

Tecnologías de Videoconferencias: Pasado, presente y futuro

**Dr. Ing. José Joskowicz
Instituto de Ingeniería Eléctrica, Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
josej@fing.edu.uy**

**Setiembre 2022
Versión 2**

Tabla de contenido

1	Introducción	2
2	Los inicios: Sistemas analógicos y mecánicos	3
3	La era electrónica analógica	12
4	La era digital	20
5	La era IP	33
6	La era de los servicios cloud	42
7	El futuro.....	45
	Referencias.....	48

1 Introducción

Los sistemas de videollamadas y videoconferencias fueron inicialmente conceptualizados y desarrollados a principios del siglo XX. La primera videollamada fue realizada en 1927, incluso antes de que se transmitiera al aire la primera señal de televisión. Sin embargo, fue necesario esperar casi un siglo, hasta el final de la segunda década del siglo XXI, para que las videollamadas y videoconferencias comiencen a formar parte de la vida cotidiana.

En este trabajo se describe el desarrollo y la evolución de los sistemas de videoconferencias, desde sus inicios mecánicos y analógicos en la década de 1920, hasta los avanzados servicios IP ofrecidos desde la nube, en la década de 2020. Para cada generación tecnológica se presentan los aspectos técnicos y funcionales y se indican las ofertas comerciales más destacadas en cada momento. Finalmente, se aventuran algunas tecnologías que serán incorporadas y desarrolladas en los años siguientes.

En todos los casos se intentó buscar las referencias originales de la época, las que se indican a lo largo del texto. Es posible que falten menciones a hechos o productos destacados. Si es así, agradezco que me lo hagan saber, y con gusto podrá ser evaluado e incluido en futuras ediciones de este documento.

¡Espero que disfruten de la lectura!

2 Los inicios: Sistemas analógicos y mecánicos

La primera presentación pública de una videollamada fue realizada el 7 de abril de 1927, entre las ciudades de Washington y Nueva York, en Estados Unidos, a través de circuitos telefónicos. Esta primera demostración tecnológica consistió en una videollamada unidireccional. En un extremo, Walter S. Gifford, presidente de la empresa American Telephone and Telegraph (AT&T) se encontraba en Nueva York, utilizando un teléfono y un equipo receptor de imágenes, mientras que Herbert Hoover, Secretario de Comercio del gobierno de Estados Unidos, estaba en Washington, frente a un equipo capaz de captar y transmitir imágenes. En la misma presentación, inmediatamente después de la comunicación Washington - Nueva York, se realizó una segunda videollamada, esta vez mediante transmisión por equipos de radio (es decir, en forma inalámbrica), entre la estación experimental 3XN en Whippany, New Jersey y el mismo receptor en Nueva York. En esta demostración se vio hablar al Ingeniero E. L. Nelson, quien explicó cómo funcionaba el dispositivo.

Al día siguiente de la demostración, el diario New York Times titulaba *“La televisión triunfa en su primera demostración; Oradores lejanos son vistos y escuchados aquí, en una prueba de televisión”* [1]. *“Es como si una foto cobrara vida”*, se indicaba en forma destacada en el mencionado periódico. La Figura 1 y la Figura 2, publicadas junto con el artículo, muestran a los protagonistas de la primera videollamada. La Figura 3 muestra los estudios de Whippany durante la segunda videollamada inalámbrica. Cabe resaltar que la primera emisión de radiodifusión de televisión sucedió en 1928, un año más tarde de esta demostración.

Respecto a la utilidad práctica y comercial de estos sistemas, Gifford mencionaba [1]: *“Cuál sea su utilidad práctica, lo dejaré a su imaginación. Sin embargo, estoy seguro de que, de muchas formas y a su debido tiempo, se verá que aumentará sustancialmente la comodidad y la felicidad humanas.”*



Figura 1: Walter S. Gifford, presidente de AT&T en la primera demostración pública de la Televisión. El Sr. Gifford habla con el Secretario Hoover en Nueva York y es capaz de ver al Sr. Hoover en la pantalla inmediatamente frente a él. (El Sr. Hoover se encuentra a la derecha). Tomado de [1].

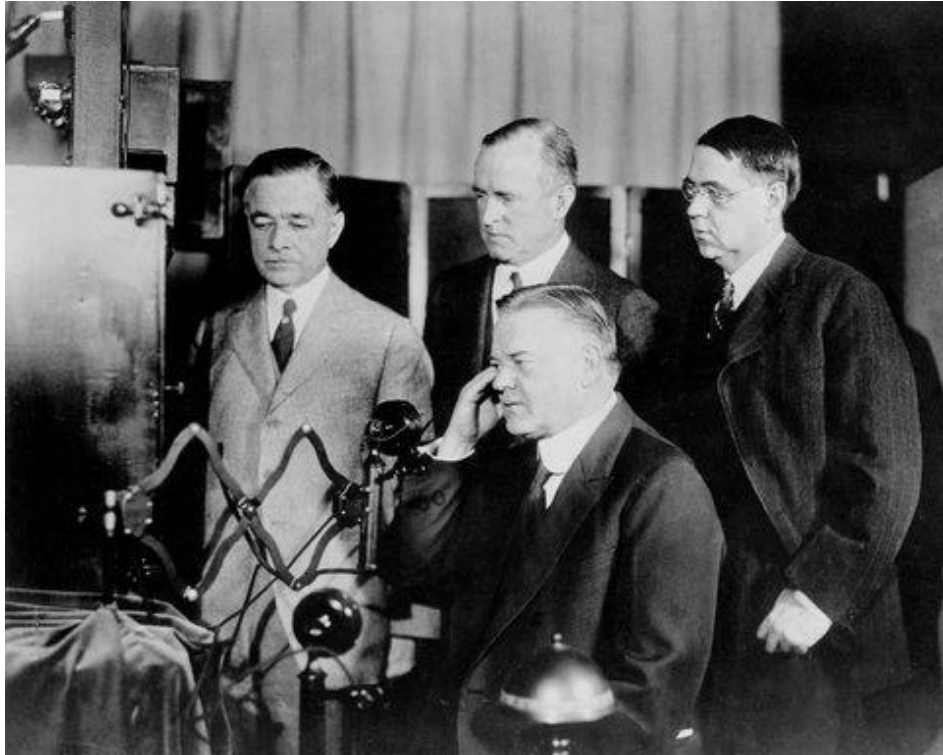


Figura 2: Herbert Hoover, Secretario de Comercio, se ha convertido en la primera personalidad de la Televisión. Se lo ve frente del dispositivo ubicado en Washington, transmitiendo su voz y su cara a la audiencia ubicada en los Laboratorios Bell en Nueva York. Tomado de [1].

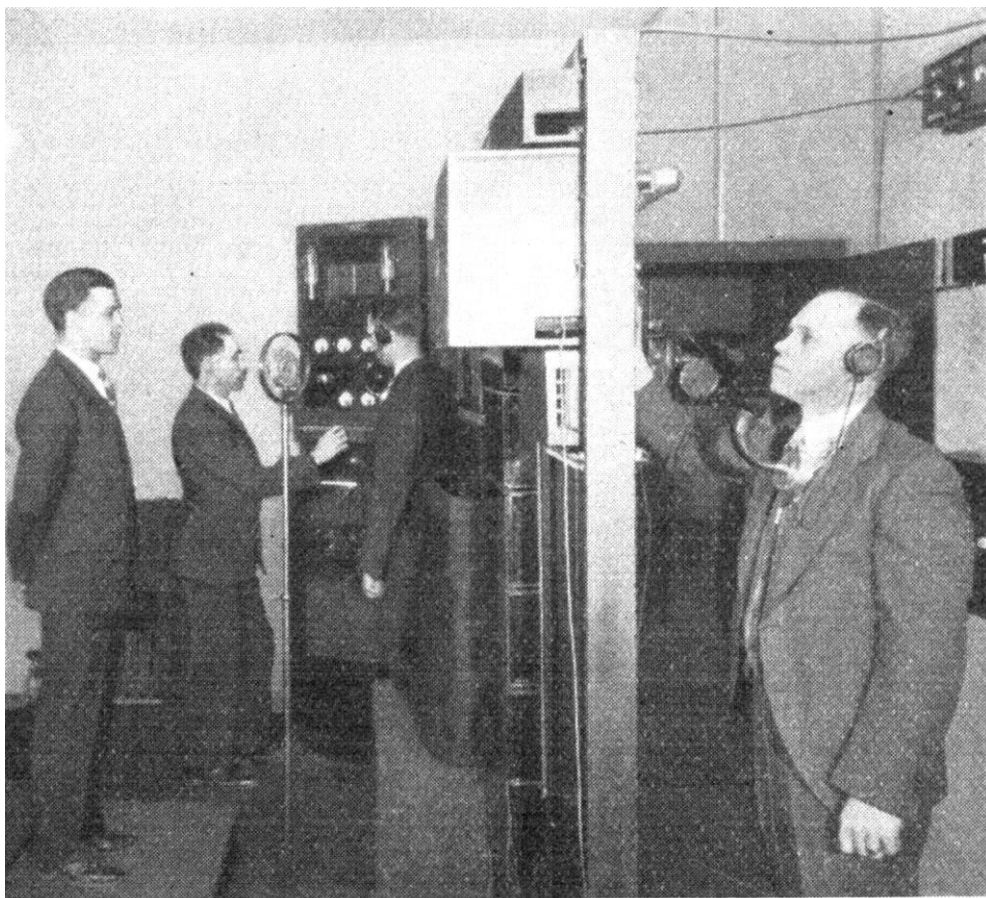


Figura 3: Equipos de transmisión en los estudios de Whippany, New Jersey. Tomado de [9].

El diseño y la construcción de este sistema de video telefonía lo realizó un equipo liderado por el Ingeniero Herbert E. Ives [2]. Si bien la mayoría de los conceptos básicos utilizados por el sistema eran ya conocidos, muchos problemas tecnológicos y prácticos tuvieron que ser resueltos para lograr un funcionamiento apropiado. La publicación "The Wireless World" de junio de 1927 presenta una muy buena descripción general del sistema [3]. El primer aspecto a considerar es la transformación de una imagen en señales eléctricas. Una solución a este problema había sido propuesta y patentada en Alemania varias décadas antes, en 1884, por Paul Gottlieb Nipkow [4], utilizando un disco perforado en forma de espiral, como se observa en la Figura 4. Para el diseño del videoteléfono, los ingenieros Frank Gray, J. W. Horton y R. C. Mathes, utilizaron un disco de este tipo [5][6]. La Figura 5 muestra un esquema conceptual. Un rayo de luz se enfoca para iluminar un área limitada sobre las aberturas móviles del disco. Un marco delante del disco permite que la luz atraviese solamente una abertura a la vez. Un lente delante del disco enfoca una imagen de esta abertura en movimiento sobre el objeto o persona cuya imagen se quiere transmitir. Como resultado, la persona u objeto es escaneado completamente en una serie de líneas paralelas sucesivas por un punto de luz que se mueve rápidamente. En el diseño de este video teléfono, se realizaba 18 escaneos completos por segundo (el equivalente a 18 cuadros por segundo). A medida que el punto de luz ilumina a la persona o el objeto, la luz es reflejada y recibida por células fotoeléctricas, colocadas frente de la persona u objeto que está siendo observado. La salida de corriente de las células fotoeléctricas es proporcional a la luz recibida, logrando el objetivo de obtener una señal eléctrica que varía en función de la intensidad de la luz que se quiere transmitir. La Figura 6 muestra una foto del equipo utilizado.

En la estación receptora, un disco similar al de la estación transmisora, también provisto de pequeñas aberturas dispuestas en forma de espiral, gira en sincronismo con el de la estación transmisora. El observador mira una pequeña abertura o marco rectangular frente al disco, como se ve en la Figura 5. Este marco solo permite ver una abertura del disco en el campo de visión. A medida que el disco gira, las aberturas atraviesan el marco una tras otra en una serie de líneas paralelas, cada una desplazada un poco respecto de la anterior, hasta que en una revolución completa del disco se cubre todo el campo. Detrás del disco hay una forma especial de lámpara de neón. En esta lámpara, el cátodo es una placa de metal plana con forma y área suficiente para llenar por completo el campo definido por el marco del frente al disco. De esta manera, la imagen se proyecta en un área pequeña, de 5 cm x 6.3 cm, que puede ser observada únicamente por una persona. La Figura 7 muestra una estación receptora y el detalle de la lámpara de neón utilizada.

Para su presentación a una gran audiencia, el sistema incluye otro dispositivo, de 61 cm x 76 cm, consistente en un tubo de neón, doblado en 50 secciones paralelas. El tubo tiene un electrodo interior y 2500 electrodos exteriores, 50 por cada tramo. Un sistema de conmutación mecánica energiza en forma secuencial a cada uno de los electrodos exteriores, generando una iluminación proporcional a la corriente recibida, la que, a su vez, es proporcional a la intensidad de luz del punto correspondiente en la persona u objeto transmitido. Esta pantalla, en términos actuales, tenía 2500 píxeles.

El reportero del New York Times mencionaba [1]: *"Cuando la imagen de la televisión se proyectaba en una pantalla de dos por tres pulgadas (5 cm x 6.3 cm), el parecido era excelente. Era como si una fotografía hubiera cobrado vida de repente y comenzara a hablar, sonreír, asentir con la cabeza y mirar de aquí para allá ... En la pantalla más pequeña el rostro y la acción*

se reproducían con perfecta fidelidad ... Cuando la pantalla se amplió a dos por tres pies (61 cm x 76 cm), los resultados no fueron tan buenos.”

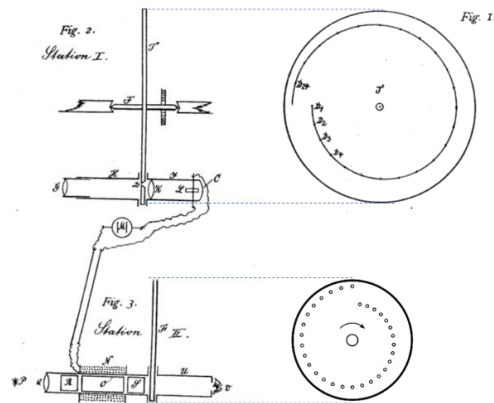


Figura 4: Esquema del disco de Nipkow, basado en [4].

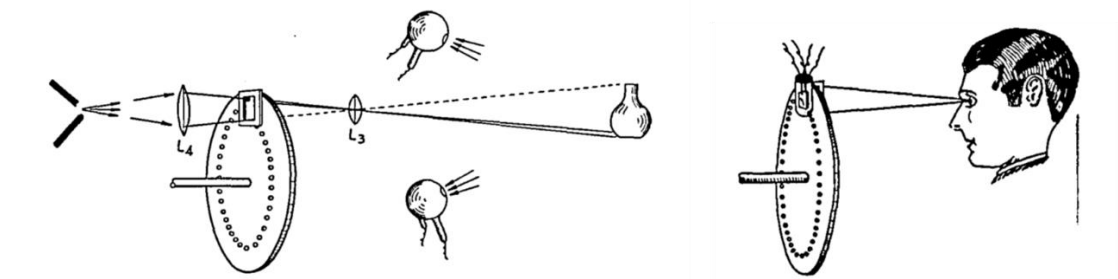


Figura 5: Izquierda: Utilización del disco de Nipkow para proyectar un haz de luz como un pequeño punto sobre el objeto a transmitir. La luz reflejada es ser recibida para varias celdas fotoeléctricas. Derecha: Aparato para reconstruir la imagen. La lámpara de neón detrás del disco se alimenta con una corriente proporcional a la generada por las celdas fotoeléctricas. Tomado de [6].

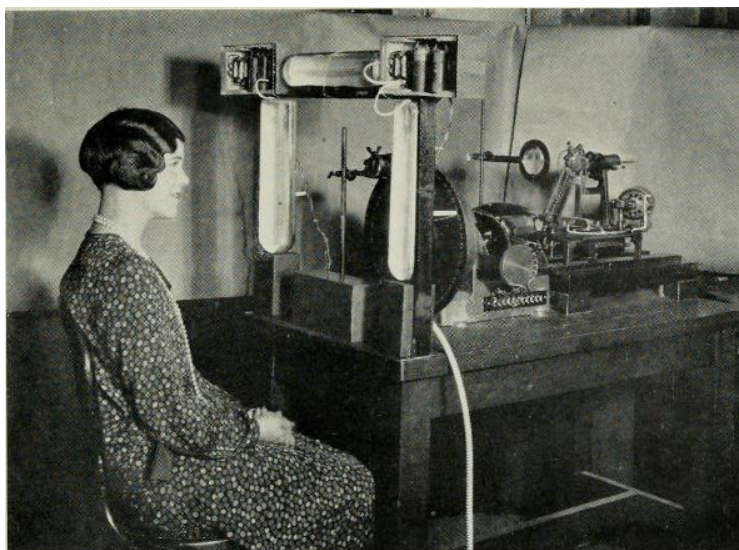


Figura 6: El disco gira 18 veces por segundo, utilizando un motor sincrónico. El disco tiene una espiral de orificios, cada uno de los cuales, permite el pasaje de la luz generada por una lámpara de arco, y lo proyecta en un punto de luz en movimiento sobre la persona. La luz reflejada por la persona es recogida por tres grandes celdas fotoeléctricas (ver detalle a la derecha de la figura), ubicadas en la parte superior y a los lados del marco. Tomado de [5].



Figura 7: Una lámpara de neón (mostrada en detalle a la derecha de la figura) funciona con la corriente de la imagen e ilumina el disco, que tiene una serie de pequeñas aberturas. A medida que las aberturas atraviesan el campo de visión, la intensidad de la lámpara varía, y el observador recibe la imagen. Tomado de [5].

