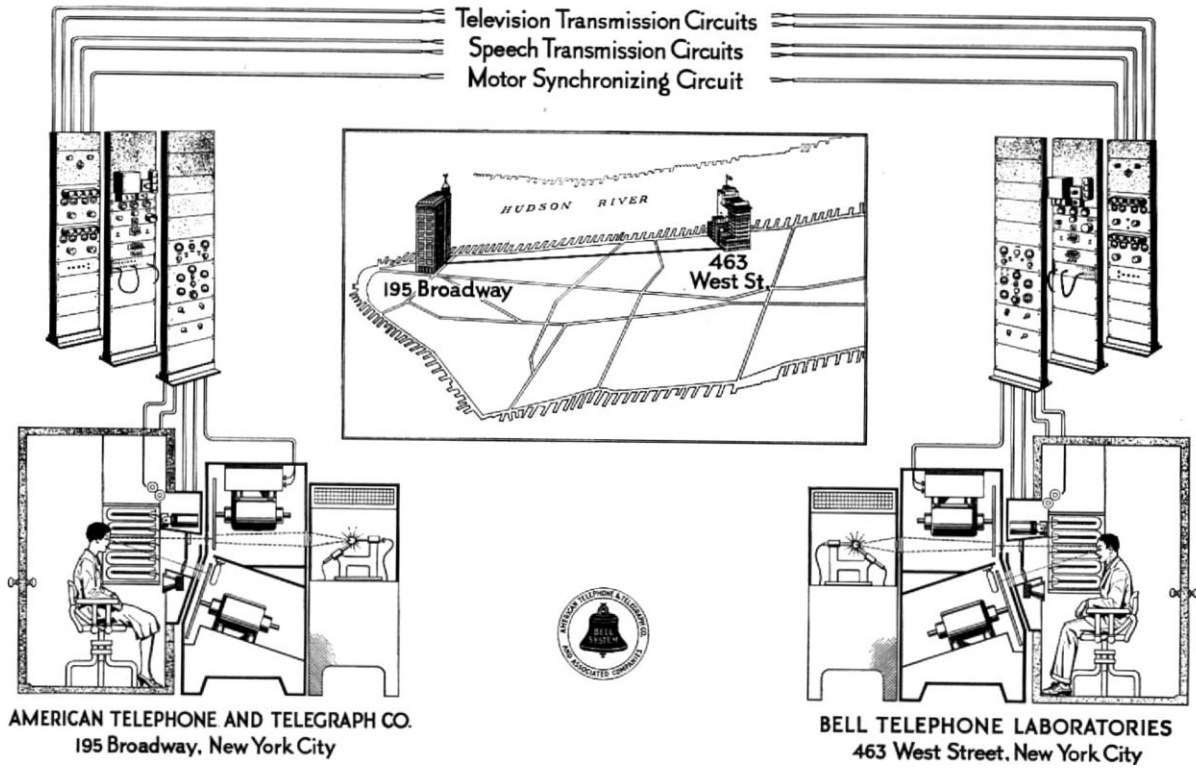
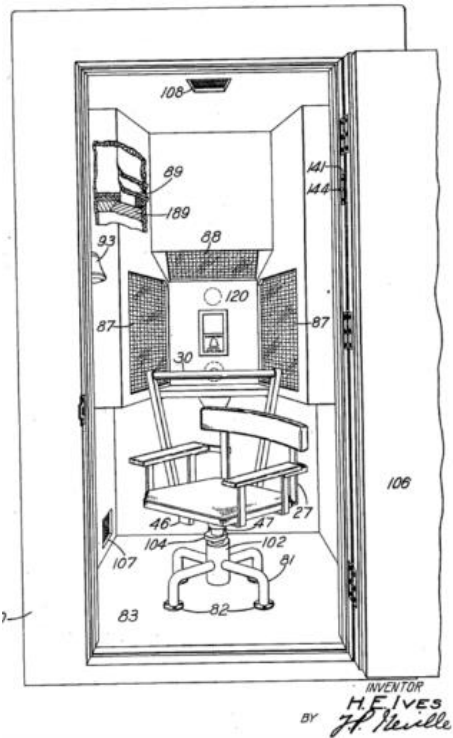Figure 12. Representation of the two-way video call system [14].

Figure 13. Left: Design of the system's user cabin, according to Ives' patent [15]. Right: Photo of the interior of the cabin [14].
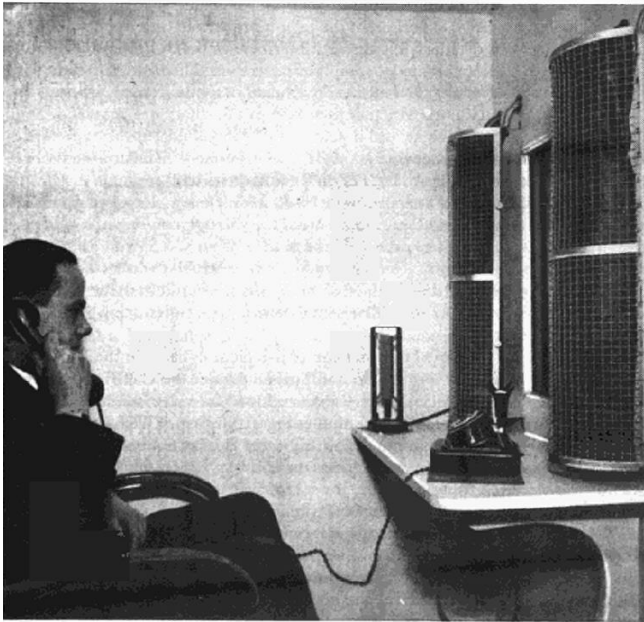
achieve these design goals, solid-state devices were used, with the exception of the capture and display image tubes. The assembly consisted of three parts: a display unit, a control unit, and a power supply. The first two were within easy reach of the user, while the power supply was out of sight, installed under the desk. The display unit contained an image tube, a camera, and a speaker. The camera was specially designed, based on a flat array of reverse-polarized silicon photodiodes, which were accessed by a low-energy scanning electron beam [24]. The control unit had a telephone handset, a speakerphone, and several buttons, including the DTMF touch-tone telephone keypad. Using these controls, the user could select whether or not to send their video, or a self-view. The screen was designed in portrait format, 11.1 cm wide by 14.6 cm high. 275 lines per frame were used, at 30 frames per second (separated into two interlaced fields). The horizontal frequency was 8250 Hz [25]. The detail of the different components of the first version of the Picturephone can be seen in Figure 17, Figure 18 and Figure 19.

Figure 14. Video telephony service in Germany, prior to World War II [19].

III. THE ANALOG ELECTRONIC AGE

Over the next thirty years, between 1930 and the 1960s, television developed and became popular. However, video phone or videoconferencing systems did not make significant advances.

Some audio conferencing experiences were carried out during this period, related to educational environments. In 1939, the first use of an audio conference with many participants was recorded in Iowa, United States. Dr. Winterstein had a system installed that allowed students to be connected from their homes to a classroom where classes were held with other students. Students who were at home could intervene in the class, talking to both the teacher and their other classmates [20]. Other similar experiences were reported in the 1940s-1950s [21].

At the 1964 "New York World's Fair", Bell Labs presented a prototype of the "Picturephone" system [22][23]. The first demonstration was made on April 20, making a video call from New York to Disneyland Park, located in California. This prototype served not only to make a public presentation of the new system that would be marketed in a short time, but also to evaluate the preferences of users. Through surveys of those who passed by the exhibition, they were asked about what specific uses potential customers considered most important, what image quality would best serve these uses, what functions are preferred, what controls are needed to activate them, how often the service would be used, among other aspects [22]. Figure 15 shows the cover of the 1964 Bell Telephone Magazine showing laboratory tests of the new device. The first call for this new commercial system was the same year, a few months later, as shown and described in Figure 16.

The overall goals of the picturephone's design included an attractive style, small enough to be used on a desk or table and with as few controls as possible to make it easy to use. It also had to be low power consumption and heat dissipation. To

Figure 15. Cover of "Bell Telephone Magazine, Spring 1964", showing laboratory experiments with the Picturephone [22].

Five years later, in 1969, a new version of the Picturephone was designed [26]. The control unit was simplified, and the format of the display unit was changed to a rectangular one, the screen size was slightly enlarged, the camera was placed just above the screen and the possibility of connecting it to the data output of a computer was included. A photo of the new style of the Picturephone is shown in Figure 20. The new equipment was launched on the market in 1970 [27]. At this new inauguration, on June 30, 1970, in the city of Pittsburgh, an executive of the Bell Telephone System indicated that the economic and social impact of this new face-to-face communication would be equal to that produced by the introduction of the telephone voice conversation in 1878 [28]. Lawrence J. Barnhorst, vice president and general manager of Bell Telephone Company of Pennsylvania, indicated at the inauguration that by 1975, 100,000 Picturephone videophones would be in use. A report, published in Bell Telephone Magazine [29], predicted that by 1985 there would be another three million Picturephones in the United States (which ultimately did not happen).

Figure 17. Display unit and camera, exterior view and interior detail [25].

Figure 16. First commercial call from the Picturephone: *"The future is here: the Picturephone service is now here"* between New York, Chicago and Washington, DC. On the opening day of the service, June 24, 1964, the first call was made from the National Geographic Society building in Washington, D.C., from Mrs. Lyndon B. Johnson to Dr. Elizabeth A. Wood, a scientist at Bell Telephone Laboratories. Above, one of the most striking benefits of the new service was demonstrated on opening day when Laura Rabinowitz, a 15-year-old deaf student, and 14-year-old Howard Mann (on screen), communicated by lip-reading during the first call with Picturephone [22].
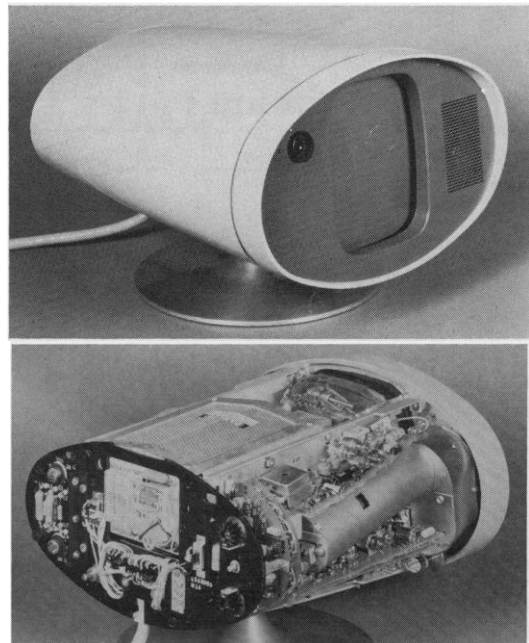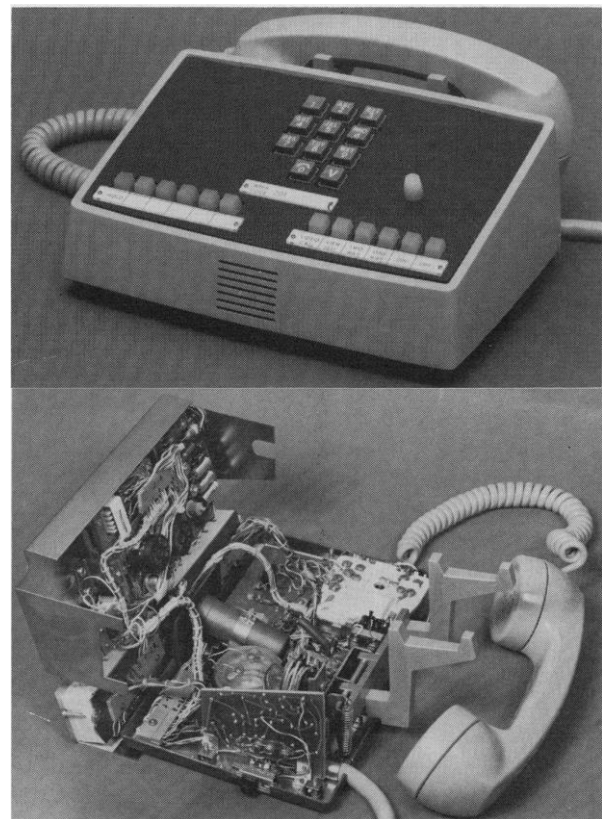

Figure 18. Control unit and telephone, exterior view and interior detail [25].
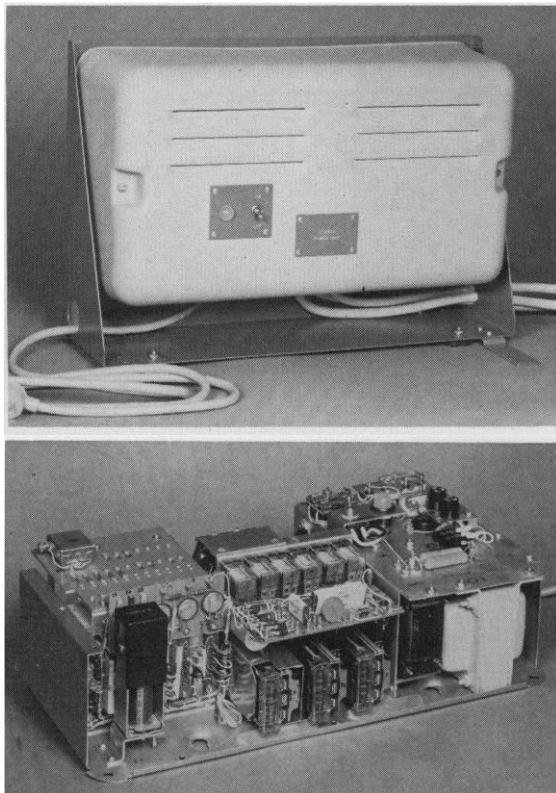
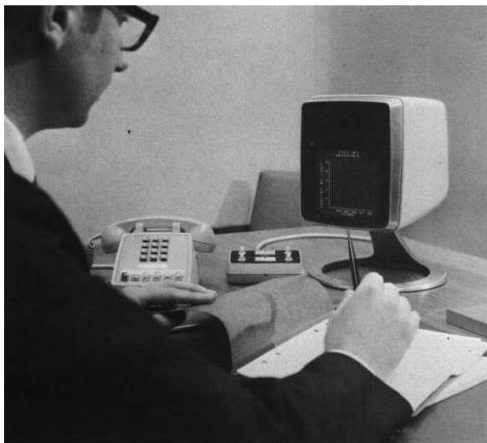Figure 19. Power Unit, exterior view and interior detail [25].

