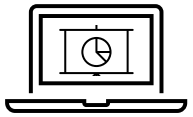
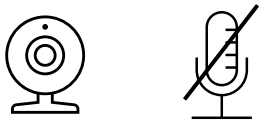
Reglas para las sesiones remotas

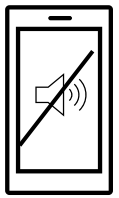


Utilizamos un PC o laptop, con pantalla que permita ver los detalles de las láminas.

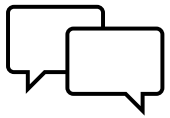
Nos conectamos desde un lugar silencioso, con buena conexión (cableada o WiFi cerca del Router).



Encendemos las cámaras. Mientras uno habla, los otros apagan sus micrófonos.

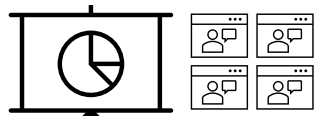


Nos quedamos conectados todo el tiempo… ¡a la sesión! Silenciamos los teléfonos celulares.



Si hay algún problema con el audio o el video, pueden usar el chat para avisar.

¡Todos participan! Pueden “levantar la mano”, o simplemente encender el micrófono e intervenir.



Pueden cambiar de foco entre la presentación y los participantes.

Codificación de voz y video

Digitalización y Codificación de Voz

CODIFICACIÓN DE VOZ Y VIDEO

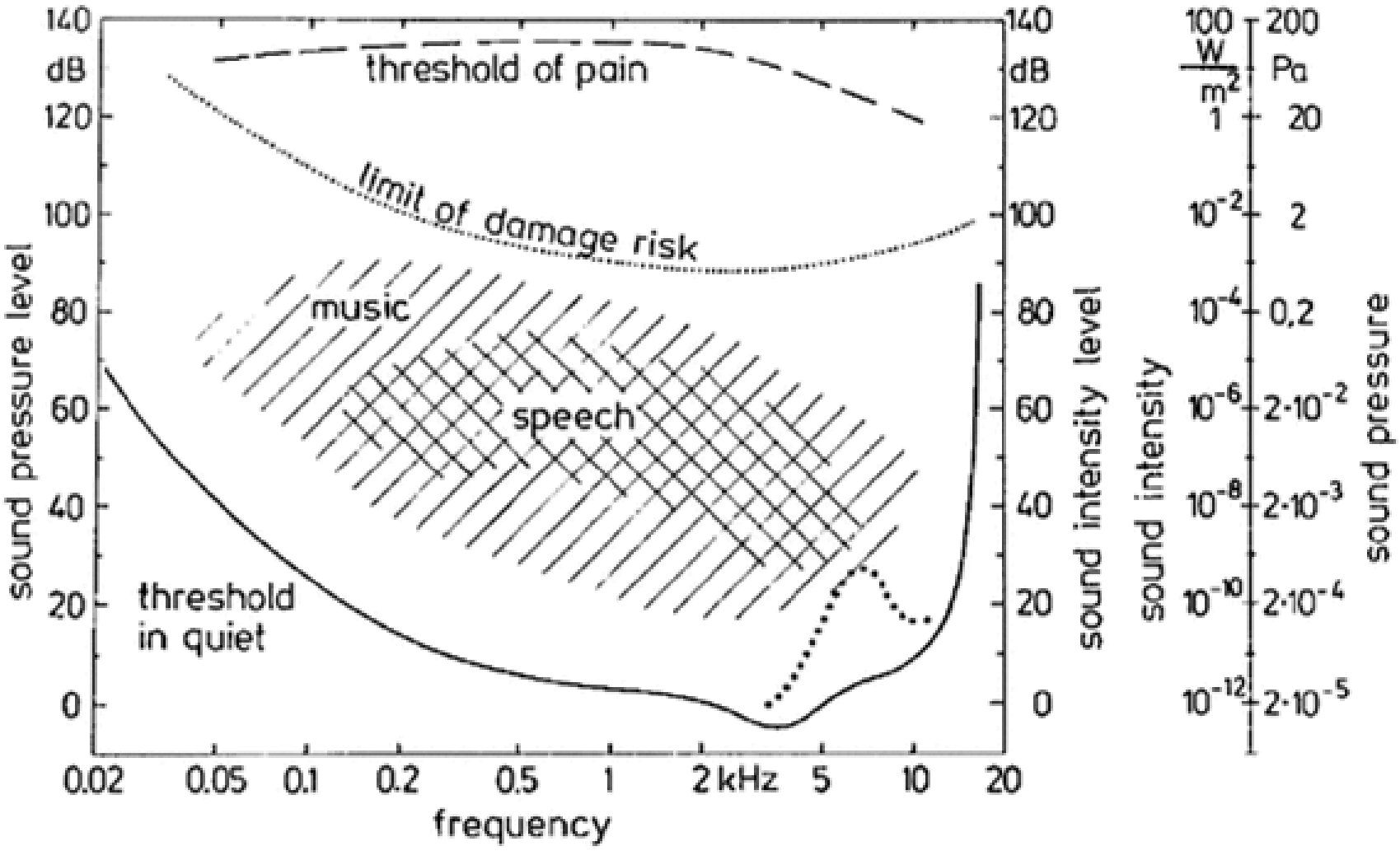
Introducción

En algún punto del sistema de telecomunicaciones la señal de audio analógica debe ser digitalizada, es decir, convertida en una secuencia de número discretos

- Este proceso puede realizarse en los propios teléfonos (cómo es el caso en los “teléfonos digitales” o en los “teléfonos IP”), en “Gateways” (o conversores de medios y señalización) o las “placas de abonados o de teléfonos analógicos” entre otros

CODECS: Codificadores / Decodificadores

Audición humana



Tomado de:
"Psychoacoustics Facts and Models", Hugo Fastl and Eberhard Zwicker, Springer, 2007

Frecuencia de muestreo y bitrate

¿Cuántas muestras por segundo son necesarias?

- Considerando que el oído puede escuchar hasta unos 20 kHz, y tomando en cuenta el *"teorema del muestreo"*, la frecuencia de muestreo debería ser > 40 kHz

¿Cuántos bits por muestra son necesarios?

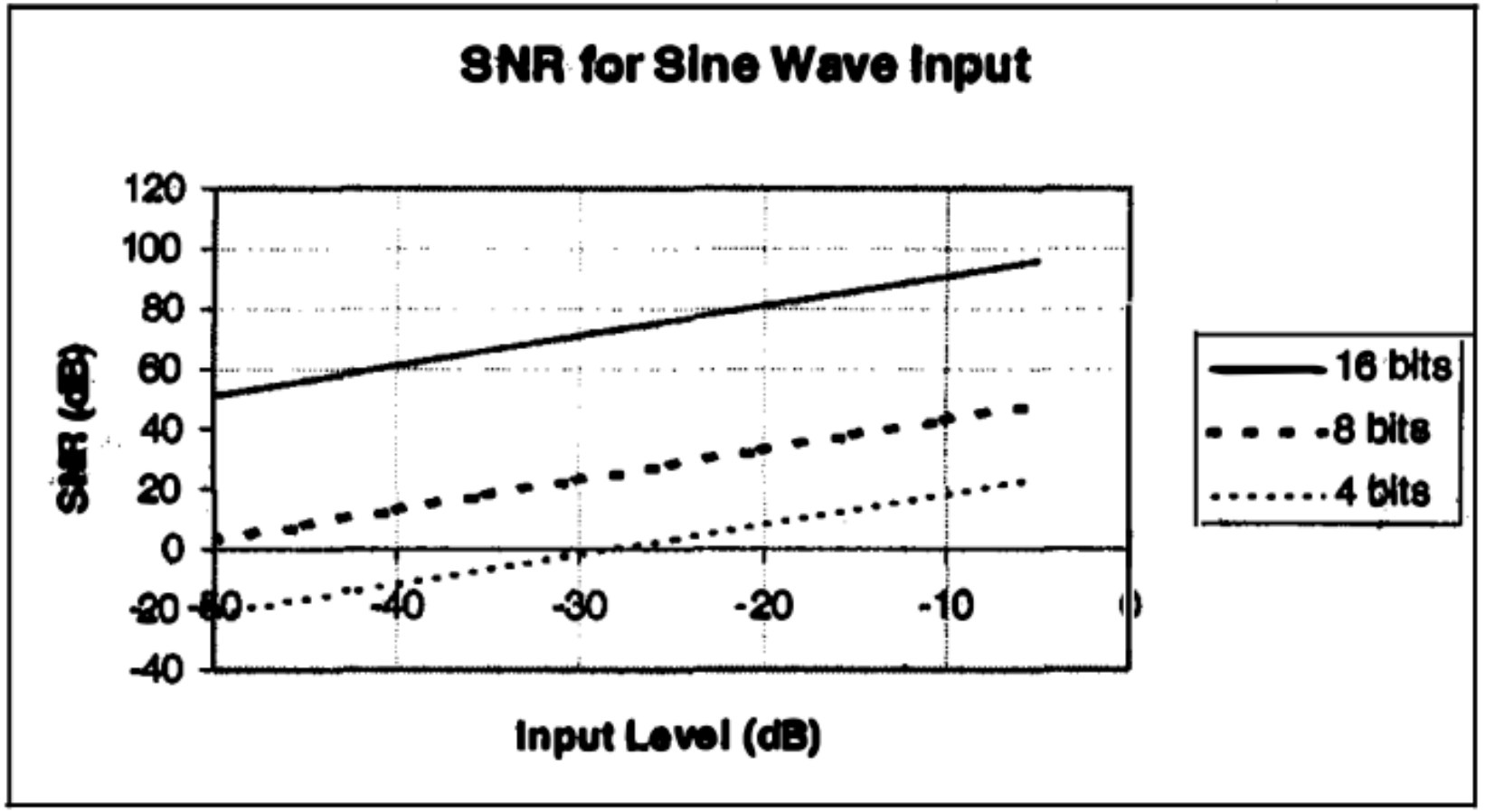
- Al digitalizar, se introduce una distorsión, propia del proceso de cuantización

$$q(t) = x_{out}(t) - x_{in}(t)$$

$$SNR = 10log_{10}\left(\frac{< x_{in}^2 >}{< q^2 >}\right)$$

- Con más bits por muestra, menor distorsión

SNR según cantidad de bits, cuantización lineal



Tomado de:
"Introduction to Digital Audio Coding and Standards", Marina Bosi and Richard Goldberg, KLUWER ACADEMIC PUBLISHERS

Ejemplo: CD (Linear PCM)

44.1 kHz x 16 bits/muestra x 2 canales (stereo) =
1.4 Mb/s

Una canción de 10 minutos:

10 x 60 x 1.4 = 840 Mbits = 105 MBytes

Un CD tiene 700 MB, podrían “entrar” unas 7 canciones

¿Cómo bajar el bitrate, y mantener una calidad aceptable?

¿Qué es aceptable?

- Depende de la aplicación…
  - Telefonía
  - Video conferencias
  - Conciertos
  - Música
  - …
  - ¿Mono, estéreo o más canales?

¿Hay otros factores?

- Retardos de codificación y de de-codificación
- Capacidad de procesamiento
- Robustez a los errores o pérdidas de información

Tipos de codecs

De voz

- Hacen uso de las características específicas de la voz humana
- Utilizados típicamente en aplicaciones de telefonía y de video conferencias
- Pueden ser de “forma de onda” o de “síntesis de voz”

De audio

- Hacen uso de las características más generales del aparato auditivo y la percepción humana del audio

Tipos de codecs

Codificación “con pérdida”

- Para comprimir “descartan” información que sea perceptualmente irrelevante o reiterativa
  - MP3 (MPEG-1 Audio Layer III), AAC (Advanced Audio Codec), Dolby AC-3, …

Codificación “sin pérdida”

- Comprimen sin perder información, se puede reconstruir exactamente el flujo original de bits
  - PCM, FLAC (Free Lossless Audio Codec), MPEG-4 ALS (Audio Lossless Coding”), ALAC (Apple Lossless Audio Coding), …

Codificación de voz por “forma de onda”

Inicialmente, los codecs se basaron en codificar de la manera más eficiente posible la “forma de onda” de la señal.

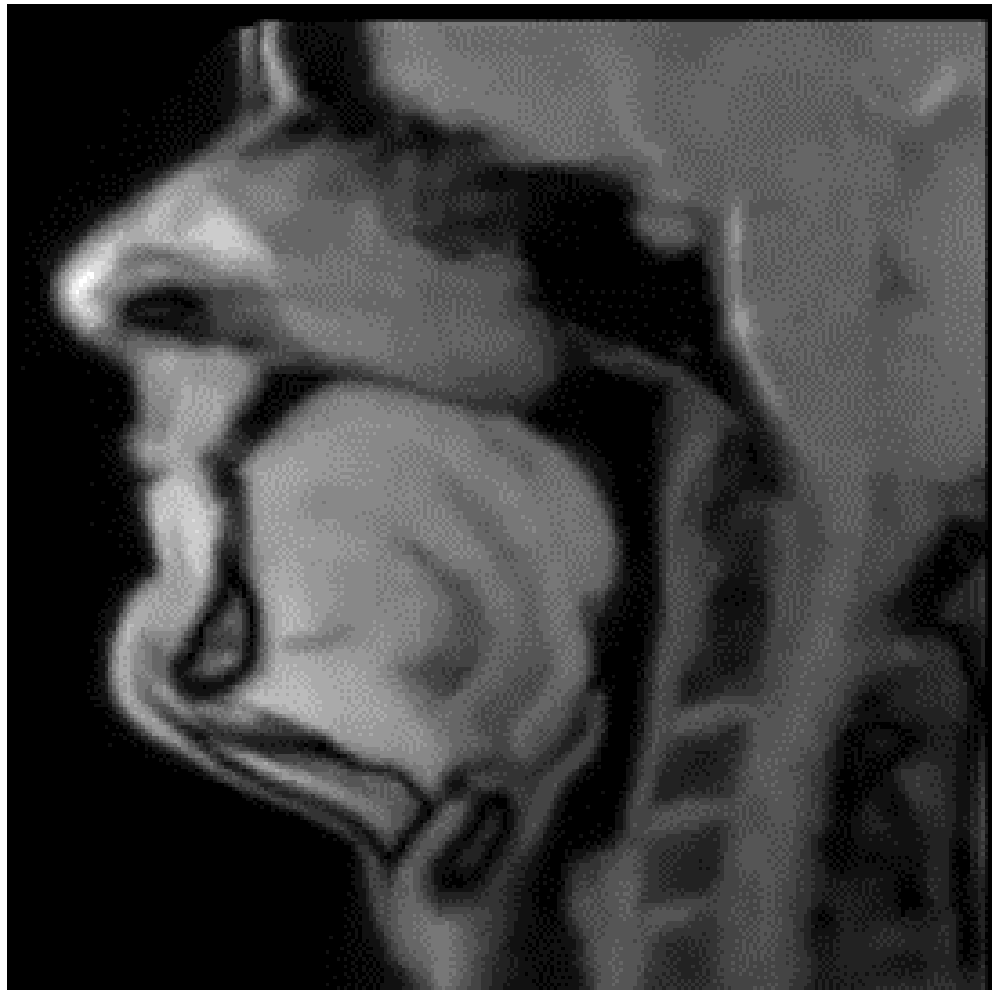
Posteriormente, para bajar la tasa de bits necesaria para la transmisión, se comenzaron a utilizar técnicas “predictivas”

- Basadas en predecir los valores de las muestras en base a la extrapolación de las muestras anteriores

Codificación de voz por “síntesis de voz”

Video de cuerdas vocales

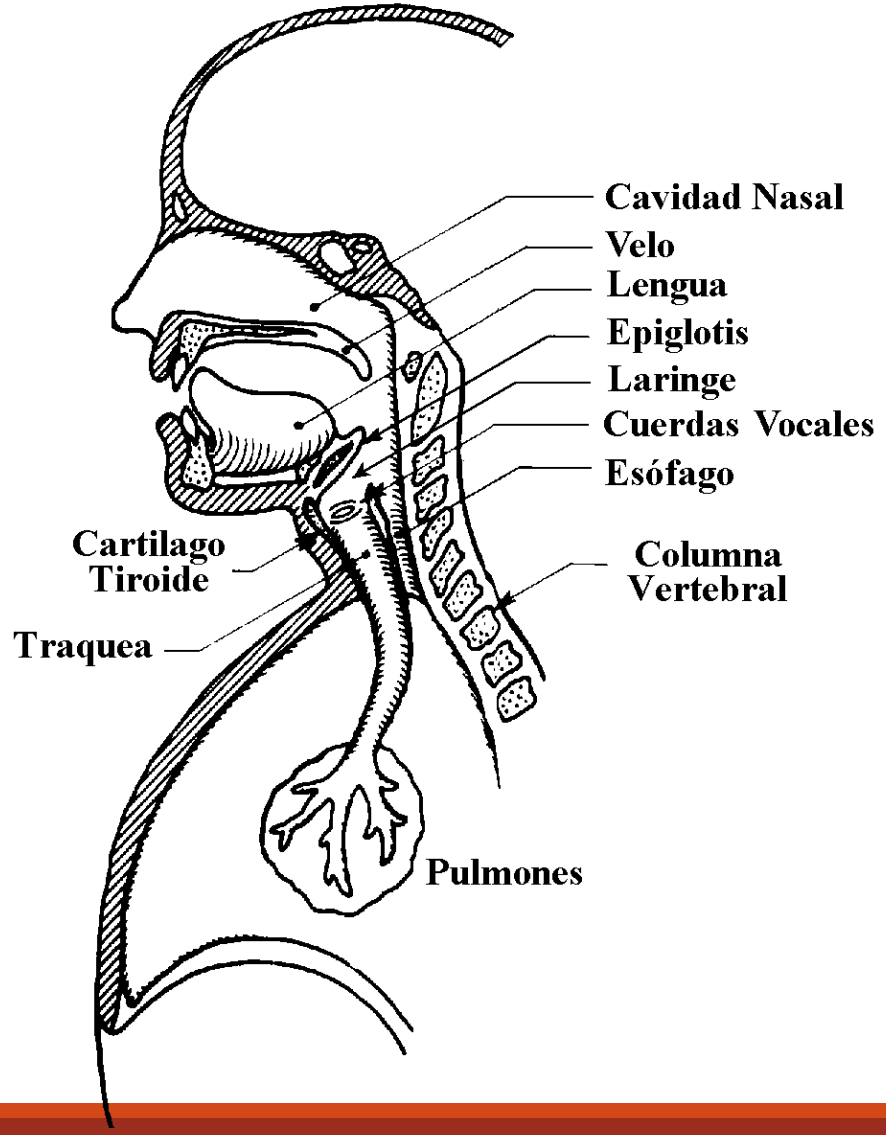
Video de cuerdas vocales



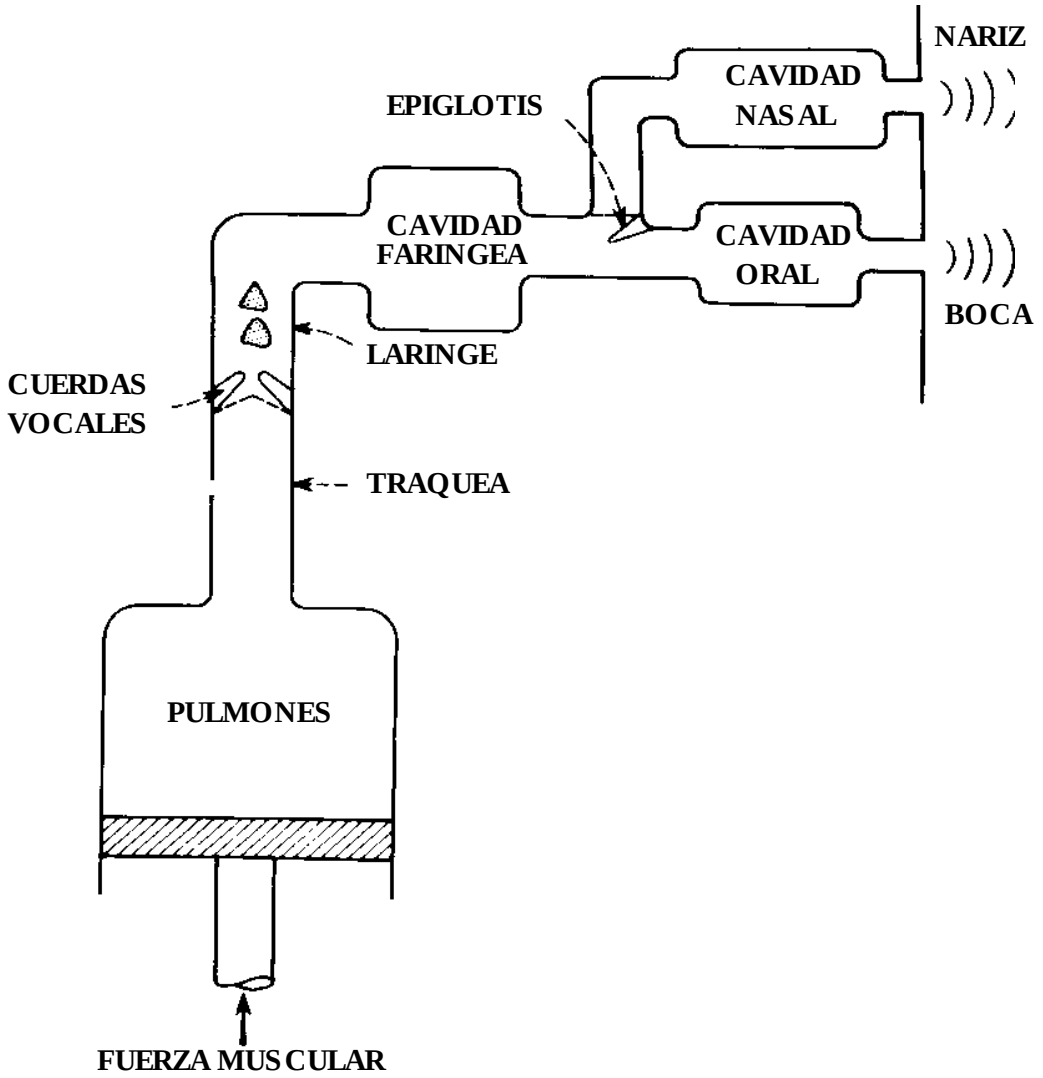
De. Jens Frahm / MPIBPC (https://www.mpinat.mpg.de/626786/real-time-mri)

Codificación de voz por “síntesis de voz”

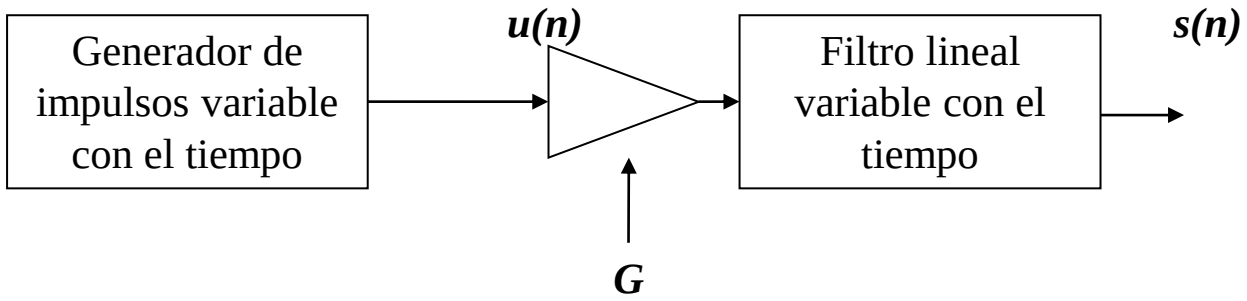
Sobre la década de 1980, se introduce la idea de generar “voz sintética”, simulando la manera en que se produce la voz humana en el conducto vocal.