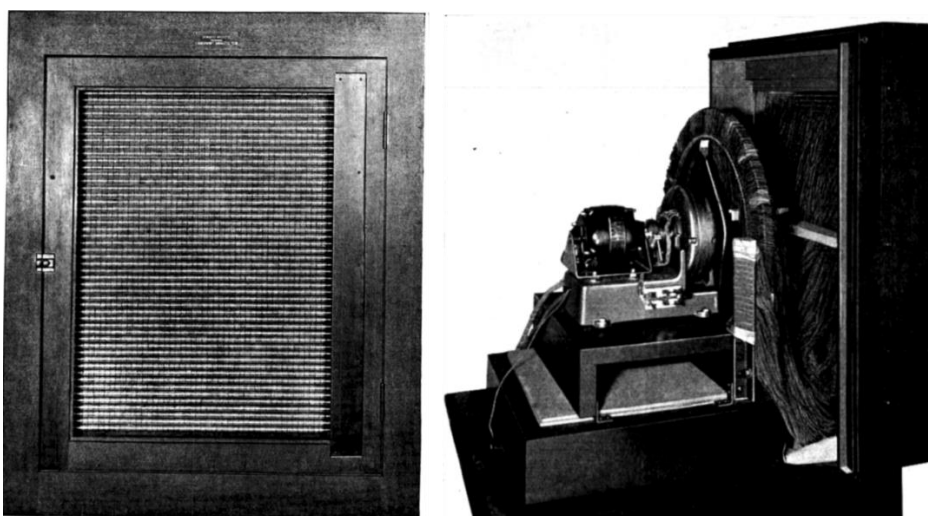


Figura 8: Izquierda: Pantalla formada por un tubo de neón, doblado en 50 filas y con 2500 electrodos. Derecha: Sistema de conmutación y cableado. La corriente se distribuye mediante 2500 cables a sucesivos electrodos, sincronizado con el giro del disco del emisor. Tomado de [5].

Otro de los problemas a resolver para el sistema fue el sincronismo entre los discos del emisor y las dos pantallas del receptor. Los ingenieros H. M. Stoller y E. R. Morton trabajaron en este aspecto, diseñando un mecanismo de control que permitió que el desplazamiento de fase entre los discos del emisor y del receptor sea el adecuado (menor de 4.3 minutos de arco, según el diseño teórico). Se decidió construir motores síncronos con 120 pares de polos, a una frecuencia de 2125 ciclos por segundo, utilizando un factor de reducción 120, logrando por tanto una

frecuencia de giro de 17.7 revoluciones por segundo [7]. La referencia de sincronismo fue tomada del equipo receptor y distribuida a la pantalla de neón (en el mismo sitio que el receptor) y al sistema emisor remoto.

Un esquema del sistema de transmisión mediante circuitos telefónicos se muestra en la Figura 9 [8]. Se utilizó una línea para enviar el video (con una línea de respaldo), una línea para las señales de sincronismo, una línea para enviar el audio y otra línea para recibir el audio. El video se enviaba en banda base, con frecuencias entre 10 Hz y 20 000 Hz. El sincronismo se transmitió en forma modulada. Para la transmisión inalámbrica se seleccionó una frecuencia portadora principal de 1575 kHz para el video, y una secundaria de 1545 kHz para el canal de audio [9].

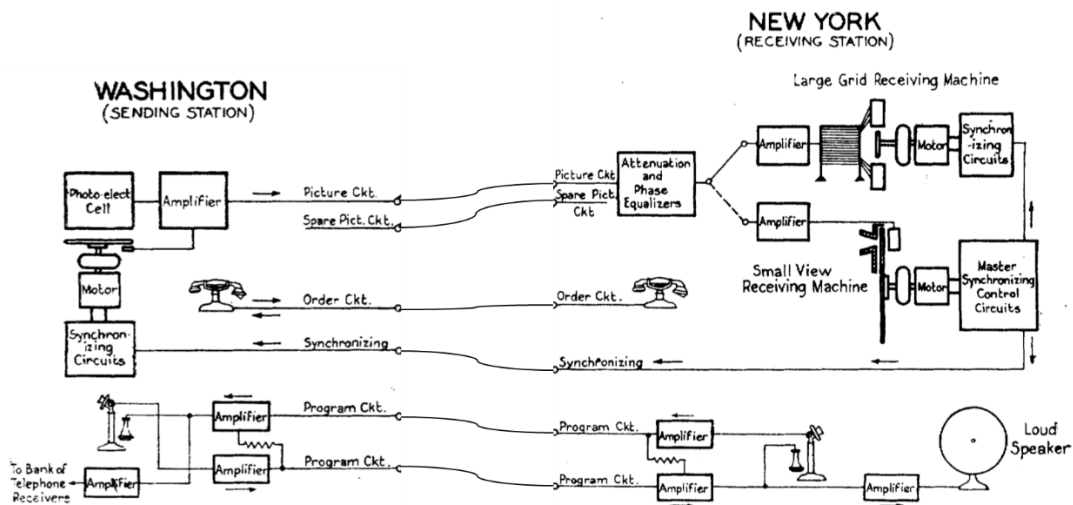


Figura 9: Esquema del sistema completo, basado en [8], incluyendo la estación emisora en Washington y la receptora en Nueva York. La comunicación entre ambos sitios fue realizada mediante líneas telefónicas: Una línea para enviar el video (con una línea de respaldo), una línea para las señales de sincronismo, una línea para enviar el audio y otra línea para recibir el audio.

El 24 de mayo del mismo año, 1927, el Ingeniero John L. Baird logra realizar una demostración similar a la de Ives, en este caso a través de líneas telefónicas entre las ciudades de Glasgow y Londres, en Inglaterra. El Profesor E. Taylor Jones, de la Universidad de Glasgow, estuvo presente durante la demostración, y publicó las siguientes notas [10]: *“El aparato receptor se instaló en una habitación semioscura, la lámpara y el obturador encerrados en una caja provista de una abertura. El observador que miraba dentro de la apertura vio al principio una banda vertical de luz en la que la luminosidad parecía viajar rápidamente hacia los lados, desapareciendo por un lado y luego reapareciendo por el otro. Cuando cualquier objeto que tenía "contraste" se colocaba a la luz en el extremo emisor, la banda se dividía en partes claras y oscuras formando una serie de "imágenes" del objeto. La impresión de movimiento lateral de la luz se perdió entonces casi por completo, y la imagen entera pareció formarse simultáneamente. La imagen estaba perfectamente estable en su posición, estaba notablemente libre de distorsiones.... El tamaño de la imagen era pequeño, no más de dos pulgadas de ancho... La cantidad de luz y sombra mostrada en la imagen fue suficiente para asegurar el reconocimiento de la persona que estaba siendo "televisada", y se vieron claramente los movimientos de la cara y los rasgos.”*

Las ideas y tecnologías usadas por Baird e Ives fueron similares, aunque la implementación fue diferente. En sus primeros sistemas, Baird utilizó un sistema con varios discos [11], como se muestra y explica en la Figura 10. En sistemas posteriores diseñados por Baird, el barrido de las líneas se realizaba en forma vertical, de arriba hacia abajo y de derecha a izquierda. En la Figura 11, tomada del video [12] se ve el punto de luz sobre la cara de la persona, la línea vertical que producía cuando el disco giraba a su velocidad nominal y la imagen obtenida en el receptor.

Cabe mencionar que, en el año siguiente, en 1928, Baird logró realizar la primera transmisión de televisión intercontinental, entre Londres y Nueva York [13].

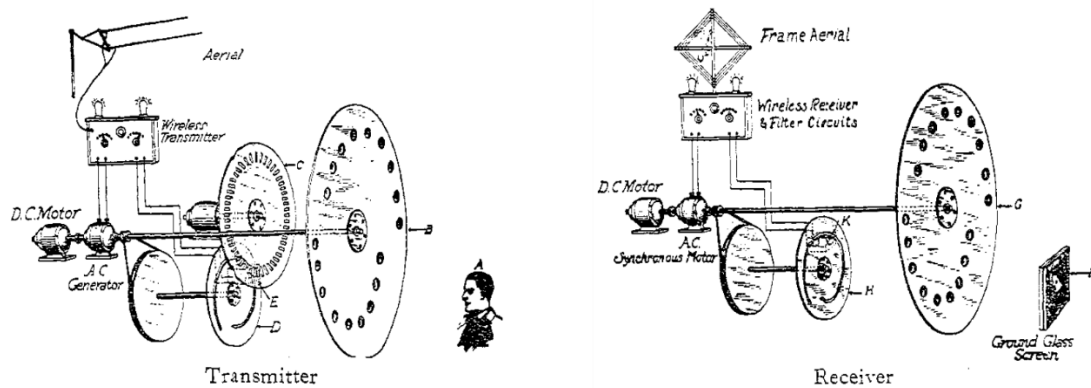


Figura 10: Sistema transmisor (izquierda) y receptor (derecha) de Bird. Por medio de un disco B que contenía lentes en formación escalonada, una serie de tiras de la imagen se hace pasar por E, la apertura de la celda sensible a la luz, después de ser interrumpida por el disco ranurado C. El disco espiral D realiza un grano más fino a la imagen subdividiendo las tiras. Tomado de [11].

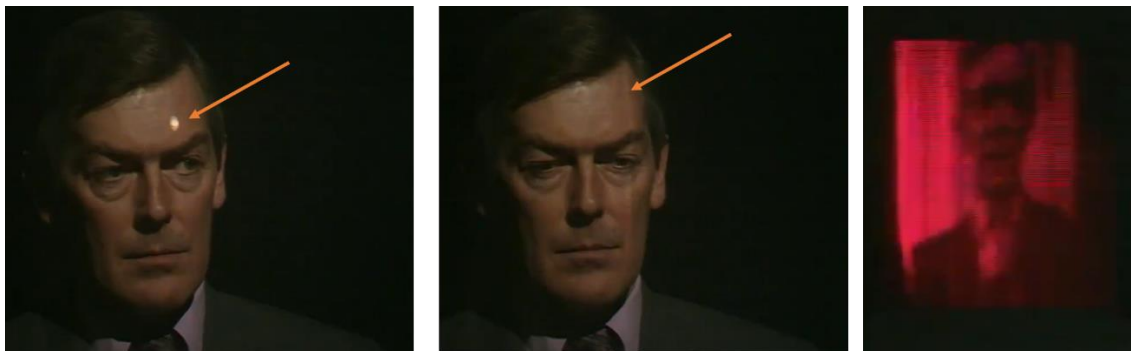


Figura 11: Escaneo realizado por el sistema de Baird. A la izquierda, punto de luz con el disco fijo. En el centro, línea de barrido vertical con el disco girando. A la derecha, imagen obtenida en el receptor. Tomado de [12].

Tres años luego de la primera videollamada unidireccional, en 1930, el equipo de Ives presenta un sistema de videollamadas bidireccional [14]. Su principio de funcionamiento es el mismo que el utilizado en la demostración de 1927, con las siguientes mejoras: Se utilizaron discos de 72 aberturas, en vez de las 50 del sistema anterior; se seleccionó una lámpara de luz azul en la estación emisora, ya que las celdas fotoeléctricas eran mucho más sensibles para esa longitud de onda; se aumentó la potencia de las lámparas de la estación receptora y se aumentó el ancho de banda de transmisión de video a 40 kHz (el doble que la utilizada en el sistema anterior). Ives patentó el sistema en abril de 1930 [15].

La transmisión entre sitios requería cinco circuitos telefónicos: dos para las señales de video, dos para las señales de audio, uno para las señales de sincronismo. Adicionalmente se disponía de un circuito de respaldo y otros dos circuitos para encender una luz de monitoreo remoto que indicaba que la cabina estaba ocupada. La Figura 12 muestra un esquema conceptual del sistema

completo. La Figura 13 presenta un detalle de la cabina especialmente diseñada para los interlocutores.

Respecto del nuevo sistema, Walter S. Gifford, mencionaba: “A pesar de que los trabajos de investigaciones y perfeccionamiento realizados en los tres años últimos han determinado una gran mejora y simplificación del equipo requerido para la televisión, éste resulta necesariamente complicado y costoso todavía. Debe ser atendido por un técnico y exige gran cantidad de aparatos. Ello se debe a los requisitos científicos indispensables a la obtención de una transmisión televisora satisfactoria. Así, pues, aunque en el aspecto técnico de la cuestión se han hecho progresos notables, las posibilidades comerciales de la televisión son inseguras todavía.” [16].

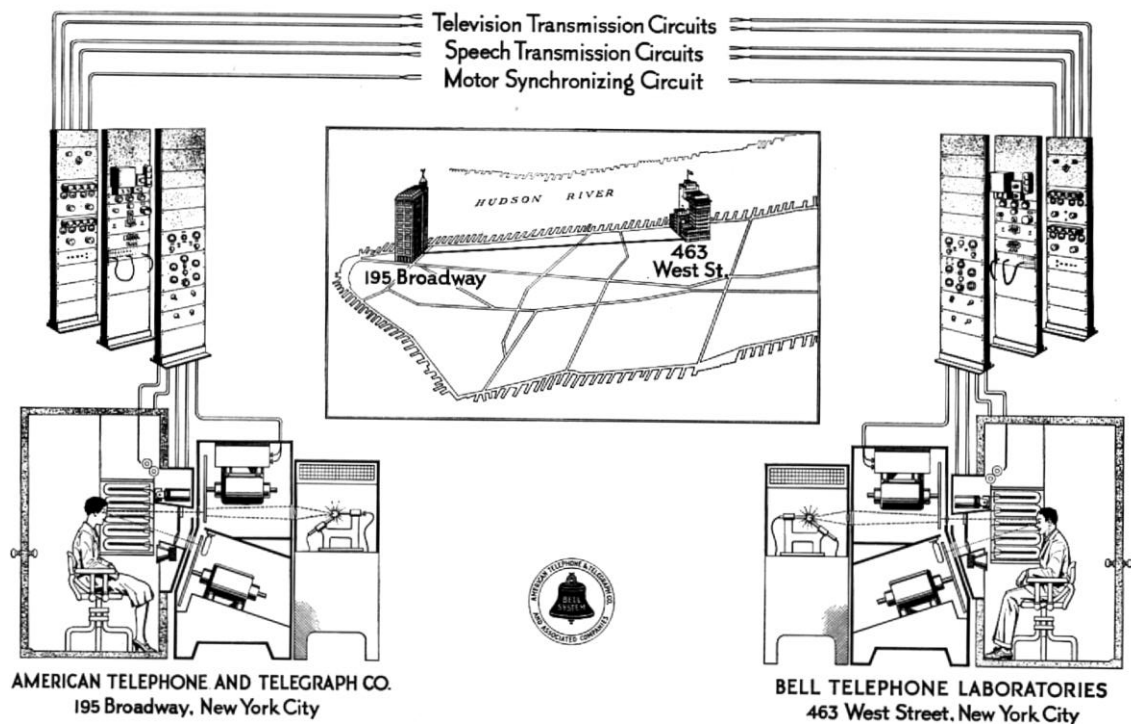


Figura 12: Representación del sistema de videollamada bidireccional. Tomado de [14].

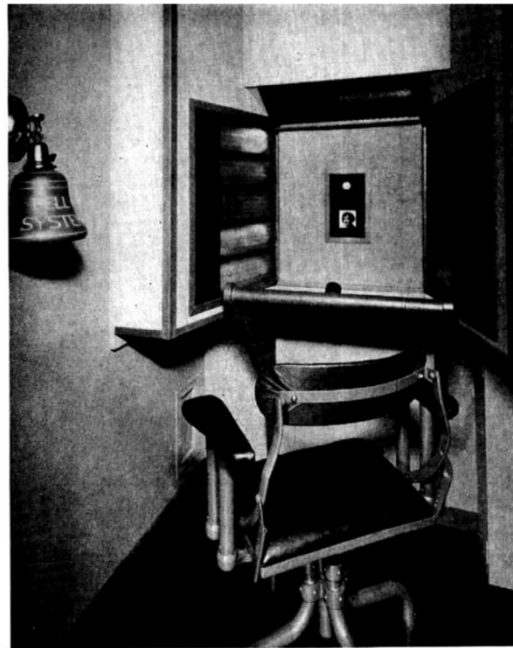
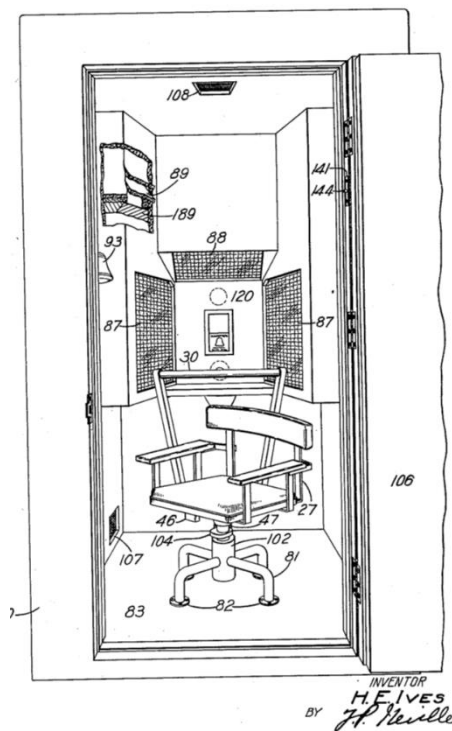


Figura 13: Izquierda: Diseño de la cabina del usuario del sistema, según la patente de Ives [15]. Derecha: Foto del interior de la cabina (Tomado de [14]).

El 19 de mayo de 1932, fue presentado en París un sistema de video bidireccional, basado en los trabajos previos de Baird. En este sistema, los discos giraban a 750 revoluciones por minuto, produciendo 12.5 cuadros por segundo. Según un reporte de la época, *“Los rasgos son fácilmente reconocibles y las expresiones en el rostro son notablemente claras. El movimiento de los labios se puede seguir fácilmente.”* [17].

El primero de de marzo de 1936, el ministro de servicios postales de Alemania, Eltz von Rubenach, inauguró el servicio de televisión y teléfono bidireccional de la Oficina Postal Alemana, que conecta Berlín y Leipzig, a una distancia de unos 190 km, por cable. El servicio fue promocionado y comercializado a 3 marcos 50 pfennigs los tres minutos de comunicación. Según la revista Nature, *“La calidad de las imágenes producidas parece ser buena. Se utiliza definición de 180 líneas y 25 fotogramas por segundo. La imagen de cabeza y hombros de una persona se produce claramente. El efecto es comparable a una proyección de tamaño pequeño de una película de cine de calidad inferior... Detalles como las correas de un reloj de pulsera o un anillo en la mano que sostiene el teléfono son claramente visibles. El aparato utilizado en Berlín fue construido por el laboratorio de la Oficina Postal Alemana y el utilizado en Leipzig por la Fernseh-Aktiengesellschaft de la cual Baird Television Ltd. posee una cuarta parte de las acciones.”* [18]. En la Figura 14 se muestra la cabina utilizada.



Figura 14: Servicio de video telefonía en Alemania, previo a la 2ª Guerra Mundial (Tomado de [19]).

3 La era electrónica analógica

Durante los siguientes treinta años, entre 1930 y la década de 1960, la televisión se desarrolló y popularizó. Sin embargo, los sistemas de video teléfono o videoconferencias no tuvieron avances significativos.