Figure 20. Picturephone "Mod II" [26].

A conceptual diagram of the Picturephone's design is shown in Figure 21. It was natural at the time to include the Picturephone service associated with existing phone services. Video calls were indicated by preceding the destination phone number with the # sign, from the DTMF telephone keypad. The resulting video signal from the Picturephone was 1 MHz. Existing telephone pairs between users' homes or offices and telephone exchanges were used for transmission. It was decided to use one pair for incoming video, one for outgoing video, and one pair for phone audio. The existing telephone exchanges were designed to switch audio, with a much more limited bandwidth of 4 kHz, so it was necessary to design a new switching and transmission system for video calls. Figure 22 shows the conceptual schema used for the switching and transmission of the Picturephone service [30]. Video could be switched on new "crossbar" public telephone exchanges, with

analog technology, designed to switch signals up to 1 MHz [31]. To achieve uniform attenuation across the bandwidth, it was necessary to include amplifiers with equalization [32] at different points in the system, as shown in Figure 23.

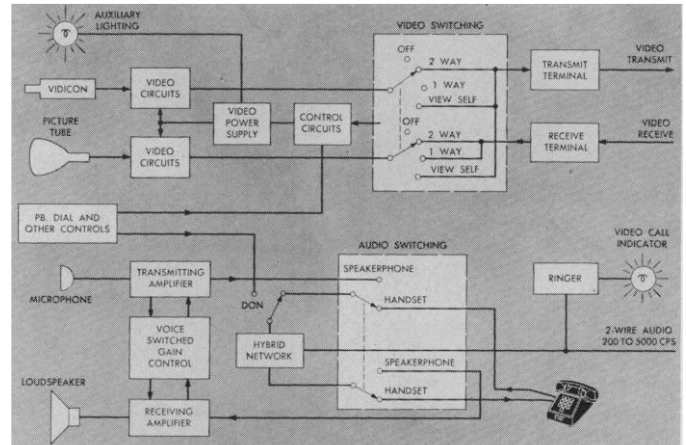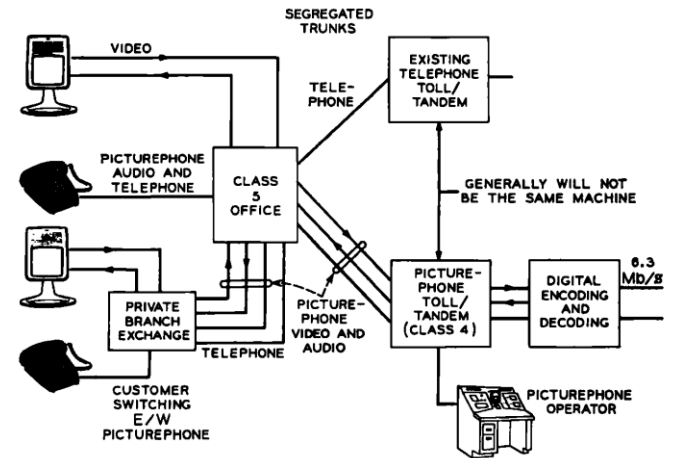Figure 21. Picturephone Conceptual Block Diagram [25]

Figure 22. Picturephone switching and transmission system. [30].
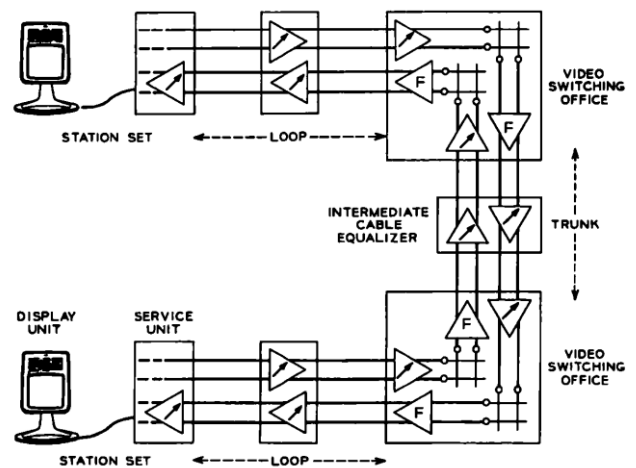

Figure 23. Picturephone Transmission System [32].

For transmission between public exchanges, generally distant, a 6.3 Mb/s digital system was used. A 3-bit Differential Pulse Code Modulation (DPCM) encoding was used for this purpose [33], which is equivalent to having up to

8 possible transmission levels. Figure 24 shows the result of the scanning on an image taken from Picturephone.

Figure 24. Left: Analog image. Right: Digitized image with 3-bit DPCM technique [33].

The technical design was challenging, but a functional final product was achieved, aesthetically very good, and with good image quality. However, the service did not have the expected commercial success. Two years after its second commercial launch, in 1972, and with an estimated investment of 500 million dollars [34], there were only 32 videophones contracted in the city of Pittsburgh. Clearly, it would not even come close to the 100,000 videophones predicted for 1975 in that city. The number of devices in Chicago peaked at 453 in early 1973, even though AT&T reduced the price of service and talk time [35]. In July 1975, only 76 customers still had the service, and by 1977 there were only nine customers left in Chicago [34]. The service was discontinued.

In the 1970s in Europe, some analogue videoconferencing systems were developed, including Confravision in the United Kingdom (basically two television studios connected by broadcasting lines) [36] and Visioconference in France. Figure 25 shows the Confravision system at work.

Figure 25. Videoconferencing with the Confravision system, in the United Kingdom [36].

IV. THE DIGITAL AGE

The techniques needed to digitize and encode the video signal efficiently had been developed since the 1940s. The first proposal for digitization, using Pulse Code Modulation (PCM) was presented in 1949 and published in 1951 [37]. In this work, the encoding with up to 5 bits per sample was evaluated, showing acceptable results in image quality. Entropic coding was proposed in 1949 [38] and developed in 1952 [39]. This technique consists of representing the most frequent values with few bits, and the less frequent values with more bits, generating a type of variable-length code (VLC). By knowing the type of signal to be encoded, and by properly selecting the form of representation, it is possible to minimize the total number of bits needed to digitize a message. Entropic encoding was first applied to video coding several years later, in 1971 [40].

From the beginning of television, it was clear that the signal to be transmitted was highly redundant, both in the internal information corresponding to each frame, and in the succession of frames. As early as 1929, a patent was filed proposing to transmit only, in analogue form, the differences between each frame and the previous one [41]. Predictive digital modulation techniques, initially known as Differential Pulse Code Modulation (DPCM), were invented in 1950 [42], and were first proposed to be applied to video in 1952 [43]. Until the early 1970s, several DPCM techniques were developed to optimize coding within each frame [44]. For example, two consecutive lines within the same frame can be very similar and therefore it is more efficient to encode only the differences. In the mid-1970s, DPCM also began to be used in the time domain, i.e., for prediction between frames [45][46][47]. Two consecutive frames of a video signal are typically very similar to each other, especially in video telephony or video conferencing applications, where the typical image is of the "head and shoulders" type and with little movement. With this in mind, various techniques make it possible to use the information in one frame to predict the information in the next frame, and to encode only the difference with the prediction, as shown in Figure 26, published in 1969.
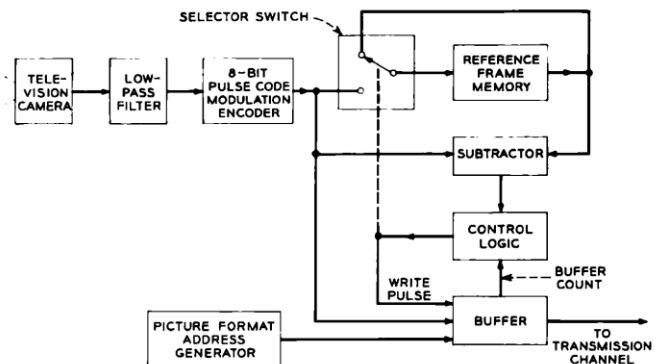

Figure 26. Differential video encoding, considering the similarity between consecutive frames [45].

Digitized video frames are encoded with a sample matrix of brightness and color information, with sample spacing capable of reproducing the greatest detail at each point. This information can also be represented in terms of spatial frequency, in a manner equivalent to the frequency representation of time-varying signals but extended to two-dimensional spaces. In this way, the image can be analyzed or processed using techniques similar to the Fourier transform. In scenes with little detail there will be a predominance of low-frequency information. Thus, while in the spatial domain the signal energy is evenly distributed among the samples, in the transform domain the energy is concentrated in a few samples. This observation led to the development of transform-based coding in the late 1960s. In 1968 it was proposed for the first time to use the Fourier transform to encode video [48]. Other types of transforms, more efficient than the Fourier transform, were proposed in the following years. In 1974 it was proposed to use the Discrete Cosine Transform (DCR) [49], which has several advantages and better performance than the previous ones. DCT continues to be used in video encoding to this day. The transform produces a matrix of almost uncorrelated coefficients. For scenes with little details, the energy or variance of the samples is not evenly distributed, but is concentrated in the lowest frequency coefficients. These coefficients are assigned a larger number of bits to encode and quantify, with appropriate precision. Coefficients that represent higher frequency can be assigned fewer bits and quantified approximately. Higher frequency coefficients can be neglected, and don't need to be coded.

Applying these transforms to full frames proved computationally inefficient and costly. But applying it to small squares (e.g., 8×8 or 16×16 pixels) produces better results and requires less computing power. The concept of using small blocks to apply transforms to them was introduced in 1969 [50] and applied from that moment on to the encoding of images and video.

Beginning in the mid-1970s, key elements of DPCM coding were merged with the transformed coding to create a hybrid coding, which began to be used in the early 1980s [51]. This type of hybrid encoding, combined with predictive techniques, marked the digital era of teleconferencing in the late 1970s and early 2000s.

In Japan, NEC developed a series of commercial video conferencing products, using digital techniques with temporal prediction between frames [52][53]. The product line became known as NETEC, and they were marketed starting in 1976. The NETEC-6 model (shown in Figure 27) required between 6 Mbps and 8 Mbps of transmission bandwidth. The NETEC 6/3 models could operate at 3 Mbps and the NETEC-X1MC model at 1.5 Mbps. Nippon Telegraph and Telephone (NTT) designed the system called TRIDEC in 1977 [54] and commercialized it in 1979.

In 1977, seven European countries (Belgium, France, Holland, Italy, Sweden, United Kingdom and West Germany) began their own developments in digital video encoding, under the group called Cooperation in Scientific and Technical Research (COST 211). This was a collaborative research project, looking for coding techniques that use redundancy reduction, with the aim of transmitting videoconferences at 2

Mbps [55]. It was hoped to have a common videoconferencing system for Europe, operating at 2 Mbps, by 1984.

In the early 1980s, the first digital transmission systems began to be deployed. In the United States, in 1982, AT&T began offering the service called High Speed Switched Digital Service (HSSDS) [56]. The service offered digital communication, at 3 Mbps, through digital switching centers deployed on the AT&T network. The 3 Mbps was achieved by combining two DS-1 links, each with a speed of 1.544 Mbps. The switching system offered was manual, by prior telephone reservation. An operator took the reservation, and the necessary interconnections were planned in the operator's internal network, to make and maintain the connection for the requested period of time. The first service offered and delivered through this new digital network was the Picturephone Meeting Service (PMS) [57]. In July 1982, the New York Times headlined *"Picturephone Service Begins"* [58]. The AT&T Company inaugurated the new video conferencing service by making the first call between New York and Washington. The note clarified: *"The new service is vastly different from the Picturephone shown at the 1964-65 New York World's Fair. That version, which never caught on, had desk-top screens for use by individuals, not groups, and its technology was less sophisticated.... With PMS, customers may either rent public rooms to use the new service or build rooms on their own premises. A Picturephone room will have color cameras, microphones and monitors. Slides and charts may be transmitted. The room will also offer a copy machine to reproduce images shown on the incoming monitor and a videotape recorder to record either incoming or outgoing picture and sound"*. Prices for the use of the service varied depending on the location of the rooms, ranging from $1300 to $2500 per meeting.
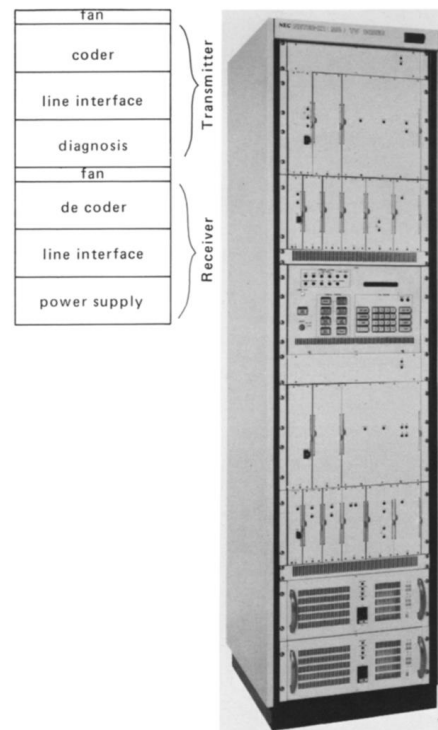

Figure 27. Codec NETEC 6 [53].