Modelo del Conducto Vocal



Modelo del Conducto Vocal



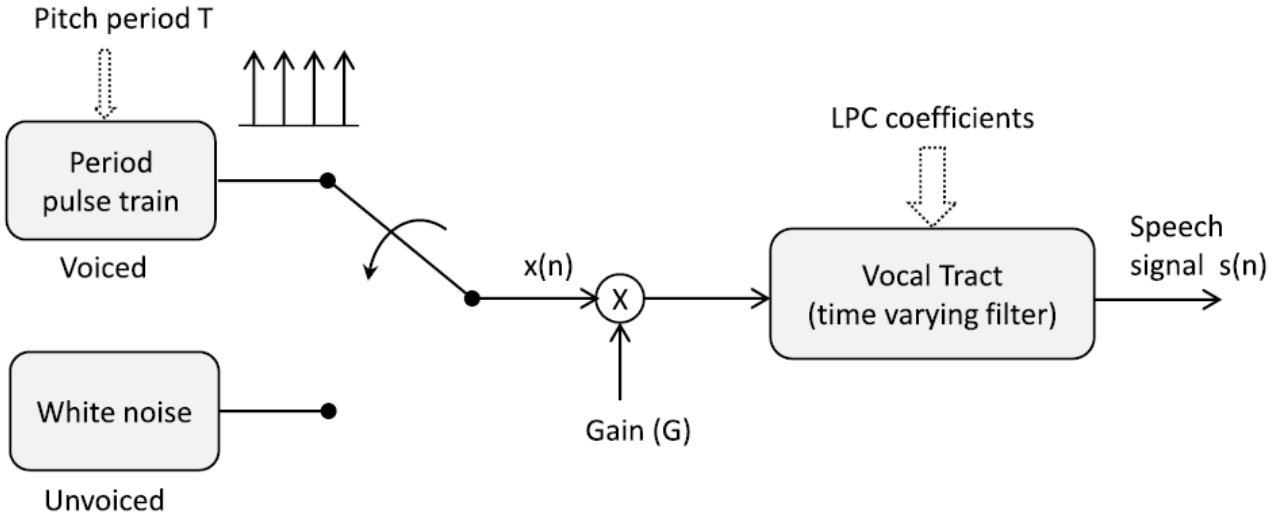
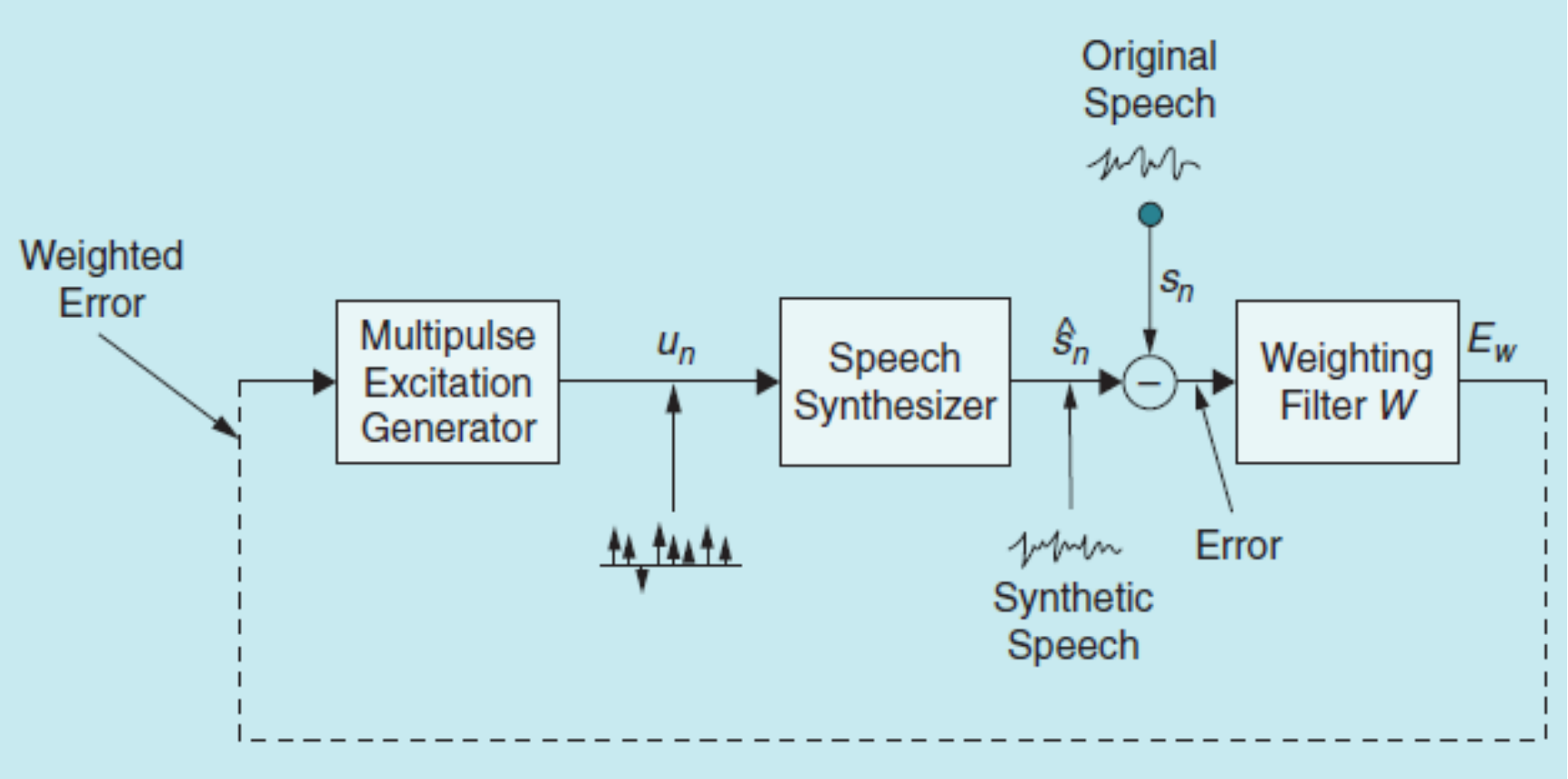


Imagen de Lingfen Sun (2013). Guide to Voice and Video over IP For Fixed and Mobile Network

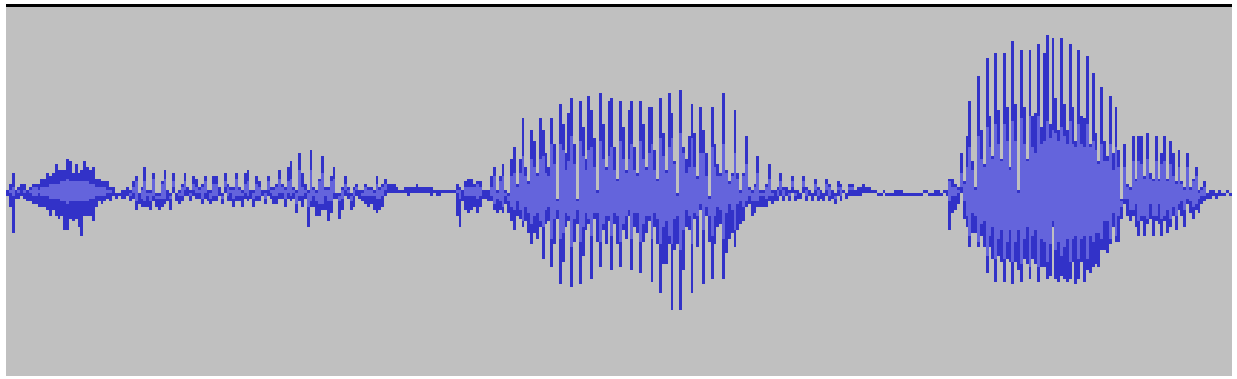
$$H(z) = \frac{1}{1 + \sum_{k=1}^{p} a_k z^{-k}}$$

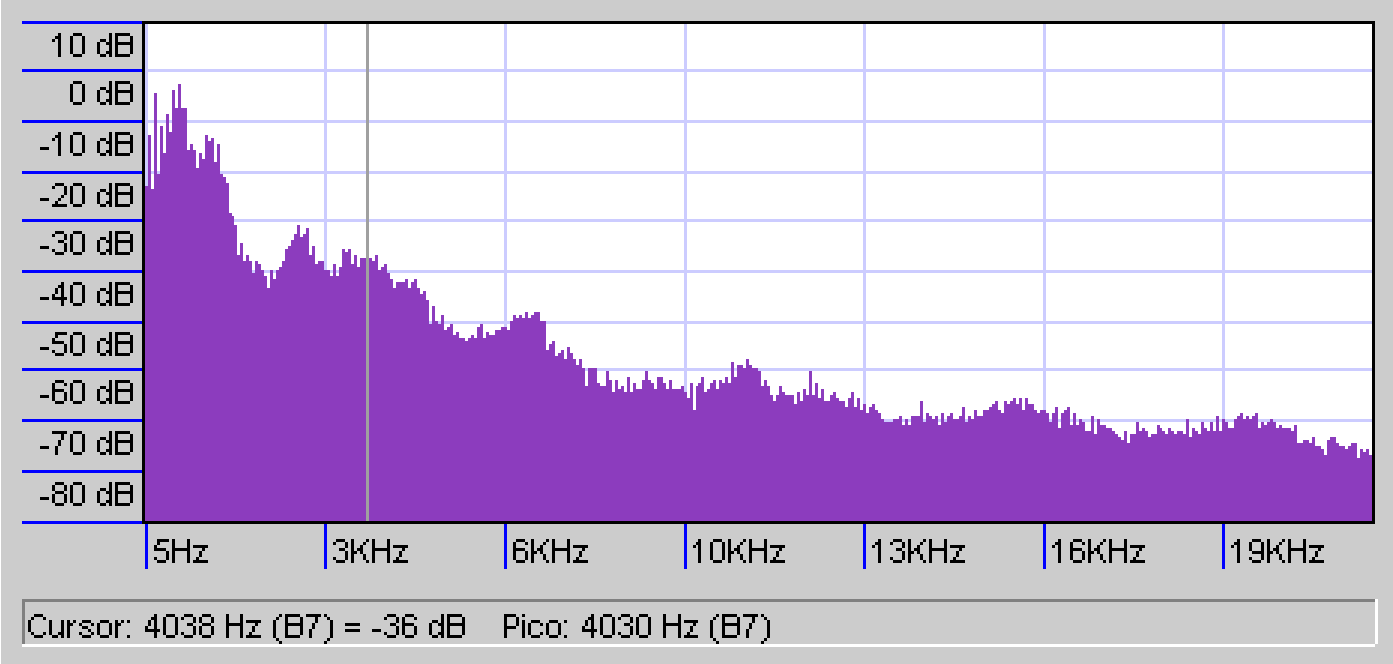
p es el orden del filtro, y a_k representan los coeficientes del filtro.

Estimación de los parámetros del modelo



Espectro típico de la voz





CODECs de Audio

CODIFICACIÓN DE VOZ Y VIDEO

CODECs

Pueden ser caracterizados por

- su tecnología (de “forma de onda”, de “síntesis de voz”)
- su tasa de bits (bit rates)
- la calidad resultante del audio codificado
- su complejidad
- el retardo que introducen

Según el ancho de banda de la señal de entrada

Banda angosta (narrowband)	300 a 3400 Hz
Banda ancha (wideband)	50 a 7000 Hz
Banda super ancha (superwideband)	50 a 14000 Hz
Banda completa (fullband)	50 a 20000 Hz

CODECs de banda angosta

Codec	Nombre	Bit rate (kb/s)	Retardo (ms)	Comentarios
G.711	PCM: Pulse Code Modulation	64, 56	0.125	Codec “base”, utiliza dos posibles leyes de compresión: μ-law y A-law
G.723.1	Hybrid MPC-MLQ and ACELP	6.3, 5.3	37.5	Desarrollado originalmente para video conferencias en la PSTN, es actualmente utilizado en sistemas de VoIP
G.728	LD-CELP: Low-Delay code excited linear prediction	40, 16, 12.8, 9.6	1.25	Creado para aplicaciones DCME (Digital Circuit Multiplex Encoding)
G.729	CS-ACELP: Conjugate Structure Algebraic Codebook Excited Linear Prediction	11.8, 8, 6.4	15	Ampliamente utilizado en aplicaciones de VoIP, a 8 kb/s
AMR	Adaptive Multi Rate	12.2 a 4.75	20	Utilizado en redes celulares GSM

CODECs de banda ancha

Codec	Nombre	Bit rate (kb/s)	Retardo (ms)	Comentarios
G.722	Sub-band ADPCM	48,56,64	3	Inicialmente diseñado para audio y videconferencias, actualmente utilizado para de telefonía de calidad en VoIP
G.722.1	Transform Coder	24,32	40	Usado en audio y videoconferencias
G.722.2	AMR-WB	6.6 a 23.85	25.9375	Estandar en común con 3GPP (3GPP TS 26.171). gran inmunidad a los ruidos de fondo en ambientes adversos (por ejemplo celulares)
G.711.1	Wideband G.711	64, 80, 96	11.875	Amplía el ancho de banda del codec G.711, optimizando su uso para VoIP
G.729.1	Wideband G.729	8 a 32 kb/s	<49 ms	Amplía el ancho de banda del codec G.729, y es “compatible hacia atrás” con este codec. Optimizado su uso para VoIP con audio de alta calidad
RtAudio	Real Time Audio	8.8, 18	40	Codec propietario de Microsoft, utilizado en aplicaciones de comunicaciones unificadas (OCS)

CODECs de banda superancha

Codec	Nombre	Bit rate (kb/s)	Retardo (ms)	Comentarios
SILK	SILK	8 a 24	25	Utilizado por Skype

CODECs de banda completa

Codec	Nombre	Bit rate (kb/s)	Retardo (ms)	Comentarios
G.719	Low-complexity, full-band	32 a 128	40	Es el primer codec “fullband” estandarizado por ITU
Opus	Opus	6 a 510	Hasta 60	Incorpora tecnología de SKYPE RFC 6716 (propuesta en set 2012)
EVS	Enhanced Voice Services	5.9 a 128	32	Diseñado para servicios de VoLTE (Voice over LTE) Es el primer códec desarrollado por 3GPP de banda completa (hasta 20 kHz)

G.711 – Pulse Code Modulation (PCM) of voice frequencies

Estandarización de “Ley A” y “Ley Mu”

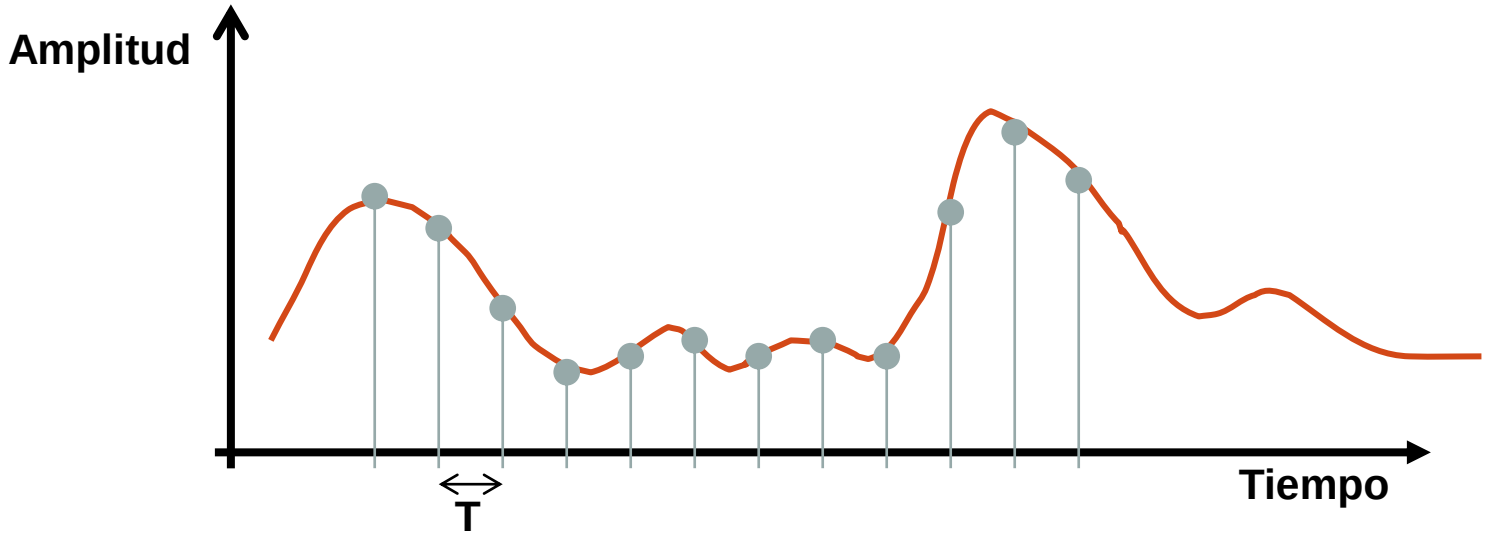
Conserva la forma de onda, codifica muestra a muestra

Tiene características “no lineales” para minimizar la cantidad de bits por muestra

Resulta en una velocidad de 64 kbit/s

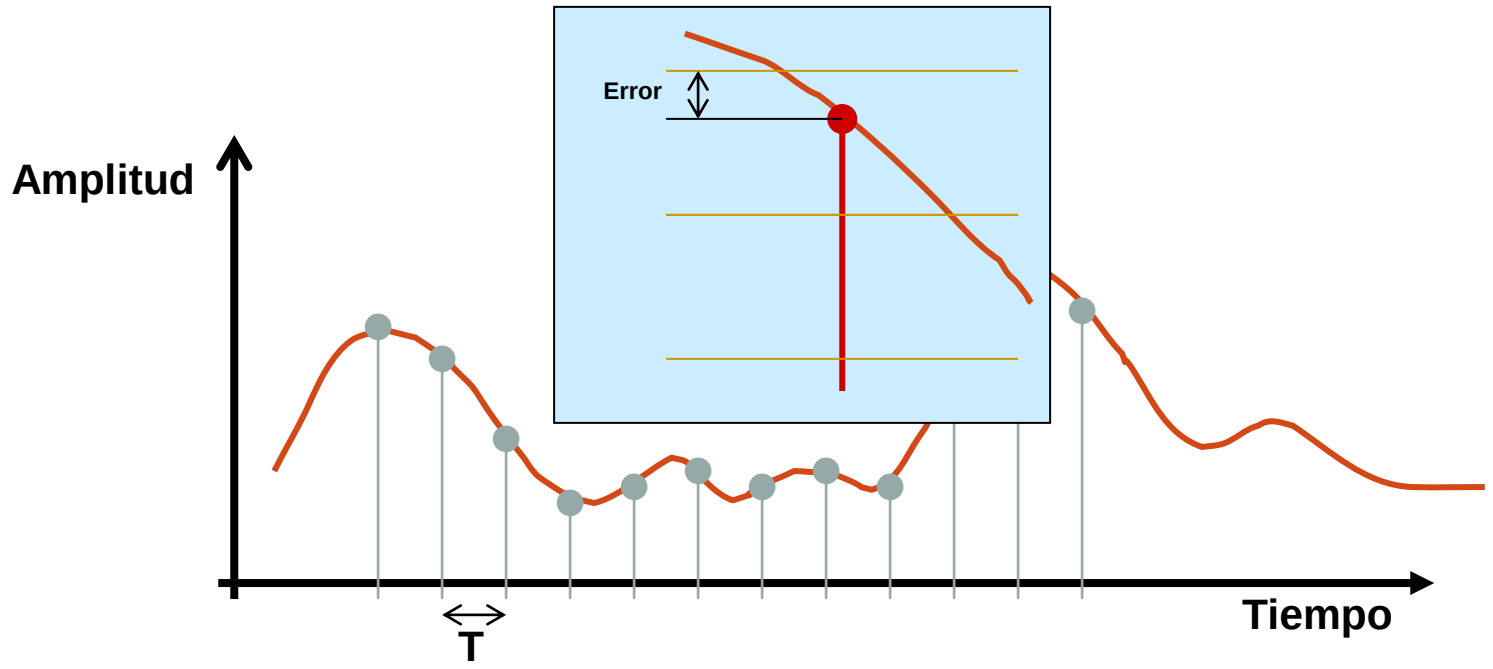
Digitalización de la voz
Proceso de digitalización

1. Muestreo
   - Se toman “muestras” de la señal a intervalos regulares. Estos intervalos deben ser tales que cumplan con el teorema de muestreo:
     - La mínima frecuencia a la que puede ser muestrada una señal y luego reconstruida es el doble de la frecuencia máxima de dicha señal



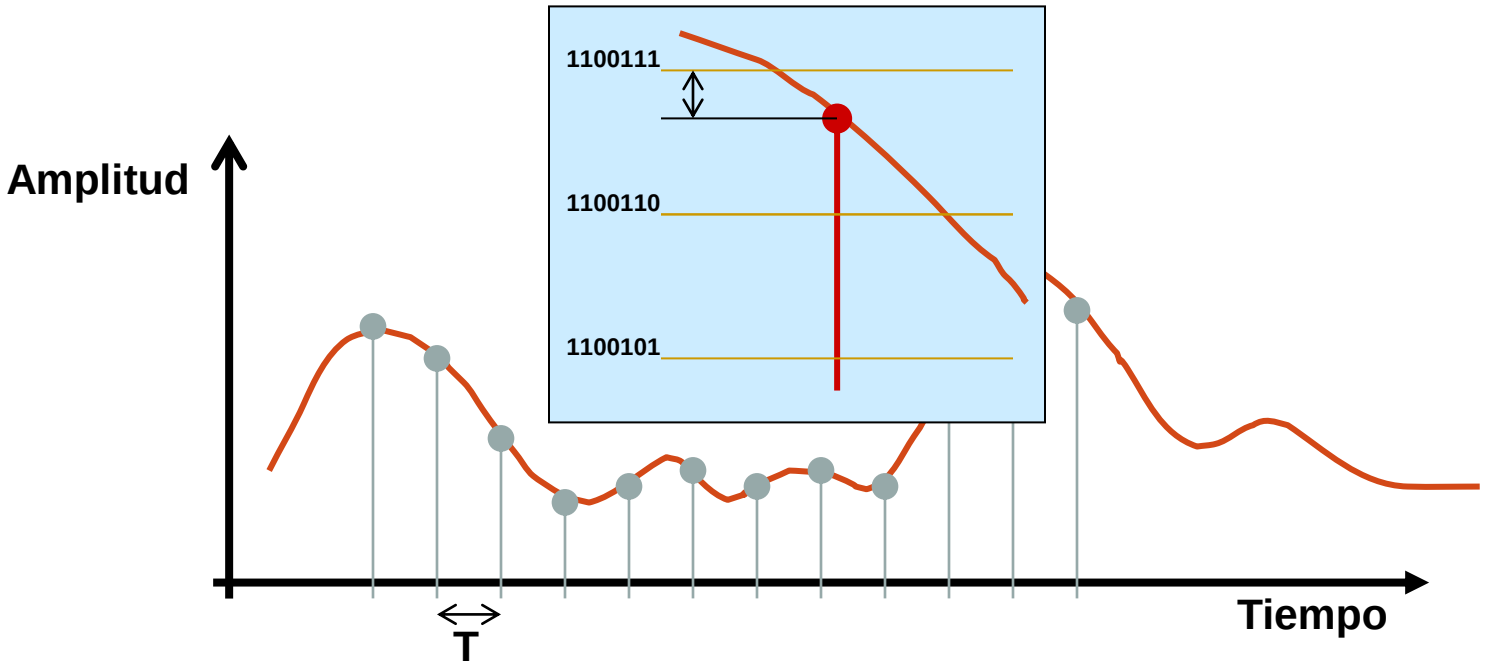
Digitalización de la voz
Proceso de digitalización

2. Cuantificación
   - Los valores de las muestras se “cuantifican” en cantidades discretas. La cantidad total de valores discretos debe introducir la menor cantidad posible de “ruido de cuantificación”



Digitalización de la voz
Proceso de digitalización

3. Codificación
   - Los valores “cuantificados” se “codifican” en números que pueden ser luego transmitidos y procesados digitalmente.