Algunas experiencias en audio conferencia fueron realizadas en ese periodo, relacionadas a ambientes educativos. En 1939 se registra el primer uso de una audioconferencia con muchos participantes, en Iowa, Estados Unidos. El Dr. Winterstein hizo instalar un sistema que permitía conectar a los alumnos, desde sus casas, con un aula donde se dictaban clases con otros alumnos. Los estudiantes que estaban en casa podían intervenir en la clase, hablando tanto con el profesor como con sus otros compañeros [20]. Otras experiencias similares se reportaron en las décadas de 1940-1950 [21].

En la “Feria Mundial de Nueva York” de 1964, los laboratorios Bell presentaron un prototipo del sistema “Picturephone” [22][23]. La primera demostración fue realizada el 20 de abril,

realizando una videollamada desde Nueva York al parque de Disneyland, ubicado en California. Este prototipo sirvió, no solamente para realizar una presentación pública del nuevo sistema que sería comercializado en poco tiempo, sino también para evaluar las preferencias de los usuarios. Mediante encuestas a quienes pasaban por la exhibición, se indagó acerca de qué usos específicos consideraban más importantes los potenciales clientes, qué calidad de imagen serviría mejor para estos usos, qué funciones se prefieren, qué controles se necesitan para activarlas, con qué frecuencia se usaría el servicio, entre otros aspectos [22]. En la Figura 15 se ve la portada de la revista "Bell Telephone Magazine", de 1964, mostrando las pruebas de laboratorio del nuevo dispositivo. La primera llamada de lanzamiento del sistema comercial fue el mismo año, unos meses más tarde, como se muestra y describe en la Figura 16.

Los objetivos generales del diseño del picturephone incluían un estilo atractivo, que fuera lo suficientemente pequeño para usarse en un escritorio o una mesa y con la menor cantidad de controles posible para que fuera fácil de usar. También tenía que ser de bajo consumo de energía y disipación de calor. Para lograr estos objetivos de diseño, se utilizaron dispositivos de estado sólido, a excepción de los tubos de captación y visualización de imágenes. El conjunto constaba de tres partes: una unidad de visualización, una unidad de control y una fuente de alimentación. Los dos primeros estaban al alcance del usuario, mientras que la fuente de alimentación estaba fuera de la vista, instalada debajo del escritorio. La unidad de visualización contenía un tubo de imagen, una cámara y un altavoz. La cámara fue diseñada especialmente, basada en una matriz plana de fotodiodos de silicio con polarización inversa, a los que se accedía mediante un haz de electrones de barrido de baja energía [24]. La unidad de control tenía un auricular de teléfono, un altavoz y varios botones, incluyendo el teclado telefónico de marcación por tonos DTMF. Mediante estos controles, el usuario podía seleccionar si enviaba o no su video, o auto visualizarse. La pantalla se diseñó en formato vertical, de 11.1 cm de ancho por 14.6 cm de alto. Se utilizaron 275 líneas por cuadro, a 30 cuadros por segundo (separados en dos campos entrelazados). La frecuencia horizontal era de 8250 Hz [25]. El detalle de los diferentes componentes de la primera versión del Picturephone se puede ver en la Figura 17, la Figura 18 y la Figura 19.

Cinco años más tarde, en 1969, una nueva versión del Picturephone fue diseñada [26]. Se simplificó la unidad de control, y se cambió el formato de la unidad de visualización por uno de aspecto rectangular, se amplió algo el tamaño de la pantalla, se ubicó la cámara justo por encima de la pantalla y se incluyó la posibilidad de conectarla a la salida de datos de una computadora. Una foto del nuevo estilo del Picturephone se muestra en la Figura 20. El nuevo equipo se lanzó al mercado en 1970 [27]. En esta nueva inauguración, el 30 de junio de 1970 en la ciudad de Pittsburgh, un ejecutivo de Bell Telephone System indicó que el impacto económico y social de esta nueva comunicación cara a cara sería igual al producido por la introducción de la conversación de voz telefónica en 1878 [28]. Lawrence J. Barnhorst, vicepresidente y gerente general de Bell Telephone Company de Pennsylvania indicó en la inauguración que, para 1975, 100 000 videoteléfonos Picturephone estarían en uso. Un informe, publicado en la revista Bell Telephone Magazine [29], predecía que en 1985 habría más tres millones de equipos Picturephone en los Estados Unidos (algo que finalmente no sucedió).



Figura 15: Portada de "Bell Telephone Magazine, Spring 1964", mostrando los experimentando de laboratorio con el Picturephone [22]



Figura 16: Primera llamada comercial del Picturephone: “El futuro se hace presente: el servicio PICTUREPHONE ya está aquí de forma comercial entre Nueva York, Chicago y Washington, DC. El día inaugural del servicio, el 24 de junio de 1964, se realizó la primera llamada desde el edificio de la National Geographic Society en Washington, D. C, de la Sra. Lyndon B. Johnson a la Dra. Elizabeth A. Wood, científica de Bell Telephone Laboratories. Arriba, uno de los beneficios más sorprendentes del nuevo servicio se demostró el día de la inauguración cuando Laura Rabinowitz, una estudiante sorda de 15 años y Howard Mann, de 14 años (en la pantalla), se comunicaron por lectura de labios durante la primera llamada con PICTUREPHONE”. Basado en [22].

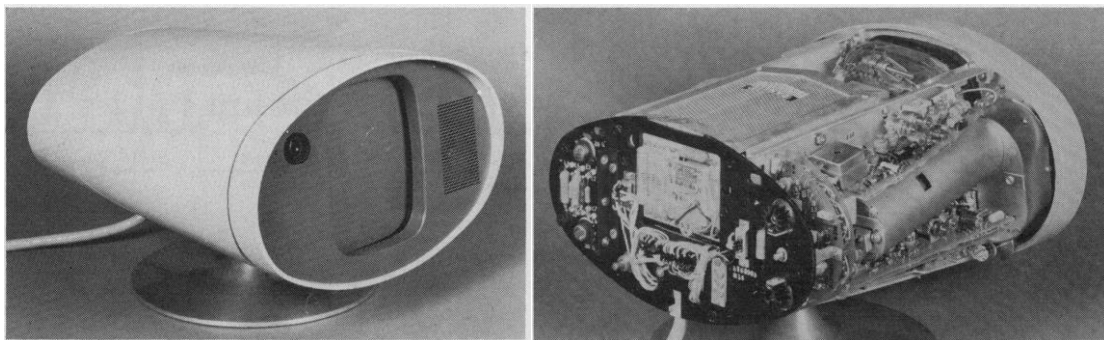


Figura 17: Unidad de visualización y cámara, vista exterior y detalle interior. Tomado de [25].



Figura 18: Unidad de control y teléfono, vista exterior y detalle interior. Tomado de [25].

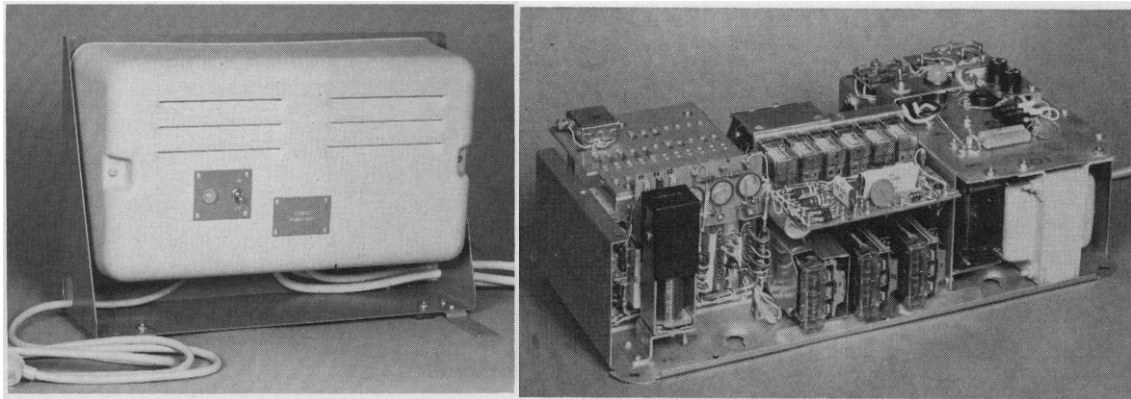


Figura 19: Unidad de energía, vista exterior y detalle interior. Tomado de [25].

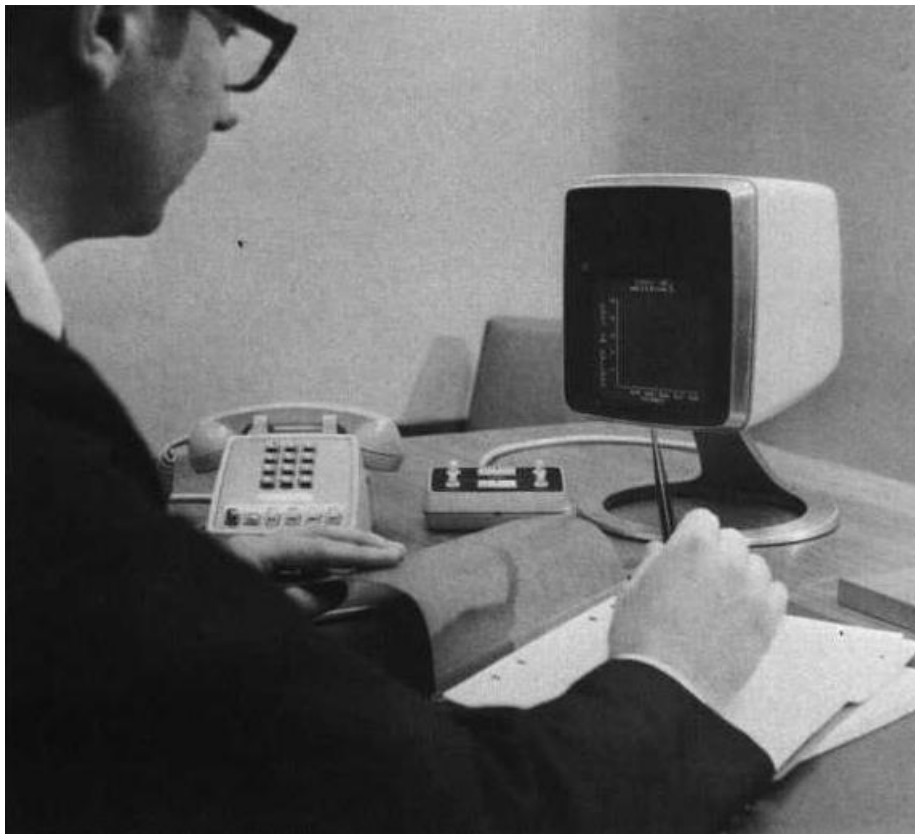


Figura 20: Picturephone "Mod II". Tomado de [26].

Un diagrama conceptual del diseño del Picturephone se muestra en la Figura 21. Fue natural en aquel momento, incluir el servicio de Picturephone asociado a los servicios telefónicos existentes. Las llamadas de video se indicaban precediendo el número telefónico de destino con el signo #, desde el teclado telefónico DTMF. La señal de video resultante del Picturphone era de 1 MHz. Para su transmisión se utilizaban los pares telefónicos existentes, entre las casas u oficinas de los usuarios y las centrales telefónicas. Se decidió utilizar un par para el video entrante, otro para el video saliente, y un par para el audio telefónico. Las centrales telefónicas

existentes estaban diseñadas para conmutar audio, con un ancho de banda mucho más acotado, de 4 kHz, por lo que fue necesario diseñar un nuevo sistema de conmutación y transmisión para las videollamadas. En la Figura 22 se muestra el esquema conceptual utilizado para la conmutación y transmisión del servicio Picturephone [30]. El video podía ser conmutado en nuevas centrales telefónicas públicas del tipo “crossbar”, con tecnología analógica, diseñadas para conmutar señales de hasta 1 MHz [31]. Para lograr una atenuación uniforme en todo el ancho de banda, fue necesario incluir amplificadores con ecualización [32] en diferentes puntos del sistema, como se muestra en la Figura 23.

Para la transmisión entre centrales públicas, generalmente distantes, se utilizó un sistema digital, de 6.3 Mb/s. Para ello se utilizó una codificación del tipo Differential Pulse Code Modulation (DPCM) de 3 bits [33], lo que equivale a tener hasta 8 posibles niveles de transmisión. En la Figura 24 se observa el resultado de la digitalización sobre una imagen tomada de Picturephone.

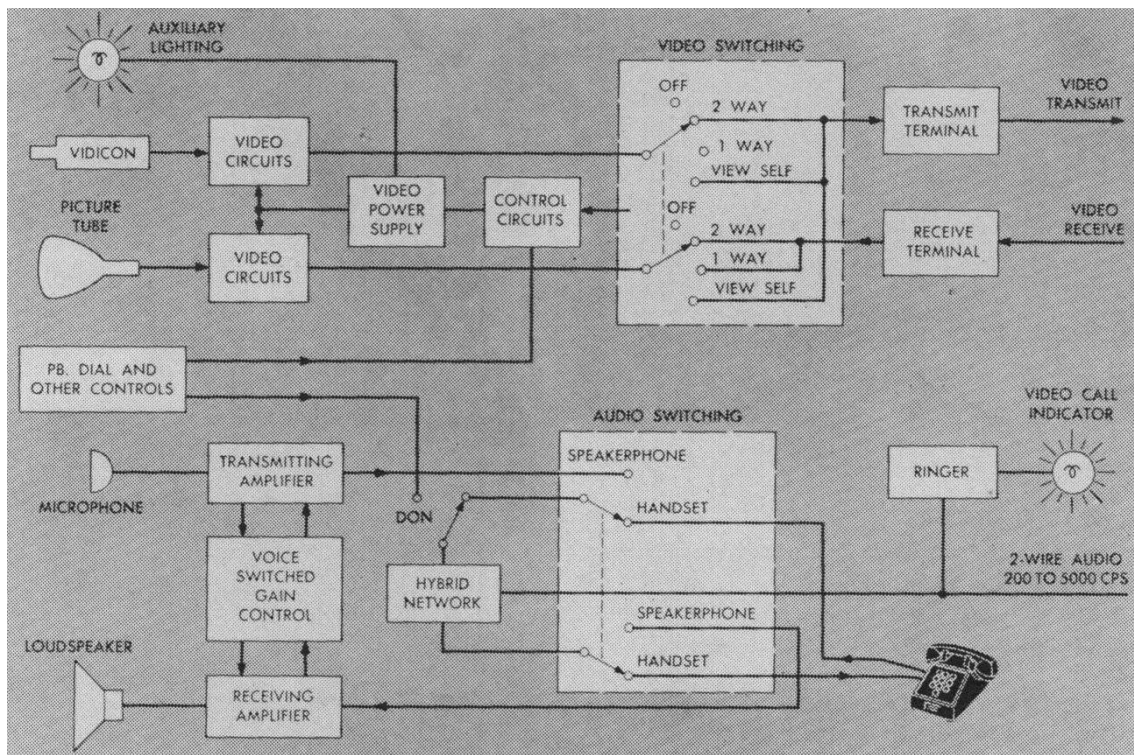


Figura 21: Diagrama de bloques conceptual del Picturephone. Tomado de [25].

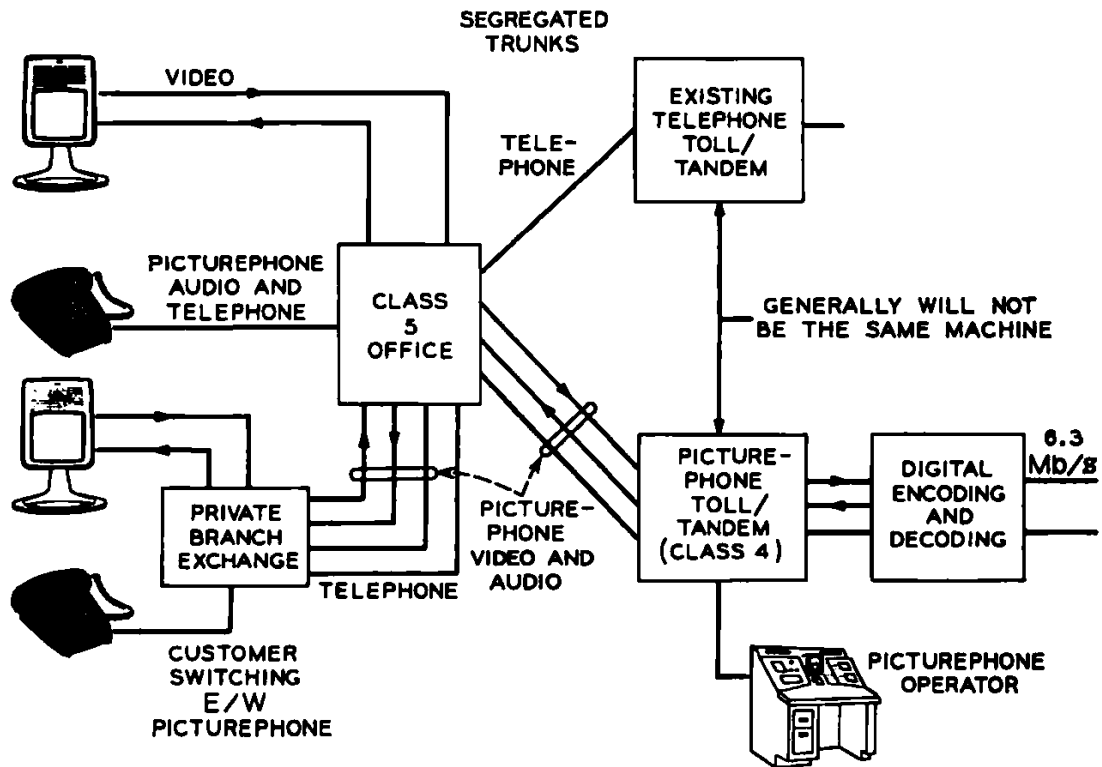


Figura 22: Sistema de conmutación y transmisión del Picturephone. Tomado de [30].