The service offered interactive, two-way video conferencing between two compatible conference rooms, connected to the HSSDS network. Conference rooms could be deployed in the private premises of companies or in convenient locations for public use. Each videoconference had to be booked in advance, due to the way the HSSDS service worked. The service used digital encoding of audio and video, at a bitrate compatible with the HSSDS transport service used. Conference rooms for the PMS service included the use of multiple microphones, speakers, monitors, and video cameras. Up to 12 participants could be seated in the conference room (6 in the front row and another 6 in the back). Video coverage of the conference room is provided with three close-up cameras, one panorama camera, two graphic cameras, and one multipurpose camera. Each close-up camera focused on a couple of speakers at the table (and the two speakers seated in the back). These cameras were automatically selected when the speakers at the table spoke. On the other hand, the panorama camera provided a wide-angle view of the conference room and participants, and could be automatically selected when multiple people, or the speakers in the second row, were speaking. Figure 28 shows the intended layout of the room. The system included a control panel for users, so they could manipulate all the equipment.

The core elements of the PMS system included a room controller and a TV processor, as seen in Figure 29. The room controller connected to all elements of the room on one side (cameras, monitors, control panel, etc.) and managed incoming and outgoing video and audio signals on the other. The TV processor interconnected the DS-1 links of the HSSDS with the analog audio and video to the room controller.

PMS services used sophisticated technology and required a very large investment to get up and running. Equipping a conference room involved an initial investment that could range from $120,000 to $500,000 [59][60]. Additionally, the fixed monthly cost for renting the service was more than $10,000, in addition to the cost per use of each videoconference, which was in the order of $1000 per hour. It is not surprising that, with these prices, the service did not become popular.
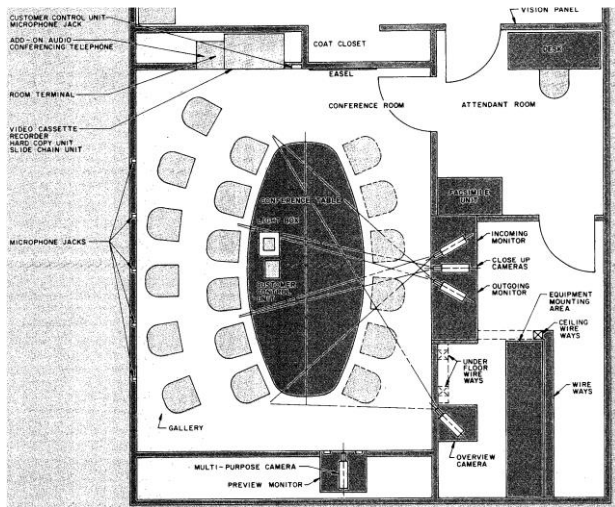

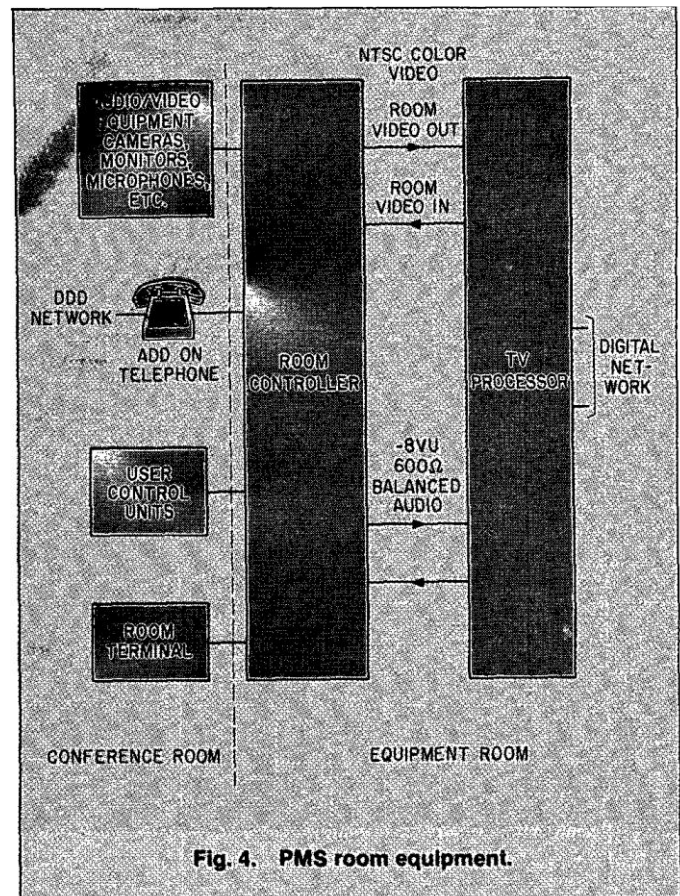Figure 28. Video conferencing room for PMS service [58].

Figure 29. PMS room equipment [58].

During the following years of the 1980s, several ventures related to videoconferencing systems were carried out. The company Compression Labs Inc (CLI) introduced a video conferencing system in 1982. The codec used intra-frame encoding using the DCT transform, applied in blocks of 16×16 pixels. They used bitrates of 1.5 Mbps and developed techniques that allowed them to operate at half this speed [60].

One of the most prominent of that decade was that of the company PicTel (later renamed PictureTel), founded in 1984 by two MIT students [61]. As part of this venture, they developed and patented a new video encoding algorithm, which they called Motion Compensated Transform (MCT) [62]. In 1986 they began marketing their first video conferencing product using this algorithm, the C-2000 model. The equipment consisted only of an encoding-decoding system (Codec), weighed more than 100 kg, and did not include the audio amplifiers, echo cancelers, video switchers, and other items needed to implement video conferencing. The digitized and encoded video signal used a bandwidth of 224 kbps.

In Japan, in 1985, the NETEC-XV model, shown in Figure 30, operated at 384 kbps, using inter- and intra-frame prediction techniques, and entropic coding [63].

These systems required investments of tens of thousands of dollars, and costs of using digital lines in the order of thousands of dollars per month. These factors prevented its popularization, and its use was restricted to a few large corporations. On the other hand, the equipment was

incompatible with each other, which meant a great restriction on its use.

Efforts to standardize the protocols and mechanisms used in videoconferencing systems began in 1984 in a project of the Consultative Committee for International Telegraphy and Telephony (CCITT)[1]. The project ended in 1988, with recommendation H.261 "Codec for audiovisual services at n × 384 kbit/s" [64] and was extended in 1990 to "p × 64 kbit/s". The standard first introduced the 352×288-pixel Common Intermediate Format (CIF) format, suitable for videoconferencing systems, and laid the groundwork for codec interoperability between various computers and manufacturers.
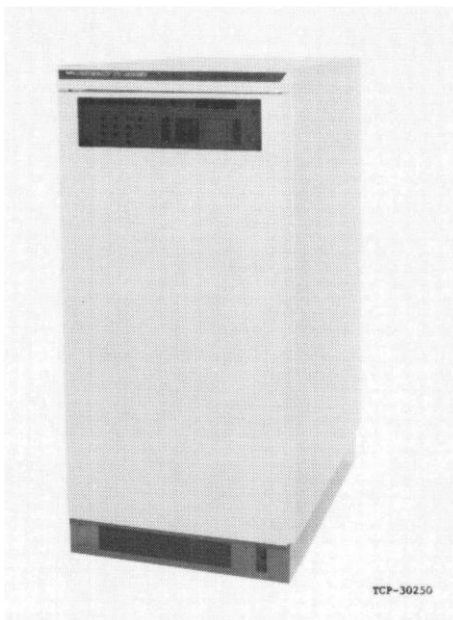

Figure 30. NETEC XV Codec [63].

In response to the unaffordable prices of both videoconferencing equipment and the contracting of digital links, personal video phones began to be developed in the late 1980s and early 1990s. At this time, new technologies were available, which made it possible to digitize the video signal in mass consumption devices. In June 1986, the Los Angeles Times published *"Say Cheese: New Phone Also Takes Pictures"* [65]. The article featured the "Luma Phone," which it defined as *"uninspiring-looking piece of equipment"*. It was a device weighing just under 4 kg, equipped with a small three-inch monitor and a small camera that allowed to take black and white images. The device used a common analog telephone line. The static images captured by the camera could be transmitted to a similar device, on the other end of a telephone line, using a "Send" button. The cost of the Luma Phone was $1450, and it was targeted for businesses or high-end residential customers. In the same article, the following year promised the launch of a similar product, cheaper, with the general public in mind. Luma Phone was a development of the Mitsubishi company, based on Atari's products called "Ataritel" [66][67], as shown in Figure 31.

[1] CCITT is currently the International Telecommunication Union (ITU)

Figure 31. Left: Ataritel. Right: Mitsubishi Luma Phone [68][69].

The cover of the 1988 "Popular Science" magazine headlined *"Video Phones: Here and Now, under $400. Uses regular phone lines at regular rates"* [70]. It was the launch of "Visitel", the promised successor to Luma Phone, at a more accessible price. The Chicago Tribune had announced the new phone a few months earlier, in November 1987, as *"the first still-image phone to hit the market at an affordable price ($399)"* [71]. A similar device, developed by Sony, was also being marketed in Japan by NTT. Both devices (shown in Figure 32) were incompatible with each other. During a phone call, the images taken by the camera could be captured and transmitted at the push of a button. When doing so, the audio was interrupted for a few seconds, and the image was transmitted.

The techniques used by these still-image video telephones consisted of digitizing the captured images and transmitting them in a modulated form over a telephone line. At that time, modems already existed, capable of transmitting data over analog telephone lines. However, these devices required a long time for the initial establishment of data communication ("handshaking") and had a typical bitrate of 1200 bps, which was not appropriate for these video phones. Since the transmission of the video occurred during the conversation, it was necessary to develop new, faster communication protocols for the establishment and transmission of the image. Visitel used a specific technology, which they patented in 1987 [72]. It consisted of a combined amplitude and phase modulation, using a carrier frequency of 1747.8 Hz. This frequency is located roughly in the middle of the audio band used by analog telephone lines (ranging from approximately 300 Hz to 3400 Hz) and can be obtained by using a standard 3.579545 MHz color TV crystal oscillator and dividing by 2048 (which is easily achieved with digital circuits, since 2048 is a power of two). Each pixel in the image was encoded with 4 bits (16 possible values) and represented by a symbol consisting of a sine wave cycle, varying amplitude and phase (between 0° and 180°), as shown in Figure 33. As each symbol represents 4 bits, a rate of 4 x 1747.8 = 6991 bps is achieved, very close to 7 kbps, and much higher than the 1.2 kbps available in conventional modems of the time.

At the beginning of the transmission, when the user presses the "send" button on an established voice call, an initial exchange of information takes place using preamble tones, detailed in Visitel's patent [73]. Using these short tones, the receiver is synchronized with the transmitter, in order to properly decode the modulated signal containing the encoded image.

The Visitel device required less than 6 seconds to transmit the full image, and they mentioned this as a competitive difference from Sony's videophone, which required a few more seconds. Other similar devices were also marketed in the late 1980s, by other brands, such as the Panasonic WG-R2 or KX-TV10 [74].

One of the interesting features of these designs was that, in some cases, images could be saved on standard telephone answering machines. Since modulation was done within the audio band, any telephone answering machine available at the time could record and play an image, as shown in the video by [75].

Figure 32. Left: Mitsubishi Visitel. Right: Sony Teleface. [70].
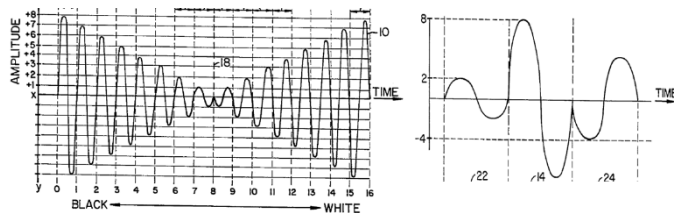

Figure 33. Left: Amplitude and phase modulation for the 16 possible gray levels (from black on the left to white on the right). Right: Example of the modulated signal for 3 pixels ("level 2 dark gray"-"black"-"level 4 light gray"). [72].

In the early 1990s, Picturetel and Compression Labs controlled more than 90% of the videoconferencing equipment market in the United States [76]. At the time, the video conferencing market was divided into two segments: In the low segment, equipment operating between 56 kbps and 384 kbps, and in the high segment, up to 2 Mbps. Picturetel was more focused on the first segment, while Compression Labs was in the second. The entire video conferencing market in the early 1990s sold a few thousand computers per year (PictureTel sold 770 devices in 1990 [77]). However, companies were trying to position themselves across all the spectrum. In 1991 PictureTel began marketing the System 4000, integrating the codec, video switcher, audio mixer, network terminal adapters and audio-video interface unit, which were previously separate, into personal computer-sized electronics. The system was the first from PictureTel to support the new H.261 standard (p × 64 kbps), and was priced

at \$40,000 [78]. Figure 35 shows the Rembrandt II/VP machine, the first CLI machine to implement the H.261 standard in 1992, with bitrates between 56 kbps and 2 Mbps [79]. A very complete list of the supply of videoconferencing equipment at the beginning of 1990 can be seen in Table 1.

Figure 34. PictureTel System 4000 with Camera, keyboard and CPU Codec [80].

Figure 35. Rembrandt II/VP Team, Compression Inc Labs (CLI) [81].