Digitalización de la voz
G.711

1. Muestreo
   - Si bien el oido humano puede llegar a escuchar sonidos de hasta 18 - 20 kHz, la mayor parte de la energía de la voz humana se encuentra por debajo de los 4 kHz.
   - El sonido resultante de filtrar la voz humana a 3.4 kHz es perfectamente inteligible, además puede distinguirse al locutor.
   - De acuerdo al teorema del muestreo, para poder reconstruir una señal de 3.4 kHz debe muestrarse a más de 6.8 kHz.
   - Originialmente se seleccionó como frecuencia de muestreo para telefonía **8 kHz** (una muestra cada **125** microseg).

T

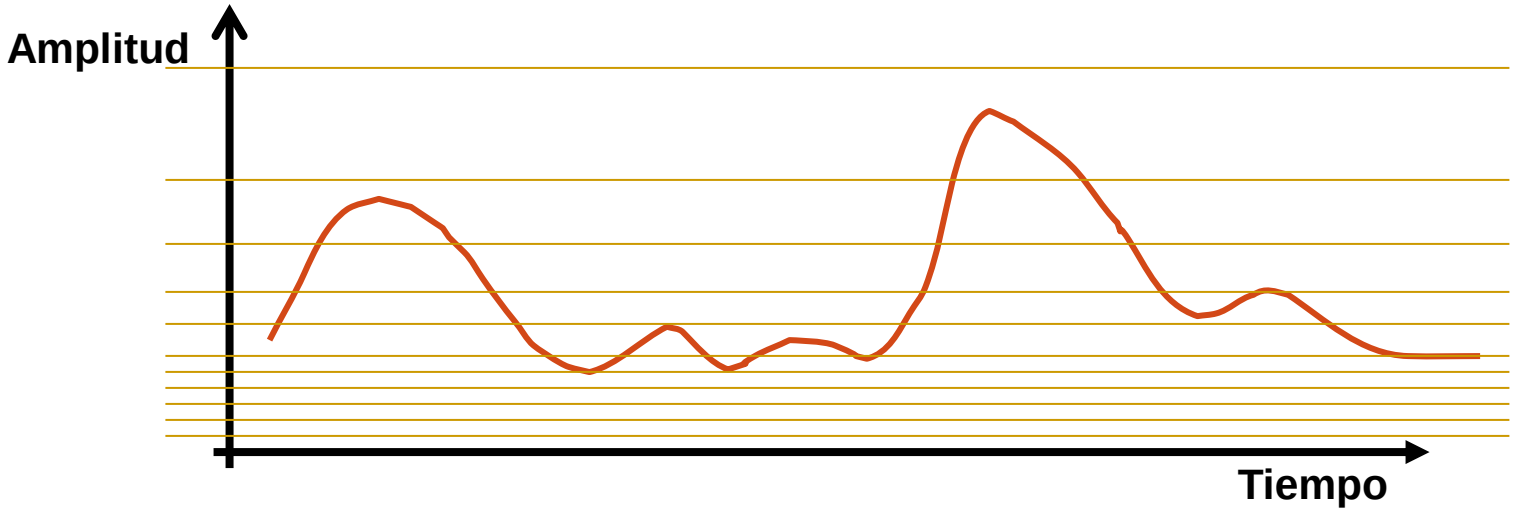
Digitalización de la voz
G.711

2. Cuantificación (1/3)

- Una cuantificación lineal genera un “error de cuantificación” constante, independiente del nivel de la señal.
- Los errores de cuantificación se traducen en “ruido” al reconstruir la señal.
- Para lograr niveles de ruido aceptables en señales de voz con cuantificadores lineales, se requieren 4096 niveles.
- El oído es mas sensible a los “ruidos” en señales bajas que en señales altas.

Digitalización de la voz
G.711

2. Cuantificación (2/3)
   - Cuantificación no lineal: Permite tener errores de cuantificación pequeños para señales pequeñas y grandes para señales grandes
   - Con menos cantidad de niveles se logra buena calidad en la señal reconstruida



Digitalización de la voz
G.711

2. Cuantificación (3/3): Leyes de Cuantificación

- Ley A (de 13 segmentos):

  $y=(1+\log(Ax) / (1 + \log(A))$ si $1/A < x < 1$

  $y= Ax/(1 + \log(A))$ si $0 < x < 1/A$

  $A = 87.6$

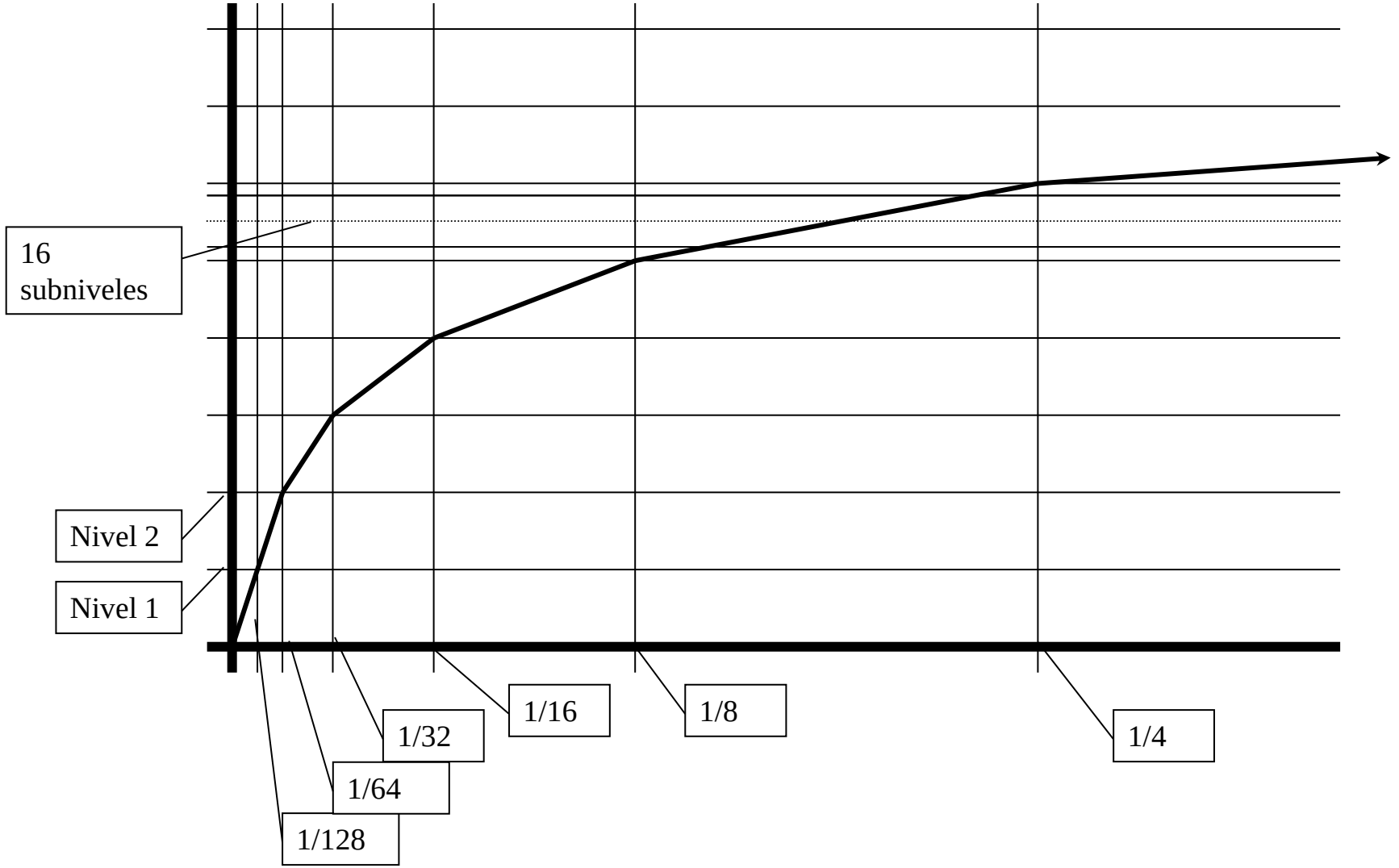
- Ley μ (de 15 segmentos):

  $y=\log(1+\mu x)/\log(1+\mu)$

  $\mu = 255$

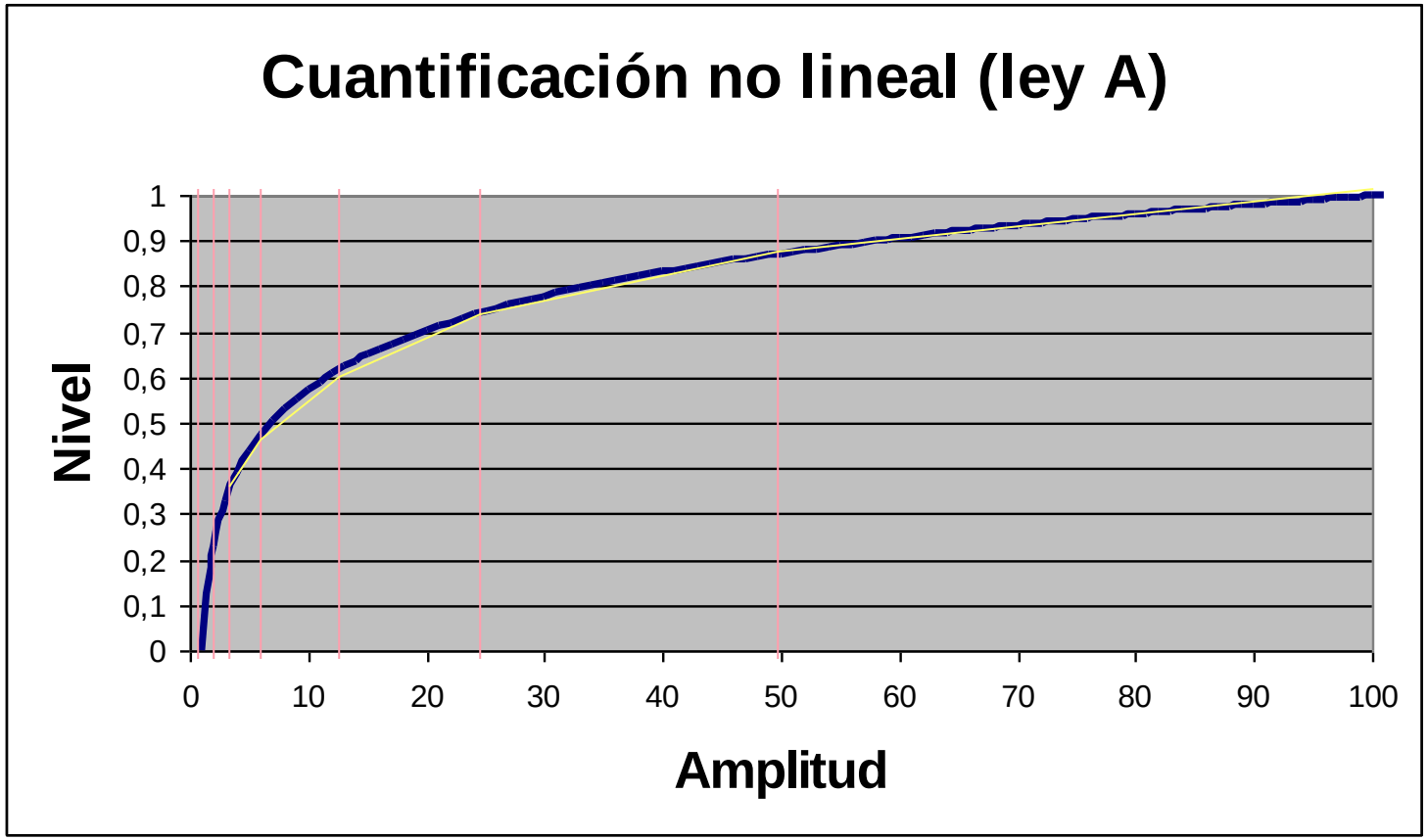
- Ambas leyes forman parte de la Recomendación ITU-T G.711

Digitalización de la voz
G.711 - Ley A



Digitalización de la voz

G.711 - Ley A



Digitalización de la voz
G.711 - Ley A

3. Codificación: Ley A o ley de los 13 segmentos

- El bit mas significativo (bit 7) indica el signo.
- Los bits 4-6 indican el numero de segmento.
- Los bits menos significativos (bits 0-3) indican el intervalo dentro del segmento.

Bit	7	6	5	4	3	2	1	0
	Signo	Segmento (0 - 7)			Intervalo (0 - 15)			

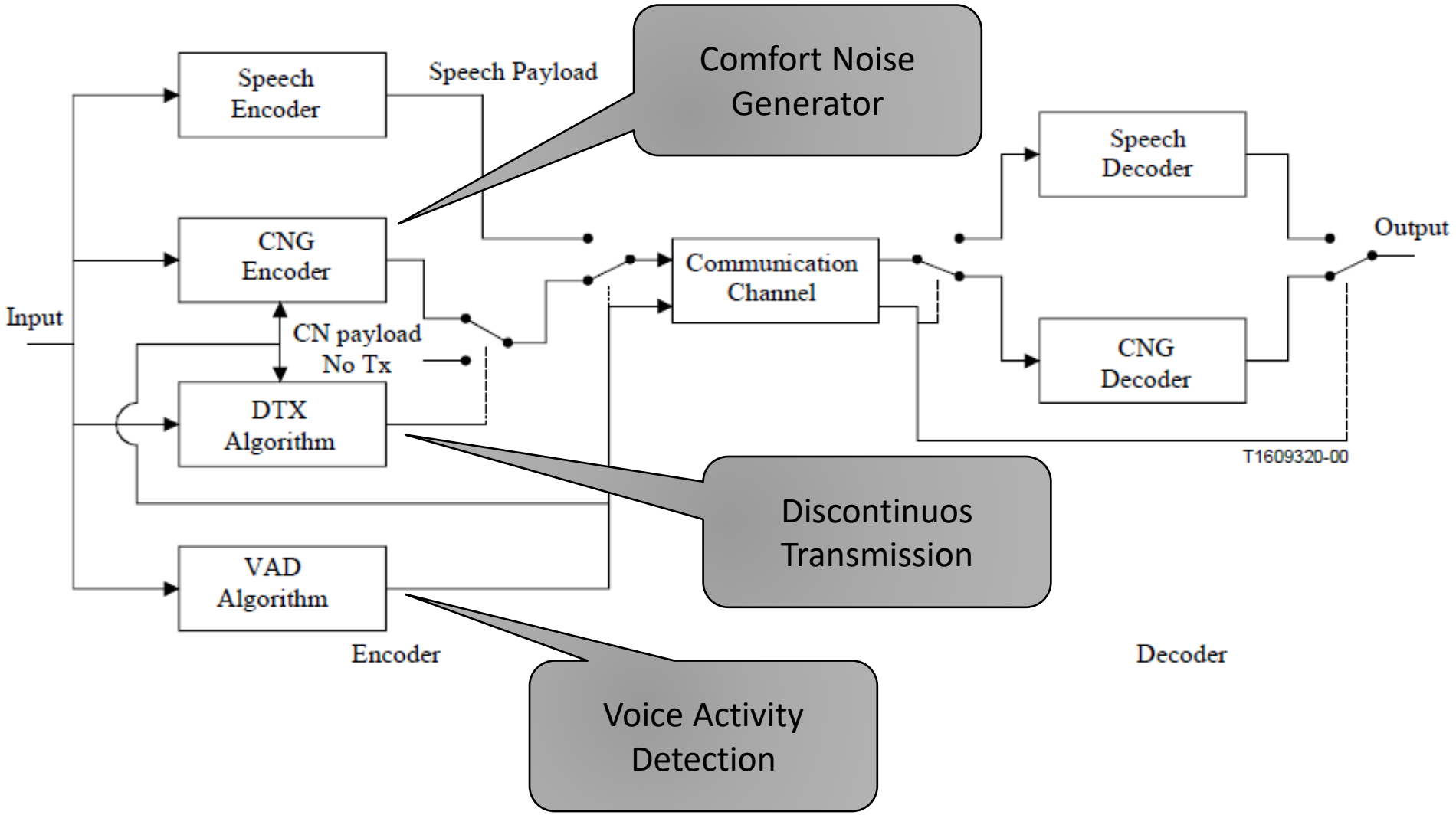
G.711 Appendix II
Comfort Noise Generation

Aproximadamente, durante el 40% de una conversación telefónica, “escuchamos sin hablar”

El apéndice II de G.711 define un “comfort noise payload format” utilizado en comunicaciones sobre redes de paquetes

Se envía con una frecuencia baja, por ejemplo, 10 veces por segundo.