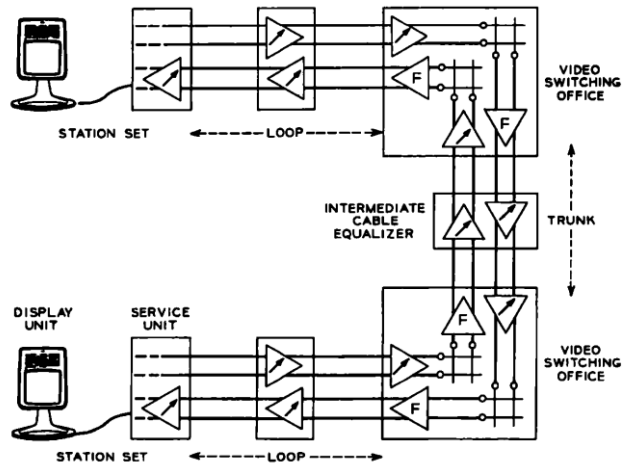


Figura 23: Sistema de transmisión del Picturephone. Tomado de [32].



Figura 24: A la izquierda: Imagen analógica. A la derecha: Imagen digitalizada con técnica DPCM de 3 bits. Tomado de [33].

El diseño técnico fue desafiante, pero se logró un producto final funcional, estéticamente muy bueno, y con buena calidad de imagen. Sin embargo, el servicio no tuvo el éxito comercial esperado. Luego de dos años de su segundo lanzamiento comercial, en 1972, y con una inversión estimada en 500 millones de dólares [34] solo se contaban con 32 videoteléfonos contratados en la ciudad de Pittsburg. Claramente, no se lograría ni siquiera estar cerca de los 100 000 videoteléfonos pronosticados para 1975 en esa ciudad. La cantidad de aparatos en Chicago alcanzó un máximo de 453 a principios de 1973, a pesar de que AT&T redujo el precio del servicio y de los minutos de conversación [35]. En julio de 1975, solo 76 clientes aún tenían el servicio, y en 1977 solo quedaban nueve clientes en Chicago [34]. El servicio fue dado de baja.

En la década de 1970 en Europa se desarrollaron algunos sistemas de videoconferencia de tecnología analógica, entre los que destacan Confravision en Reino Unido (básicamente dos estudios de televisión conectados por líneas de broadcasting) [36] y Visioconference en Francia.



Figura 25: Videoconferencia con el sistema Confravision, en Reino Unido. Tomado de [36]

4 La era digital

Las técnicas necesarias para digitalizar y codificar la señal de video de manera eficiente se venían desarrollando desde la década de 1940. La primera propuesta de digitalización, utilizando técnicas de modulación por pulsos codificados, o Pulse Code Modulation (PCM) fue presentada en 1949 y publicada en 1951 [37]. En este trabajo se evaluó la codificación con hasta 5 bits por muestra, mostrando resultados aceptables en la calidad de la imagen. La codificación entrópica se propuso en 1949 [38] y se desarrolló en 1952 [39]. Esta técnica consiste en representar los valores más frecuentes con pocos bits, y los valores menos frecuentes con más bits, generando un tipo de codificación de largo variable, o Variable-Length Code (VLC). Conociendo el tipo de señal a codificar, y seleccionando apropiadamente la forma de representación, es posible minimizar la cantidad total de bits necesarios para digitalizar un mensaje. La codificación entrópica se aplicó por primera vez a la codificación de video varios años después, en 1971 [40].

Desde el inicio de la televisión, estaba claro que la señal a transmitir era altamente redundante, tanto en la información interna correspondiente a cada cuadro, como en la sucesión de cuadros. Ya en 1929 se presentó una patente donde se proponía transmitir únicamente, en forma analógica, las diferencias de cada cuadro con el anterior [41]. Las técnicas de modulación digital predictiva, conocidas inicialmente como Differential Pulse Code Modulation (DPCM) se inventaron en 1950 [42], y se propuso su aplicación por primera vez al video en 1952 [43]. Hasta comienzos de la década de 1970, varias técnicas de DPCM se desarrollaron para optimizar la codificación dentro de cada cuadro [44]. Por ejemplo, dos líneas consecutivas dentro de un mismo cuadro pueden ser muy similares y por lo tanto, es más eficiente codificar únicamente las diferencias. A mediados de la década de 1970, DPCM también se comenzó a usar en el dominio temporal, es decir, para la predicción entre cuadros [45][46][47]. Dos cuadros consecutivos de una señal de video son típicamente muy similares entre sí, especialmente en aplicaciones de videollamadas o videoconferencias, donde la imagen típica es del tipo “cabeza y hombros” y con poco movimiento. Con esta idea, diversas técnicas permiten utilizar la información de un cuadro para predecir la del cuadro siguiente, y codificar únicamente la diferencia con la predicción, como se muestra en la Figura 26, publicada en 1969.

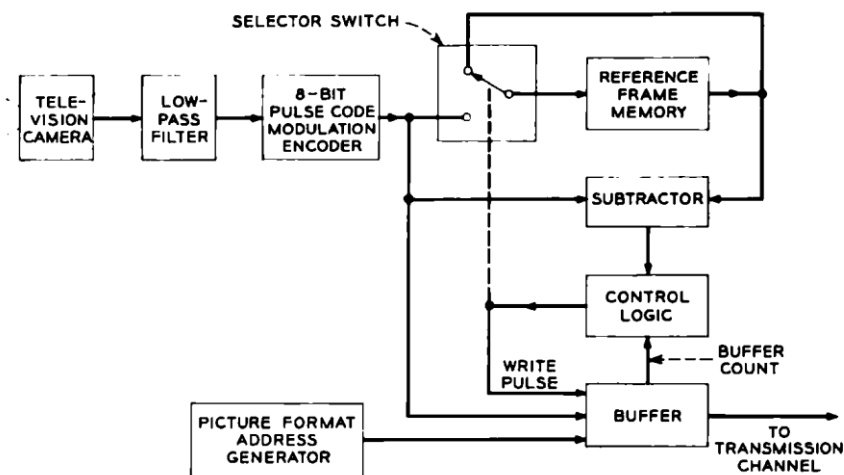


Figura 26: Codificación de video diferencial, considerando la similitud entre cuadros consecutivos. Tomado de [45].

Los cuadros de video digitalizados se codifican con una matriz de muestras de información de brillo y color, con un espaciado entre muestras capaz de reproducir el mayor detalle en cada punto. Esta información también se puede representar en términos de *frecuencia espacial*, de una manera equivalente a la representación en frecuencia de señales variables en el tiempo, pero extendida a espacios de dos dimensiones. De esta forma, la imagen puede analizarse o procesarse mediante el uso de técnicas similares a la transformada de Fourier. En escenas con pocos detalles habrá un predominio de información de baja frecuencia. Por lo tanto, mientras que en el dominio espacial la energía de la señal se distribuye uniformemente entre las muestras, en el dominio de la transformada la energía se concentra en unas pocas muestras. Esta observación condujo al desarrollo de la codificación basada en transformadas a fines de la década de 1960. En 1968 se propuso por primera vez utilizar la transformada de Fourier para codificar video [48]. Otros tipos de transformadas, más eficientes que la de Fourier, fueron propuestos en los siguientes años. En 1974 se propone utilizar la transformada discreta del coseno o Discrete Cosine Transform (DCR) [49], que presenta varias ventajas y mejor desempeño que las anteriores. La DCT continúa siendo utilizada en codificación de video hasta el día de hoy. La transformada produce una matriz de coeficientes casi no correlacionados. Para escenas con poco detalle, la energía o varianza de las muestras no se distribuye uniformemente, sino que se concentra en los coeficientes de más baja frecuencia. A estos coeficientes se les asigna un número mayor de bits para codificar y cuantificar, con una precisión apropiada. A los coeficientes que representan mayor frecuencia se les pueden asignar menos bits y cuantificarse de manera aproximada. Los coeficientes de mayor frecuencia pueden descartarse, y no necesitan codificarse.

Aplicar estas transformadas a cuadros completos resultó ineficiente y costoso desde el punto de vista computacional. Pero aplicarlo en pequeños cuadrados (por ejemplo, de 8x8 o 16x16 pixeles) produce mejores resultados y requiere menor capacidad de cómputo. El concepto del uso de pequeños bloques para aplicar en ellos las transformadas fue introducido en 1969 [50], y aplicado a partir de ese momento a la codificación de imágenes y video.

A partir de mediados de la década de 1970, los elementos clave de la codificación DPCM se fusionaron con la codificación que utilizaba transformadas para crear una codificación híbrida, la que se comenzó a utilizar a principios de la década de 1980 [51]. Este tipo de codificación híbrida, combinado transformadas con técnicas predictivas, marcó la era digital de las teleconferencias, entre fines de los 1970 y comienzos del 2000.

En Japón, la empresa NEC desarrolló una serie de productos comerciales para videoconferencias, utilizando técnicas digitales con predicción temporal entre cuadros [52][53]. La línea de productos se conoció como NETEC, y se comercializaron a partir de 1976. El modelo NETEC-6 (mostrado en la Figura 27) requería entre 6 Mbps y 8 Mbps de ancho de banda de transmisión. Los modelos NETEC 6/3 podían funcionar a 3 Mbps y el modelo NETEC-X1MC a 1.5 Mbps. Por su parte, Nippon Telegraph and Telephone (NTT) diseñó el sistema llamado TRIDEC en 1977 [54] y lo comercializó en 1979.

En 1977, siete países europeos (Bélgica, Francia, Holanda, Italia, Suecia, Reino Unido y Alemania Occidental) comienzan sus propios desarrollos en codificación digital de video, bajo el grupo llamado Co-Operation in Scientific and Technical Research (COST 211). Se trató de un proyecto de investigación colaborativo, buscando técnicas de codificación que utilicen reducción de redundancia, con el objetivo de transmitir videoconferencias a 2 Mbps [55]. Se esperaba tener un sistema de videoconferencias común para Europa, funcionando a 2 Mbps, para 1984.

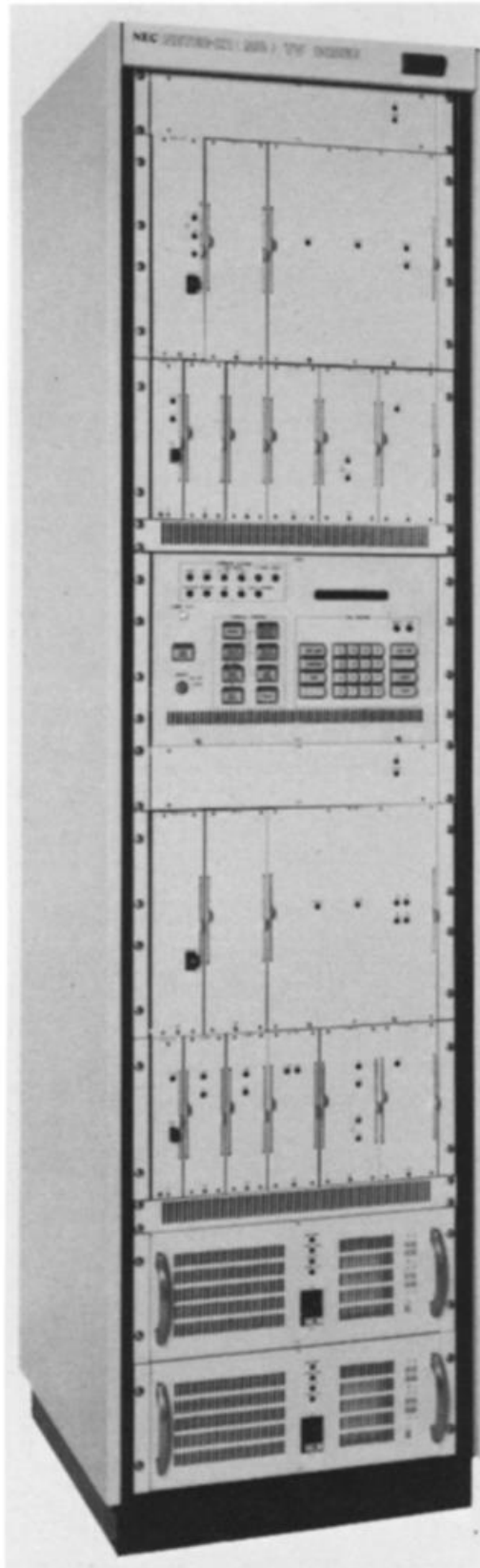
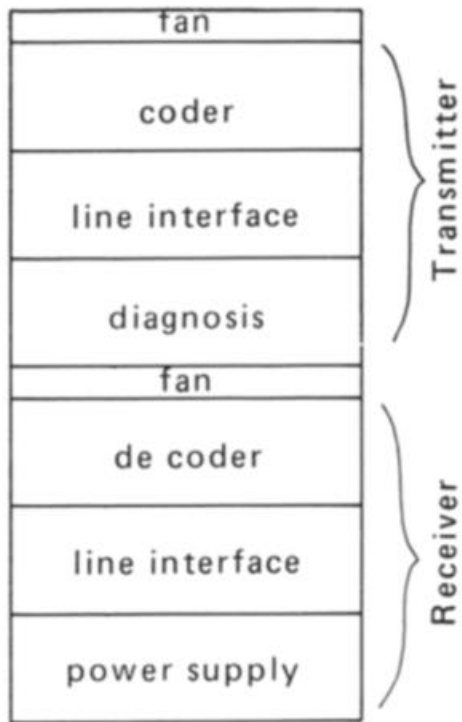


Figura 27: Codec NETEC 6. Tomado de [53].

A comienzos de la década de 1980 se comenzaron a desplegar los primeros sistemas de transmisión digital. En Estados Unidos, en 1982, AT&T comenzó a ofrecer el servicio llamado High Speed Switched Digital Service (HSSDS) [56]. El servicio ofrecía comunicación digital, a 3 Mbps, a través de centros de conmutación digital desplegados en la red de AT&T. Los 3 Mbps se lograban mediante la combinación de dos enlaces del tipo DS-1, cada uno con una velocidad de 1.544 Mbps. El sistema de conmutación ofrecido era manual, mediante reserva telefónica previa. Una operadora tomaba la reserva, y se planificaban las interconexiones necesarias en la red interna del operador, para realizar y mantener la conexión durante el periodo de tiempo solicitado. El primer servicio ofrecido y brindando a través de esta nueva red digital fue el Picturephone Meeting Service (PMS) [57]. En julio de 1982, el New York Times titulaba *“Comienza el servicio Picturephone”* [58]. La Compañía AT&T inauguró el nuevo servicio de videoconferencias realizando la primera llamada entre Nueva York y Washington. La nota aclaraba *“El nuevo servicio es muy diferente del Picturephone que se mostró en la Feria Mundial de Nueva York de 1964-65. Esa versión, que nunca se popularizó, tenía pantallas de escritorio para uso individual, no grupal, y su tecnología era menos sofisticada... Con PMS los clientes pueden alquilar salas públicas para utilizar el nuevo servicio o construir salas en sus propias instalaciones. Una sala de Picturephone tendrá cámaras a color, micrófonos y monitores. Se pueden transmitir diapositivas y gráficos. La sala también ofrecerá una fotocopidora para reproducir las imágenes que se muestran en el monitor entrante y una grabadora de cinta de video para grabar la imagen y el sonido”*. Los precios del uso del servicio variaban según la ubicación de las salas, entre 1300 y 2500 dólares por reunión.

El servicio ofrecía la realización de videoconferencias interactivas y bidireccionales entre dos salas de conferencias compatibles, conectadas a la red HSSDS. Las salas de conferencias se podían implementar en las instalaciones privadas de empresas o en lugares convenientes para el uso público. Cada videoconferencia debía ser reservada previamente, por la forma de trabajar del servicio HSSDS. El servicio utilizaba codificación digital del audio y video, a una tasa de bits compatible con el servicio de transporte HSSDS utilizado. Las salas de conferencias para el servicio de PMS incluían el uso de varios micrófonos, parlantes, monitores y cámaras de video. Hasta 12 participantes podían estar sentados en la sala de conferencias (6 en primera fila, y otros 6 detrás). La cobertura de video de la sala de conferencias se proporciona con tres cámaras de primer plano, una cámara de visión general, dos cámaras gráficas y una cámara multipropósito. Cada cámara de primeros planos enfocaba a un par de conferencistas en la mesa (y a los dos conferencistas sentados detrás). Estas cámaras se seleccionaban automáticamente cuando hablan los conferenciantes de la mesa. Por otra parte, la cámara de visión general proporcionaba una vista de gran angular de la sala de conferencias y los participantes, y podía seleccionarse automáticamente cuando hablan varias personas, o los conferencistas de la segunda fila. La Figura 28 muestra la disposición prevista de la sala. El sistema incluía un panel de control para los usuarios, de manera que puedan manipular todo el equipamiento.

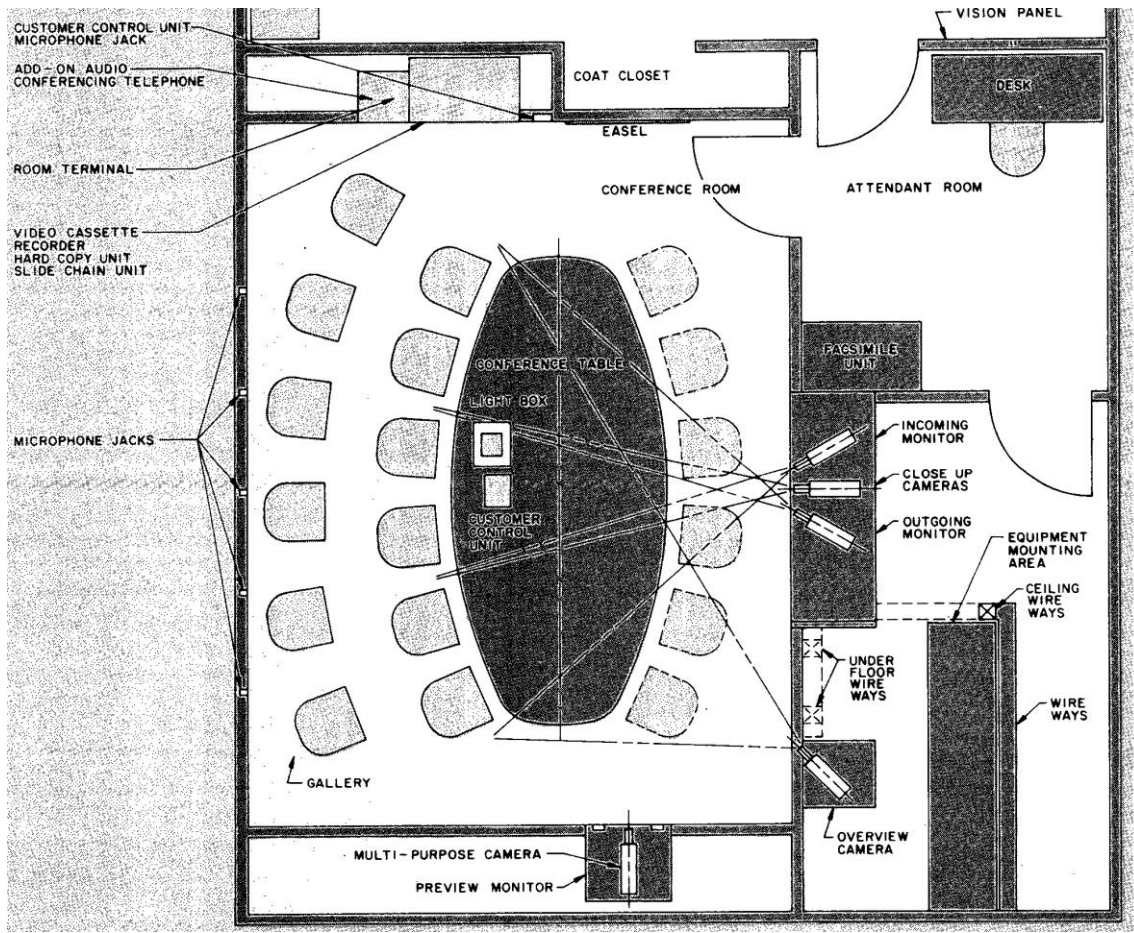


Figura 28: Sala de videoconferencias para el servicio PMS. Tomado de [57].