In 1992 AT&T, along with CLI, designed and began marketing a videophone that worked with analog telephone lines, the Model 2500, shown in Figure 36. This videophone was the first to encode color video in digital format, and transmit it at 19.2 kbps, over analog telephone lines, using a built-in modem. The device had a 3.3-inch color LCD screen and ran at 10 frames per second. The camera allowed to capture up to 3 people in front of the device. The audio was encoded with CELP techniques at 6.8 kbps. The video had a resolution of 127×112 pixels in luminance, and 32×28 in chrominance (or color). It used predictive coding techniques and the DCT transform, generating a stream of 10 kbps. Audio, video, and signaling were multiplexed into a data channel, using the X.25 protocol, encapsulated in LAPB frames (a protocol used in the ISDN standard) [82]. It took approximately 10 seconds for the video to set up [83]. Its value was \$1500.

Table 1 – Videoconferencing equipment market in 1990 (from [76]).

Vendor	Model	Room requirement	Transmission (bit/sec)	Video resolution	Frame refresh rate (frame/sec)	Audio operation	Interfaces	Price range	Warranty
American Lightwave Systems, Inc. Wallingford, Conn. (203) 265-8880	FT 1310/1550 System, FN6000 System	Rack-mounted, portable	Full-motion video (16 channels)	1,000 pixels by 1,000 pixels	30 or 60	Up to 4 audio channels	NTSC, SuperNTSC, RGB, HDTV, PAL	$4,000 to $100,000	1 year
Colorado Video, Inc. Boulder, Colo. (303) 530-9580	Digital Transceiver Model 286, Video Transceiver Model 250 and 290	Rack-mounted, portable, desktop, conference room	1.9K to 560K	Up to 512 pixels by 512 pixels	Slow-scan	External	NTSC, CCIR, RS-170, GPIO	$3,000 to $20,000 per site	1 year
Compression Labs, Inc. San Jose, Calif. (408) 435-3000	Rembrandt II/06	Rack-mounted, desktop	56K to 384K	256 pixels by 240 pixels	15	Internal or external	Video: NTSC, RGB, PAL; Data: RS-232, RS-449, V.35, X.21	$31,500	1 year
Concept Communications, Inc. Dallas (214) 746-3888	Image 30	Rack-mounted, portable, desktop, conference room	56K to 768K	256 pixels by 200 pixels	30	Internal, external	Video: NTSC, RGB; Data: RS-449	$5,995 to $12,000	6 months free, $75/month thereafter
Datacube, Inc. Peabody, Mass. (508) 535-6644	128 Series	Multibus-, PC bus-, VME bus-compatible	N/A	384 pixels by 512 pixels	30	N/A	NTSC, RGB, RS-170	$2,000	1 year
Datapoint Corp. San Antonio, Texas (512) 699-7000	MINX	Desktop; also configurable for conference rooms or roll-abouts	Analog video over local video or broadband LAN networks	330 lines analog video	30	Half-duplex audio with workstation; half- or full-duplex with network interface	NTSC, PAL, VCR and PC	Local net connectivity with workstations: $6,995 for server plus $5,995 per workstation	1 year
EyeTel Communications, Inc. New Westminster, B.C., Canada (604) 525-2511	VISTACOM Videophone	Integrated desktop system	56K or 64K	240 pixels by 256 pixels	4 to 30	Integrated 9.6K bit/sec, echo cancellation	Data: RS-232; Network: V.35	$16,500 integrated video system including codec	1 year
GPT Video Systems Stamford, Conn. (203) 348-6600	Madison	Portable	56K/64K to 2.048M	352 pixels by 288 lines	30	Internal	Video: NTSC, RGB, PAL; Data: RS-449, DS1	$45,000	1 year
Grass Valley Group Grass Valley, Calif. (916) 478-3803	4500 Video Codec	Rack-mounted	45M	NTSC-quality	59.94	Stereo audio	Video: NTSC; Data: RS-232	$8,995 for codec; $12,950 for multiplexer	2 years
Interand Corp. Chicago (312) 478-1700	System 900 and 950 Imagephone 150, Med 185T, and 185T	Portable, desktop, conference room	Up to 64K	640 pixels by 480 pixels	30	External	Video: RGB; Data: RS-232, RS-449	$10,000 to $20,000	1 to 2 years
Mitsubishi Electronics America, Inc. Torrance, Calif. (213) 217-5732	MVC-7000	Portable	Variable from 64K to 384K	384 pixels by 240 pixels	Variable from 10 to 15	Internal	Video: NTSC; Data: RS-422/449, X.21, V.35	$40,000 to $50,000	1 year
NEC America, Inc. Herndon, Va. (703) 834-4000	VisuaLink 1000 and 3000	Portable	56K to 2.048M	Up to 352 pixels by 288 pixels	10 to 30	Internal (external optional)	Video: NTSC, PAL; Data: V.35, RS-449, RS-232, RS-232C	$30,000 for VisuaLink 1000; $62,100 for VisuaLink 3000	2 years
PictureTel Corp. Peabody, Mass. (508) 977-9500	V-3100 Videoconferencing System, C-3000 Codec	Portable	56K to 384K	256 pixels by 240 pixels	10	Internal	Video: NTSC, PAL; Data: RS-449, V.35, X.21	$30,000 for codec; $55,000 for system	1 year
Videophone, Inc. Houston (713) 741-1971	Videophone line	Rack-mounted, portable, desktop or conference room	Up to 19.2K	Either 256 pixels by 256 pixels or 512 pixels by 512 pixels	Still frame	External	NTSC, RGB, RS-232, RS-449	$8,500 to $35,000	1 year
VideoTelecom Corp. Austin, Texas (512) 834-9734	Conference System 200 and 300	Rack-mounted, portable, desktop, conference room	56K to 768K	256 pixels by 240 pixels; still frame up to 512 pixels by 480 pixels	12 to 15	Internal	NTSC, RGB, RS-232, RS-449, fax (standard Group III), VCR	$30,000 to $60,000	1 year, plus free software updates

GPIO = General purpose I/O
HDTV = High-definition television
MINX = Multimedia Information Network Exchange
N/A = Not applicable
NTSC = National Television Standard Code
PAL = Phase-alternating line
RGB = Red, green, blue

In addition to the standardization of codecs, signaling standards were necessary for the interoperability of videoconferencing systems. In 1990, the CCITT published the first standard for video telephones or videoconferencing terminals, Recommendation H.320 "Narrow-Band visual telephone systems and terminal equipment" [84]. This recommendation describes the technical requirements for terminal equipment for the videotelephone service, with a data rate of up to 1920 kbps. The high-level architecture of the terminal equipment, and its logical blocks, are defined as shown in Figure 37. The characteristics of each functional block are described in other ITU recommendations (e.g., H.221, H.242, H.230, H.261, and the I.400 series). The H.320 terminals could operate connected to the new Integrated Services Digital Networks (ISDN), which could operate at multiples of 64 kbps (n×64 kbps), with end-to-end digital technology.

The first commercial implementation of H.320 was developed by British Telecom (BT), the VC2100 codec [85]. This codec could operate at all data rates from 64 kbps to 2 Mbps, implementing many of the optional H.320 operating modes, and became an industry benchmark at the time. In the first half of the 1990s, BT also marketed other H.320-compatible equipment, such as the VC 7000. It was basically a phone with a monitor, as shown in Figure 38. It connected to ISDN services, included a color charge-coupled device (CCD) camera, and it was possible to connect an external camera for more mobility. The display consisted of a 10-inch color monitor [86].
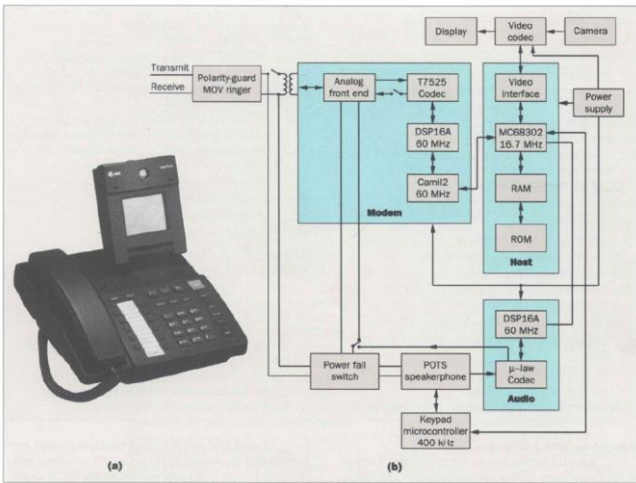
Figure 36. Left: Picture from AT&T analog videophone 2500 and CLI. Right: Diagram of how it works [82].
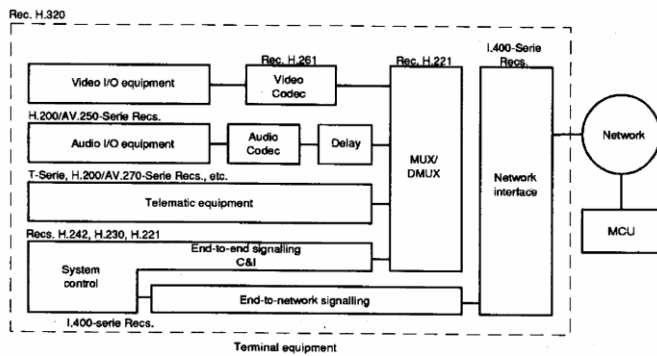

Figure 37. Conceptual diagram of a videoconferencing terminal equipment, according to the H.320 standard [84].

Figure 38. VC7000 Videophone [86].

PC video conferencing systems began to be developed in the early 1990s and were commercialized a few years later. BT developed the VC8000 system, which made it possible to turn a PC into a multimedia video conferencing terminal. Images, graphics and text could be shared through an application. The VC 8000 kit consisted of a multimedia communications card, a video camera, an audio unit and software, supplied by IBM, Olivetti or ICL. The audio unit was a regular analog phone, connected to the PC.

In 1993, PictureTel announced the PCS100 PC video conferencing system (later renamed Live100), compatible with PC-AT and the H.320 standard. To connect it, it used the ISDN network. It worked with a resolution of 352×288 pixels and up to 15 frames per second. The launch price was $6000

[87]. Two years later, in 1995, the PCS50 model (later renamed Live50) was launched, simpler, and priced at $2500 [88]. Picturetel's kits included a PC card, a camera, a phone, and the software, compatible with Windows 3.1 and then with Windows 95 [89]. A schematic of these elements can be seen in Figure 39.

Several other PC video conferencing brands and products were marketed in the mid-1990s. These include AT&T Telemedia, Intel ProShare [90] (see Figure 40), InVision, MRA VidCall, and Northern Telecom Visit Video. A comparison between them can be seen in [91]. They typically had a resolution of 352×288 (or similar) and 15 frames per second, worked over ISDN lines, allowed for video telephony, file sharing, applications, and whiteboard areas. The specific hardware was compatible with ISA or EISA slots. Prices were around $2500.
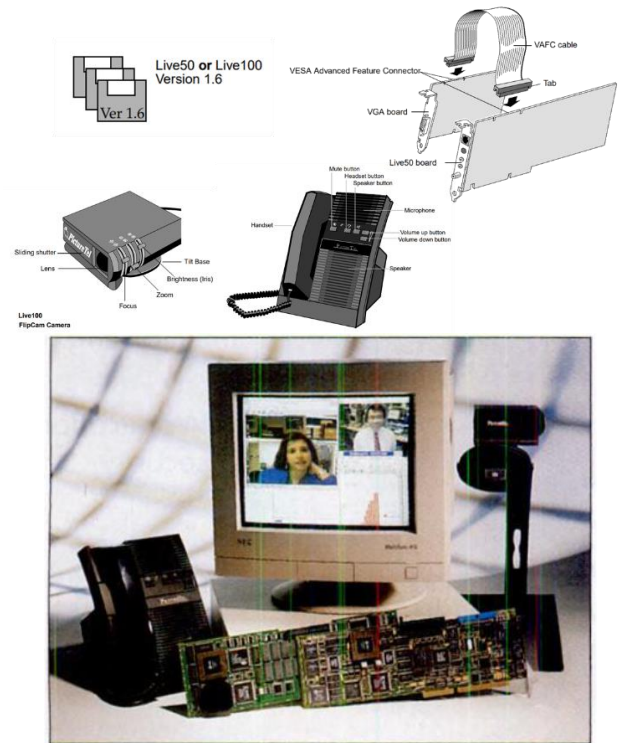

Figure 39. Up: Schematic of the PictureTel Live50 kit [89]. Bottom: Live PCS 100 kit, [91].

Figure 40. Intel Proshare [90].

V. THE IP AGE

Communication and packet switching technologies had begun to be developed in the late 1960s, with academic studies. In the early 1970s, a new computer-to-computer communication technology was developed, which its designer, Bob Metcalfe, called Ethernet. It was so successful that in 1980 several companies adopted it. Digital, Intel and Xerox started using it, at speeds of 10 Mbps, making it a "de facto standard". In February 1980 the IEEE Computer Society held the first meeting of the "Local Network Standards Committee", and the Ethernet protocol was standardized as IEEE 802.3. In the late 1980s, in 1989, the concept of the "World Wide Web" (www) was created by Tim Barners Lee, giving birth to the Internet.

In the early 1990s, Ethernet was beginning to become popular as the local area network protocol, and the Internet was beginning to expand. Developers of videoconferencing systems, based until that time on transmission and digital signaling over ISDN networks or point-to-point links, began to put their attention and efforts into these new data networks.

The audio from the March 1992 meeting of the Internet Engineering Task Force (IETF) held in San Diego, California, was broadcast for the first time live, via a nascent Internet, to participants at 20 different sites, spread across three continents [92]. Multicast technology was used. The audio was encoded in PCM at 64 kbps, using a Sparc Station running the "Visual Audio Tool" (VAT) application, developed by Van Jacobson and Steve McCanne. The system was called the Internet Multicast Backbone or MBone. Between 1993 and 1994, the ability to transmit video was added to the system, at a bit rate of 128 kbps and 1 to 4 frames per second, with the Network Video (NV) application developed by R. Frederick at Xerox Parc. VAT, NV, and other applications used are shown in Figure 41. The quality of the transmission, both audio and video, was not good, and variable in the time, depending on the quality and saturation of the Internet links used. However, the system became popular and by 1994 the MBone network had 750 subnets connected [93].