G.711 Appendix II
Comfort Noise Generation



G.711.1

Wideband embedded extension for G.711 PCM

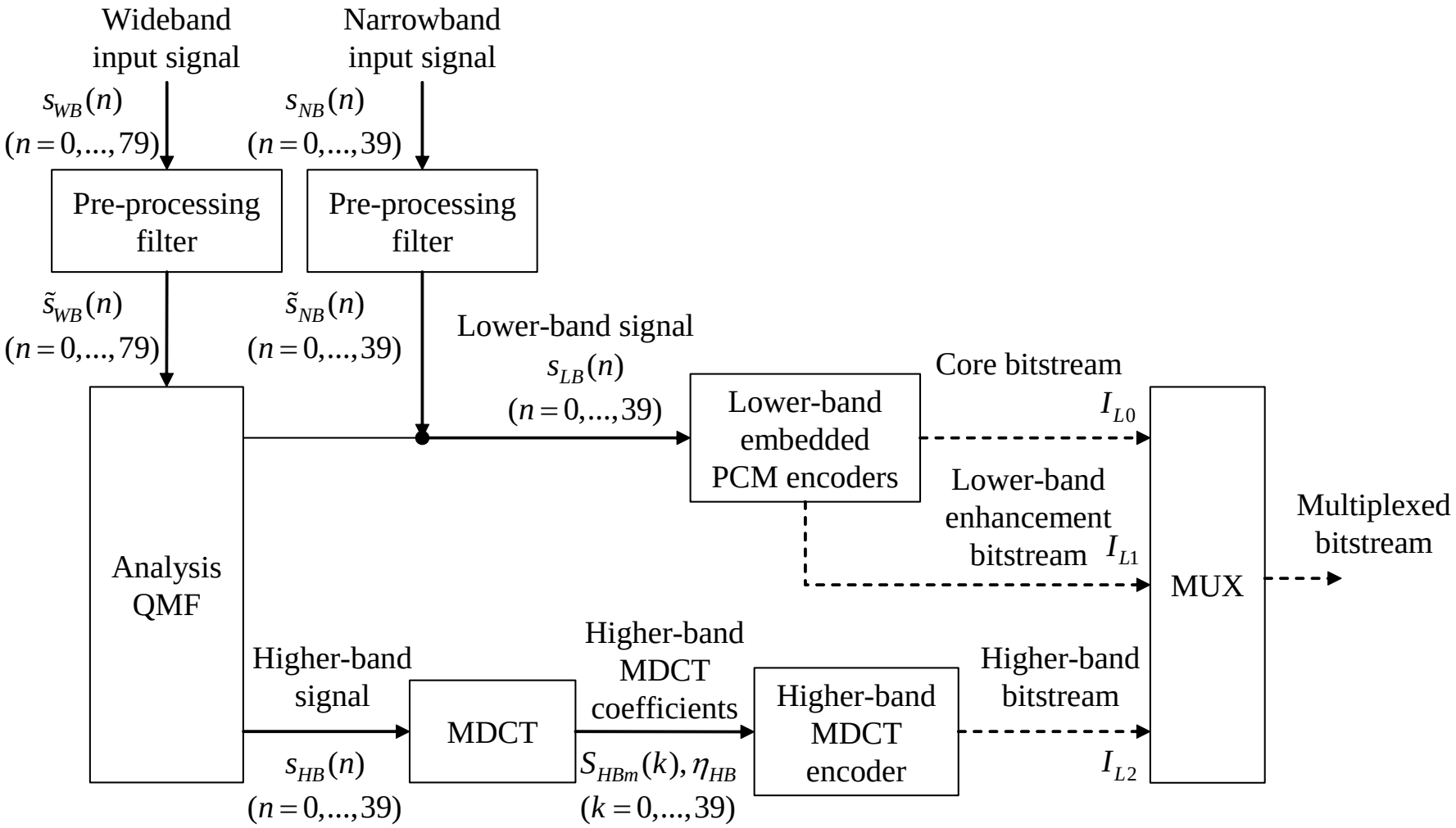
Aprobado en Marzo de 2008, como una extensión de G.711 para banda ancha (7 kHz)

Trabaja en 64, 80 y 96 kb/s

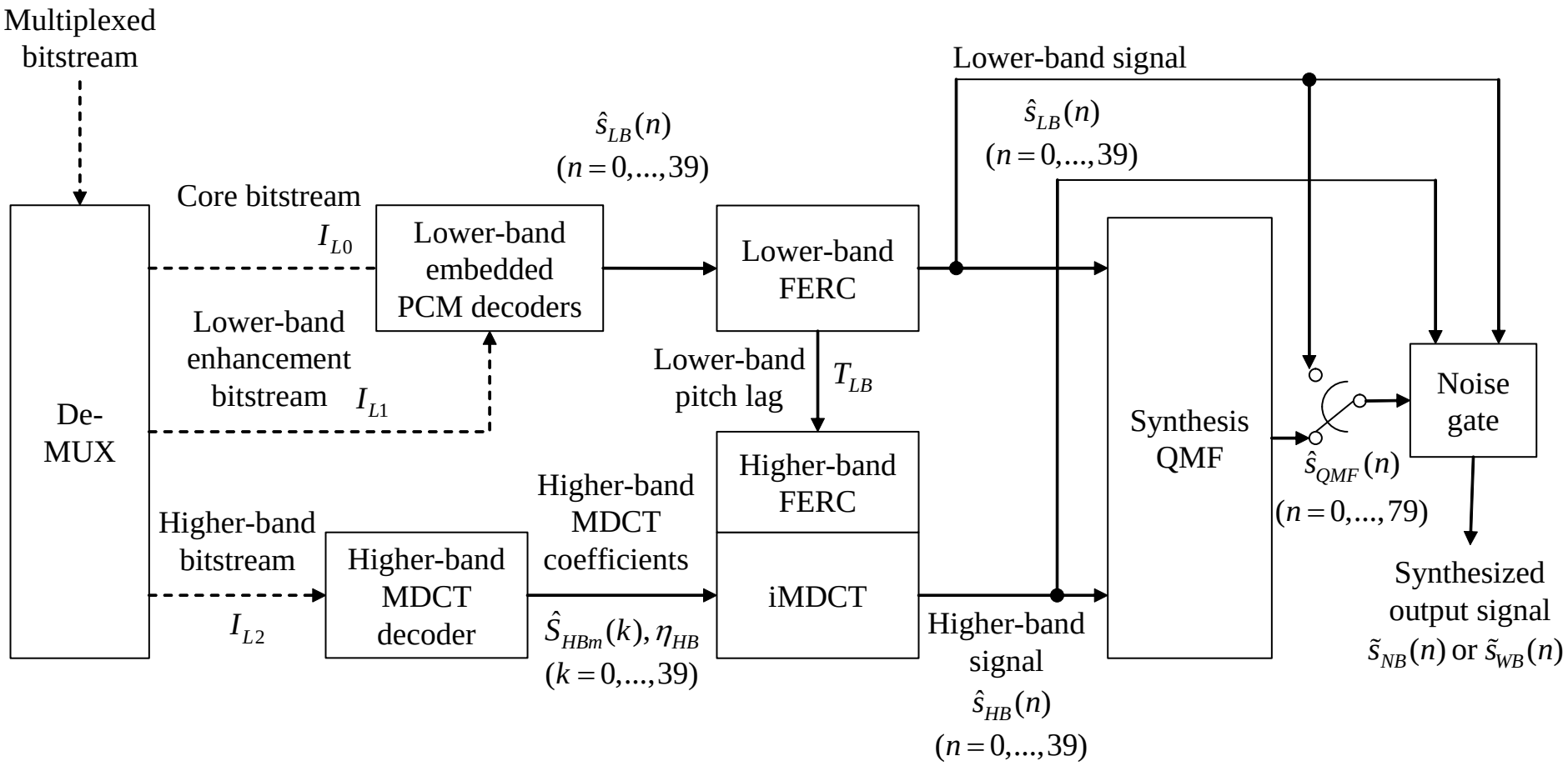
Las muestras codificadas pueden ser convertidas en G.711 por medio de un simple truncado

Las muestras de entrada son tomadas cada 16 kHz, pero también está soportada la frecuencia de muestreo de 8 kHz (compatibilidad con G.711)

Codificador G.711.1



Decodificador G.711.1



Modos de operación G.711.1

Mode	Sampling rate (kHz)	Core layer (Layer 0, I_{L0})	Lower-band enhancement layer (Layer 1, I_{L1})	Higher-band enhancement layer (Layer 2, I_{L2})	Overall bit rate (kbit/s)
		64 kbit/s	16 kbit/s	16 kbit/s	
R1	8	x	–	–	64
R2a	8	x	x	–	80
R2b	16	x	–	x	80
R3	16	x	x	x	96

Tramas G.711.1

Son de 5 ms y tienen un total de 480 bits por trama

- 320 bits de la capa 0 (G.711), correspondientes a 8 bits x 40 muestras
- 80 bits de la capa 1
- 80 bits de la capa 2

La demora total del algoritmo lleva un total de 11.875 ms

- 5 ms para la información de la trama
- 5 ms extras necesarios para el análisis MCDT (“lookahead”)
- 1.875 ms para la implementación del filtro QMF

G.729 - Coding of speech at 8 kbit/s using conjugate structure algebraic-code-excited linear prediction (CS-ACELP)

No conserva la forma de onda, sino que utiliza técnicas de “síntesis de voz”

El modelado de la boca y la garganta se hace por medio de filtros lineales y la voz se genera a partir de una vibración periódica de aire que los excita

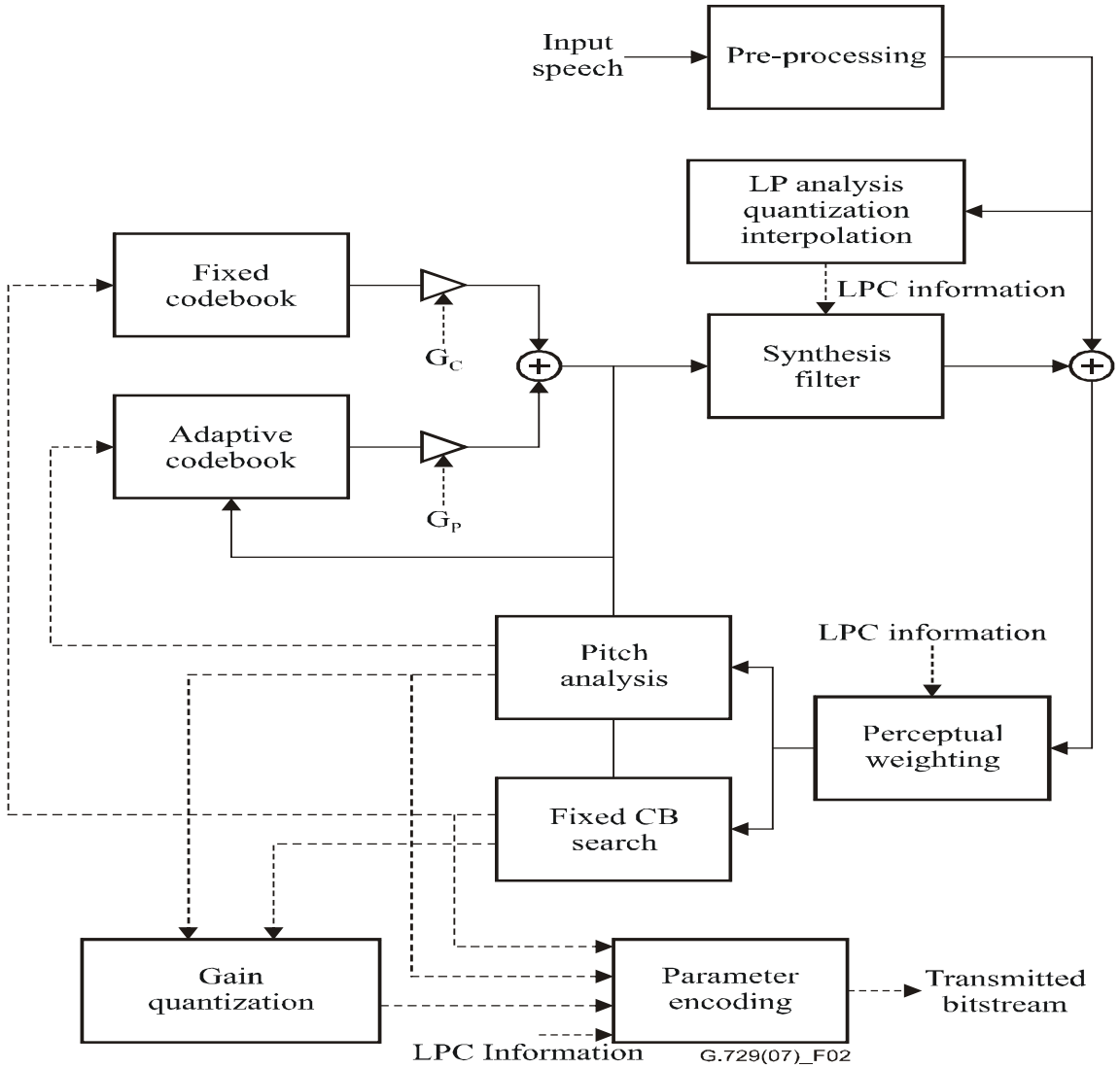
Utiliza “ventanas” de 10 ms para obtener los parámetros y se usan 80 bits (10 bytes) para representarlos

- Resulta en una velocidad de 8 kbit/s

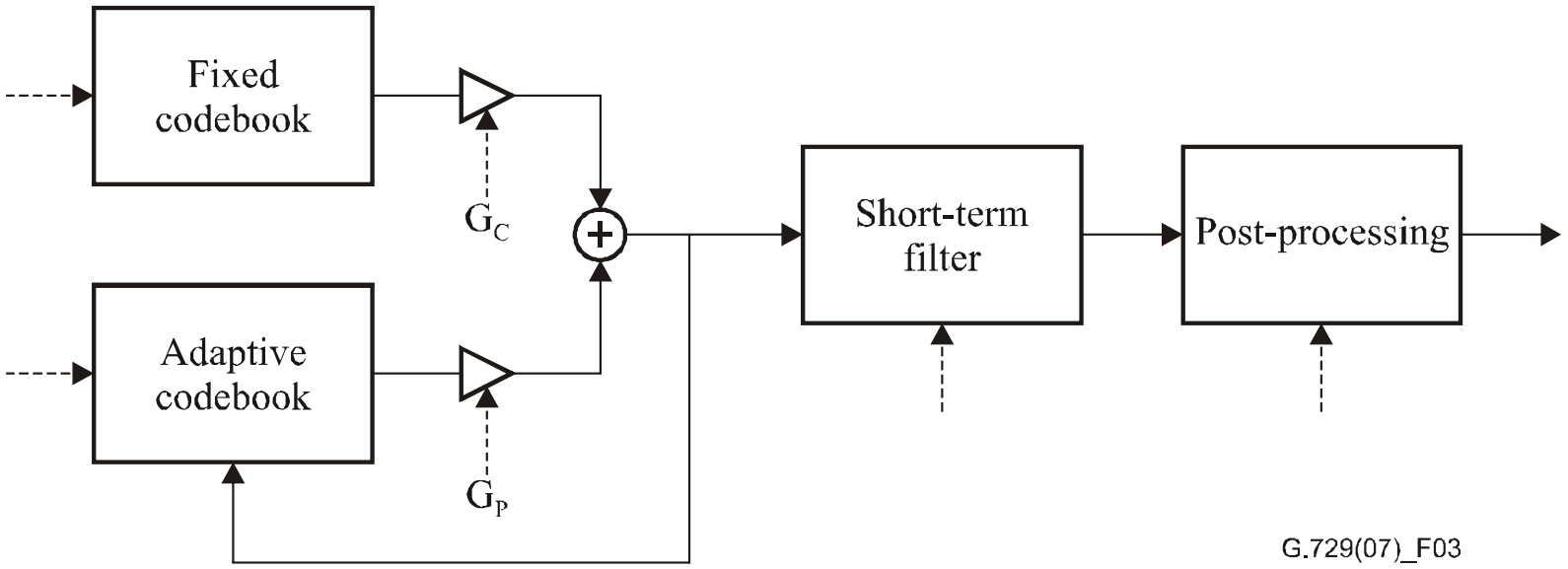
Tiene 5 ms de “look ahead”, resultando en una demora total de 15 ms

Utiliza técnicas CS-ACELP (Conjugate-Structure Algebraic-Code-Excited Linear Prediction)

Codificador G.729



Decodificador G.729



G.729

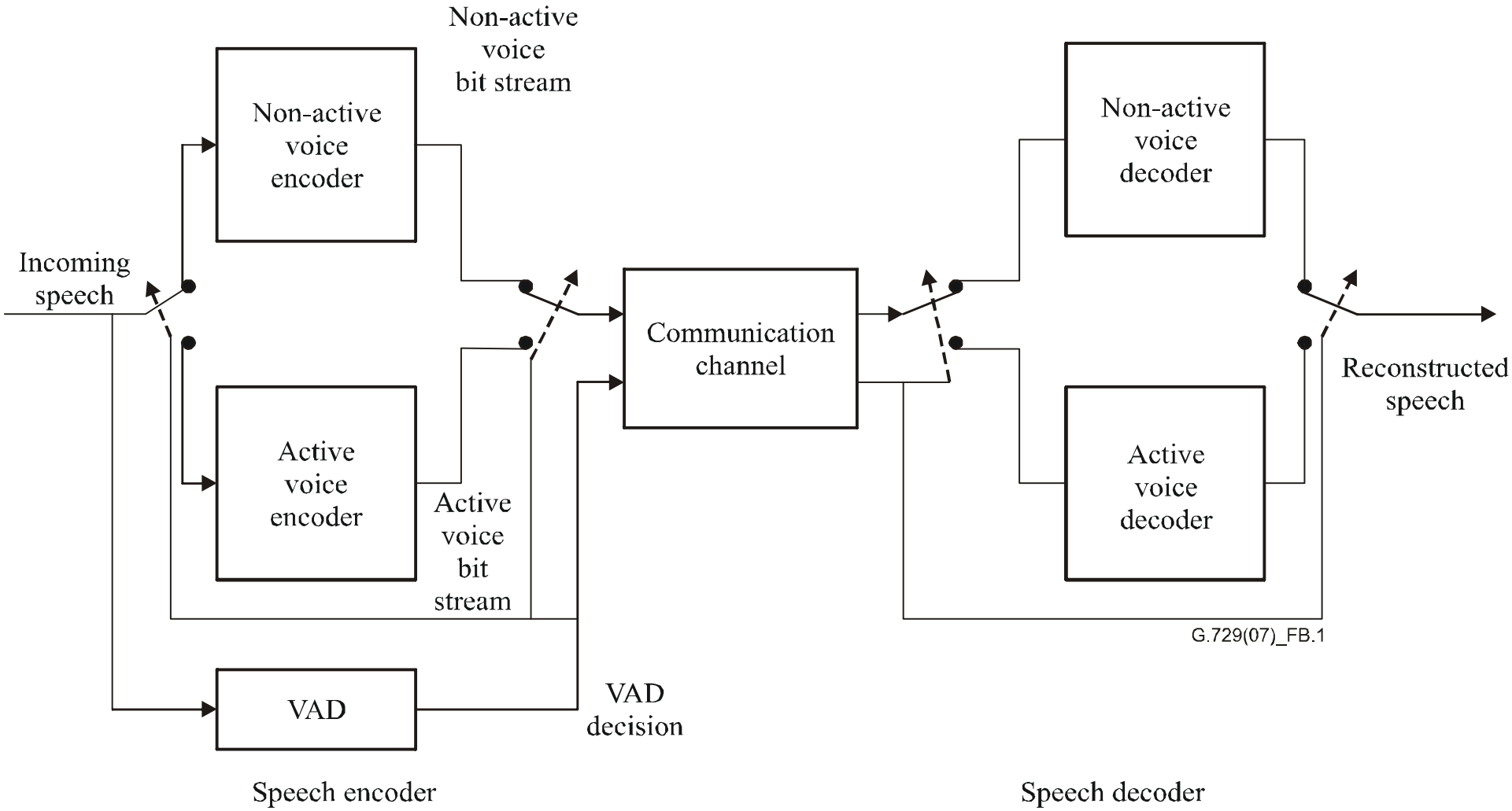
G.729 A

- Variante del codec para lograr menor complejidad
- Es interoperable con G.729

G.729 B

- Detección de actividad de voz y silencios
- Modelado y regeneración del “ruido de fondo” (CNG = Confort Noise Generation)
- Menor ancho de banda en la LAN

G.729 B VAD (Voice Activity Detection)



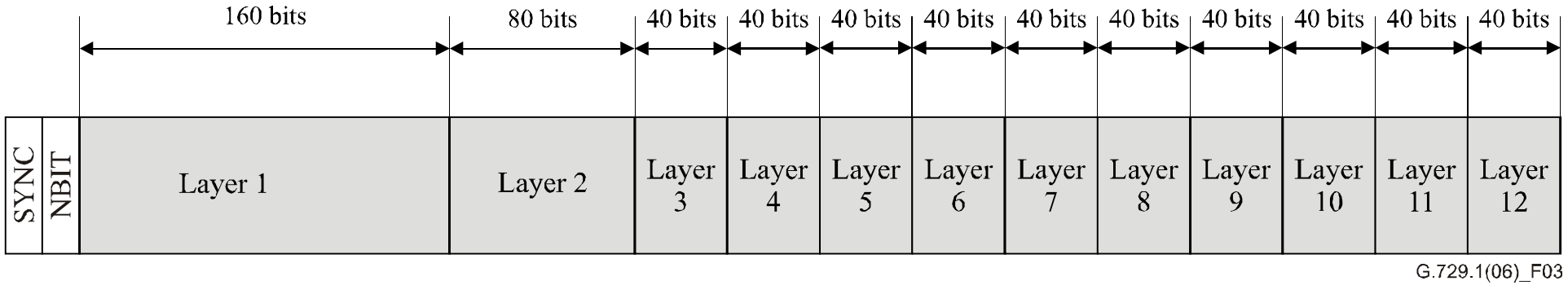
G.729.1 - An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729

Aprobado en mayo de 2006

Diseñado para proveer una transición sencilla en el mundo de la telefonía entre sistemas que utilizan banda angosta (300 a 3400 Hz) y nuevos sistemas que soporten banda ancha (50 a 7000 Hz)

Inter operable con la recomendación G.729 y sus anexos A y B, los que tienen amplia difusión en el mundo de VoIP

Trama G.729.1



Capa 1: Codificación basada en CELP, de 8kb/s y compatible con G.729

Capa 2: Mejoras en las frecuencias de la banda baja (50 a 4000 Hz), de 4 kb/s

Capas siguientes: Agregan progresivas mejoras en la banda alta, 2 kb/s adicionales cada una

G.723.1

6.4 kb/s

- Utiliza un algoritmo MPC-MLQ (Multi-Pulse Maximum Likelihood Quantization), generando 24 bytes por cada ventana de 30 ms.

5.3 kb/s

- Utiliza ACELP (Algebraic Code Excited Linear Prediction), generando 20 bytes por cada ventana de 30 ms

El retardo total (latencia) es de 37.5 ms

- El algoritmo requiere de 7.5 msegundos de muestras adicionales (“look ahead”).