Los elementos centrales del sistema PMS incluían un controlador de sala y un procesador de TV, como se ven en la Figura 29. El controlador de sala conectaba con todos los elementos de la sala por un lado (cámaras, monitores, panel de control, etc.) y gestionaba las señales de video y audio entrantes y salientes por el otro. El procesador de TV interconectaba los enlaces DS-1 del HSSDS con el audio y video analógico hacia el controlador de sala.

Los servicios de PMS utilizaban una sofisticada tecnología y requerían una inversión muy grande para su puesta en funcionamiento. Equipar una sala de conferencias implicaba una inversión inicial que podía variar entre 120 mil y 500 mil dólares [59][60]. Adicionalmente, el costo mensual fijo, por el arrendamiento del servicio, era superior a los 10 mil dólares, además del costo por uso de cada videoconferencia, que era del orden de los mil dólares por hora. El servicio no se popularizó.

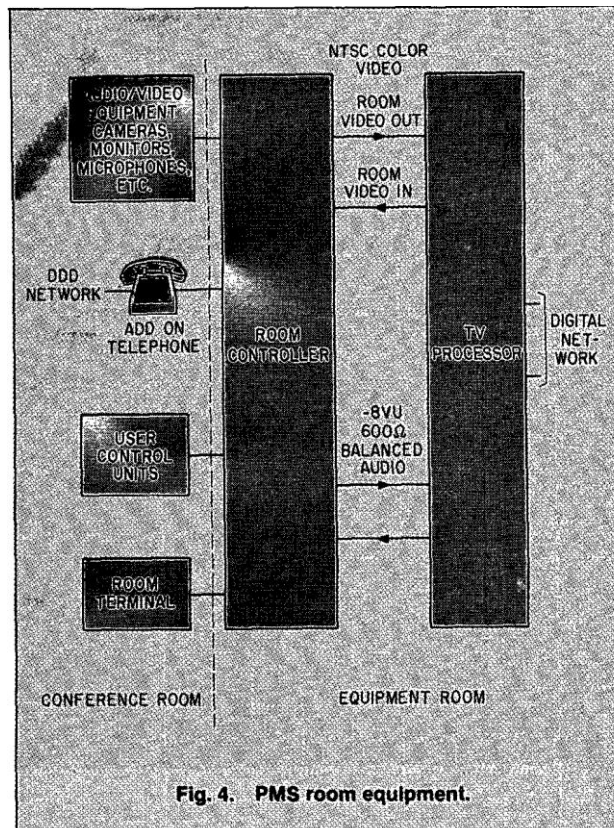


Figura 29: Equipamiento de la sala de PMS. Tomado de [57].

Durante los siguientes años de la década de 1980, se realizaron varios emprendimientos relacionados a sistemas de videoconferencias. La empresa Compression Labs Inc (CLI) presentó un sistema de videoconferencias en 1982. El códec utilizaba codificación intra-cuadros mediante la transformada DCT, aplicada en bloques de 16x16 pixeles. Utilizaron tasas de bits de 1.5 Mbps y lograron desarrollar técnicas que permitían funcionar a la mitad de esta velocidad [60].

Uno de los más destacados de esa década fue el de la empresa PicTel (luego renombrada como PictureTel), fundada en 1984 por dos estudiantes del MIT [61]. En el marco de este emprendimiento, desarrollaron y patentaron un nuevo algoritmo de codificación de video, al que llamaron Motion Compensated Transform (MCT) [62]. En 1986 comenzaron a comercializar su primer producto de videoconferencias utilizando este algoritmo, el modelo C-2000. El equipo consistía únicamente en un sistema de codificación-decodificación (Codec), pesaba más de 100 kg, y no incluía los amplificadores de audio, canceladores de eco, conmutadores de video y otros elementos necesarios para implementar videoconferencias. La señal de video digitalizada y codificada utilizaba un ancho de banda de 224 kbps.

En Japón, en 1985, el modelo NETEC-XV, mostrado en la Figura 30, funcionaba a 384 kbps, utilizando técnicas de predicción inter e intra-cuadros, y codificación entrópica [63].

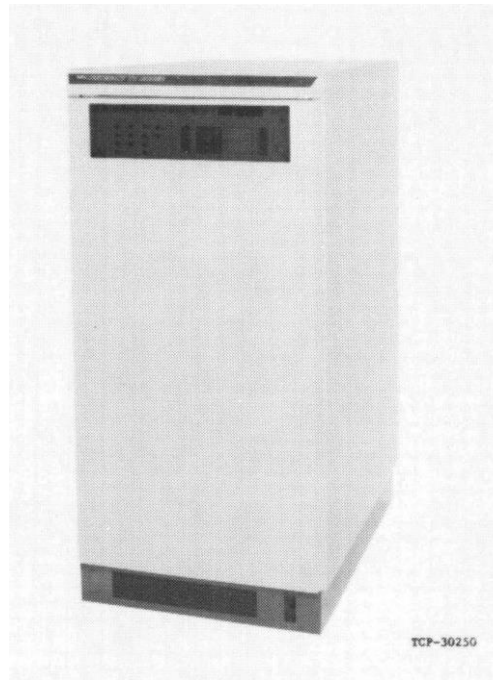


Figura 30: Codec NETEC XV. Tomado de [63].

Estos sistemas requerían inversiones de decenas de miles de dólares, y costos de utilización de líneas digitales del orden de los miles de dólares por mes. Estos factores impidieron su popularización, y su uso quedó restringido a unas pocas grandes corporaciones. Por otra parte, los equipos eran incompatibles entre sí, lo que suponía una gran restricción en su uso.

En 1984 comenzaron los esfuerzos por estandarizar los protocolos y mecanismos utilizados en los sistemas de videoconferencias, en un proyecto del Comité Consultivo Internacional Telegráfico y Telefónico (CCITT)¹. El proyecto terminó en 1988, con la recomendación H.261 “Código para servicios audiovisuales a $n \times 384$ kbit/s” [64], y fue extendido en 1990 a “ $p \times 64$ kbit/s”. El estándar introdujo por primera vez el formato Common Intermediate Format (CIF), de 352×288 píxeles, apropiado para sistemas de videoconferencia, y sentó las bases para la interoperabilidad de códecs entre diversos equipos y fabricantes.

Como respuesta a los precios inaccesibles, tanto de los equipos de videoconferencia, como de la contratación de enlaces digitales, los video teléfonos personales comenzaron a desarrollarse a fines de la década del 1980 y principios de la década de 1990. Para esta época se contaba con nuevas tecnologías, que permitían digitalizar la señal de video en dispositivos de consumo masivo. En junio de 1986, el diario Los Angeles Times publicaba “*El nuevo teléfono también toma fotos*” [65]. El artículo presentaba el “Luma Phone”, al que definía como “*un equipo de aspecto poco inspirador*”. Se trataba de un aparato de algo menos de 4 kg de peso, equipado con un pequeño monitor de tres pulgadas y una pequeña cámara que permitía tomar imágenes en blanco y negro. El equipo utilizaba una línea telefónica analógica común. Las imágenes estáticas que capturaba la cámara podían ser transmitidas a un aparato similar, del otro lado de una línea telefónica, utilizando un botón de “Send”. El costo del Luma Phone era de 1450 dólares americanos, y estaba dirigido a empresas o clientes residenciales de alto nivel. En el mismo artículo se prometía para el año siguiente el lanzamiento de un producto similar, más barato,

¹ CCITT es actualmente la Unión Internacional de Telecomunicaciones, o ITU

pensando para el público en general. Luma Phone fue un desarrollo de la empresa Mitsubishi, basado en los productos de Atari llamados “Ataritel” [66][67], como se muestra en la Figura 31.



Figura 31: Izquierda: Ataritel. Derecha: Luma Phone de Mitsubishi. Tomado de [68] y [69] respectivamente.

La portada de la revista “Popular Science” de 1988 titulaba “*Video Teléfonos: Aquí y ahora, por debajo de 400 dólares. Utiliza líneas telefónicas comunes y tarifas de llamadas comunes*” [70]. Se trataba del lanzamiento de “Visitel”, el prometido sucesor de Luma Phone, a un precio más accesible. El diario Chicago Tribune había anunciado el nuevo teléfono pocos meses antes, en noviembre de 1987 como “*el primer teléfono de imágenes fijas que llega al mercado a un precio asequible (\$399)*” [71]. Un dispositivo similar, desarrollado por Sony, también estaba siendo comercializado en Japón por NTT. Ambos dispositivos eran incompatibles entre sí. Durante una llamada telefónica, se podían capturar y transmitir las imágenes captadas por la cámara presionando un botón. Al hacerlo, por unos segundos se interrumpía el audio, y la imagen se transmitía en forma modulada por la misma línea telefónica.



Figura 32: Izquierda: Mitsubishi Visitel. Derecha: Sony Teleface. Tomado de [70].

Las técnicas utilizadas por estos video teléfonos de imágenes fijas consistían en digitalizar las imágenes capturadas, y transmitir las en forma modulada sobre una línea telefónica. En esa época ya existían los *módems*, capaces de transmitir datos a través de líneas telefónicas analógicas. Sin embargo, estos dispositivos requerían un tiempo prolongado para el establecimiento inicial de la comunicación de datos (“handshaking”) y tenía una tasa de bits típica de 1200 bps, lo que no era apropiada para estos video teléfonos. Dado que la transmisión del video se producía durante la conversación, fue necesario desarrollar nuevos protocolos de comunicación, más rápidos para el establecimiento y la transmisión de la imagen. Visitel utilizaba una tecnología específica, la que patentaron en 1987 [72]. Consistía en una modulación combinada de amplitud y fase, utilizando una frecuencia portadora de 1747.8 Hz. Esta frecuencia se ubica aproximadamente en el medio de la banda de audio utilizada por las líneas telefónicas analógicas (que van de 300 Hz a 3400 Hz aproximadamente) y se puede obtener utilizando un oscilador de cristal estándar de TV en color de 3,579545 MHz y dividiendo por 2048 (lo que se logra fácilmente con circuitos digitales, ya que 2048 es una potencia de dos).

Cada pixel de la imagen era codificado con 4 bits (16 posibles valores), y representada por un símbolo consistente en un ciclo de onda sinusoidal, variando la amplitud y la fase (entre 0° y 180°), como se muestra en la Figura 33. Como cada símbolo representa 4 bits, se logra una tasa de $4 \times 1747.8 = 6991$ bps, muy cercana a los 7 kbps, y muy superior a los 1.2 kbps disponibles en los módems convencionales de la época.

Al inicio de la transmisión, cuando el usuario oprime el botón “send” sobre una llamada de voz establecida, se realiza un intercambio inicial de información mediante tonos de preámbulo, detallados en la patente de Visitel [73]. Mediante estos breves tonos, se sincroniza el receptor con el emisor, para poder decodificar apropiadamente la señal modulada que contiene la codificación de la imagen.

El equipo Visitel requería menos de 6 segundos para transmitir la imagen completa, y mencionaba esto como una diferencia competitiva respecto del videoteléfono de Sony, que requería algunos segundos más. Otras marcas y modelos similares también se comercializaron sobre fines de ‘1980, como el Panasonic WG-R2 o KX-TV10 [74].

Una de las características interesantes de estos diseños era que, en algunos casos, las imágenes podían ser guardadas en contestadores telefónicos estándar. Dado que la modulación se realizaba dentro de la banda de audio, cualquier contestador telefónico disponible en la época podía grabar y reproducir una imagen, como se muestra en el video de [75].

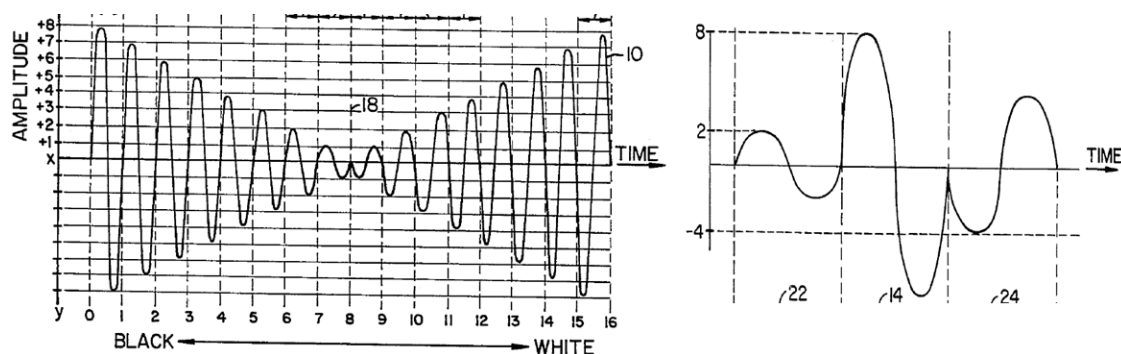


Figura 33: Izquierda: Modulación de amplitud y fase para los 16 posibles niveles de gris (desde negro a la izquierda a blanco a la derecha). Derecha: Ejemplo de la señal modulada para 3 pixeles (“gris oscuro de nivel 2”-“negro”-“gris claro de nivel 4”). Basado en [72].

A comienzos de 1990, entre Picturetel y Compression Labs controlaban más del 90 % del mercado de equipos de videoconferencias en Estados Unidos [76]. En ese momento, el mercado de videoconferencias se dividía en dos segmentos: En el segmento bajo, equipos que funcionaban entre 56 kbps y 384 kbps, y en el segmento alto, hasta 2 Mbps. Picturetel estaba más centrada en el primer segmento, mientras que Compression Labs en el segundo. Todo el mercado de videoconferencias a principios de los 1990 comercializaba unos pocos miles de equipos al año (PictureTel comercializó 770 equipos en 1990 [77]). Sin embargo, las empresas trataban de posicionarse en todo el espectro. En 1991 PictureTel comienza a comercializar el System 4000, integrando en una electrónica del tamaño de una computadora personal, el códec, el conmutador de video, el mezclador de audio, los adaptadores de terminal de red y la unidad de interfaz de audio y video, los que previamente estaban separados. El sistema fue el primero de PictureTel en soportar el nuevo estándar H.261 (p x 64 kbps), y tenía un precio de 40 mil dólares [78]. La Figura 34 muestra el nuevo equipo. En la Figura 35 se muestra el equipo Rembrandt II/VP, el primer equipo de CLI en implementar el estándar H.261 en 1992, con tasas de bits entre 56 kbps y 2 Mbps [79]. Un listado muy completo de la oferta de equipos de videoconferencia a comienzos del año 1990 se puede ver en la Tabla 1.



Figura 34: Equipo PictureTel System 4000, con Cámara, teclado y CPU Codec, tomado de [80].



Figura 35: Equipo Rembrandt II/VP, de Compression Inc Labs (CLI), tomado de [81].

En 1992 AT&T, junto con CLI, diseñan y comienzan a comercializar un videoteléfono que funcionaba con líneas telefónicas analógicas, el modelo 2500. Este videoteléfono fue el primero en codificar video a color en formato digital, y transmitirlo a 19.2 kbps, sobre líneas telefónicas analógicas, utilizando un módem incorporado. El aparato tenía una pantalla a color LCD de 3.3 pulgadas, y funcionaba a 10 fps (cuadros por segundo). La cámara permitía captar hasta 3 personas frente al aparato. El audio era codificado con técnicas CELP a 6.8 kbps. El video tenía una resolución de 127 x 112 pixeles en luminancia, y 32 x 28 en crominancia (o color). Utilizaba técnicas de codificación predictivas y la transformada DCT, generando un flujo de 10 kbps. Audio, video y señalización eran multiplexadas en un canal de datos, utilizando el protocolo X.25, encapsulado en tramas LAPB (protocolo utilizado en el estándar ISDN) [82]. El establecimiento del video demoraba aproximadamente 10 segundos [83]. Su valor era de 1500 dólares.