The key concepts that made MBone possible were multicast over IP and the use of a new protocol that began to be developed in 1992: the Real Time Protocol (RTP), proposed by the IETF's Audio and Video Transport Working Group [94]. The new RTP protocol provided timing and sequencing services, allowing audio and video applications to adapt to latencies and errors introduced by packet networks. In addition, by the early 1990s, several Internet-connected workstations supported sound and video cards, with prices at a level that made their large-scale deployment possible.

Tim Dorcey, from Cornell University's Department of Information Technology, developed the CU-SeeMe application in 1992. It was first developed for the Macintosh and then, in 1994, for Windows. In 1993 the app was introduced as part of an educational project funded by the National Science Foundation (NSF) in the United States, called "Global Schoolhouse" [95]. It was the first Internet-based multipoint video conferencing to connect schools in the United States and, potentially, schools around the world. An image can be seen in Figure 42. In 1995, CU-SeeMe described itself as follows: *"Outgoing and incoming video is displayed in small, 4-bit grayscale windows providing surprisingly acceptable clarity and refresh-rates. An additional slide window can transmit and receive larger-sized graphics. Audio is presently available only on the Macintosh platform, however, a plug-in module for both Windows and Macintosh provides text-based "chatting" capabilities. Minimal hardware is required for such sophisticated capabilities. As a simple receiver, any Macintosh with a 68020 processor or higher running System 7 will work. PCs require a 386SX processor or higher running Windows 3.1 in Enhanced Mode. Each system also requires an IP (Internet Protocol) connection with a minimum bandwidth of 28.8 kbps. Internet connections made through a 14.4 kbps modem lack audio features as well as acceptable video refresh rates. In order to send video as an origination site, your computer will need a video digitizing board and a video camera. Fortunately, video peripherals have recently become quite affordable"* [96]. The CU-SeeMe application was marketed by the company White Pine Software.
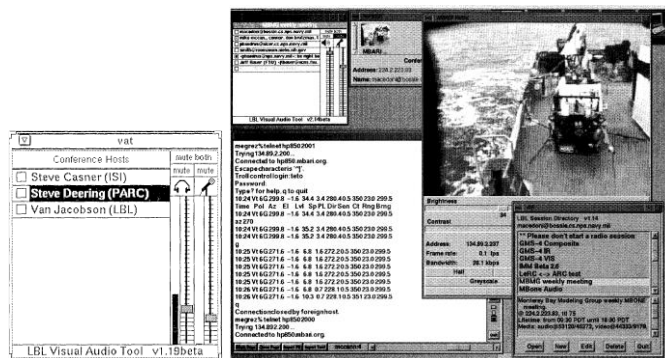

Figure 41. Left: Visual Audio Tool (VAT) application, running in an "X Window", used in the IETF's first multicast audio transmission, taken from [92]. Right: Network Video (NV), VAT, Whiteboard (WB), and Session Directory (SD) applications used in an MBone session broadcast from the Monterey Bay Aquarium, taken from [93].
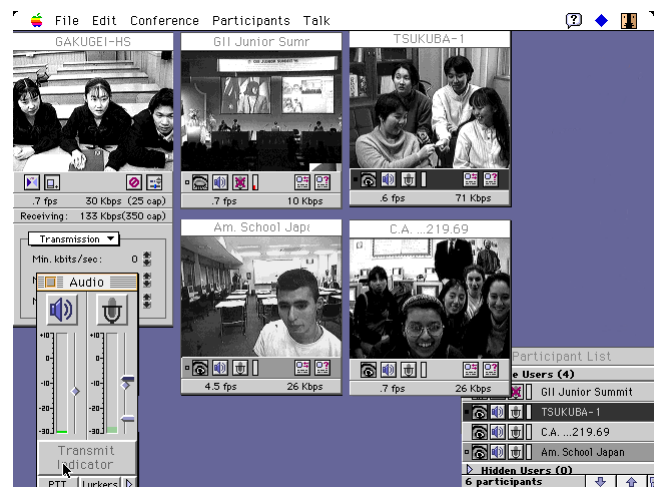

Figure 42. CU-SeeMe used in educational environments [95].

In France, at the Institut National de Recherche en Informatique et en Automatique (INRIA), Thierry Turletti developed the INRIA Videoconferencing System (IVS) application in 1994 [97]. This video conferencing application used PCM and ADPCM to encode audio, and the H.261 video codec, initially developed for video-over-ISDN systems. It was necessary to adapt it to packet networks, for which Turletti himself proposed how to do it to the IETF [98]. The proposal used the RTP over UDP protocol to provide timing and sequencing and added a specific header for H.261 packets [99]. Figure 43 shows the IVS application.
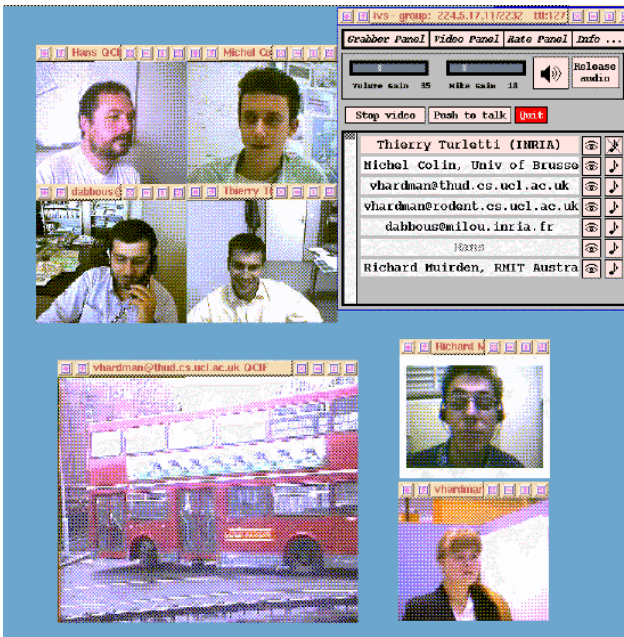

Figure 43. IVS [100].

By 1995, there were numerous videoconferencing systems operating over data networks. Among them, Avistar, Cameo Personal Video System (from Compression Labs), MediaPhone (from Fiber and Wireless), Communique! (from Insoft), BeingThere (from Intelligence at Large), InVision, VidCall (from MRA), Livelan-V (from PictureTel), InPerson (from Silicon Graphics), ShowMe (from Sun Microsystems), and VTel, among others. Prices ranged from $1000 to $5000. A full comparison between these systems can be seen in [101].

While all packet-network videoconferencing systems operated on the same concepts, they were mostly incompatible with each other. Similar to what happened with ISDN, it was necessary to have standards that would allow videoconferencing sessions over IP to be established between different systems from different manufacturers or developers. Thus, in October 1996, the first version of the H.323 standard was ratified by ITU-T Study Group 16 [102]. H.323 was the first standard for the transmission of multimedia (voice, video, and data) over packet networks. This first version was relatively basic and was successively improved over the next ten years with various improvements and additions. H.323 established not only the coding requirements for audio, video, and data, but also the signaling protocols for the establishment of sessions, as outlined in Figure 44. These new signaling protocols basically extended the existing ITU protocols in ISDN to be used over packet networks.
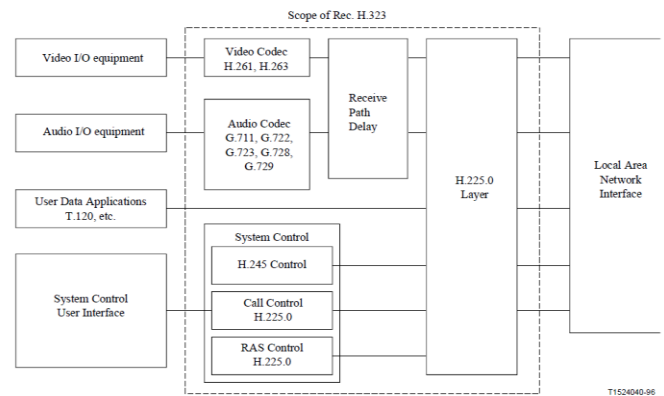

Figure 44. Scope of recommendation H.323, setting standards for the control system, in addition to multimedia encoding [102].

Also in 1996, ITU standardized the H.324 recommendation "Terminal for low bit rate multimedia communication" [103]. This recommendation describes videoconferencing terminals that can be connected to the analog public switched telephone network, using modems.

Around the second half of the 1990s, several videoconferencing systems began to adopt the new standards, H.323 and H.324, which promised universality and interoperability, both over classic telephone networks and over packet networks. In 1996, the New York Times announced "Intel Plans PC Video Phone Technology: After years of false starts by AT&T and others, the Intel Corporation hopes to bring video-phone technology to the masses. The company plans to announce at a meeting with analysts in New York today that starting later this year, most new home personal computers will be capable of making and receiving video-phone calls over standard telephone lines. Frank Gill, executive vice president of Intel's Internet Communications Group, said he expected that hundreds of thousands of video-phone ready computers would be sold this year and millions more in 1997. That contrasts sharply with Intel's two-year-old Proshare video-conferencing product, the industry leader, which has only 50,000 users" [104]. The system, called the Intel Video Phone, could operate over packet networks or over telephone lines, achieving in the latter case between 7 and 11 frames per second (adjustable with a control that allowed the image quality to be lowered and the number of frames per second to be increased) [105]. Also in 1996, Microsoft introduced the first version of the Netmeeting application, which allowed audio conferencing. Microsoft announced at its launch: "NetMeeting is the Internet's first real-time communications client that includes support for international conferencing standards and provides true multiuser application-sharing and data-conferencing capabilities. NetMeeting makes voice and data communications over the Internet as easy as a phone call" [106]. The application, based on the H.323 standard, could be downloaded together with Internet Explorer (beta version 3.0). At the end of the same year, it was announced that video would be added to the application, running at up to 10 frames per second, over 10 Mbps Ethernet networks. It could run on a 166 MHz Pentium

processor [107]. Figure 45 shows the Intel Video Phone and Microsoft Netmeeting applications.
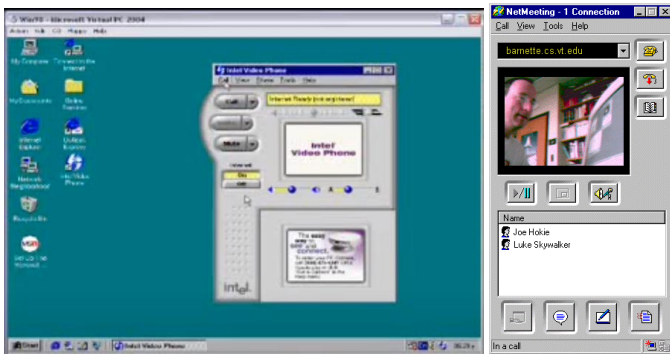

Figure 45. Left: Intel Video Phone. Right: Microsoft Netmeeting.

In addition to Intel and Microsoft applications, PC video conferencing systems in the late 1990s included the Boca Video Phone Kit, Diamond Supra Video Phone Kit 3000, Gallant Intervision Pro, Panasonic EggCam GP-KR0011, Tekram How-R-U Video Conferencing Kit, and 3Com BigPicture Video Kit 1622 (see Figure 46). Prices for these kits ranged from $200 to $400 in 1997 (nearly one-tenth of the ISDN PC systems in existence a few years earlier). A comparative analysis between them can be seen in [108].

In 1997, Henning Schulzrinne presented at the IETF the first draft of a new protocol designed to manage videoconferencing, which he called the Session Initation Protocol (SIP) [109]. Unlike H.323, SIP emerged from the world of the Internet, based on the experiences of the MBone system. As the protocol's author mentions, *"a text-based approach was chosen for the design of SIP and RTSP. Rather than inventing a new protocol representation from whole cloth, reusing the most successful Internet protocol, HTTP, seemed the more appropriate choice. By using HTTP as a base, the protocols can immediately re-use a number of evolving protocols for electronic commerce, authentication, content labels and client-side access control, protocol extensions, state management and content negotiation. Also, servers, proxies and firewalls, all already tuned for high performance, manageability and reliability, can be easily modified to accommodate these new protocols"* [110]. The basic operation of the protocol is shown in Figure 47. The protocol was formally standardized in 1999, and began to compete strongly with its ITU predecessor, H.323.

Figure 46. PC video conferencing kits available in 1997. From left to right: Boca Video Phone Kit, Diamond Supra Video Phone Kit 3000, Gallant Intervision Pro, Panasonic EggCam GP-KR0011, Tekram How-R-U Video Conferencing Kit and 3Com BigPicture Video Kit 1622 [108].
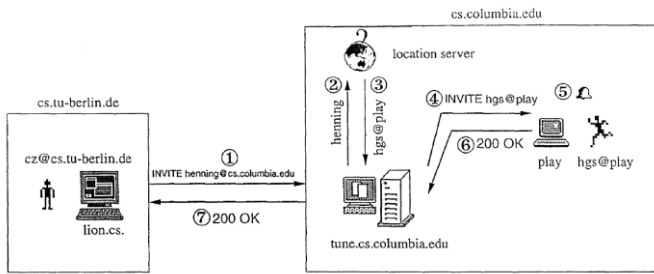
Figure 47. SIP Protocol Scheme of Operation [110].