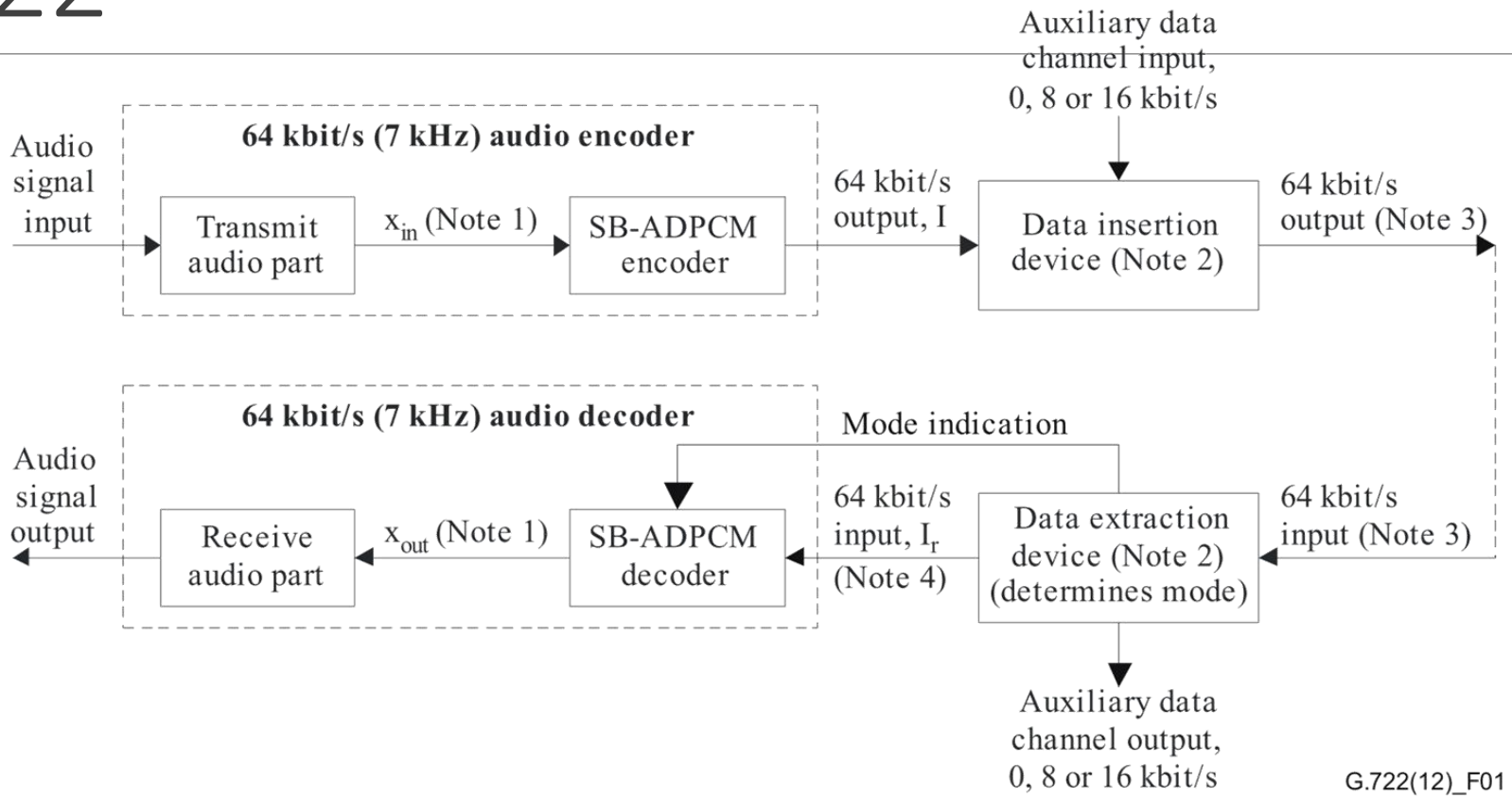
G.722 - 7 kHz audio-coding within 64 kbit/s

Codec de Banda Ancha

Utiliza técnicas de ADPCM, separando la señal en dos sub-componentes (banda baja y banda alta)

Opera en tres posibles modos, en 64, 56 o 48 kb/s

G.722



NOTE 1 – X_{in} and X_{out} are digital signals uniformly coded with 14 bits and 16 kHz sampling.
NOTE 2 – These devices are only necessary for applications requiring an auxiliary data channel within the 64 kbit/s.
NOTE 3 – Comprises 64, 56 or 48 kbit/s for audio coding and 0, 8 or 16 kbit/s for data.
NOTE 4 – 64 kbit/s signal comprising 64, 56 or 48 kbit/s for audio coding depending on the mode of operation.

Figure 1 – Simplified functional block diagram

G.722

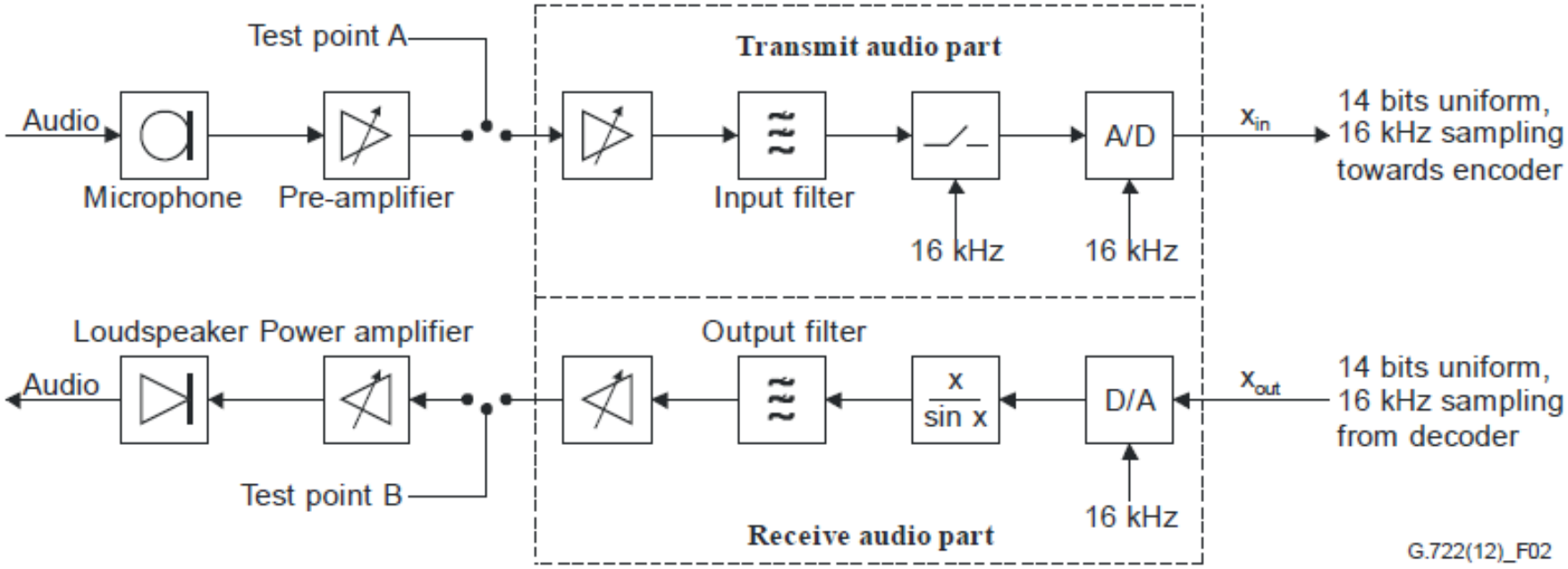


Figure 2 – Possible implementation of the audio parts

G.722

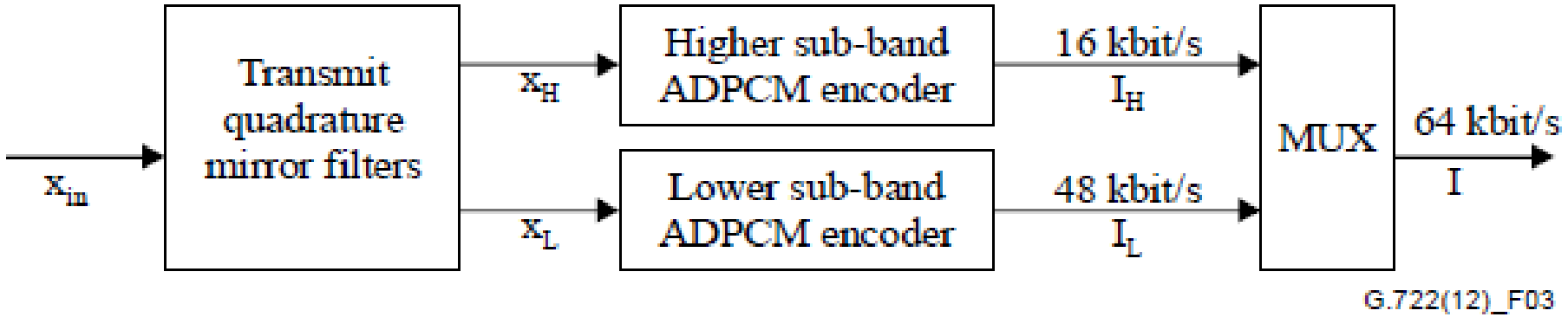


Figure 3 – Block diagram of the SB-ADPCM encoder

G.722

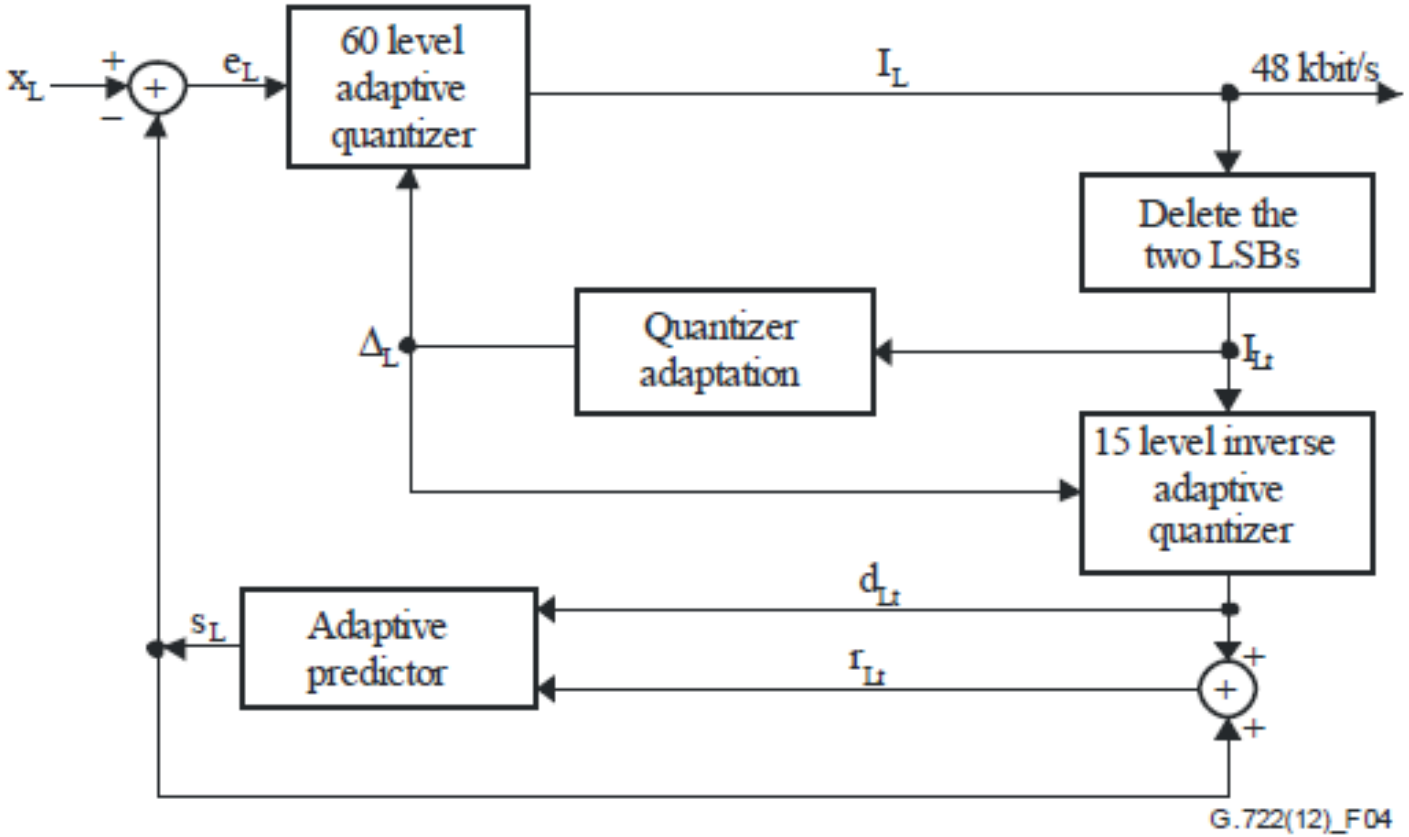


Figure 4 – Block diagram of the lower sub-band ADPCM encoder

G.722

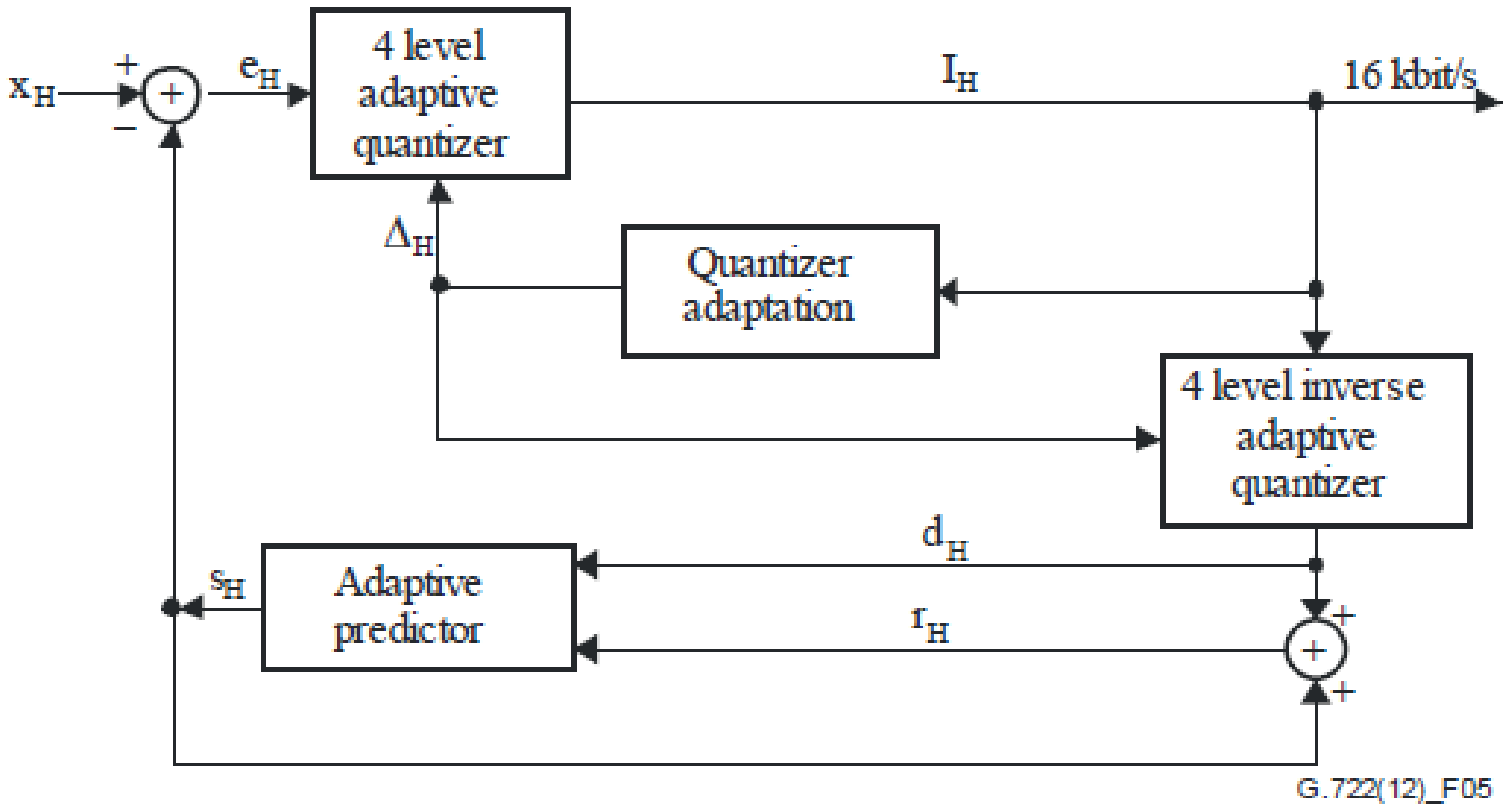


Figure 5 – Block diagram of the higher sub-band ADPCM encoder

G.722

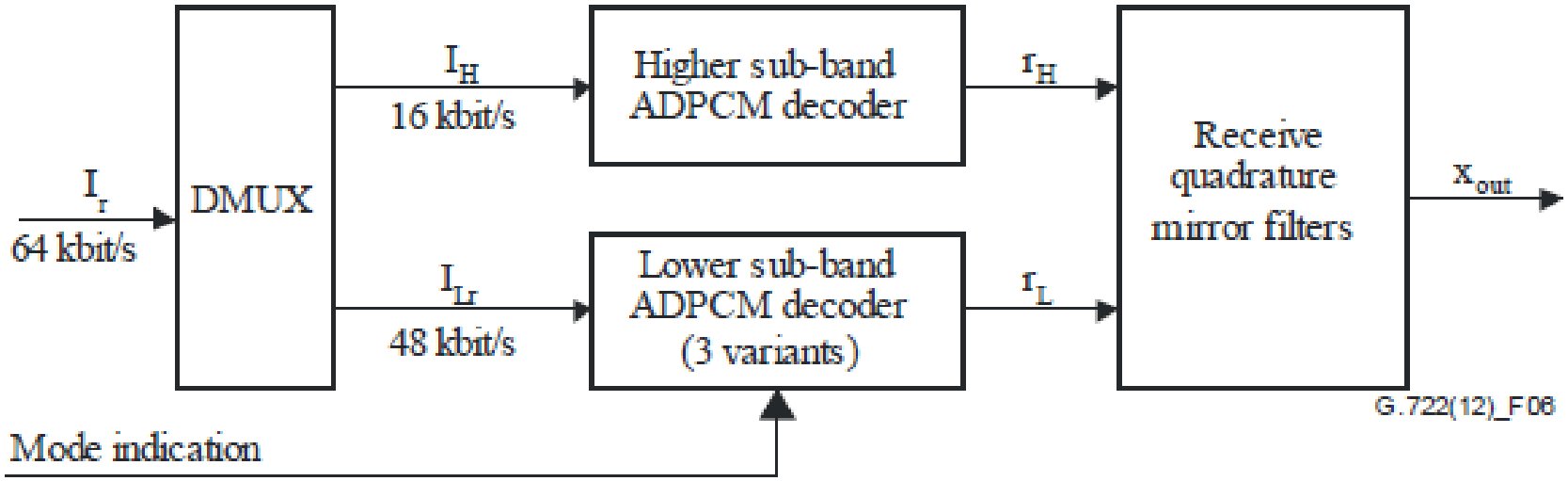


Figure 6 – Block diagram of the SB-ADPCM decoder

RTAudio (Real-time Audio)

Utiliza técnicas de codificación VBR (Variable Bit Rate)

- No todas las ventanas o cuadros de voz se codifican con la misma cantidad de bytes.

El retardo total (latencia) del algoritmo es menor a 40 ms

Nuevo “default” de Microsoft

- “RTAudio is the preferred Microsoft® Real-Time audio codec and is the default codec for Microsoft’s Unified Communications platforms” (1)

(1) http://en.wikipedia.org/wiki/RTAudio

AMR (Adaptive Multi Rate)

Utilizado típicamente en redes celulares GSM

Hace uso de tecnologías DTX (Discontinous Transmition), VAD (Voice Activity Detection) para detección de actividad vocal y CNG (Confort Noise Generation).

De forma similar a G.729, se basa en el modelo ACELP

- Ventanas de audio de 20 ms (160 muestras)
- Cada ventana de 20 ms es a su vez dividida en 4 sub-ventanas, de 5 ms (40 muestras) cada una.
- Pora cada ventana se extraen los parámetros LP del modelo CELP (los coeficientes de los filtros LP)
- Por cada sub-ventana se obtienen los índices de los “codebooks” fijos y adaptivos y las ganancias.

AMR (Adaptive Multi Rate)

Según la forma en que se cuanticen los parámetros (de acuerdo a cuantos bits se utilicen para cada parámetro) se obtienen tramas de 95, 103, 118, 134, 148, 159, 204 o 244 bits, las que corresponden a velocidades de transmisión que varían entre 4.75 y 12.2 kb/s.

AMR es licenciado

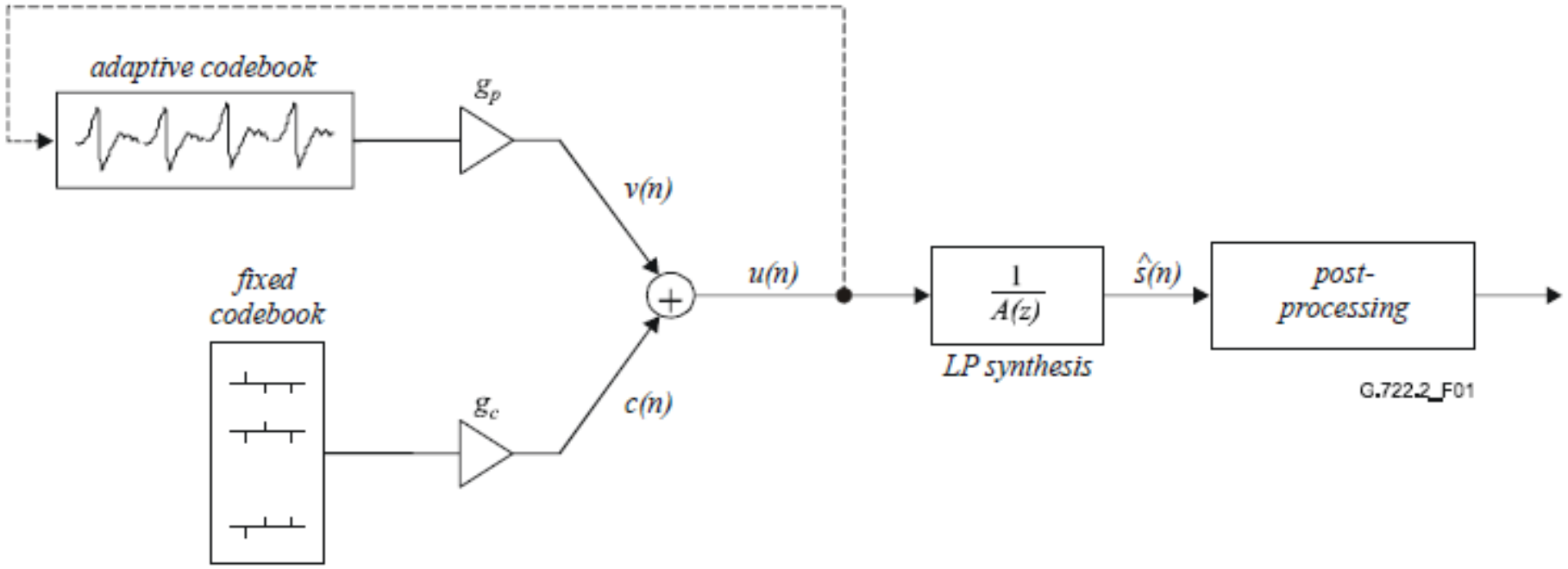
AMR-WB (G.722.2)

Codec de Banda Ancha (16 kHz), basado en un muestreo inicial de 14 bits por muestra

9 posibles velocidades entre 6.6 y 23.85 kb/s

Basado en CELP, utilizando un filtro de orden 16

AMR-WB (G.722.2)


adaptive codebook
g_p
$v(n)$
fixed codebook
g_c
$c(n)$
$u(n)$
$\frac{1}{A(z)}$
LP synthesis
$\hat{s}(n)$
post-processing
G.722.2_F01

SILK

Utilizado por Skype.