Tabla 1 – Oferta de equipos de videoconferencia en 1990. Tomado de [76]

Videoconferencing equipment									
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) 530-5580	Digital Transceiver Model 288, 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) 741-1871	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

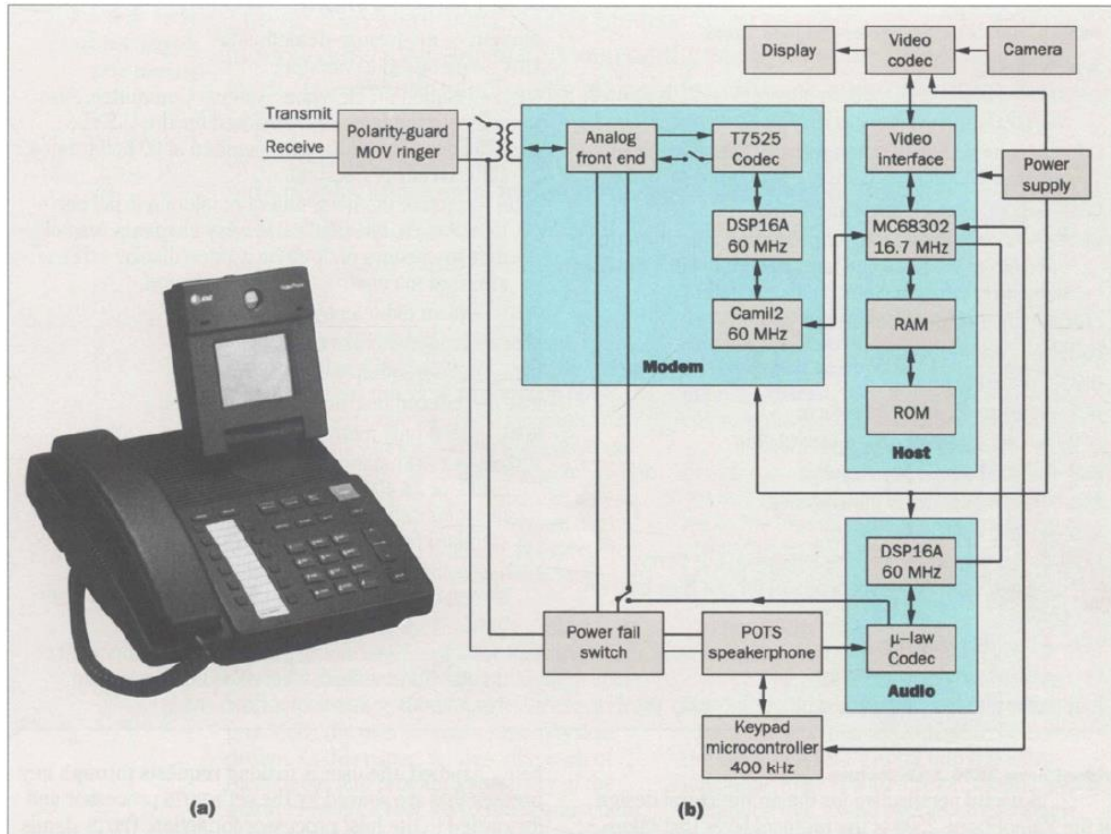


Figura 36: Izquierda: Foto del videoteléfono analógico 2500, de AT&T y CLI. Derecha: Diagrama de bloques de su funcionamiento. Tomado de [82].

Además de la estandarización de los códecs, para la interoperabilidad de los sistemas de videoconferencias, era necesario tener normas de señalización. En 1990, la CCITT publicó el primer estándar para video teléfonos o terminales de videoconferencia, la recomendación H.320 "Sistemas y equipos terminales de telefonía visual de banda estrecha" [84]. Esta recomendación describe los requisitos técnicos de los equipos terminales para el servicio de videoteléfono, con una velocidad de transmisión de datos de hasta 1920 kbps. Se define la arquitectura de alto nivel del equipo terminal, y sus bloques lógicos, como se muestra en la Figura 37. Las características de cada bloque funcional se describen en otras recomendaciones de CCITT (por ejemplo, H.221, H.242, H.230, H.261 y la serie I.400). Los terminales H.320 podían operar conectados a la nuevas Redes Digitales de Servicios Integrados o "Integrated Services Digital Networks" (ISDN), las que podían funcionar a múltiplos de 64 kbps ($nx64$ kbps), con tecnología digital de punta a punta.

La primera implementación comercial de H.320 la desarrolló British Telecom (BT) y se conoció como el códec VC2100 [85]. Este códec podía operar a todas las velocidades de datos desde 64 kbps hasta 2 Mbps implementando muchos de los modos operativos opcionales de H.320 y se convirtió en un punto de referencia para la industria del momento. En la primera mitad de la década de 1990, BT comercializó también otros equipos compatibles con H.320, como el VC 7000. Era básicamente un teléfono con un monitor, como se muestra en la Figura 38. Se conectaba a servicios ISDN, incluía una cámara a color CCD y era posible conectar una cámara externa para tener más movilidad. La pantalla consistía en un monitor a color de 10 pulgadas [86].

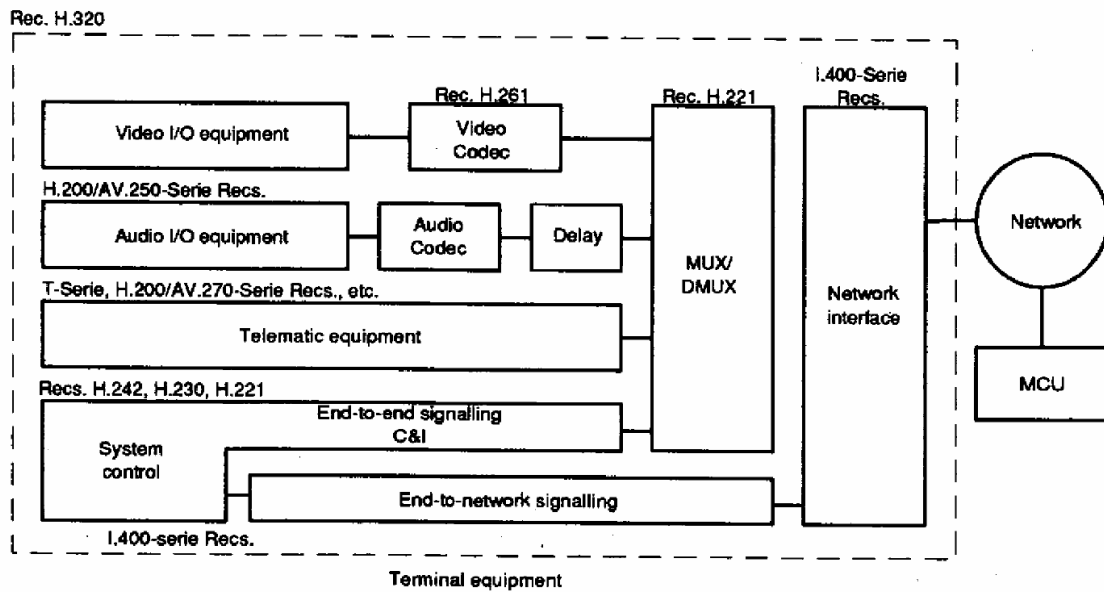


Figura 37: Diagrama conceptual de un equipo terminal de videoconferencias, según el estándar H.320. Tomado de [84].



Figura 38: Videoteléfono VC7000. Tomado de [86].

Los sistemas de videoconferencia para PC comenzaron a desarrollarse a comienzos de la década de 1990, y a comercializarse pocos años después. BT desarrolló el sistema VC8000, que permitía convertir un PC en un terminal multimedia de videoconferencias. Mediante una aplicación se podían compartir imágenes, gráficos y textos. El kit VC 8000 consistía en una tarjeta de comunicaciones multimedia, una cámara de video, una unidad de audio y el software, suministrado por IBM, Olivetti o ICL. La unidad de audio era un teléfono normal, conectado a la PC.

En 1993, PictureTel anunciaba el sistema de videoconferencias para PC PCS100 PictureTel, (luego renombrado Live100), compatible con PC-AT y con el estándar H.320. Para su conexión utilizaba la red ISDN. Trabajaba con una resolución de 352 x 288 píxeles y hasta 15 cuadros por segundo. El precio de lanzamiento fue de 6000 dólares [87]. Dos años luego, en 1995, se lanzaba el modelo PCS50 (luego renombrado como Live50), más sencillo, y a un precio de 2500 dólares

[88]. Los kits de Picturetel incluían una tarjeta para PC, una cámara, un teléfono y el software, compatible con Windows 3.1 y luego con Windows 95 [89]. Un esquema de estos elementos se puede ver en la Figura 39.

Varias otras marcas y productos de videoconferencia para PC se comercializaron a mitad de la década de 1990. Entre ellos, AT&T Telemedia, Intel ProShare [90] (ver Figura 40), InVision, MRA VidCall y Northern Telecom Visit Video. Una comparación entre ellos se puede ver en [91]. Típicamente tenían una resolución de 352 x 288 (o similar) y 15 fps, funcionaban sobre líneas ISDN, permitían realizar una videollamada, compartir archivos, aplicaciones y áreas de pizarras. El hardware específico era compatible con slots ISA o EISA. Los precios rondaban los 2500 dólares.

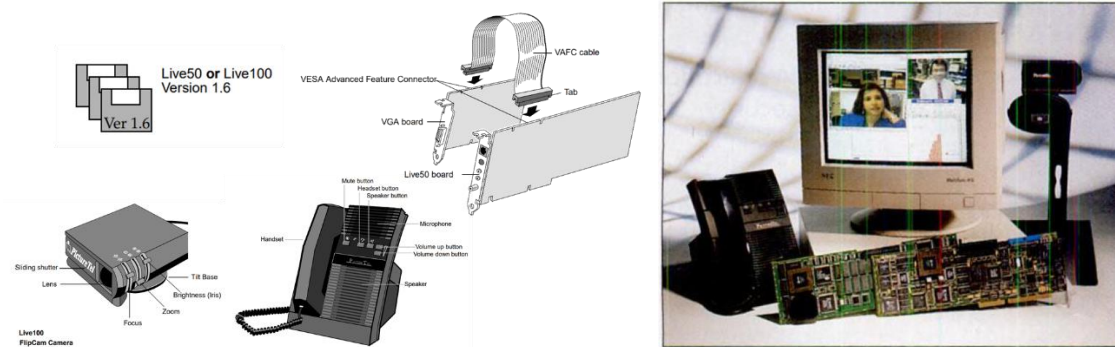


Figura 39: Izquierda: esquema del kit PictureTel Live50, Tomado de [89]. Derecha: kit Live PCS 100, tomado de [91].



Figura 40: Intel Proshare. Tomado de [90].

5 La era IP

Las tecnologías de comunicación y conmutación de paquetes habían comenzado a desarrollarse a fines de la década de 1960, con estudios académicos. A comienzos de la década de 1970, se desarrolló una nueva tecnología de comunicación entre computadores, a la que su diseñador,

Bob Metcalfe, llamó Ethernet. Fue tan exitosa, que en 1980 varias compañías la adoptaron. Digital, Intel y Xerox comenzaron a usarla, a velocidades de 10 Mbps, convirtiéndola en un “estándar de hecho”. En febrero de 1980 la Sociedad de Computación del IEEE realizó la primera reunión del “comité de estandarización de redes de área local” (“Local Network Standards Committee”), y el protocolo Ethernet fue estandarizado como IEEE 802.3. A fines de la década de 1980, en 1989, el concepto de “World Wide Web” (www) fue creado por Tim Barner Lee, dando origen a Internet.

A principios de la década de 1990 Ethernet comenzaba a popularizarse como el protocolo de área local e Internet comenzaba a expandirse. Los desarrolladores de sistemas de videoconferencia, basados hasta ese momento en transmisión y señalización digital sobre las redes ISDN o enlaces punto a punto, comenzaron a poner su atención y esfuerzos en estas nuevas redes de datos.

El audio de la reunión de marzo de 1992 del Grupo de Trabajo de Ingeniería de Internet (IETF) realizada en San Diego, California, fue transmitido por primera vez en vivo, a través de una naciente Internet, a participantes en 20 sitios diferentes, dispersos en tres continentes [92]. Se utilizó tecnología de “multicast”. El audio era codificado en PCM a 64 kbps, utilizando una Sparc Station donde corría la aplicación “Visual Audio Tool” (VAT), desarrollada por Van Jacobson y Steve McCanne. El sistema fue llamado Internet Multicast Backbone o Mbone. Entre 1993 y 1994 se agregó al sistema la posibilidad de transmitir video, a un bitrate de 128 kbps y 1 a 4 cuadros por segundos, con la aplicación Network Video (NV) desarrollada por R. Frederick en Xerox Parc. VAT, NV y otras aplicaciones utilizadas se muestran en la Figura 41. La calidad de la transmisión, tanto de audio como de video, no era buena, y variaba según el momento, en función de la calidad y saturación de los enlaces de Internet utilizados. Sin embargo, el sistema se hizo popular y en 1994 la red Mbone tenía 750 subredes conectadas [93].

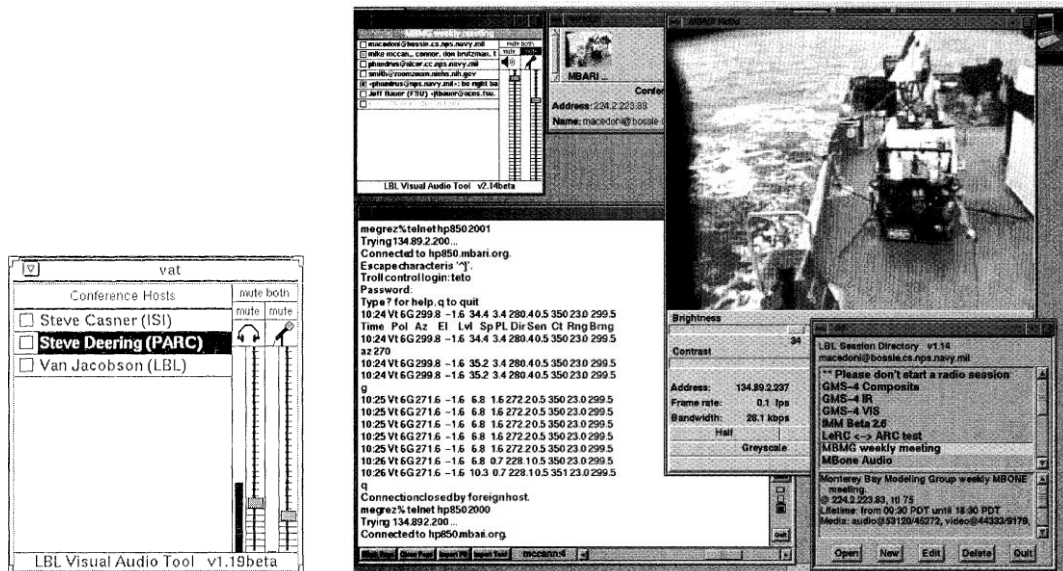


Figura 41: Izquierda: Aplicación Visual Audio Tool (VAT), corriendo en un “X Window”, usada en la primera transmisión multicast de audio del IETF, tomado de [92]. Derecha: Aplicaciones Network Video (NV), VAT, Whiteboard (WB) y Session Directory (SD) utilizadas en una sesión de Mbone transmitida desde el Acuario de Monterey Bay, tomado de [93].

Los conceptos claves que hicieron posible el Mbone fueron el “multicast” sobre IP y el uso de un nuevo protocolo que se comenzó a desarrollar en 1992: el Real Time Protocol (RTP), propuesto por el grupo de trabajo de transporte de audio y video del IETF [94]. El nuevo protocolo RTP brindaba servicios de temporización y secuenciación, lo que permitió que las aplicaciones de audio y video se puedan adaptar a las latencias y los errores introducidos por las redes de paquetes. Adicionalmente, sobre comienzos de la década de 1990, varias estaciones de trabajo conectadas a Internet soportaban tarjetas de sonido y de video, con precios a un nivel que hacía posible su implementación a gran escala.

Tim Dorcey, del departamento de Tecnología de la Información de la Universidad de Cornell, desarrolló en 1992 la aplicación CU-SeeMe. Primero fue desarrollada para Macintosh y luego, en 1994, para Windows. En 1993 se presentó la aplicación como parte de un proyecto educativo financiado por la National Science Fundation (NSF) de Estados Unidos, llamado “Global Schoolhouse” [95]. Fue la primera videoconferencia multipunto basada en Internet para conectar escuelas en los Estados Unidos y, potencialmente, escuelas en todo el mundo. Una imagen se puede ver en la Figura 42. En 1995, CU-SeeMe se describía de la siguiente manera: *“El video saliente y entrante se muestra en pequeñas ventanas en escala de grises de 4 bits que brindan una claridad y frecuencias de actualización sorprendentemente aceptables. Una ventana deslizante adicional puede transmitir y recibir gráficos de mayor tamaño. Actualmente, el audio solo está disponible en la plataforma Macintosh; sin embargo, un módulo complementario para Windows y Macintosh proporciona capacidades de "chat" basadas en texto. Se requiere un hardware mínimo para capacidades tan sofisticadas. Funciona en cualquier Macintosh con un procesador 68020 o superior que ejecute System 7. Las PC requieren un procesador 386SX o superior con Windows 3.1. Cada sistema también requiere una conexión IP con un ancho de banda mínimo de 28,8 kbps. Las conexiones a Internet realizadas a través de módems de 14,4 kbps carecen de funciones de audio, así como de frecuencias de actualización de video aceptables. Para enviar video la computadora necesita una placa de digitalización de video y una cámara de video. Afortunadamente, los periféricos de video se han vuelto bastante asequibles recientemente.”* [96]. La aplicación CU-SeeMe fue comercializada por la empresa White Pine Software.



Figura 42: CU-SeeMe utilizado en ambientes educativos. Tomado de [93].

En Francia, en el Institut National de Recherche en Informatique et en Automatique (INRIA), Thierry Turlotti desarrolló en 1994 la aplicación INRIA Videoconferencing System (IVS) [97]. Esta aplicación de videoconferencias utilizaba PCM y ADPCM para codificar el audio, y el códec H.261 de video, desarrollado inicialmente para sistemas de video sobre ISDN. Fue necesaria su adaptación a redes de paquetes, para lo que el propio Turlotti propuso un formato al IETF [98]. Utilizó también el protocolo RTP sobre UDP, para brindar temporización y secuenciación, y agregó un cabezal específico para los paquetes H.261 [99]. En la Figura 43 se ve la aplicación IVS.



Figura 43: IVS. Tomado de [100].

Para 1995 existían numerosos sistemas de videoconferencias que funcionaban sobre redes de datos. Entre ellos, Avistar, Cameo Personal Video System (de Compression Labs), MediaPhone (de Fiber and Wireless), Communique! (de Insoft), BeingThere (de Intelligence at Large), InVision, VidCall (de MRA), Livelan-V (de PictureTel), InPerson (de Silicon Graphics), ShowMe (de Sun Microsystems) y VTel, entre otros. Los precios oscilaban entre los 1000 y los 5000 dólares. Una completa comparación entre estos sistemas se pueden ver en [101].

Si bien todos los sistemas de videoconferencia sobre redes de paquetes funcionaban sobre los mismos conceptos, eran mayoritariamente incompatibles entre sí. De manera similar a lo que sucedió con ISDN, se hacía necesario tener estándares que permitieran establecer sesiones de videoconferencias sobre IP entre diversos sistemas de diferente fabricantes o desarrolladores. Es así que, en octubre de 1996 es ratificada la primera versión del estándar H.323, por el grupo de estudio 16 de la ITU-T [102]. H.323 fue el primer estándar para la transmisión de multimedia (voz, video y datos) a través de redes de paquetes. Esta primera versión era relativamente básica, y fue mejorada sucesivamente en los siguientes diez años con diversas mejoras y agregados. H.323 estableció no solamente los requerimientos de codificación para audio, video y datos, sino también los protocolos de señalización para el establecimiento de sesiones, como se esquematiza en la Figura 44. Estos nuevos protocolos de señalización, básicamente

extendieron los protocolos de ITU ya existentes en ISDN, para poder ser utilizados sobre redes de paquetes.