Over the next few years, H.323 and SIP competed, trying to win the market. A 2000 PC Magazine report entitled *"Many Voices, Little Consensus"* [111] indicated that very few of the Internet telephony equipment available at the time was interoperable. During the first decade of the 2000s, H.323 had several revisions, ending in 2009 with revision 7. Several years later, in 2022, Revision 8 was published. SIP, on the other hand, was quickly accepted by the 3GPP group, standardizer of cellular systems, who incorporated it into their core protocols of their IP Multimedia Subsystem (IMS) in 2001 [112]. This fact, and its easy implementation at the software level, meant that SIP ended up being the most widely used videoconferencing session initiation protocol in the 2010s, and even the most used at the time of writing, in 2023.

In 2003 Niklas Zennström and Janus Friis created Skype, based on the concept of decentralized "Peer-to-Peer" used by Kazaa (a system previously created by the same people). The name originally intended had been "Sky peer-to-peer", which was abbreviated "Skyper". However, the Internet domain "skyper" was already being used, so the last "r" was removed, creating the name "Skype". At its launch it allowed audio calls, two years later, in 2005, the possibility of making video calls was added [113]. The working principle of Skype was very different from the video conferencing applications of the time. It did not use any of the signaling protocols, nor the standardized audio codecs. Instead, it was based on a peer-to-peer architecture, where the only central element was an authentication server ("Login Server"). Signaling and media (audio and then video) were done directly between peers (user applications). Some of these applications were defined as "Super Nodes" for functional purposes, as shown in Figure 48. Any node with a public IP address that had sufficient CPU capacity, memory, and network bandwidth was a candidate to become a Super Node. Communications with some of the users were done through these Super Nodes [114]. The audio used the iLBC, iSAC or iPCM codecs, developed by Global IP Solutions [115]. These audio codecs had been specially designed for audio transmission over packet networks, and included mechanisms that were highly resilient against packet loss. Skype used both UDP and TCP for the transport of the encoded media, making it robust to work behind firewalls. The service became popular. By 2005, Skype had more than 2.8 million users in the U.S. and 30.6 million worldwide, with 155,000 new users added every day. *"A reason Skype is so popular is that it is free. Another is that it works. That may not seem like much, but it matters when calls with other free VoIP programs sound more like walkie-talkie conversations than phone calls",* said a 2005 New York Times article [116]. That same year, Skype was bought by eBay in a $2.6 billion

transaction [117]. In 2009 eBay sold Skype to the Silver Lake group and in 2011 it was sold again, this time to Microsoft, for $8.5 billion [118].
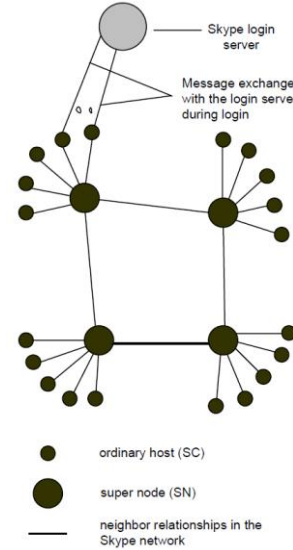

Figure 48. Skype Connectivity Architecture [114].

Around the early 2000s, other instant messaging apps began to add the ability to make video calls. In 2002, Yahoo Messenger, version 5.5, incorporated the possibility of making video calls, up to 20 frames per second and with a resolution of 320×240 pixels [119]. In 2004, the popular instant messaging app ICQ also introduced video [120].

In parallel to the proliferation of desktop videoconferencing applications and kits, in the mid-2000s the first professional videoconferencing systems were developed with the concept of "telepresence", working on packet networks. In these types of systems, the images on the screens are displayed in life-size, clear enough to capture the small nuances of facial expressions and body language. The audio comes from the direction of the person speaking, achieving a much more natural feel. Hewlett-Packard launched its telepresence line of products in 2005, with the name Halo Collaboration Studio. This system made it possible to organize meetings between four sites, through the use of rooms built with strict standards for sound and lighting. Each room was ideal for six people but could accommodate up to fourteen. Equipping a room for a Halo Studio was priced at approximately $350,000. A similar but smaller system was announced in 2008. In this case, the hardware for a room included two or four chairs, movable front and back walls, a table, a high-definition video camera, a large monitor, and audio and switching equipment. The price of these rooms was around $120,000 [121]. For its operation, HP offered a dedicated data network, which it called the Halo Video Exchange Network (HVEN). It consisted of a fiber optic network, specifically designed to provide the service without relying on the internet or other data networks. The service of this network and its operation had a monthly cost of $12,000 per site [122].

Cisco entered the world of telepresence in 2006 with its Cisco TelePresence 1000 and 3000 products. The TelePresence 1000 included a single screen, for small meetings, and was priced at $80,000. The Telepresence 3000 system was a full-room, 12-participant system that included three 65-inch high-definition plasma screens, at a cost of $300,000 [123]. Unlike HP's strategy, Cisco's products did not require the hiring of a new dedicated network, instead operated over local area networks.

Also in 2006, Polycom announced the launch of its RPX RealPresence Experience products, available in different configurations, supporting rooms from 4 to 28 people, and priced from $250,000 per room [124].

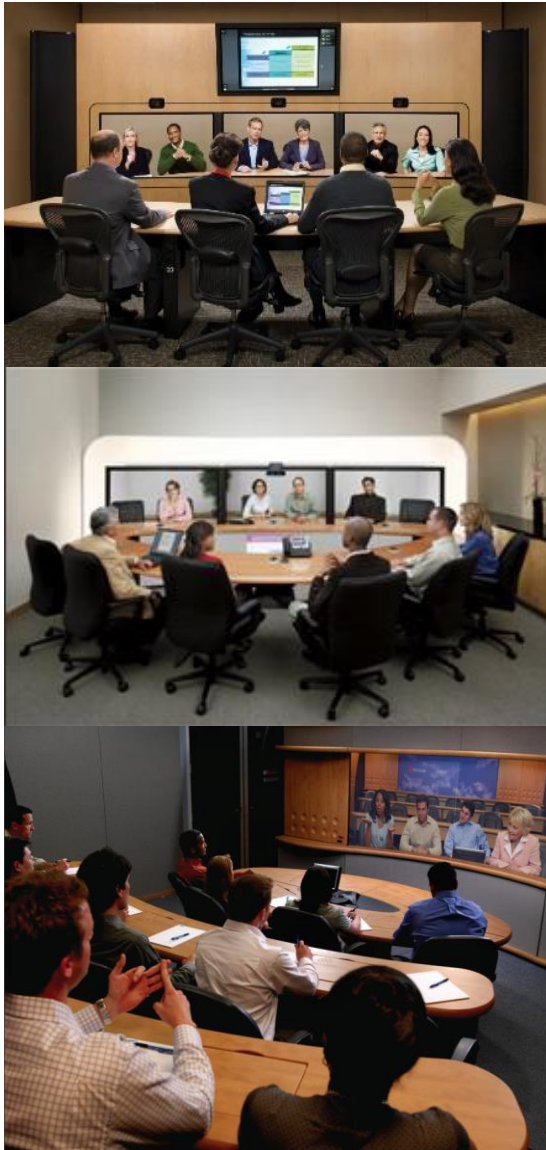Figure 49 shows the telepresence systems of HP, Cisco, and Polycom, operating in the second half of the 2000s.

Figure 49. Up: HP Halo Collaboration Studio. Center: Cisco Telepresence 3000. Bottom: Polycom RPX. [125][126][127].

VI. THE CLOUD SERVICES AGE

In 2011, Gary Sullivan, one of the main people responsible for video encoding developments, published an article titled *"Video Telephony Has Finally Arrived"* [128]. In the article, he acknowledged that for decades there was an attempt to promote video telephony and videoconferencing services, both in the mass consumption and corporate markets. Sullivan mentioned that while many companies invested in video conferencing equipment in the 1990s, the equipment usually "ended up gathering dust, unused." On the other hand, in mass consumption, *"for most of us, video telephony is not yet part of our daily lives". "It seemed like people didn't want it anywhere",* said Sullivan, who finally ventured that this would change quickly, and the technology would become popular within just a couple of years. According to the aforementioned article, the obstacles that prevented video conferencing from prospering up to that point were due to limitations in hardware, networks, and video compression. By 2011, the hardware needed for video processing was within the reach of any smartphone, data networks and the Internet supported the necessary bandwidths, with the required reliability, and video encoding systems had advanced enough to achieve appropriate compression rates (the ITU-T H.264 standard had been available for some years, and commercial systems had begun to use it).

Around the beginning of the 2010s, videoconferencing systems based on web services began to be developed, hosted in the cloud and marketed on a "pay-per-use" basis, without requiring large investments in local equipment. These applications were focused on the corporate market.

In 2010, Adobe launched its Adobe Connet 8 product, a web-based video conferencing system. It was presented as a lightweight system, optimized for low bandwidths, and not intended to compete with high-end telepresence systems [129].

In 2011, the Blue Jeans service was launched to the market. Krish Ramakrishnan, co-founder and CEO of Blue Jeans Network said in a press release, *"What was once an elite boardroom technology has moved to the cloud".* In the first two months of service, they had 4000 subscribers, from more than 500 companies with a presence in more than 100 countries [130].

The FaceTime app began shipping with the iPhone 4 in 2010 as a video chat app between Apple smartphones [131]. A few months later, it was incorporated into Mac desktop computers. Steve Jobs said, *"FaceTime makes video telephony to or from mobile devices easy for the first time. We've sold more than 19 million FaceTime-ready iPhone 4 and iPod touch devices in the past four months, and now those users can make FaceTime calls with tens of millions of Mac users"* [132]. Several years later, in 2018, the ability to video conferencing between multiple users was introduced, with the Group FaceTime feature, which allowed up to 32 participants [133], as seen in Figure 50.

Figure 50. Apple Group Facetime [133].

Whatsapp was created in 2009, but it was only at the end of 2016 that it incorporated the video call function [134] and in 2018 the possibility of group calls, of up to four participants [135]. It was expected that, with more than 1 billion users, this feature would quickly be included in daily habits and routines as a form of communication between its users. However, this did not happen at the time of its introduction.

In 2011 Microsoft bought Skype [118] and relaunched it as Skype for Business in 2015 [136], and then incorporated it as part of Microsoft Teams in 2017 [137]. In 2019, Microsoft Teams allowed video conferencing of up to four participants.

Eric Yuan, who worked on the Cisco WebEx product, founded the company Saasbee in 2011. The following year, it changed the name to Zoom. "*At the time, Zoom was just an idea in a seemingly very competitive video conferencing space and most investors incorrectly thought that the existing products like Skype, Webex and others were solving this problem. But, what most people didn't see at that time was that Eric had a vision to build a new type of video communications experience that would solve a massive need*", said Jim Scheinman (Eric's partner, who proposed the company's name change, inspired by the children's book "Zoom City") [138]. As of May 2013, the new video conferencing services company had one million users [139].

Along with cloud services, it was necessary to develop technologies that would allow Internet browsers to be able to play multimedia content in a standardized way. Until that time, in order to be able to use video conferencing sites from browsers, it was necessary to download and install "plugins" or extensions, which made it difficult to use and interoperate. In 2011, Google made the source code of the WebRTC project available to the developer community [140]. This technology had initially been developed by Global IP Solutions (or GIPS), a company founded in 1999 in Sweden and acquired by Google in 2011. WebRTC made audio and video available across browsers, through a standard, uniform set of features. By gradually being incorporated into the various browsers, WebRTC made it possible to achieve the "click and enter the conference" paradigm. That is, being able to start a video conferencing session from a browser without the need to run external programs or download extensions dependent on each platform and operating system.

The rise of cloud services did not make video conferencing services widely accepted and used, and Gary Sullivan's predictions of the early 2010s once again did not come true. By the end of the 2010s it was clear that the problem was not technology, but simply that people were not interested in using video calls or videoconferences massively and daily to communicate, neither in their private lives nor in their work. However, this changed drastically at the beginning of the 2020s, due to the Covid-19 pandemic. With the mass confinement of the population worldwide, video conferencing became almost the only way to maintain visual interactions with family and work colleagues. On a mass level, Zoom quickly positioned itself as one of the dominant options. Between January and April 2020, the number of minutes of Zoom meetings increased 2500-fold, as shown in Figure 51 [141]. Other platforms also increased notably they use and subscribers. Microsoft Teams went from having 20 million users in 2019 to 270 million in 2022 [142]. All companies that offered video conferencing services had to implement improvements and new features quickly. For example, at the beginning of the Pandemic, Teams supported video conferencing of only 4 people. He then expanded it to 9 participants, and in July 2020 he allowed up to 49 participants to be displayed at the same time. Zoom was initially criticized for security aspects, which it quickly corrected. The use of video conferencing was quickly incorporated into daily life, both at a personal and family level, as well as at a business level.
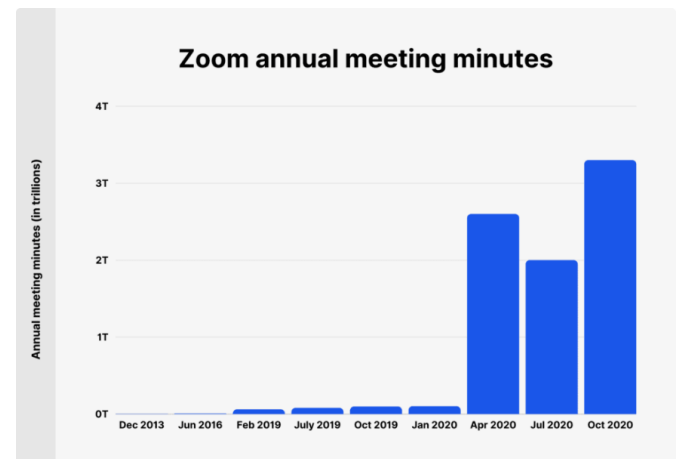
Figure 51. Zoom Usage Minutes [141].