Ancho de banda variable, entre 6 a 40 kb/s, trabajando entre las bandas angostas (8 kHz) y las bandas super anchas (superwideband) (24 kHz)

Utiliza tramas de 20 ms y tiene un retardo de 25 ms.

Desde marzo de 2009 las licencias de uso de SILK son gratuitas.

En marzo de 2010 el codec fue enviado como borrador de RFC al IETF

SILK fue reemplazado por el codec OPUS, el que finalmente fue aceptado con el RFC 6716 en setiembre de 2012

En mayo de 2011, Skype fue comprado por Microsoft por 8.500 millones de dólares…

OPUS

Soporta VBR (Variable Bit Rate) y CBR (Constant Bit Rate).

- El “default” es VBR

	Ancho de banda del audio	**Bit rate (kb/s)**
NB (Narrowband)	4 kHz	8 – 12 kb/s
WB (Wide Band)	8 kHz	16 – 20 kb/s
FB (Full Band)	20 kHz	28 – 40 kb/s para voz 48 - 64 kb/s para música “mono” 64 – 128 kb/s para música estereo

OPUS

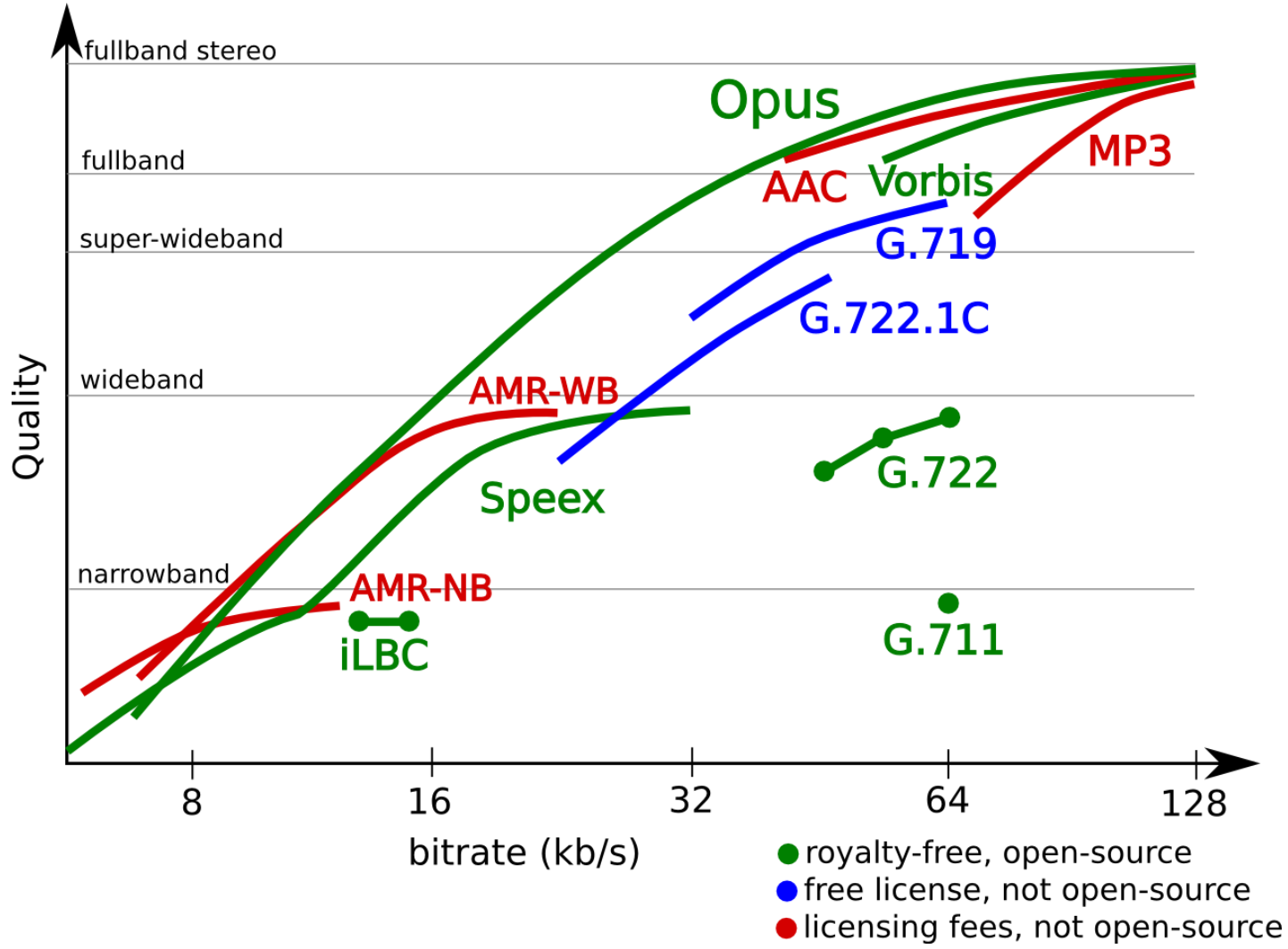
Utiliza “ventanas” de 2.5, 5, 10, 20, 40, o 60 ms.

- Típicamente se utiliza 20 ms

Permite combinar múltiples ventanas en paquetes de hasta 120 ms

“promete” mejor calidad, a igual bitrate, que otros codecs

OPUS



Tomado de: http://www.opus-codec.org/comparison/

EVS: Enhanced Voice Services

Diseñado para servicios de VoLTE (Voice over LTE)

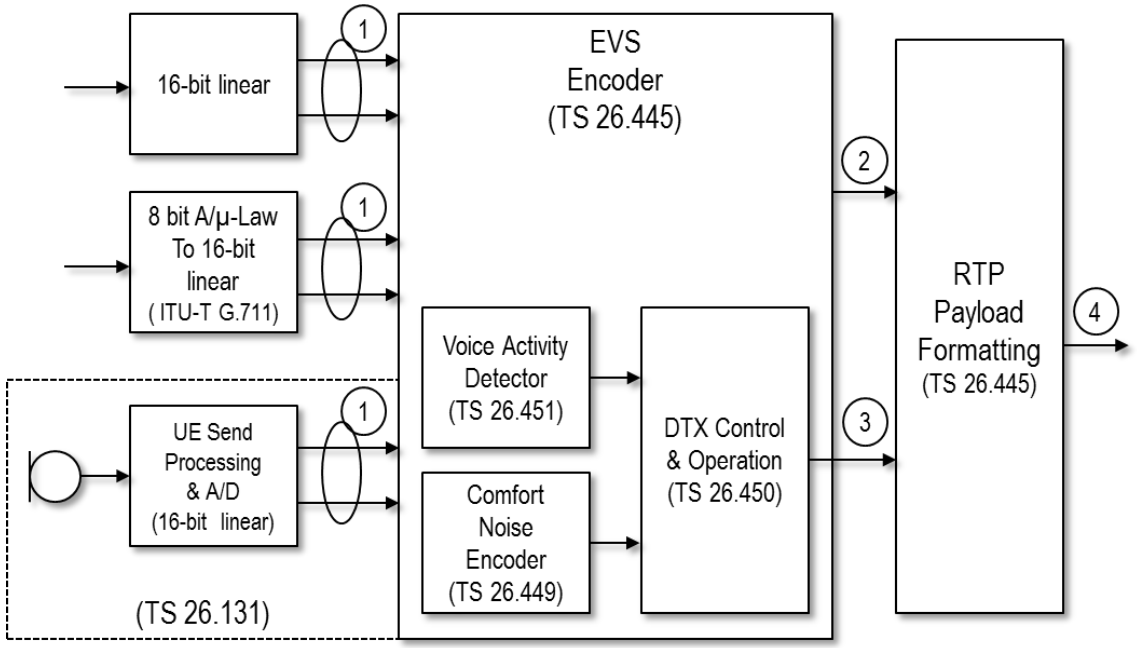
Es el primer códec desarrollado por 3GPP de banda completa (hasta 20 kHz)

Provee interoperabilidad con AMR-WB

Es de velocidad variable (VBR)

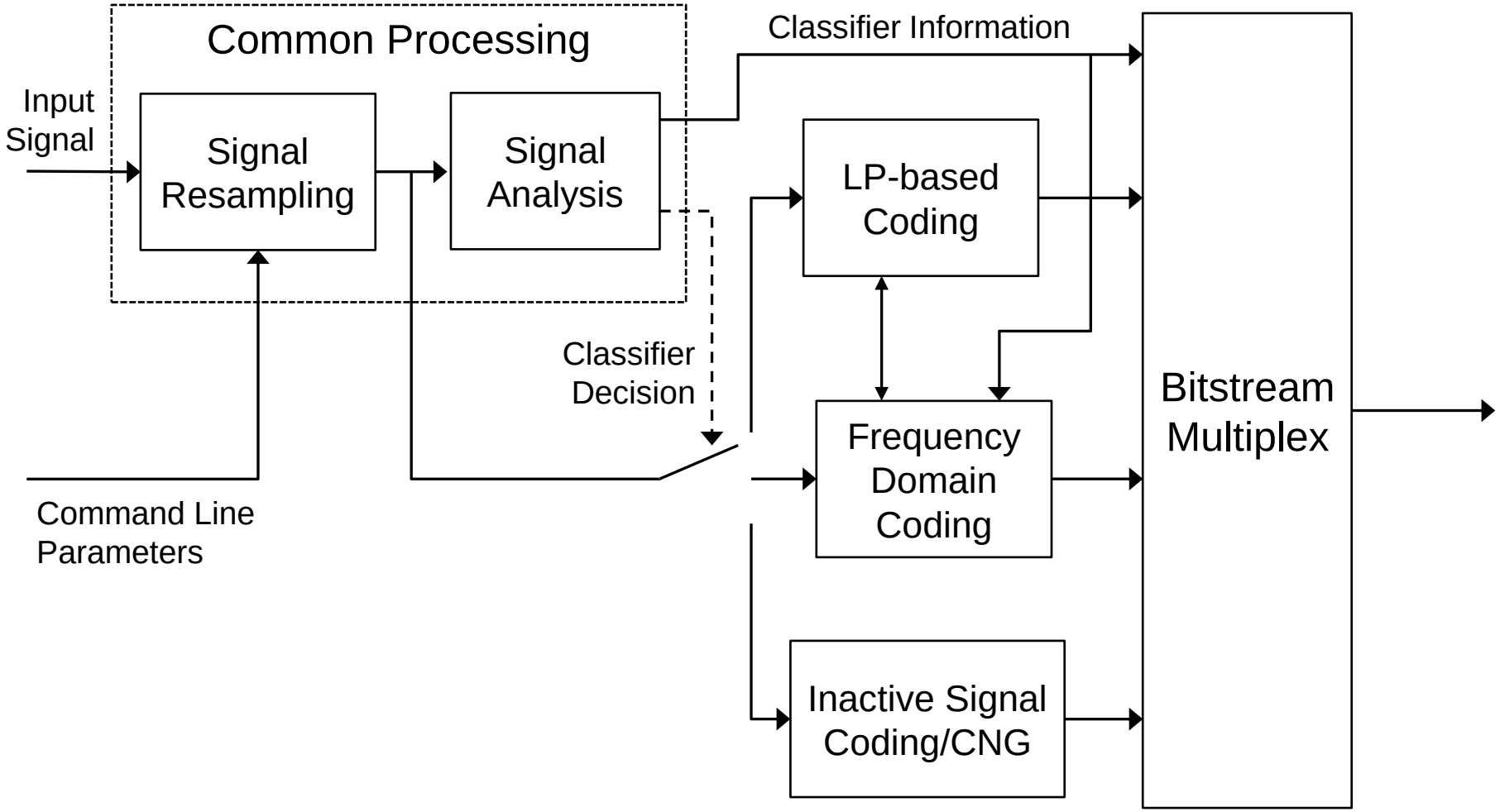
Bandwidth	Bit Rate (kbps)
Narrowband (NB)	5.9, 7.2, 8, 9.6, 13.2, 16.4, 24.4
Wideband (WB)	5.9, 7.2, 8, 9.6, 13.2, 16.4, 24.4, 32, 48, 64, 96, 128 (6.6 ~ 23.85 for AMR-WB IO)
Super-wideband (SWB)	9.6, 13.2, 16.4, 24.4, 32, 48, 64, 96, 128
Fullband (FB)	16.4, 24.4, 32, 48, 64, 96, 128

EVS: Enhanced Voice Services

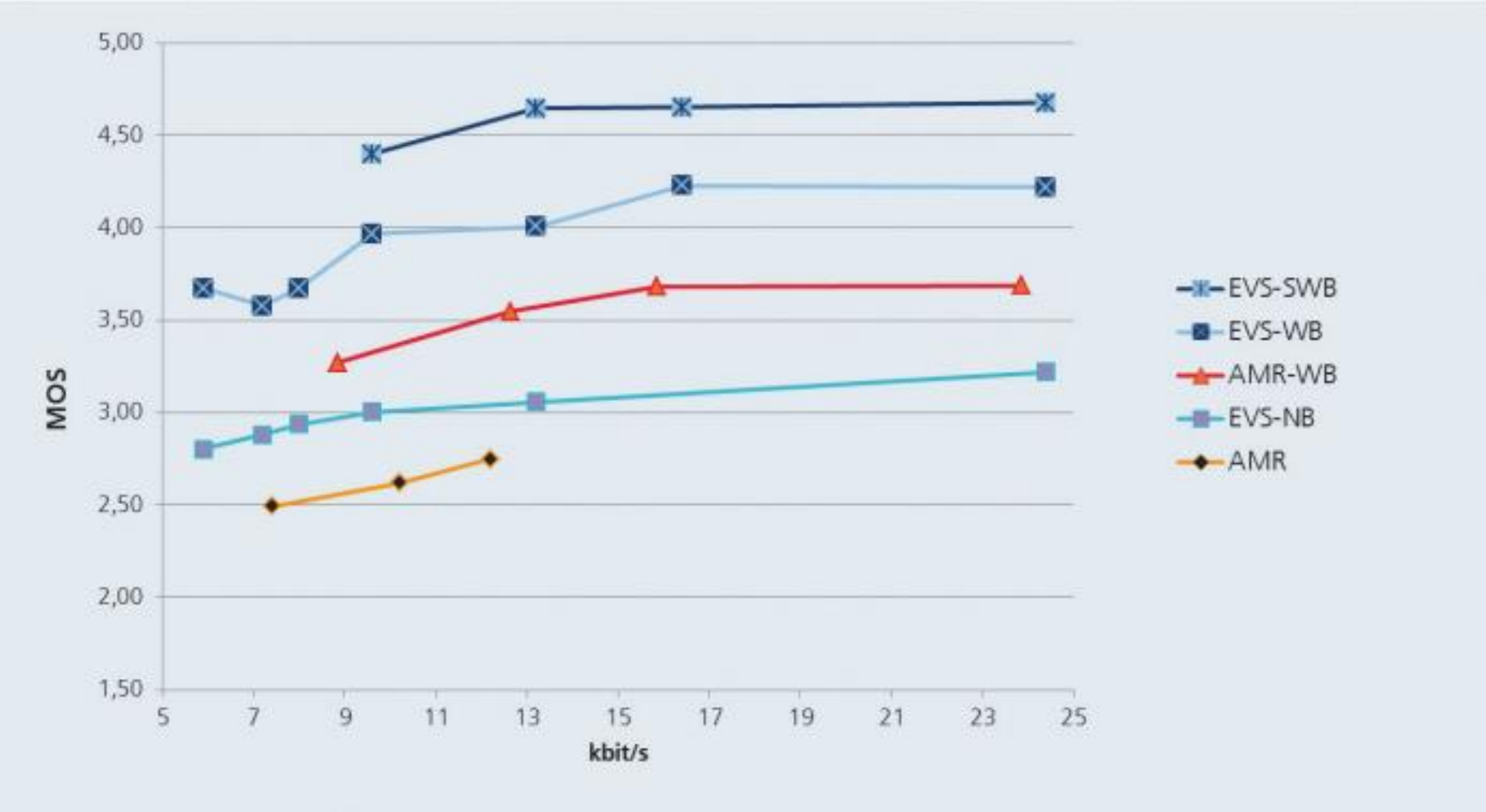


1. 16-bit Linear PCM Samples and Sample Rate (8, 16, 32 or 48 kHz)
2. Encoded audio frame, 50 frames/s, number of bits/frame depending on the EVS codec mode
3. Encoded Silence Descriptor frames (variable frame rate)
4. RTP Payload Packets

EVS: Encoder

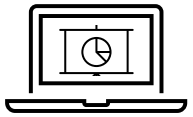

Common Processing
Input Signal
Signal Resampling
Signal Analysis
Classifier Information
LP-based Coding
Classifier Decision
Frequency Domain Coding
Bitstream Multiplex
Command Line Parameters
Inactive Signal Coding/CNG

Comparación de codecs de voz



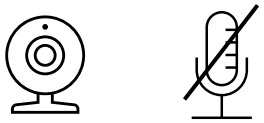
Tomado de The Future of Communication: Full-HD Voice powered by EVS and the AAC-ELD Family, Fraunhofer

Reglas para las sesiones remotas

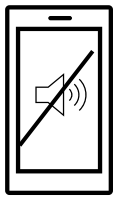


Utilizamos un PC o laptop, con pantalla que permita ver los detalles de las láminas.

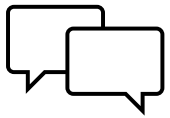
Nos conectamos desde un lugar silencioso, con buena conexión (cableada o WiFi cerca del Router).



Encendemos las cámaras. Mientras uno habla, los otros apagan sus micrófonos.

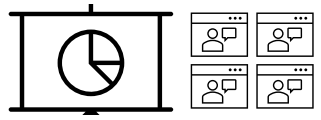


Nos quedamos conectados todo el tiempo… ¡a la sesión! Silenciamos los teléfonos celulares.



Si hay algún problema con el audio o el video, pueden usar el chat para avisar.

¡Todos participan! Pueden “levantar la mano”, o simplemente encender el micrófono e intervenir.

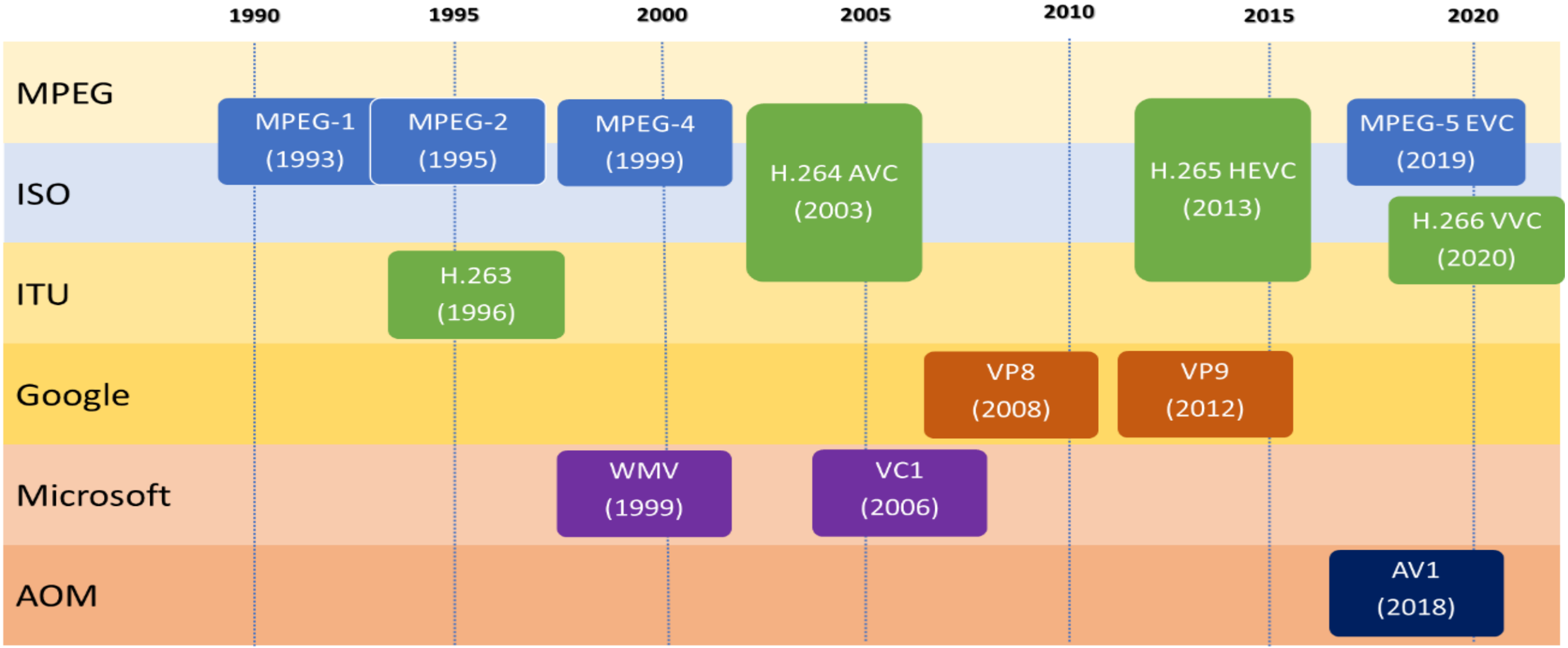


Pueden cambiar de foco entre la presentación y los participantes.