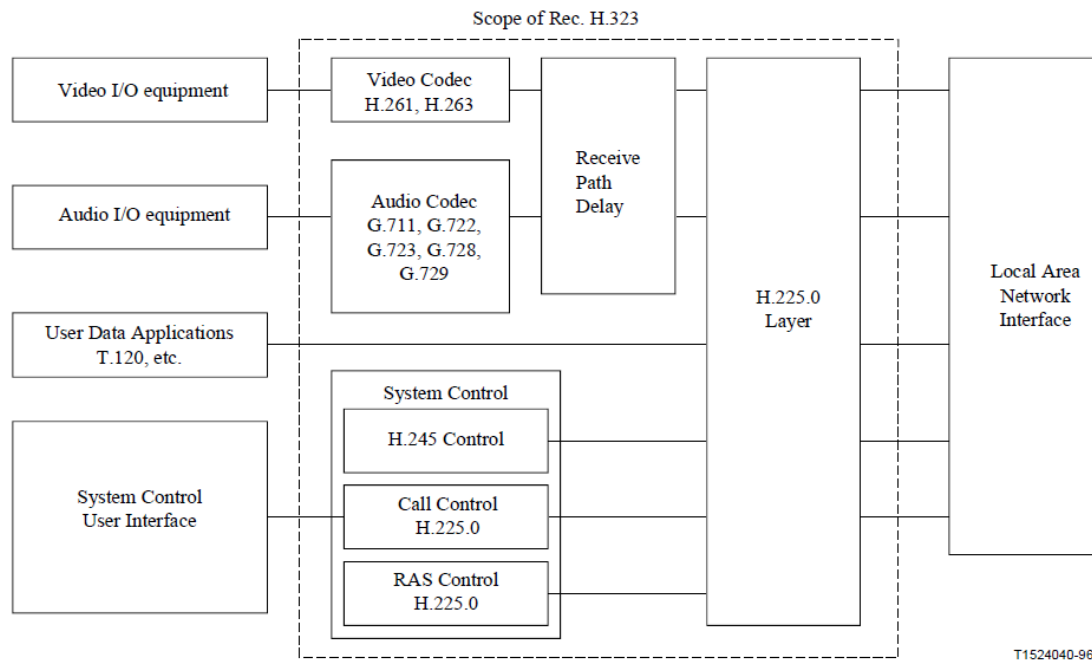


Figura 44: Alcance de la recomendación H.323. estableciendo estándares para el sistema de control, además de la codificación multimedia. Tomado de [102].

También en 1996, ITU estandarizó la recomendación H.324 “Terminal para comunicación multimedia a baja velocidad” [103]. Esta recomendación describe terminales de videoconferencia que pueden ser conectados a la red telefónica pública conmutada analógica, utilizando módems.

Sobre la segunda mitad de la década de 1990, varios sistemas de videoconferencias comenzaron a adoptar los nuevos estándares, H.323 y H.324, que prometían universalidad e interoperabilidad, tanto sobre redes telefónicas clásicas, como sobre redes de paquetes. En 1996, el New York Times anunciaba “Intel planea tecnología de video teléfono para PC: Después de años de falsos comienzos por parte de AT&T y otros, Intel espera llevar la tecnología de videoteléfonos a las masas... Frank Gill, vicepresidente ejecutivo del Grupo de Comunicaciones por Internet de Intel, dijo que esperaba que este año se vendieran cientos de miles de computadoras preparadas para videoteléfonos y millones más en 1997. Eso contrasta marcadamente con el producto de videoconferencia Proshare de Intel de dos años de antigüedad, el líder de la industria, que tiene sólo 50 000 usuarios” [104]. El sistema, llamado Intel Video Phone, podía funcionar sobre redes de paquetes o sobre líneas telefónicas, logrando en este último caso entre 7 a 11 cuadros por segundo (regulables con un control que permitía bajar la calidad de imagen y subir la cantidad de cuadros por segundo) [105]. También en 1996 Microsoft presentó la primera versión de la aplicación Netmeeting, que permitía realizar audio conferencias. Microsoft anunciaba en su lanzamiento: “NetMeeting es el primer cliente de comunicaciones en tiempo real de Internet que incluye compatibilidad con los estándares de conferencias internacionales y brinda verdaderas capacidades para compartir aplicaciones y conferencias de datos entre múltiples usuarios. NetMeeting hace que las comunicaciones de voz y datos a través de Internet sean tan fáciles como una llamada telefónica” [106]. La aplicación,

basada en el estándar H.323, podía ser descargada junto con Internet Explorer (en su versión beta 3.0). A fines del mismo año, se anunció el agregado de video a la aplicación, funcionando hasta con 10 cuadros por segundo, sobre redes Ethernet de 10 Mbps. Podía correr sobre un procesador Pentium de 166 MHz [107]. En la Figura 45 se muestran las aplicaciones Intel Video Phone y Microsoft Netmeeting.

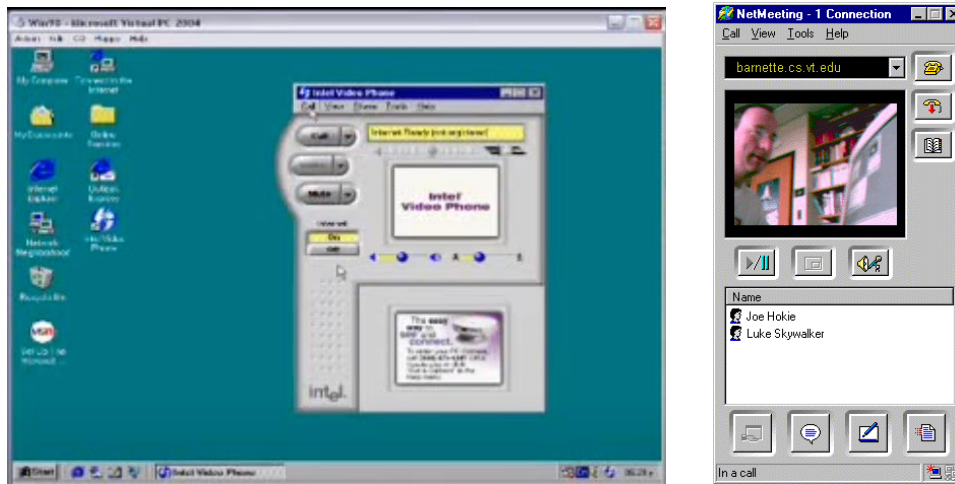


Figura 45: Izquierda: Intel Video Phone. Derecha: Microsoft Netmeeting

Además de las aplicaciones de Intel y Microsoft, entre los sistemas de videoconferencia para PC a fines de la década de 1990 se encontraban Boca Video Phone Kit, Diamond Supra Video Phone Kit 3000, Gallant Intersvision Pro, Panasonic EggCam GP-KR0011, Tekram How-R-U Video Conferencing Kit y 3Com BigPicture Video Kit 1622 (ver Figura 46). Los precios de estos kits oscilaban entre los 200 y los 400 dólares en 1997 (prácticamente la décima parte de los sistemas ISDN para PC existentes pocos años antes). Un análisis comparativo entre ellos se puede ver en [108].



Figura 46: Kits de videoconferencia para PC disponibles en 1997. De izquierda a derecha: Boca Video Phone Kit, Diamond Supra Video Phone Kit 3000, Gallant Intersvision Pro, Panasonic EggCam GP-KR0011, Tekram How-R-U Video Conferencing Kit y 3Com BigPicture Video Kit 1622. Tomado de [108].

En 1997, Henning Schulzrinne presenta en el IETF el primer borrador de un nuevo protocolo diseñado para gestionar videoconferencias, al que llamó Session Initiation Protocol (SIP) [109]. A diferencia de H.323, SIP surgió del mundo de Internet, basado en las experiencias del sistema Mbone. Tal como menciona el autor del protocolo, “se eligió un enfoque basado en texto para el diseño de SIP. En lugar de inventar una nueva representación de protocolo de la nada, reutilizar el protocolo de Internet más exitoso, HTTP, parecía la opción más adecuada. Al usar HTTP como base, los protocolos pueden reutilizar inmediatamente una serie de protocolos en evolución para el comercio electrónico, autenticación, etiquetas de contenido y control de acceso del lado del cliente, extensiones de protocolo, gestión de estado y negociación de contenido. Además, los servidores, proxy y firewalls, todos ya ajustados para un alto rendimiento, capacidad de administración y confiabilidad, se pueden modificar fácilmente para adaptarse a este nuevo protocolo” [110]. El funcionamiento básico del protocolo se muestra en la Figura 47. El protocolo fue formalmente estandarizado en 1999, y comenzó a competir fuertemente con su antecesor de ITU, H.323.

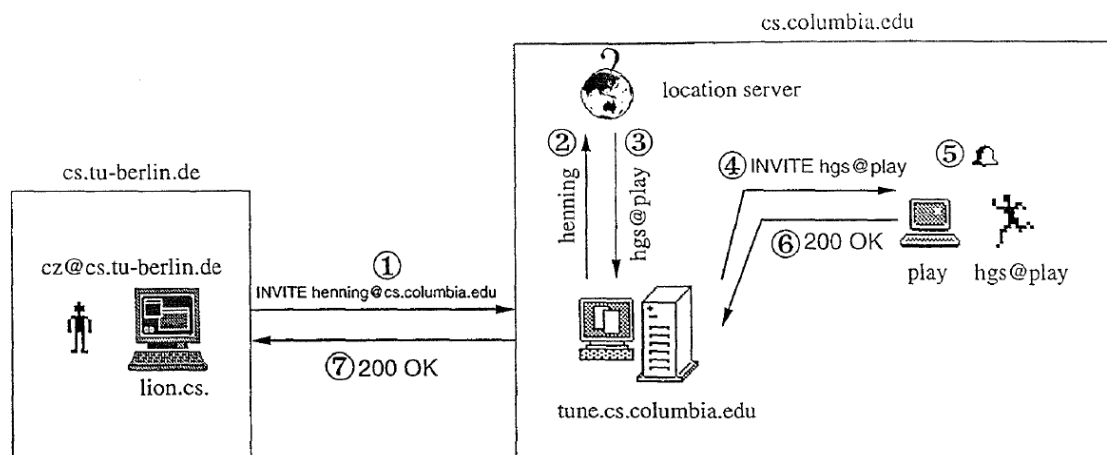


Figura 47: Esquema de funcionamiento del protocolo SIP. Tomado de [110].

Durante los siguientes años, H.323 y SIP compitieron, tratando de ganar el mercado. En un reporte de PC Magazine de 2000, titulado “Muchas voces, pocos consensos” [111] se indicaba que muy pocos de los equipos de telefonía por Internet disponibles en ese momento eran interoperables. Durante la primera década de 2000, H.323 tuvo varias revisiones, terminando en 2009 con la revisión 7. Varios años más tarde, en 2022, se publicó la revisión 8. Por su parte, SIP fue aceptado rápidamente por el grupo 3GPP, estandarizador de los sistemas celulares, quien lo incorporó a sus protocolos del núcleo de su IP Multimedia Subsystem (IMS) en 2001 [112]. Este hecho, y su fácil implementación a nivel de software, hizo que SIP haya terminado siendo el protocolo de inicio de sesiones de videoconferencias más utilizado en la década de 2010, y aun el más utilizado al momento de escribir este trabajo, en 2022.

En 2003 Niklas Zennström y Janus Friis crearon Skype, basado en el concepto de “Peer-to-Peer” descentralizado que utilizaba Kazaa (sistema creado anteriormente por las mismas personas). El nombre originalmente pensado había sido “Sky peer-to-peer”, que se abreviaba “Skyper”. Sin embargo, el dominio de Internet “skyper” ya estaba siendo utilizado, por lo que le quitaron la última “r”, creando el nombre “Skype”. En su lanzamiento permitía realizar llamadas de audio, dos años más tarde, en 2005, se agregó la posibilidad de realizar videollamadas [113]. El principio de funcionamiento de Skype fue muy diferente a las aplicaciones de videoconferencia

de la época. No utilizó ninguno de los protocolos de señalización, ni los códecs estandarizados. Por el contrario, se basó en una arquitectura peer-to-peer, donde el único elemento central era un servidor de autenticación (“Login Server”). La señalización y el medio (audio y luego video), se realizaban directamente entre pares (aplicaciones de usuarios). Algunas de estas aplicaciones eran definidas como “Super Nodos” a efectos funcionales, como se muestra en la Figura 48. Cualquier nodo con una dirección IP pública que tuviera suficiente capacidad de CPU, memoria y ancho de banda de red, era candidato para convertirse en un Super Nodo. Las comunicaciones con algunos de los usuarios se realizaban a través de estos Super Nodos [114]. El audio utilizaba los códecs iLBC, iSAC o iPCM, desarrollados por Global IP Solutions [115]. Estos códecs de audio habían sido diseñados especialmente para transmisión de audio sobre redes de paquetes, e incluían mecanismos de alta resiliencia frente a las pérdidas de paquetes. Skype utilizaba tanto UDP como TCP para el transporte del medio codificado, lo que lo hacía robusto para funcionar detrás de firewalls. El servicio se popularizó. Para 2005, Skype tenía más de 2.8 millones de usuarios en los Estados Unidos y 30.6 millones en todo el mundo y se sumaban 155 mil usuarios nuevos por día. *“Una de las razones por las que Skype es tan popular es que es gratis. Otra es que funciona. Puede que no parezca mucho, pero importa cuando las llamadas con otros programas de servicio VoIP gratuitos suenan más como conversaciones de walkie-talkie que como llamadas telefónicas”*, mencionaba un artículo del New York Times de 2005 [116]. Ese mismo año, Skype fue comprada por eBay, en una transacción de 2600 millones de dólares [117]. En 2009 eBay vendió Skype al grupo Silver Lake y en 2011 fue nuevamente vendido, en este caso a Microsoft, por 8500 millones de dólares [118].

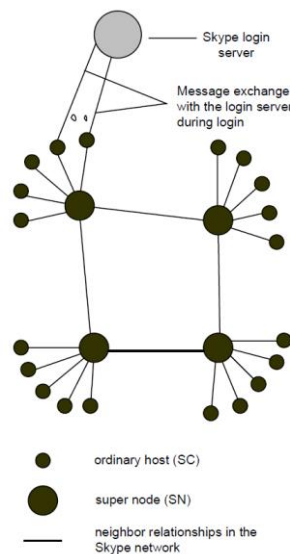


Figura 48: Arquitectura de conectividad de Skype. Tomado de [110].

Sobre comienzos de 2000, otras aplicaciones de mensajería instantánea comenzaron a agregar la posibilidad de videollamadas. En 2002 Yahoo Messenger, en su versión 5.5, incorporó la posibilidad de realizar videollamadas, de hasta 20 cuadros por segundo y con una resolución de 320 x 240 píxeles [119]. En 2004, la popular aplicación de mensajería instantánea ICQ, incorporó también el video [120].

En paralelo a la proliferación de aplicaciones y kits de videoconferencia de escritorio, a mitad de la década 2000 se desarrollaron los primeros sistemas de videoconferencia profesional con el concepto de “telepresencia”, trabajando sobre redes de paquetes. En este tipo de sistemas, las

imágenes en las pantallas se muestran en tamaño real, lo suficientemente claras como para captar los pequeños matices de las expresiones faciales y el lenguaje corporal. El audio proviene de la dirección de la persona que habla, logrando una sensación mucho más natural. La empresa Hewlett-Packard lanzó su línea de telepresencia en 2005, a la que llamó Halo Collaboration Studio. Este sistema permitía organizar reuniones entre cuatro sitios, mediante el uso de salas construidas con estándares exigentes de sonido e iluminación. Cada sala era ideal para seis personas, pero podía acomodar hasta 14. Equipar a una sala para un Halo Studio tenía un precio de 350 mil dólares, aproximadamente. En 2008 se anunció un sistema de similares características, pero más pequeño. En este caso, el hardware para una sala incluía dos o cuatro sillas, paredes delanteras y traseras móviles, una mesa, una cámara de video de alta definición, un monitor grande y equipos de audio y conmutación. El precio de estas salas era superior a 120 mil dólares [121]. Para su funcionamiento, HP ofrecía una red de datos dedicada, a la que llamó Halo Video Exchange Network (HVEN). Consistía en una red de fibra óptica, diseñada específicamente para brindar el servicio sin depender de Internet u otras redes de datos. El servicio de esta red y su operación tenía un costo mensual de 12 mil dólares por sitio [122].

La empresa Cisco ingresó en el mundo de la telepresencia en 2006, con sus productos Cisco TelePresence 1000 y 3000. El TelePresence 1000 incluía una sola pantalla, para reuniones pequeñas, y tenía un precio de 80 mil dólares. El sistema TelePresence 3000 era una de sala completa, para 12 participantes, que incluía tres pantallas de plasma de alta definición de 65 pulgadas, a un costo de 300 mil dólares [123]. A diferencia de la estrategia de HP, los productos de Cisco no requerían la contratación de una nueva red dedicada, sino que operaban sobre las redes de área local.

También en 2006, Polycom anunció el lanzamiento de sus productos RPX RealPresence Experience, disponibles en diferentes configuraciones, soportando salas de 4 a 28 personas, y con precios desde 250 mil dólares por sala [124].

En la Figura 49 se ven los sistemas de Telepresencia de HP, Cisco y Polycom, funcionando en la segunda mitad de la década de 2000.



Figura 49: Izquierda: HP Halo Collaboration Studio. Centro: Cisco Telepresence 3000. Derecha: Polycom RPX. Tomado de [125], [126] y [127] respectivamente.