After two years of pandemic, in 2022, video conferencing technologies reached their maturity and popularization. Participating in a video call or videoconference became for many people the usual way to meet, both personally and professionally. Expensive hardware-based videoconferencing systems were completely replaced by cloud services, many of them free, or with very affordable costs per use for any person or company. In 2002, Zoom offered free video conferencing services for up to 100 people and up to 40 minutes, and for $150 annually, similar services for unlimited time [143]. Similarly, Microsoft Teams offers free video conferencing services for up to 100 people and up to 60 minutes, and for $4 per user per year, similar services for up to 300 people and 30 hours per meeting [144].

VII. The Future

The next few years of the 2020s present new challenges for video conferencing systems, several of which are described below.

With the end of the pandemic, several companies have returned to the face-to-face work modality, but many others are opting for a hybrid scheme, where employees have days of face-to-face work and others of teleworking. This has led to the proliferation of hybrid conferences, where some of the participants are in person in a room, and part are located remotely. Typical company meeting rooms are not set up for these types of meetings. Although the well-known huddle rooms began to be incorporated into companies even before the pandemic, the return to offices has generated a new and growing demand for this type of rooms. In this type of conference, there is a clear asymmetry between face-to-face and remote participants. Remote participants have difficulty listening and actively participating in meetings. When it comes to audio, echo, reverb, and distance to microphones in the room are some of the main issues. The general view of the room often does not allow to visualize the details of the person who is speaking. Classic whiteboards are not properly viewed by remote participants, and they do not allow interactivity. Several of these points already have technical solutions at reasonable costs. These include interactive monitors or televisions that can function as shared whiteboards, cameras that automatically focus on the person speaking, and echo cancellation and compensation systems. However, even with the latest technologies available in 2023, the experience of remote participants is poor. On the other hand, face-to-face participants often have difficulty initiating the video conference. Camera kits, microphones, monitors or TVs and other accessories are often not integrated, and starting a new conference session can be very difficult for non-technical users.

The first approach to hybrid rooms is being developed with the incorporation of technologies that integrate room components (cameras, microphones, monitors, whiteboards) with web videoconferencing services. Zoom Rooms or Teams Rooms are examples of this type [145], as shown in Figure 52. Even so, several aspects still require improvements in technology and investments by companies: acoustic panels, multiple microphones, better echo cancellation and reverberation mitigation systems, automatic approach to the speaker, etc. It is expected that in the near future these technologies will be consolidated and popularized, significantly improving the quality of the hybrid conference experience.

When incorporating video calls and video conferencing into our daily lives, both personal and work, it is very useful to have virtual backgrounds, which avoid showing what we have behind our face. Almost all video conferencing systems have this function, allowing to blur the background, or insert a predefined image. However, the detection of the contour or boundaries of the face, or the parts of the body that fall within the visual field of the video call, is inaccurate. When there are multiple people in the scene, many times one of them is omitted and replaced by the virtual background. Image processing systems will need to substantially improve contour detection and the use of virtual backgrounds in these systems.

Figure 52. Huddle Room with Teams Rooms [145].

The use of avatars as a graphic representation of people can be used in videoconferences. A stylized image of each participant can replace the actual image during the video call. This technology is beginning to be offered on different platforms. Zoom and Teams already has it built-in, although with simple and unelaborate avatars. Whatsapp has announced it for future versions. Companies like Loom.ai have more elaborate avatar offerings, compatible with various video conferencing applications. An example is shown in Figure 53. Soon, more realistic and sophisticated avatars will surely be available, which faithfully represent the images of the participants. It may be difficult to distinguish whether what you are watching is a live video, or a digital representation of your interlocutor.
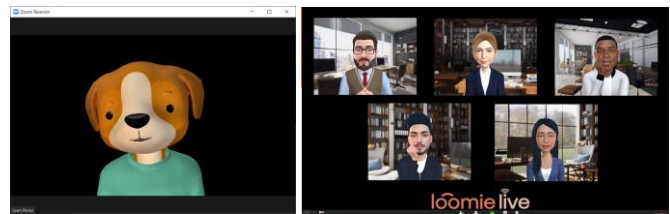

Figure 53. Left: Zoom avatar. Right: Avatars of loomie.ia [146].

When several people who speak different languages participate in a video conference, it often becomes difficult to understand them. A common language is usually selected for the dialogue (e.g., English). However, speaking in a non-native language makes communication difficult. Several of the companies that offer video conferencing services are starting to offer real-time simultaneous translation to text applications. For example, Cisco Webex has been offering the Machine Translation to Text service since 2021, as shown in Figure 54 [147]. The possibility of having simultaneous oral translations, in real time, will help to maintain much more fluid conversations, with each participant speaking in their language, which will be translated into another language and synthesized, using their same voice pitch.

At the end of the meetings, these systems should leave minutes, next actions, and/or summaries of the content, both for participants and for those who could not attend. In the future, applications such as GPT or similar may be applied to corporate videoconferencing, providing tools for automatic

summaries, automatic reports of pending issues or next actions.

Figure 54. Simultaneous translation to text [147].

With the announcement of the Metaverse by Facebook in 2021, immersive meetings are starting to develop. To participate in these meetings, participants must wear virtual reality headsets. The meetings in this virtual space are highly collaborative, where avatars of each participant interact, walk through the virtual space, write on shared whiteboards, see in three dimensions, among other interesting aspects. Figure 55 shows an example of such experiences.

A recent interview between Lex Fridman and Mark Zuckerberg (in September 2023) was conducted via virtual reality video conferencing. The profiles of Fridman's and Zuckerberg's heads were scanned and used to create extremely real avatars of each of them. The sensation was described by Fridman as follows: *"This is so great. And doesn't feel awkward to be really close to you... This is incredible. The realism here is incredible. We are surrounded by darkness with ultrarealistic face and just feels like we are in the same room"* [148]. Some scenes from the interview can be seen in Figure 56.

Several aspects are still in research and development so that this type of videoconferencing or virtual sessions can become popular. These include the design of lighter headsets, aspects of coexistence between the real and virtual worlds (e.g., how to include real-world items such as tables, chairs, etc. in the virtual world), aspects of physical safety of people (e.g., how to prevent a person from trying to lean on a virtual object that does not exist in the real world), fatigue and dizziness that many people often experience.

Figure 55. Participant in an immersive meeting, writing on a shared whiteboard. Left: The real-world participant. Right: View of the participant in the virtual world.

Figure 56. An immersive videoconference with ultrarealistic faces [148].

VIII. CONCLUSIONS

From its beginnings in the early 20th century, to the current state in the second decade of the 21st century, video conferencing systems were invented and reinvented for almost a hundred years. Each technological generation improved the quality and usability and lowered the prices of investment and use, to the point where video calls and videoconferencing are free in the second decade of the 21st century. With each technological change, various companies launched new products on the market, forecasting their wide and rapid adoption and popularization. Each time, the predictions failed. For decades, the use of videoconferencing systems was reduced to a few occasions and to specialized corporate services. It took a pandemic and the need for the forced confinement of people for these systems to finally become popular.

To maintain personal, family, and work relationships, during the pandemic people were forced to use video conferencing technology. Companies that already offered these technologies abruptly increased the number of users and developed new features in record time.

At the end of the lockdown, the use of video conferencing has been established. But even so, several people went back to sending messages or making audio calls, instead of using the video features. Everything seems to indicate that, beyond price, usability, quality and other factors, in most cases, people are simply not interested in being seen…

REFERENCES

[1] The New York Times, April 8, 1927

[2] Television, Herbert E. Ives, The Bell System Technical Journal, October 1927, pp 551-559

[3] A Successful Public Demonstration of Television Between Washington and New York, A. Dinsdale, Wireless World, June 1st, 1927.

[4] German patent no. 30105, Nipkow, P., Published January 15, 1884

[5] The Production and Utilization of Television Signals, Frank Gray, J. W. Horton y R. C. Mathes, Transactions of the American Institute of Electrical Engineers, June 1927

[6] The Use of a Moving Beam of Light to Scan a Scene for Television, F. Gray, J. Opt. Soc. Am. 16, 177-190, 1928

[7] Synchronization of Television, H. M. Stoller and E. R. Morton, Transactions of the American Institute of Electrical Engineers, June 1927

[8] Wire Transmission System for Television, D. K. Gannett and E. I. Green, Transactions of the American Institute of Electrical Engineers, June 1927

[9] Radio Transmission System for Television, Edward L. Nelson, Transactions of the American Institute of Electrical Engineers, June 1927

[10] Television, Prof. E. Taylor Jones, University of Glasgow, NATURE , No. 3007, Vol. 119, p.896, June 18 1927

[11] Television, J L Baird, Journal of Scientific Instruments, Vol. 4, Number 5, 1927

[12] A picture like that, rather trembly and strange, BBC Twitter Archive, https://twitter.com/i/status/1415211963815137286

[13] Persons in Britain Seen Here by Television As They Pose Before Baird's Electric 'Eye', The New York Times, February 9, 1928

[14] Two-Way Television, Herbert E. Ives, Frank Gray and M. W. Baldwin, Transactions of the American Institute of Electrical Engineers, October 1930

[15] Communications System, Herbert E. Ives, US Patent 2,099,115, Application April 8, 1930

[16] La Televisión obtuvo otro triunfo, URUGUAY Radiotelefonía, Montevideo, Año 1, No, 3, Julio 1930

[17] NATURE, No. 3265, Vol. 129, p. 788, May 28, 1932

[18] NATURE, Vol. 137, p. 39, March 7, 1936

[19] Issues in Telecommunication and Disability, Stephen von Tetzchner, COST 219, Commission of the European Communities, 1991

[20] Telephone-Based Instructional Systems, Palagu V. Rao, Bruce L. Hicks, Audiovisual Instruction, pp. 19-22, April 1972

[21] El medio telefónico y la educación, Domingo Gallego, Revista de educación, ISSN 0034-8082, No. 263, pp. 165-175, 1980

[22] Developing Picturephone Service, Arthur D. Hall, Bell Telephone Magazine, pp. 14-21, Spring 1964

[23] Ballad for the Fair, https://youtu.be/47xTOTFoTfc (Minute 7:07), AT&T, 1964

[24] The Silicon Diode Array Camera Tube, Merton H. Crowell and Edward F. Labuda, The Bell System Technical Journal, Vol. 48, Issue 5, 1969

[25] The Picturephone Set, Larned A. Meacham, IEEE Transaction on Broadcasting, Vol. BC-12, No. 1, June 1966

[26] Getting the Picture, C. G. Davis, Bell Laboratories Record, Vol. 47, No. 5, May/June 1969

[27] Debut of the First Picturephone, https://youtu.be/BQMnlKMFD8M, AT&T, 1970

[28] Picture-Telephone Service Is Started in Pittsburgh, Donald Janson, New York Times, July 1, 1970

[29] A long look ahead, Bell Telephone Magazine, Vol. 51, No. 1, p. 4, January/February 1972

[30] The Picturephone System: Switching Plan, J. F. Urich, The Bell System Technical Journal, Vol. 50, Issue 2, 1971

[31] The Picturephone System: Central Office Switching, P. N. Burgess and J. E. Stickel, The Bell System Technical Journal, Vol. 50, Issue2, 1971

[32] The Picturephone System: Baseband Video Transmission on Loops and Short-Haul Trunks, J. M. Brown, The Bell System Technical Journal, Vol. 50, Issue 2, 1971

[33] The Picturephone System: Digital Encoding of the Video Signal, J. B. Millard and H. I. Maunsell, The Bell System Technical Journal, Vol. 50, Issue2, 1971

[34] Anatomy of a failure: picturephone revisited, A. Michael Noll, TELECOMMUNICATIONS POLICY, May/June 1992

[35] Picturephone , Engineering and Technology History Wiki https://ethw.org/Picturephone

[36] British Movietone: Confravision, https://www.youtube.com/watch?v=EVC8RPJfjEk

[37] Television by Pulse Code Modulation, W M Goodall, The Bell System Technical Journal, Vol. 30, Issue 1, 1951

[38] The transmission of information, Robert M Fano, Technical Report No. 65, Research Laboratory of Electronics, MIT, March 17, 1949

[39] A Method for the Construction of Minimum Redundancy Codes, David A Huffman, Proceedings of the I.R.E., Vol. 40, No. 10, September 1952

[40] Image Coding by Adaptive Block Quantization, Manfred Tasto, Paul A Wintz, IEEE Transactions on Communication Technology, Vol. Com-19, No.6, December 1971

[41] Transmission and Reception of Pictures, Ray D Kell, US Patent 1,796,030, Application April 25, 1929

[42] Differential Quantization of Communication Signals, Cassius C Cuttler, US Patent 2,605,361, Application June 29, 1950

[43] Experiments with Linear Prediction in Television, C W Harrison, The Bell System Technical Journal, Vol. 31, Issue 4, 1952

[44] Television Coding Using Two-Dimensional Spatial Prediction, D. J. Connor R. F. W. Pease and W. G. Scholes, The Bell System Technical Journal, Vol. 50, Issue 3, 1971

[45] A video encoding system with conditional picture-element replenishment, F. W. Mounts, The Bell System Technical Journal, Vol. 48, Issue 7, 1969

[46] Interframe coding of videotelephone pictures, B.G. Haskell, F.W. Mounts and J.C. Candy, Proceedings of the IEEE, Vol. 60, Issue 7, 1972

[47] 1.544-Mbits/s Transmission of TV Signals by Interframe Coding System, Hiroshi Yasuda, Fumio Kanaya and Hisashi Kawanishi, IEEE Transactions on Communications, Vol. 24, Issue 10, October 1976

[48] Television Bandwidth Reduction by Encoding Spatial Frequencies, H. C. Andrews;W. K. Pratt, Journal of the SMPTE, Vol. 77, Issue: 12, 1968

[49] Diskrete Cosine Transform, N. Ajmed, T. Natarajan and K. R. Rao, IEEE Transactions on Computers, January 1974

[50] Picture Bandwidth Compression by Linear Transformation and Block Quantization, J W Woods and T S Huang, Symposium on Picture bandwidth Compression, MIT, Cambridge Mass., April 1969.