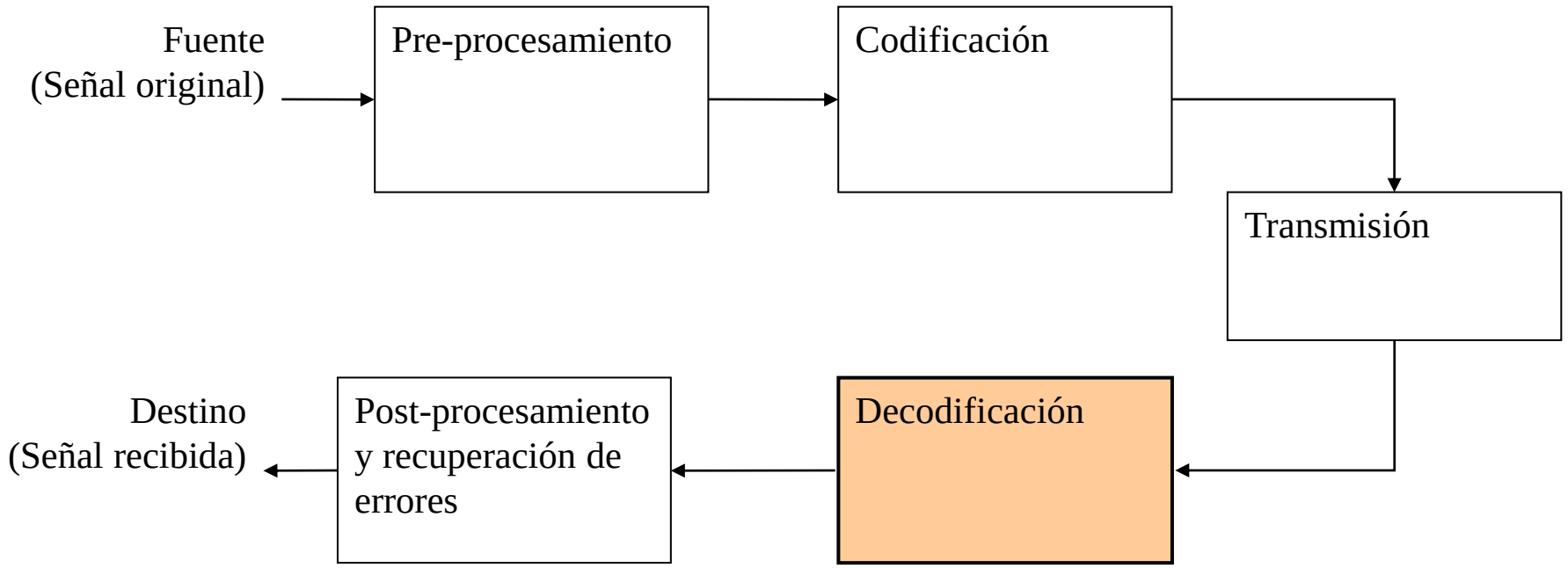
Digitalización y Codificación de Video

CODIFICACIÓN DE VOZ Y VIDEO

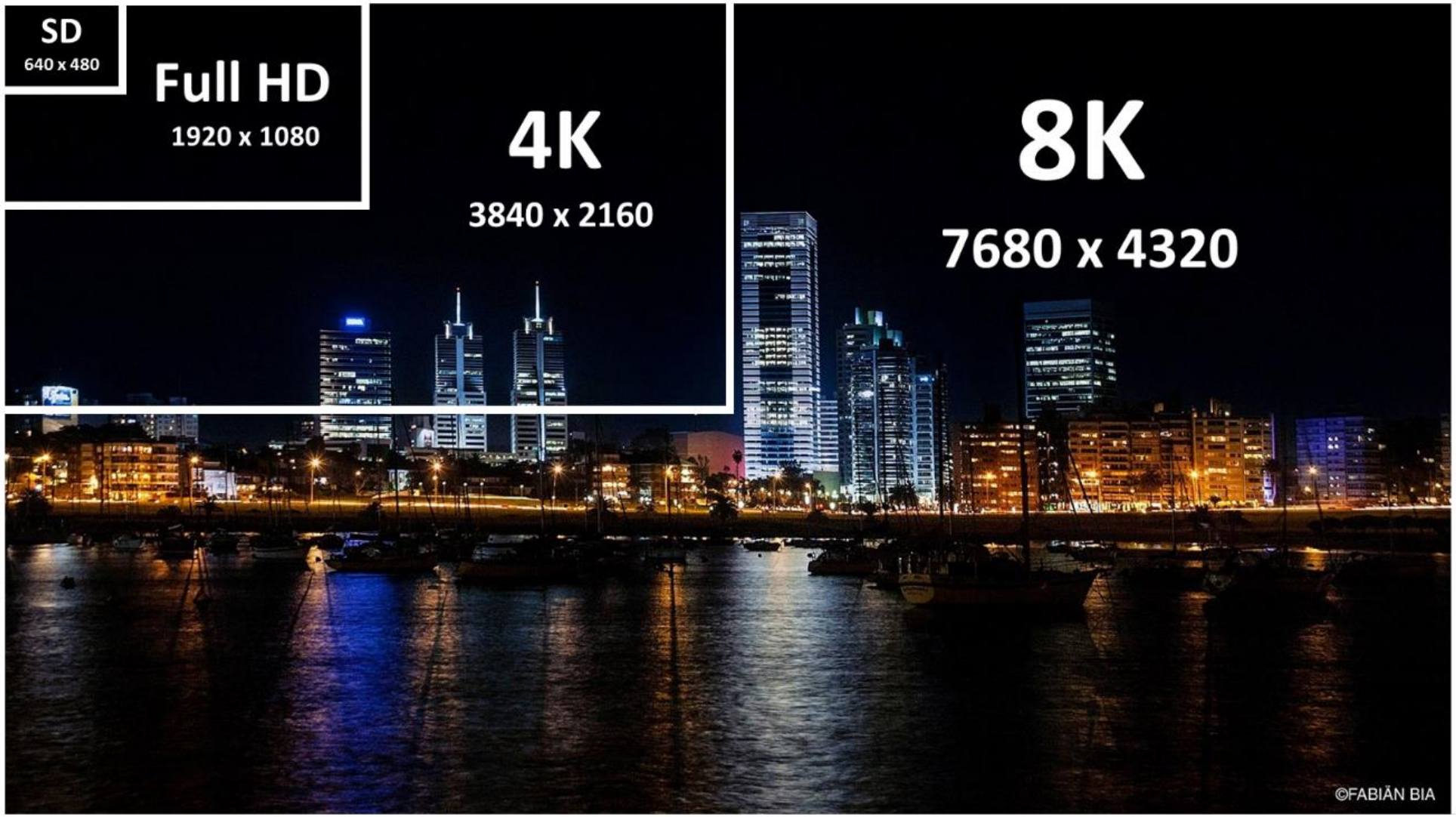
Evolución de la codificación de video



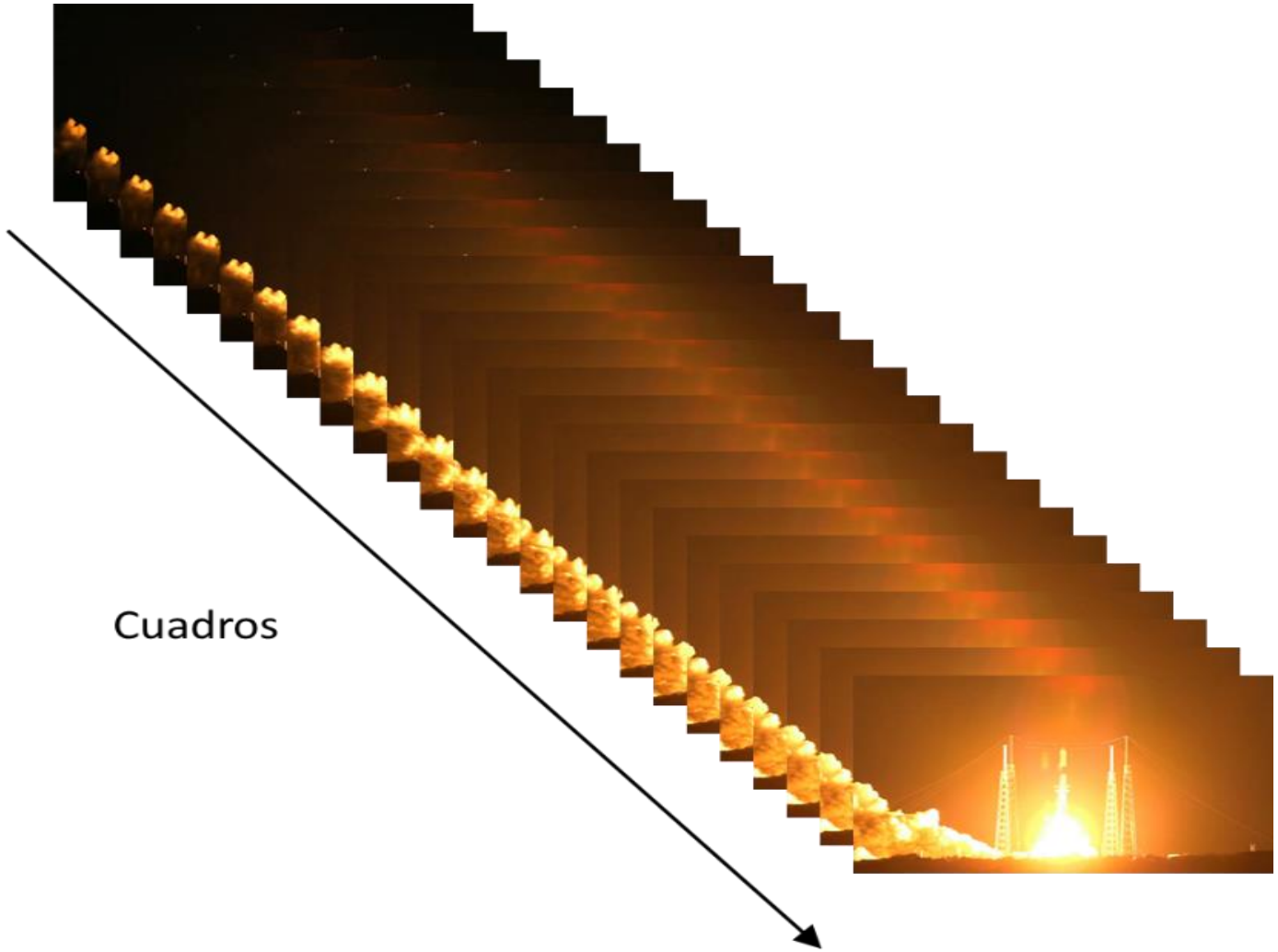
Estandarizaciones



Resolución


SD
640 x 480
Full HD
1920 x 1080
4K
3840 x 2160
8K
7680 x 4320
©FABIÃN BIA

Frame Rate



Entrelazado

Cuadro = Campo 1 + Campo 2

Filtros de “desentrelazado”

Imagen original, entrelazado

Imagen con filtro de desentrelazado

Ancho de banda

¿Cuánta información por segundo es necesario enviar para la transmisión de video?

- Cada píxel de cada imagen se puede asociar a 3 valores, correspondientes a la intensidad de los tres colores básicos: Rojo, Verde y Azul.
- Cada uno de estos valores se puede codificar con un mínimo de 8 bits (1 byte), lo que corresponde a 256 valores posibles. O sea, cada píxel se puede codificar con 3 bytes.

Ancho de banda

En una resolución HD, cada cuadro de imagen contiene 1920×1080 píxeles, lo que lleva a 6 220 880 bytes por cuadro (aproximadamente 6 MBytes por cuadro).

Considerando que en televisión digital es habitual enviar 30 cuadros por segundo (30 fps), la cantidad de información a transmitir llega a **1,4 Gbits/s**.

Un plan típico de acceso a internet hogareño por fibra óptica tiene 400 Mb/s de bajada.

¿¿Cómo podemos ver streamming en HD??

¿Cómo es posible bajar esta tasa de bits a valores más “razonables”?

Técnicas utilizadas para la digitalización del video

Submuestreo del color
- El sistema visual humano es más sensible a la luz que al color

Transformación
- Los valores relacionados a las muestras pueden ser transformados en otro conjunto de valores equivalentes, que representan la misma información de manera diferente
- En video se utiliza típicamente la “Transformada Discreta del Coseno” o DCT por sus siglas en inglés

Compresión
- En función de la “escala de cuantización” utilizada, el proceso puede reducir la cantidad de información, a costa de distorsionar la señal original

Predicción
- “Predecir” el valor de ciertas muestras en función de otras, de manera de poder enviar únicamente como información la diferencia

Codificación entrópica (Entropy Coding)
- Representa los valores cuantizados tomando ventaja de las frecuencias relativas con las que aparece cada símbolo
- Códigos de largo variable (o “VLC” por sus siglas en inglés)

Percepción de luz y color

Las células denominadas “conos” son los responsables de la visión del color dentro del ojo humano. Hay tres tipos de conos, sensibles a los colores rojo, verde y azul, respectivamente.

Las células denominadas “bastones” son las responsables de la visión de la luminosidad general (intensidad de la luz), y no son sensibles al color.

Uno ojo típico tiene del orden de **100 millones de bastones** y solo **6 millones de conos**.

Como consecuencia, el sistema visual humano es mucho más sensible a la intensidad de luz que a su color.

$$Y = Y_r R + Y_g G + Y_b B$$

$$Y_r = 0.299$$

$$Y_g = 0.587$$

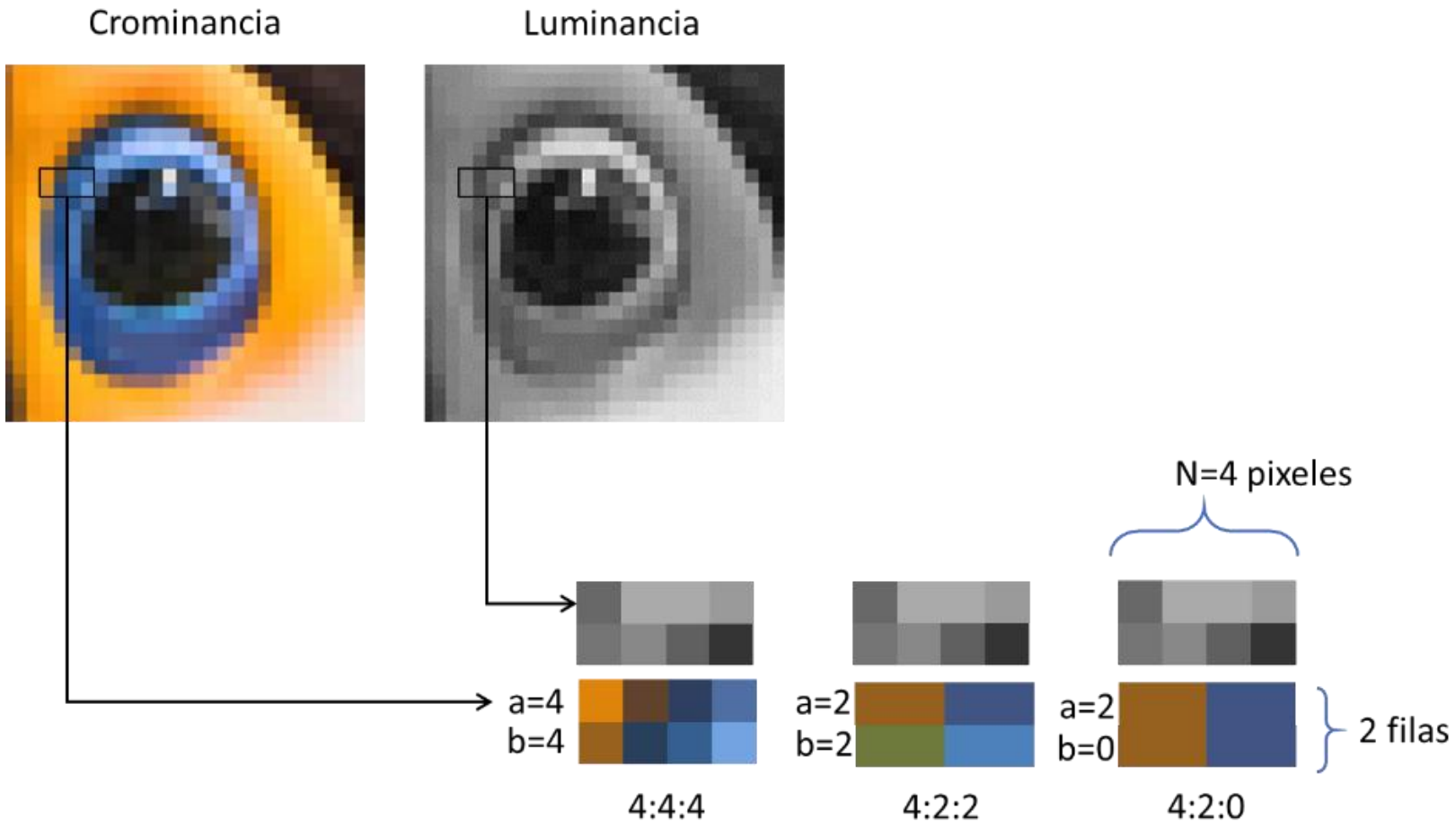
$$Y_b = 0.114$$

$$C_r = R - Y$$

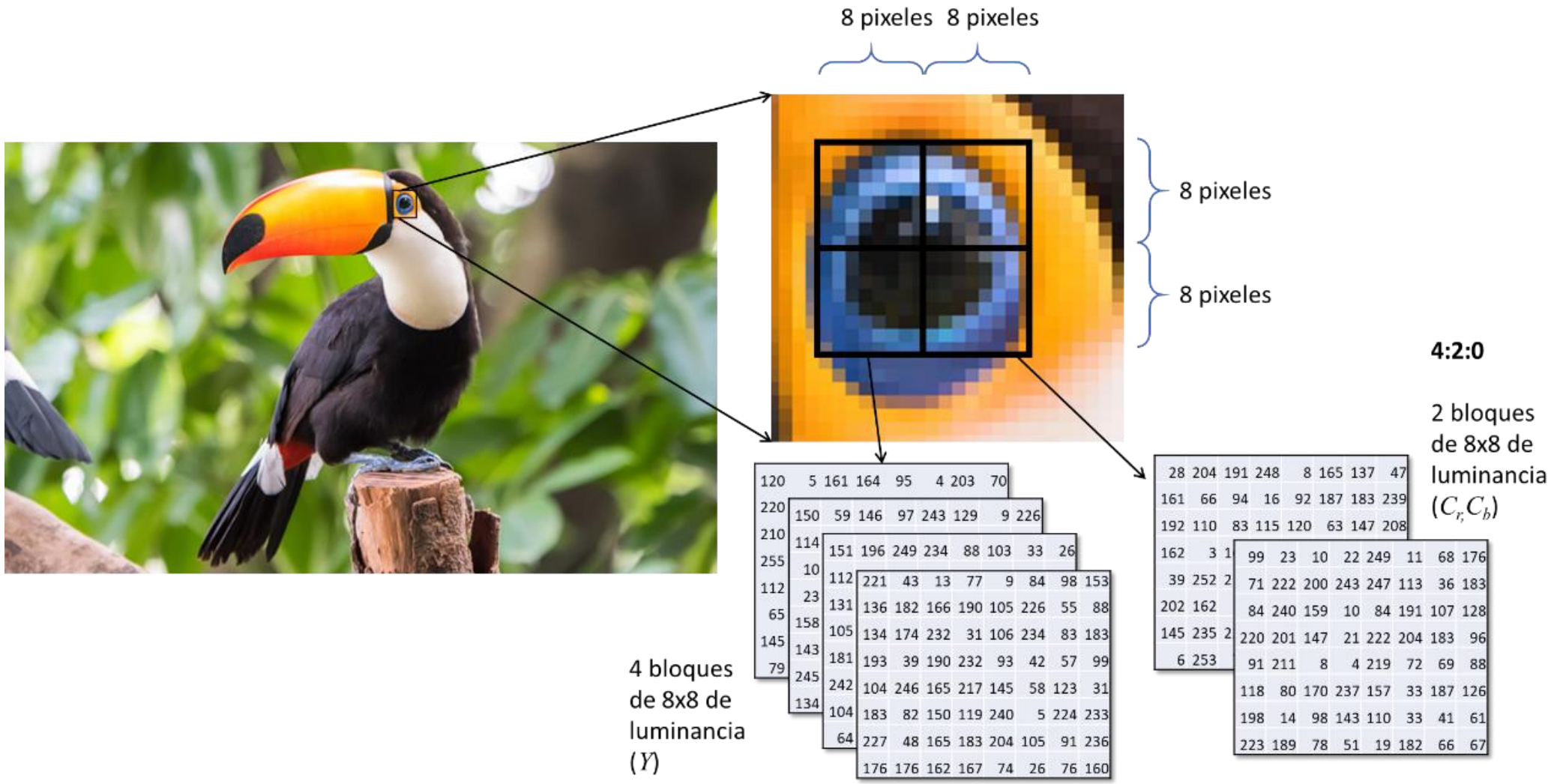
$$C_b = B - Y$$

Submuestreo del color

Formato: **N:a:b** (por ejemplo, 4:2:0).