6 La era de los servicios cloud

En 2011, Gary Sullivan, uno de los principales responsables de los desarrollos de codificación de video, publicó un artículo titulado “La video telefonía por fin ha llegado” [128]. En el artículo reconocía que durante décadas se trató de impulsar los servicios de video telefonía y videoconferencia, tanto en los ámbitos de consumo masivo como empresariales. Sullivan mencionaba que, si bien muchas empresas invirtieron en equipos de videoconferencias en la década de 1990, por lo general, *“terminaban acumulando polvo, sin usar”*. Por otra parte, en el consumo masivo, *“para la mayoría de nosotros, la videotelefonía todavía no forma parte de nuestra vida diaria... Parecía que la gente no lo quería en ninguna parte”*, indicaba Sullivan, quien finalmente aventuraba que esto cambiaría rápidamente y la tecnología se popularizaría *“dentro de solo un par de años”*. Según el mencionado artículo, los obstáculos por los que las videoconferencias no habían prosperado hasta ese momento se debían a carencias en el hardware, las redes, y la compresión del video. Para 2011, el hardware necesario para el procesamiento de video estaba al alcance de cualquier smartphone, las redes de datos e Internet soportaban los anchos de banda necesarias, con la confiabilidad requerida, y los sistemas de codificación de video habían avanzado lo suficiente como para lograr tasas de compresión apropiadas (el estándar ITU-T H.264 estaba disponible hacía algunos años, y los sistemas comerciales habían comenzado a utilizarlo).

Sobre los comienzos de la década de 2010 comenzaron a desarrollarse los sistemas de videoconferencias basados en servicios web, alojados en la nube y comercializados en modalidad “pago por utilización”, sin requerir grandes inversiones en equipamiento local. Estas aplicaciones estaban enfocadas en el mercado corporativo.

En 2010 Adobe lanzó al mercado su producto Adobe Connet 8, un sistema de videoconferencias sobre la web. Se presentaba como un sistema “liviano”, optimizado para bajos anchos de banda, y sin pretender competir con los sistemas de gama alta de telepresencia [129].

En 2011 fue lanzado al mercado el servicio de la empresa Blue Jeans. Krish Ramakrishnan, cofundador y director ejecutivo de Blue Jeans Network indicaba en un comunicado de prensa: *“Lo que alguna vez fue una tecnología de sala de reuniones de élite, ahora se ha trasladado a la nube”*. En los primeros dos meses de servicio, tuvieron 4000 suscriptores, de más de 500 empresas con presencia en más de 100 países [130].

La aplicación FaceTime se comenzó a distribuir con el iPhone 4, en 2010, como una aplicación de video chat entre los smartphones de Apple [131]. Unos meses más tarde, fue incorporada a las computadoras de escritorio Mac. Steve Jobs mencionaba: *“FaceTime hace que las videollamadas hacia o desde dispositivos móviles sean fáciles por primera vez. Hemos vendido más de 19 millones de dispositivos iPhone 4 y iPod Touch compatibles con FaceTime en los últimos cuatro meses, y ahora esos usuarios pueden hacer llamadas FaceTime con decenas de millones de usuarios de Mac.”* [132]. Varios años más tarde, en 2018, se incorporó la posibilidad de hacer videoconferencias entre varios usuarios, con la función Group FaceTime, que permitía hasta 32 participantes [133], como se ve en la Figura 50.



Figura 50: Apple Group Facetime. Tomado de [133].

Whatsapp fue creada en 2009, pero recién a fines de 2016 incorporó la función de videollamadas [134] y en 2018 la posibilidad de llamadas grupales, de hasta cuatro participantes [135]. Se esperaba que, con más de 1000 millones de usuarios, esta función fuera rápidamente incluida en los hábitos y rutinas diarias como forma de comunicación entre sus usuarios. Sin embargo, esto no sucedió al momento de su introducción.

En 2011 Microsoft compra Skype [118], y lo relanza al mercado como Skype for Business en 2015 [136], para luego incorporarlo como parte de Microsoft Teams en 2017 [137]. En 2019, Microsoft Teams permitía realizar videoconferencias de hasta cuatro participantes.

Eric Yuan, quien trabajaba en el producto Cisco WebEx, fundó la empresa SaaSbee en 2011. Al año siguiente, le cambió el nombre a Zoom. *“En ese momento, Zoom era solo una idea en un mercado de videoconferencia aparentemente muy competitivo y la mayoría de los inversores pensaron incorrectamente que los productos existentes como Skype, Webex y otros estaban resolviendo este problema. Pero lo que la mayoría de la gente no vio en ese momento fue que Eric tenía la visión de crear un nuevo tipo de experiencia de comunicaciones por video que resolvería una necesidad masiva.”*, mencionaba Jim Scheinman [138] (socio de Eric, y quien propuso el cambio de nombre de la empresa, inspirado en el libro infantil “Zoom City”). En mayo de 2013, la nueva compañía de servicios de videoconferencia tenía un millón de usuarios [139].

Asociado a los servicios cloud, era necesario desarrollar tecnologías que permitieran dotar a los navegadores de internet de la posibilidad de reproducir contenido multimedia, en forma estandarizada. Hasta el momento, para poder utilizar sitios de videoconferencia desde los navegadores, era necesario descargar e instalar “plugins” o extensiones, lo que dificultaba el uso y la interoperabilidad. En 2011, Google deja disponible a la comunidad de desarrolladores el código fuente del proyecto WebRTC [140]. Esta tecnología había sido desarrollada inicialmente por Global IP Solutions (o GIPS), una empresa fundada en 1999 en Suecia y adquirida por Google en 2011. WebRTC permitió que el audio y el video estén disponibles en todos los navegadores, a través de un conjunto estándar y uniforme de funciones. Al ser paulatinamente incorporado en los diversos navegadores, WebRTC permitió alcanzar el paradigma de “hacer clic y entrar en la conferencia”. Es decir, poder iniciar una sesión de videoconferencia desde un navegador sin

necesidad de ejecutar programas externos o descargar extensiones dependientes de cada plataforma y sistema operativo.

El auge de los servicios cloud no hicieron que los servicios de videoconferencias fueran ampliamente aceptados y utilizados, y las predicciones de Gary Sullivan de comienzos de la década de 2010, una vez más, no se cumplieron. Para fines de la década de 2010 era claro que el problema no era la tecnología, sino, simplemente, que las personas no estaban interesadas en utilizar masiva y diariamente las videollamadas o videoconferencias para comunicarse, ni en su vida privada, ni en su trabajo.

Sin embargo, esto cambió drásticamente al comienzo de la década de 2020, debido a la Pandemia de Covid-19. Con el confinamiento masivo de la población a nivel mundial, las videoconferencias pasaron a ser casi la única manera de mantener interacciones visuales con familiares y colegas de trabajo. A nivel masivo, Zoom se posicionó rápidamente como una de las opciones dominantes. Entre enero y abril de 2020, la cantidad de minutos de reuniones de Zoom se multiplicaron por 2500, como se muestra en la Figura 51 [141]. Por su parte, Microsoft Teams pasó de tener 20 millones de usuarios en 2019 a 270 millones en 2022 [142]. Todas las compañías que ofrecían servicios de videoconferencia tuvieron que implementar mejoras y nuevas funciones rápidamente. Por ejemplo, al comienzo de la Pandemia, Teams soportaba videoconferencias solamente de 4 personas. Luego lo amplió a 9 participantes, y en julio de 2020 permitió visualizar hasta 49 participantes. Zoom fue inicialmente criticado por aspectos de seguridad, los que rápidamente corrigió. El uso de las videoconferencias se incorporó rápidamente en la vida diaria, tanto a nivel personal y familiar, como empresarial.

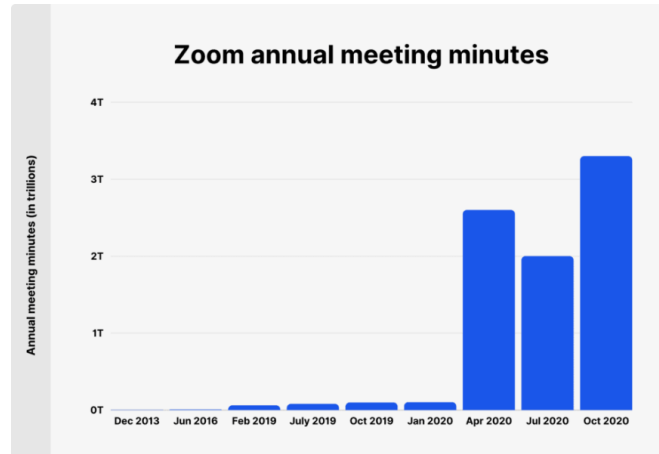


Figura 51: Minutos de uso de Zoom. Tomado de [141].

Luego de dos años de pandemia, en 2022, las tecnologías de videoconferencia llegaron a su madurez y popularización. Participar de una videollamada o videoconferencia se convirtió para muchas personas, en la manera habitual de reunirse, tanto a nivel personal como a nivel laboral. Los costosos sistemas basados en hardware de videoconferencias fueron totalmente reemplazados por servicios cloud, muchos de ellos gratuitos, o con costos por utilización muy accesibles por cualquier persona o empresa. En 2002, Zoom ofrece servicios gratuitos de videoconferencias de hasta 100 personas y hasta 40 minutos, y por 150 dólares anuales, servicios similares de tiempo ilimitado [143]. De forma similar, Microsoft Teams ofrece servicios gratuitos de videoconferencias de hasta 100 personas y hasta 60 minutos, y por 4 dólares anuales por usuario, servicios similares de hasta 300 personas y 30 horas por reunión [144].

7 El futuro

Los próximos años de la década 2020 presentan nuevos desafíos para los sistemas de videoconferencias, varios de los cuales se describen a continuación.

Con el fin de la pandemia, varias empresas han vuelto a la modalidad de trabajo presencial, pero muchas otras están optando por un esquema híbrido, donde los empleados tienen días de trabajo presencial y otros de teletrabajo. Esto ha llevado a la proliferación de *conferencias híbridas*, donde parte de los participantes se encuentran en forma presencial en una sala, y parte se encuentran en forma remota. Las salas de reuniones de las empresas típicas no están preparadas para este tipo de reuniones. Si bien los conocidos “huddle rooms” comenzaron a incorporarse en las empresas desde antes de la pandemia, el retorno a las oficinas ha generado una nueva y creciente demanda de este tipo de salas. En este tipo de conferencias, existe una clara asimetría entre los participantes presenciales y los remotos. Los participantes remotos tienen dificultades para escuchar y participar activamente en las reuniones. En lo que respecta al audio, el eco, la reverberación y la lejanía a los micrófonos en la sala son algunos de los principales problemas. La vista general de la sala muchas veces no permite visualizar los detalles de la persona que está hablando. Las pizarras clásicas no son apropiadamente visualizadas por los participantes remotos, además de que no permiten interactividad. Varios de estos puntos ya disponen de soluciones técnicas. Entre ellas, monitores o televisores interactivos que permiten funcionar como pizarras compartidas, cámaras que enfocan automáticamente a la persona que habla, y sistemas de cancelación y compensación de eco. Sin embargo, aun con las últimas tecnologías disponibles en 2022, la experiencia de los participantes remotos resulta pobre. Por otra parte, los participantes presenciales muchas veces tienen dificultades para iniciar la videoconferencia. Los kits de cámaras, micrófonos, monitores o televisores y demás accesorios suelen no estar integrados, y comenzar una nueva sesión de conferencia puede llegar a ser muy difícil para los usuarios no técnicos.

La primera aproximación a las salas híbridas se está desarrollando con la incorporación de tecnologías que integran los componentes de las salas (cámaras, micrófonos, monitores, pizarras) con los servicios de videoconferencia web. Zoom Rooms o Teams Rooms son ejemplos de este tipo [145]. Aun así, varios aspectos todavía requieren mejoras en la tecnología e inversiones por parte de las empresas: paneles acústicos, varios micrófonos, mejores sistemas de cancelación de eco y mitigación de reverberaciones, aproximación automática al locutor, etc. Es de esperar que en un futuro cercano estas tecnologías se consoliden y popularicen, mejorando notoriamente la calidad de la experiencia de las conferencias híbridas.



Figura 52: Huddle Room con Teams Rooms. Tomado de [145].

Al incorporar las videollamadas y videoconferencias a nuestra vida diaria, tanto personal como laboral, resulta muy útil tener *fondos virtuales*, que eviten mostrar lo que tenemos por detrás de nuestro rostro. Casi todos los sistemas de videoconferencias tienen esta función, permitiendo difuminar el fondo, o insertar alguna imagen predefinida. Sin embargo, la detección del contorno o límites del rostro, o de las partes de cuerpo que quedan dentro del campo visual de la videollamada, es imprecisa. Cuando hay varias personas en la imagen, muchas veces una de ellas es omitida y reemplazada por el fondo virtual. Los sistemas de procesamiento de imágenes deberán mejorar sustancialmente la detección de contornos y el uso de fondos virtuales en estos sistemas.

El uso de *avatares* como representación gráfica de las personas podrá ser usado en las videoconferencias. Una imagen estilizada de cada participante reemplazará la imagen real durante la videollamada. Esta tecnología comienza a ofrecerse sobre diferentes plataformas. Zoom ya la tiene incorporado, aunque con íconos sencillos y poco elaborados. Whatsapp lo ha anunciado para próximas versiones. Empresas como Loom.ai disponen de ofertas de avatares más elaborados, compatibles con diversas aplicaciones de videoconferencias. En breve seguramente se dispondrá de avatares más realistas y sofisticados, que representen fielmente a las imágenes de los participantes. Posiblemente será difícil distinguir si lo que se observa es un video en vivo, o una representación digital de nuestro interlocutor.

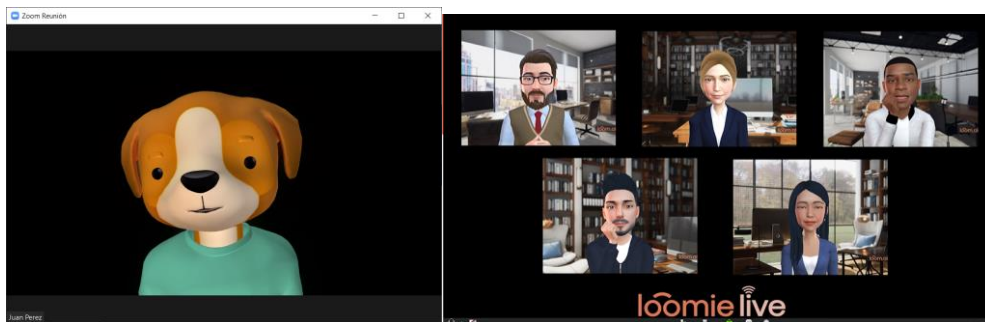


Figura 53: Izquierda: Avatar de Zoom. Derecha: Avatares de loomie.ia. Tomado de [146].

Cuando varias personas que hablan diferentes idiomas participan de una videoconferencia, muchas veces se hace difícil comprenderlos. Normalmente se selecciona un idioma común para el diálogo (por ejemplo, el inglés). Sin embargo, hablar en un idioma no nativo dificulta la comunicación. Varias de las empresas que ofrecen servicios de videoconferencia están comenzando a ofrecer aplicaciones de traducción simultánea a texto en tiempo real. Por ejemplo, Cisco Webex ofrece el servicio de traducción automática a texto desde 2021, como se muestra en la Figura 54 [147]. La posibilidad de tener *traducciones simultáneas verbales*, en tiempo real, ayudará a mantener conversaciones mucho más fluidas, con cada participante hablando en su idioma, la que será traducida a otro idioma y sintetizada, utilizando su mismo timbre de voz en el futuro.

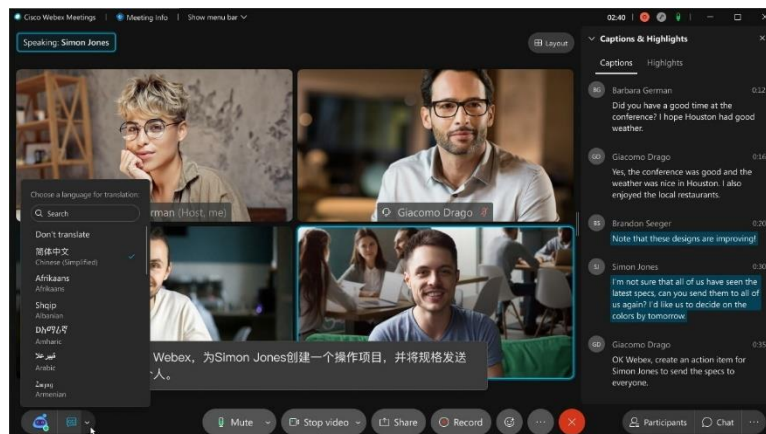


Figura 54: Traducción simultánea a texto. Tomado de [147].

Con el anuncio del Metaverso por Facebook en 2021, las reuniones inmersivas están comenzando a desarrollarse. Para participar de estas reuniones, los participantes deben utilizar cascos de realidad virtual. Las reuniones en este espacio virtual son altamente colaborativas, donde avatares de cada participante interactúan, caminan por el espacio virtual, escriben sobre pizarras compartidas, ven en tres dimensiones, entre otros interesantes aspectos. La Figura 55 muestra un ejemplo de este tipo de experiencias.

Varios aspectos aún están en investigación y desarrollo para que este tipo de videoconferencias o sesiones virtuales se puedan popularizar. Entre ellas, el diseño de cascos más livianos, aspectos de coexistencia entre el mundo real y el virtual (por ejemplo, cómo incluir en el mundo virtual los elementos del mundo real, como mesas, sillas, etc.), aspectos de seguridad física de las personas (por ejemplo, como evitar que una persona intente apoyarse en un objeto virtual que no existe en el mundo real), fatiga y sensación de mareo que muchas personas suelen experimentar.



Figura 55: Participante de una reunión inmersiva, escribiendo en una pizarra compartida. Izquierda: El participante del mundo real. Derecha: Vista del participante en el mundo virtual.

Al finalizar las reuniones, estos sistemas deberán dejar minutas, próximas acciones, y/o resúmenes del contenido, tanto para los participantes como para quienes no pudieron asistir. En el futuro, aplicaciones como GPT-3 o similares podrán ser aplicadas a las videoconferencias corporativas, brindando herramientas de *resúmenes automáticos*, reportes automáticos de los temas pendientes o próximas acciones.

Referencias

- [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/BQMnIKMFD8M>
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] Discrete 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 Ishiguro
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: Códex 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 Announcement, 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/photostream/>
- [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] Videophones 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] Picturitel Corp's new systems 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=l6El_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 Turlletti
Connexions, Vol. 8, No. 10, 1994
- [98] RTP Paylad Format for H.261 Video Streams
T. Turlletti
raft-ietf-avt-h261-00, IETF, 1995
- [99] RTP Paylad Format for H.261 Video Streams
T. Turlletti
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>