[51] History of Video Compression, Cliff Reader, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), 4th Meeting: Klagenfurt, Austria, 22-26 July 2002

[52] Digital Television Transmission Using Bandwidth Compression Techniques, Hisashi Kaneko and Tatsuo lshiguro, IEEE Communications Magazine, July 1980

[53] Data reduction of Picture Signals Review on the Studies in Japan Takahiko Fukinuki, Future Generation Computer Systems, Vol. 1, Issue 5, September 1985

[54] Transmitting 4-MHz TV Signals by Combinational Difference Coding H. Yasuda; H. Kuroda; H. Kawanishi; F. Kanaya and H. Hashimoto IEEE Transactions on Communications, Vol. 25, Issue 5, 1977

[55] European Collaboration on Picture Coding Research for 2 Mbit/s Transmission, John E Thompson, IEEE Transactions on Communications, Vol. 29, Issue 12, 1981

[56] High speed switched digital service, H. London, T. Giuffrida IEEE Communications Magazine, Vol. 21, Issue 2, March 1983

[57] The design of picturephone meeting service conference centers for video teleconferencing, Bernard A. Wright, IEEE Communications Magazine, Vol. 21, Issue 2, March 1983

[58] Picturephone service begins, The New York Times, July 9, 1982

[59] Bell trots out 'picturephone' service,Frank T. Csongos, UPI ARCHIVES July 8, 1982

[60] Teleconferencing, Ram Srinivasan, SPIE Digital Image Processing, Vol. 528, 1985

[61] PictureTel Corp. History, Funding Universe, http://www.fundinguniverse.com/company-histories/picturetel-corp-history/

[62] Method and apparatus for efficiently communicating image sequences, Brian L. Hinman, US Patent 4,703,350, Application Jun. 3, 1985

[63] A Motion-Compensated Interframe CODEC, K. Iinuma et al, SPIE International Technical Symposium/Europe - Cannes, France, 25 November 1985

[64] H.261: Códec para servicios audiovisuales a n x 384 kbit/s, CCITT, 1988

[65] Say Cheese: New Phone Also Takes Picture, L.A. Times Archives, June 22, 1986

[66] Ataritel; The Wraps Are About To Come Off, Ataritel Announcment, 1983

[67] Ataritel. http://www.atarimuseum.com/ataritel/index.html (accesible desde https://archive.ph/szBh0)

[68] Atari Oddities, https://www.technologizer.com/2012/02/12/atari-oddities/9/

[69]   https://www.flickr.com/photos/jeepersmedia/9723937345/in/photo stream/

[70] Popular Science, March 1988

[71] Visitel phone callers can reach out and see somebody, Jonathan Takiff and Knight-Ridder, Chicago Tribune, November 13, 1987

[72] Video-optimized modulator-demodulator with adjacent modulating amplitudes matched to adjacent pixel gray values, Charles S. Meyer, US Patent 4,739,413, April 19, 1988

[73] Conversational Video Phone, Lawrence D. Emmons, James S. Mackley, David H. Stokes, Rudolph S. Stefenel, William E., Brisko, Jr. and A. Cooper-Hart, US Patent 4,985,911, January 15, 1991

[74] Videohones como of age, Frank Vizard, Popular Mechanics, January 1989, pp. 72-74

[75] Mitsubishi VisiTel Used As a Still Camera, Japhy Riddle, https://youtu.be/yFwxhc9d3Zc

[76] Videoconference will fast forward in 1990, Salvatore Salamone Network World, Vol. 7, No. 5, January 29, 1990

[77] Market Place; Signals of a Boom In Video Meetings, Glenn Rifkin The New York Times, October 8, 1991

[78] Picturetel Corp´s new sysdems crash cost, equipment needed, for videoconferencing, CBR Staff Writer, February 10, 1991

[79] Compression Labs Adds H.261 video support to Rembrandt, CBR Staff Writer, July 13, 1992

[80] https://www.ccapitalia.net/galeria/main.php?g2_itemId=14039

[81] Video Conferencing: Silicon Valley's 50+ Year History, Eric Dorsey IEEE Silicon Valley Technology History Committee Event, July 22, 2020

[82] The videophone 2500 — Video telephony on the public switched telephone network, Scott H. Early, Andrew Kuzma and Eric Dorsey, AT&T Technical Journal, Vol. 72, Issue 1, 1993

[83] Face to face calling, Popular Science, February 1993, pp.31

[84] Narrow-Band visual telephone systems and terminal equipment CCITT, Recommendation H.320, 1990

[85] The evolution of conferencing, I Parke, BT Technol J, Vol. 15, No. 4, October 1997

[86] Retro Tech: Videophone '93, https://www.youtube.com/watch?v=I6El_iKwQsY

[87] PictureTel offers up PC-based video, Ellen Messmer Network World, Vol.10, No. 29, July 19, 1993

[88] New Products, ComputerWorld, February 6, 1995

[89] PictureTel Live50 & Live100 Installation Guide, Release 1.6, PictureTel Corporation, 1996

[90] Video System 200: Pictures Worth a Thousand Words, PC Magazine, May 17, 1994

[91] Desktop Videoconferencing, PC Magazine, June 14, 1994

[92] First IETF Internet Audiocast, Stephen Casner, ACM SIGCOMM Computer Communication Review, Vol. 22, Issue 3, July 1992

[93] MBone provides audio and video across de Internet, Michael R. Macedonia and Donald P. Brutzman, Computer, April 1994

[94] A Transport Protocol for Real-Time Applications (Draft), https://datatracker.ietf.org/doc/html/draft-ietf-avt-rtp-00, Audio-Video Transport WG, December 15, 1992

[95] Global SchoolNet It´s Not About Technology, https://youtu.be/PC33Y4Jbiys?t=47, 1993

[96] CU-SeeMe, Web Weaver, and WebSTAR 1.1, D Surry, TechTrends, Springer, 1995

[97] The INRIA videoconferencing system (IVS), Thierry Turletti Connexions, Vol. 8, No. 10, 1994

[98] RTP Paylad Format for H.261 Video Streams, T. Turletti, raft-ietf-avt-h261-00, IETF, 1995

[99] RTP Paylad Format for H.261 Video Streams, T. Turletti, raft-ietf-avt-h261-00, IETF, 1995

[100] IVS Home Page, http://planete.inria.fr/ivs/

[101] Drawing a bead on Desktop Conferencing, James Kobielus, Network World Collaboration, May/June 1995

[102] Recommendation H.323: "Visual telephone systems and equipment for local area networks which provide a non-guaranteed quality of service", ITU-T, November 1996

[103] Recommendation H.324: "Terminal for low bit rate multimedia communication", ITU-T, March 1996

[104] Intel Plans PC Video-Phone Technology, Laurence Zuckerman, The New York Times, May 30, 1996

[105] Intel Video Phone: Does it steal the show?, Ted Needleman, PC Magazine, October 7, 1997

[106] Microsoft NetMeeting Conferencing Software Provides Easy Voice, Data Internet Communications, Microsoft, May 29, 1996

[107] Netmeeting gets videoconferencing, Justin Hibbard, Computerworld, December 9, 1996

[108] Desktop Videoconferencing Kits, PC Magazine, October 7, 1997

[109] SIP: Session Initiation Protocol, ietf-mmusic-sip-02.txt M. Handley, H. Schulzrinne, E. Schooler, September 25, 1997

[110] A comprehensive multimedia control architecture for the Internet, Henning Schulzrinne, Proceedings of 7th International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV '97), 1997

[111] Voice Over IP, PC Magazine, October 3, 2000

[112] IP Multimedia Call Control Protocol based on SIP and SDP, 3GPP TS 24.229 V0.8.01.0.0, November, 2001

[113] Gadgets of the week: Skype adds video calling, The New York Times, December 1, 2005

[114] An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol, Salman A. Baset and Henning G. Schulzrinne, IEEE Infocom 2006

[115] GIPS Codecs, Global IP Solutions, http://www.gipscorp.alcatrazconsulting.com/files/english/datasheets/Codecs.pdf

[116] Internet Phone Service Creating Chatty Network, Ethan Todras-Whitehill, The New York Times, March 24, 2005

[117] EBay to Buy Skype, Internet Phone Service, for $2.6 Billion, Ken Belson, The New York Times, September 13, 2005

[118] For Microsoft, Skype Opens Vast New Market in Telecom, Steve Lohr The New York Times, May 10, 2011

[119] Yahoo's latest messenger boosts webcam quality, Brian Osborne, August, 2002

[120] ICQ offers video chat, Evan Hansen, CNet, October 25, 2004

[121] HP launches smaller Halo telepresence system, Matt Hamblen, Computerworld, March 18, 2008

[122] HP Halo Telepresence Solutions, Fact Sheet, https://www.hp.com/hpinfo/newsroom/press_kits/2008/halo/fs_telepresence.pdf

[123] Cisco Steps Into Telepresence With New HD Videoconferencing Line Jennifer Follett, CRN, October 23, 2006

[124] Polycom Announces Polycom RPX - First Immersive Environment With EyeConnect Technology For a RealPresence Experience Polycom Press Room, May 22, 2006

[125] HP Press Kit: HP Halo: Expanding Access, https://www.hp.com/hpinfo/newsroom/press_kits/2008/halo/products.html

[126] Cisco TelePresence System 3000, http://cisco-telepresence.blogspot.com/2011/05/cisco-telepresence-system-3000.html

[127] Polycom® RealPresence™ Experience High Definition (RPX™ HD) http://spectralink.polycom.com/global/documents/support/sales_marketing/products/video/rpx_hd_brochure_hi_res.pdf, Polycom, 3726-17534-001, 2007

[128] Video Telephony has finally arrived, Thomas Wiegand and Gary J. Sullivan, IEEE Spectrum, 25 August, 2011

[129] Adobe unveils Connect 8 Web and videoconferencing tool set, Matt Hamblen, Computerworld, November 1, 2010

[130] Blue Jeans launches "any(ware) video conferencing", Regina Hope Sinsky, Venture Beat, June 29, 2011

[131] Apple Introduces FaceTime Video Chat For iPhone 4, Jay Yarow Insider, Jun 7, 2010, https://www.businessinsider.com/apple-introduces-facetime-video-chat-for-iphone-4-2010-6

[132] Apple Brings FaceTime to the Mac, Apple Press Release, October 20, 2010

[133] iOS 12.1 brings Group FaceTime and new emoji to iPhone and iPad Apple Press Release, October 29, 2018

[134] WhatsApp Rolls Out Video Calling for a Billion-Plus Users, David Pierce, Wired, November 15, 2016, https://www.wired.com/2016/11/whatsapp-rolls-video-calling-billion-plus-users/

[135] Llamadas y videollamadas grupales, https://blog.whatsapp.com/group-calling-for-voice-and-video-is-here

[136] Introducing Skype for Business, Microsoft, November 11, 2014, https://www.microsoft.com/en-us/microsoft-365/blog/2014/11/11/introducing-skype-business/

[137] Microsoft Teams rolls out to Office 365 customers worldwide, Microsoft, March 14, 2017, https://news.microsoft.com/2017/03/14/microsoft-teams-rolls-out-to-office-365-customers-worldwide/

[138] Where did Zoom come from?, Jim Scheinman, Maven Ventures, April 18, 2019, https://medium.com/maven-ventures/where-did-zoom-come-from-96793bff6df0

[139] Zoom Video Communications Reaches 1 Million Participants, Robbie Pleasant, TMCNET FEATURE, May 23, 2013, https://www.tmcnet.com/topics/articles/2013/05/23/339279-zoom-video-communications-reaches-1-million-participants.htm

[140] Google release of WebRTC source code, https://lists.w3.org/Archives/Public/public-webrtc/2011May/0022.html

[141] Zoom User Stats: How Many People Use Zoom in 2022?, Brian Dean Backlink, January 6, 2022, https://backlinko.com/zoom-users

[142] Microsoft Teams Revenue and Usage Statistics (2022), David Curry Business Of Apps, June 30,2022, https://www.businessofapps.com/data/microsoft-teams-statistics/#:~:text=Microsoft%20Teams%20saw%20a%20huge,Zoom%20from%20February%20to%20June.

[143] Zoom Planes y Servicios, https://zoom.us/pricing

[144] Encuentra el plan de Microsoft Teams adecuado para tus necesidades https://www.microsoft.com/es-ww/microsoft-teams/compare-microsoft-teams-options

[145] Microsoft Teams Rooms, https://www.microsoft.com/en-us/microsoft-teams/microsoft-teams-rooms

[146] LoomieLive, https://loomai.com/loomielive

[147] Cisco Adds Real-Time Translation to Webex Video Meetings Matthew Humphries, PC Magazine, March 11, 2021, https://www.pcmag.com/news/cisco-adds-real-time-translation-to-webex-video-meetings

[148] Conversation between Lex Fridman and Mark Zuckerber, https://twitter.com/lexfridman/status/1707453830344868204, 28 September 2023

José Joskowicz was born in Montevideo, Uruguay in 1969. He is an Electronics Engineer specialized in Telecommunications from the University of the Republic of Uruguay, in 1995. He received the Ph.D. degree in Telematics Engineering from the University of Vigo, Vigo, Spain, in 2012. He is internationally certified as a Project Management Professional (PMP) by the Project Management Institute (PMI). He is an Associate Professor at the Faculty of Engineering of the University of the Republic of Uruguay, a member of the National Researchers System, and Principal Engineer at the company ISBEL. Throughout his professional career, he has conducted various consultancy projects and has led the design and implementation of various national and international projects in the field of Information and Communication Technologies (ICT). His academic research activities focus on the study of Quality of Experience in multimedia applications.