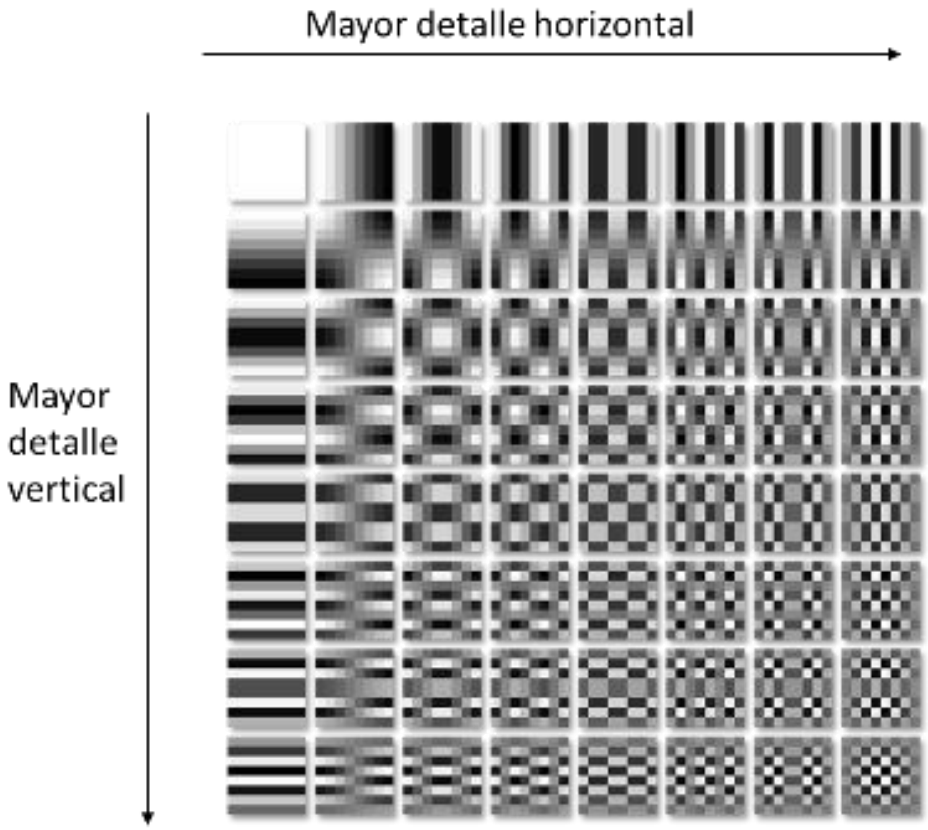


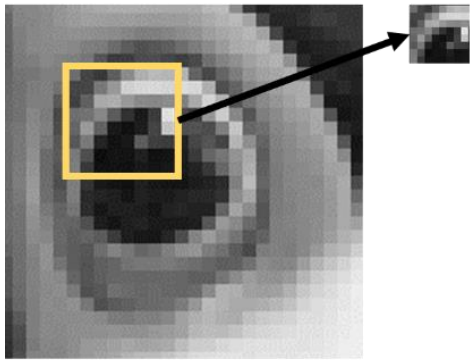
Submuestreo del color

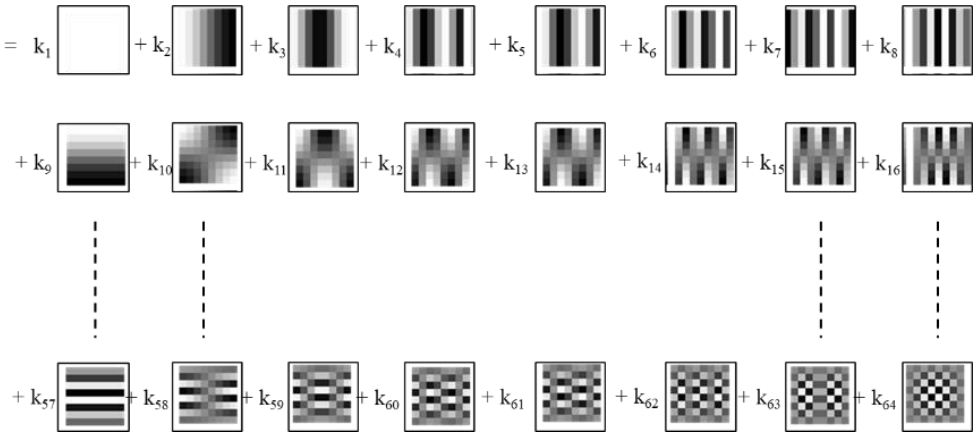


Codificación de Imágenes con DCT

Discrete Cosine Transformation (DCT)







Codificación de Imágenes con DCT

Discrete Cosine Transformation (DCT)

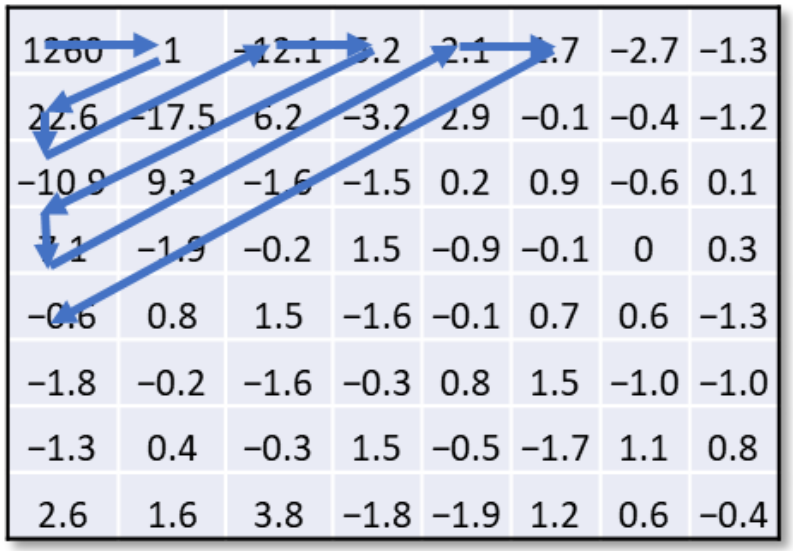
Valores originales

158	158	158	163	161	161	162	162
157	157	157	162	163	161	162	162
157	157	157	160	161	161	161	161
155	155	155	162	162	161	160	159
159	159	159	160	160	162	161	159
156	156	156	158	163	160	155	150
156	156	156	159	156	153	151	144
155	155	155	155	153	149	144	139

Valores DCT

1260	1	−12.1	5.2	2.1	1.7	−2.7	−1.3
22.6	−17.5	6.2	−3.2	2.9	−0.1	−0.4	−1.2
−10.9	9.3	−1.6	−1.5	0.2	0.9	−0.6	0.1
7.1	−1.9	−0.2	1.5	−0.9	−0.1	0	0.3
−0.6	0.8	1.5	−1.6	−0.1	0.7	0.6	−1.3
−1.8	−0.2	−1.6	−0.3	0.8	1.5	−1.0	−1.0
−1.3	0.4	−0.3	1.5	−0.5	−1.7	1.1	0.8
2.6	1.6	3.8	−1.8	−1.9	1.2	0.6	−0.4

Recorrido en zig-zag

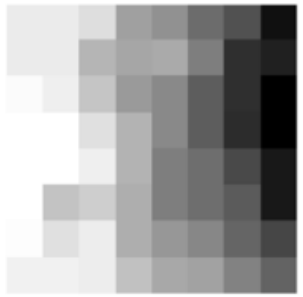


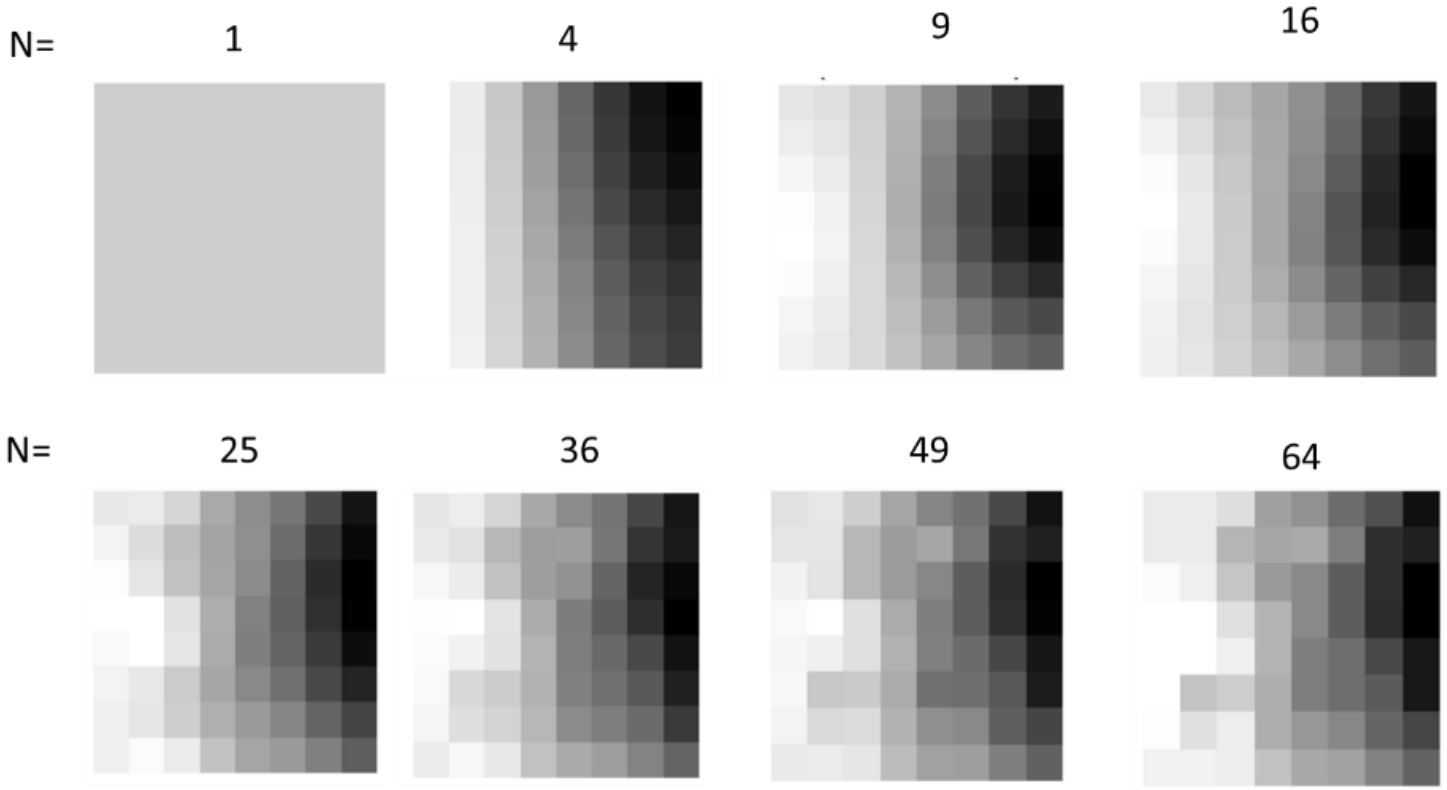
Compresión

Una manera de *comprimir* la información es quedarnos únicamente con los primeros coeficientes DCT, y descartar los últimos.

Se perderá algo de detalle del bloque, pero se podría tener una representación bastante buena

Bloque original





Compresión

Figura original

N= 1

4

9

16

N= 25

36

49

64

Compresión

Una técnica que permite comprimir aún más la información consiste en dividir los coeficientes DCT por un factor de escala llamado **matriz de cuantización** o **Quantization Matrix**.

Los valores de escalado pueden ser diferentes para cada coeficiente (de allí que se hable de una matriz).

Cada coeficiente de DCT se divide por un factor, y se trunca a valores enteros, que puedan ser fácilmente representados por bytes

Valores DCT

1260	1	−12.1	5.2	2.1	1.7	−2.7	−1.3
22.6	−17.5	6.2	−3.2	2.9	−0.1	−0.4	−1.2
−10.9	9.3	−1.6	−1.5	0.2	0.9	−0.6	0.1
7.1	−1.9	−0.2	1.5	−0.9	−0.1	0	0.3
−0.6	0.8	1.5	−1.6	−0.1	0.7	0.6	−1.3
−1.8	−0.2	−1.6	−0.3	0.8	1.5	−1.0	−1.0
−1.3	0.4	−0.3	1.5	−0.5	−1.7	1.1	0.8
2.6	1.6	3.8	−1.8	−1.9	1.2	0.6	−0.4

Valores escalados DCT

126	0	-1	1	0	0	0	0
2	-2	1	0	0	0	0	0
-1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

¡Notar que, en el último cuadro, solamente hay 10 valores diferentes de cero

Compresión

Ajustando el factor de escala, es posible reducir aún más la cantidad de valores, a costa de perder más detalles finos de la imagen.

Este factor de escala es conocido como el **Quantization Parameter (QP)**.

Cuanto más alto el **QP**, peor es la calidad de la imagen resultante (recordar que los valores se ***dividen*** entre el **QP**).

Codificación entrópica

Una vez que se obtienen los valores finales de los coeficientes DCT, se puede aplicar una técnica de **codificación entrópica** o **entropy coding**.

Esta técnica consiste en presentar los valores más frecuentes con pocos bits, y los valores menos frecuentes con más bits.

Dado que el cero es uno de los valores más frecuentes, bastaría representarlo con un solo bit 0, y utilizar más bits para otros valores.

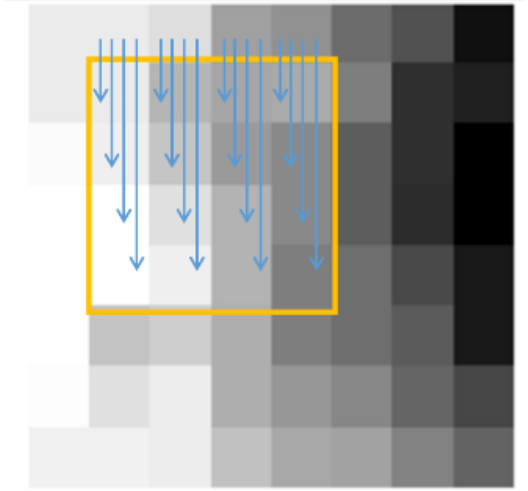
Combinando el *submuestreo de crominancia*, la *transformación DCT* y la *codificación entrópica*, se puede obtener una representación mucho más compacta de cada macrobloque de cada imagen del video.

Predicción dentro de cuadros

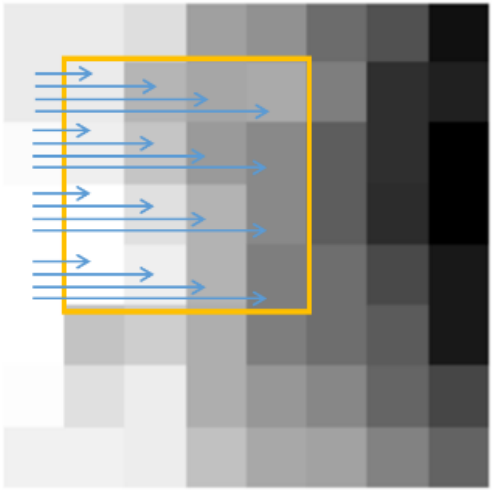
Las partes cercanas dentro de un mismo cuadro pueden ser muy parecidas.

Por ello es posible intentar predecir el valor de la luminancia y crominancia de cada píxel en función de los pixeles cercanos.

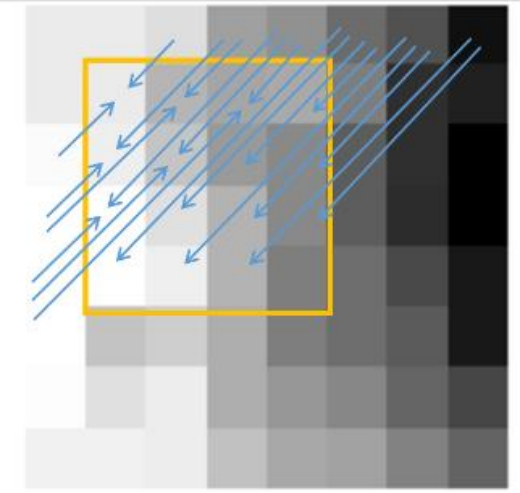
Predicción en base a los píxeles superiores

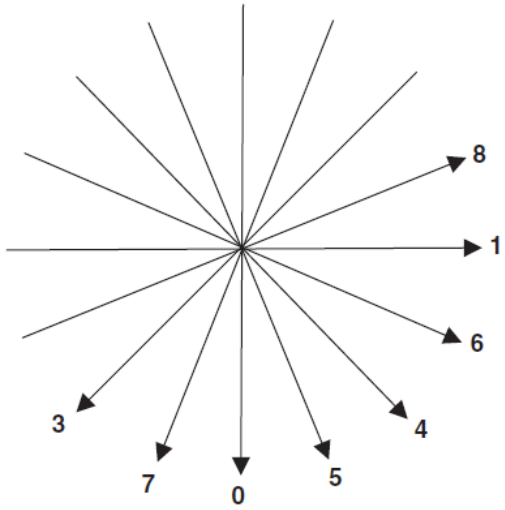


Predicción en base a los píxeles laterales



Predicción en base a los píxeles superiores y laterales





Predicción entre cuadros

Cada cuadro en un video es tomado a pocos milisegundos del cuadro anterior.

Por ejemplo, si se utiliza una tasa de 25 cuadros por segundo, cada cuadro se toma cada 40 milisegundos.

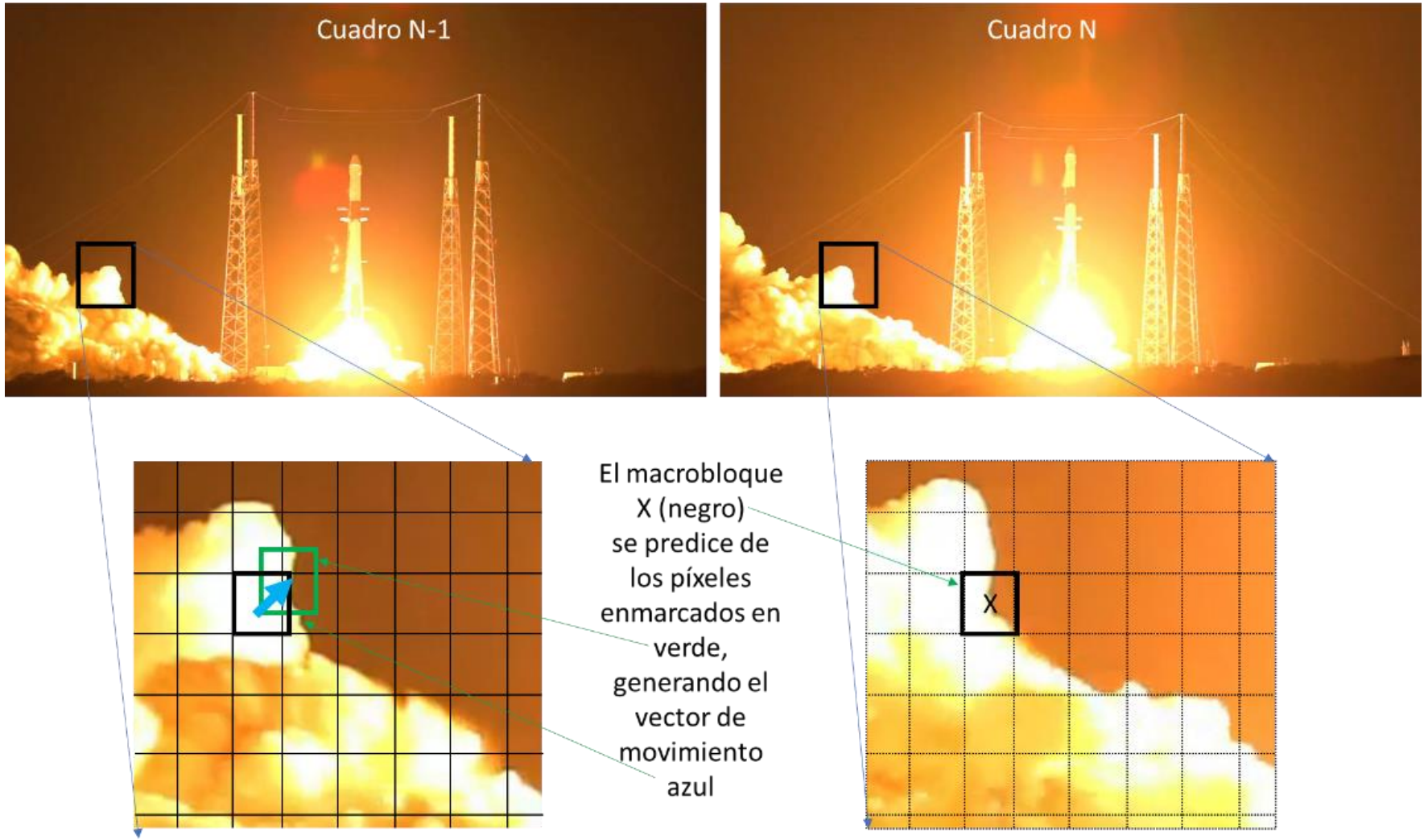
Es de esperar que gran parte de estos cuadros sean muy parecidos a los cuadros anteriores.

La información de cada cuadro puede estar altamente correlacionada con los cuadros anteriores y también con los cuadros futuros (es decir, los que serán tomados en los próximos instantes).

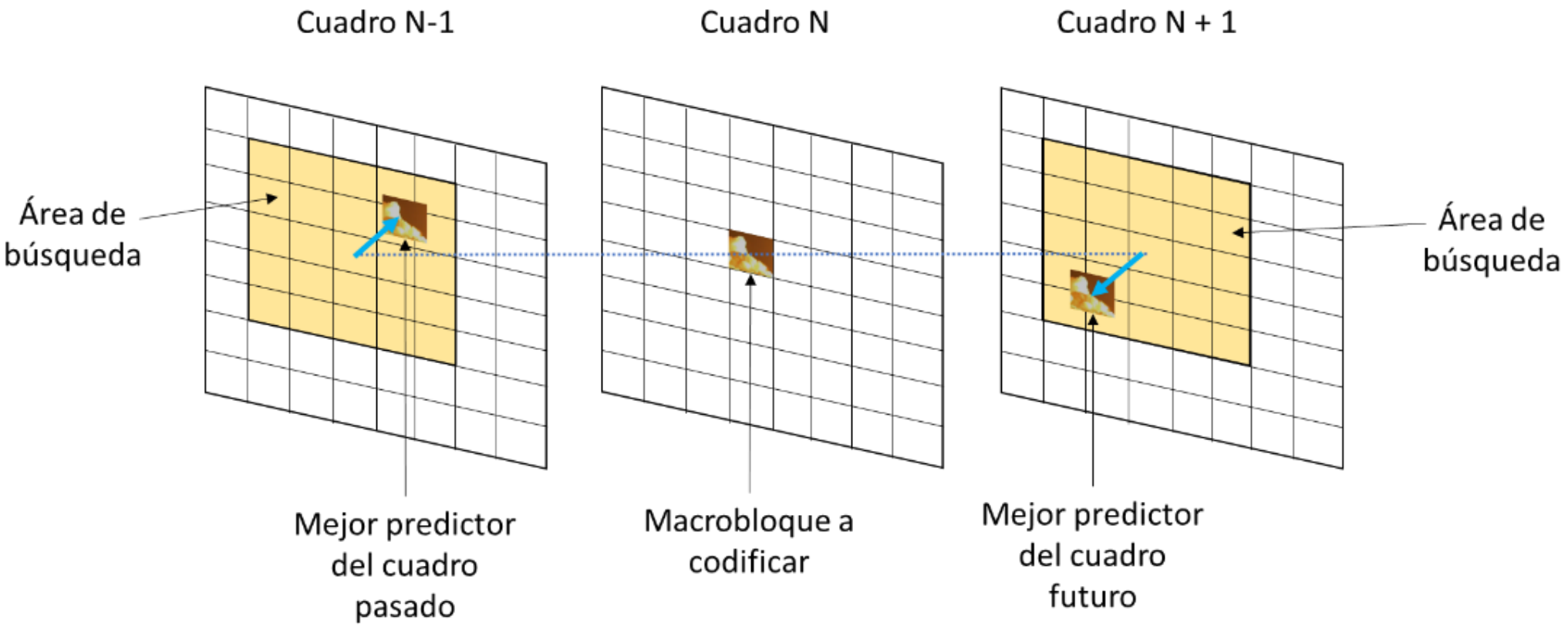
Esto permite utilizar técnicas que eliminen información redundante temporal:

- **Estimación del movimiento o Motion Estimation (ME)**
- **Compensación del movimiento o Motion Comensation (MC).**

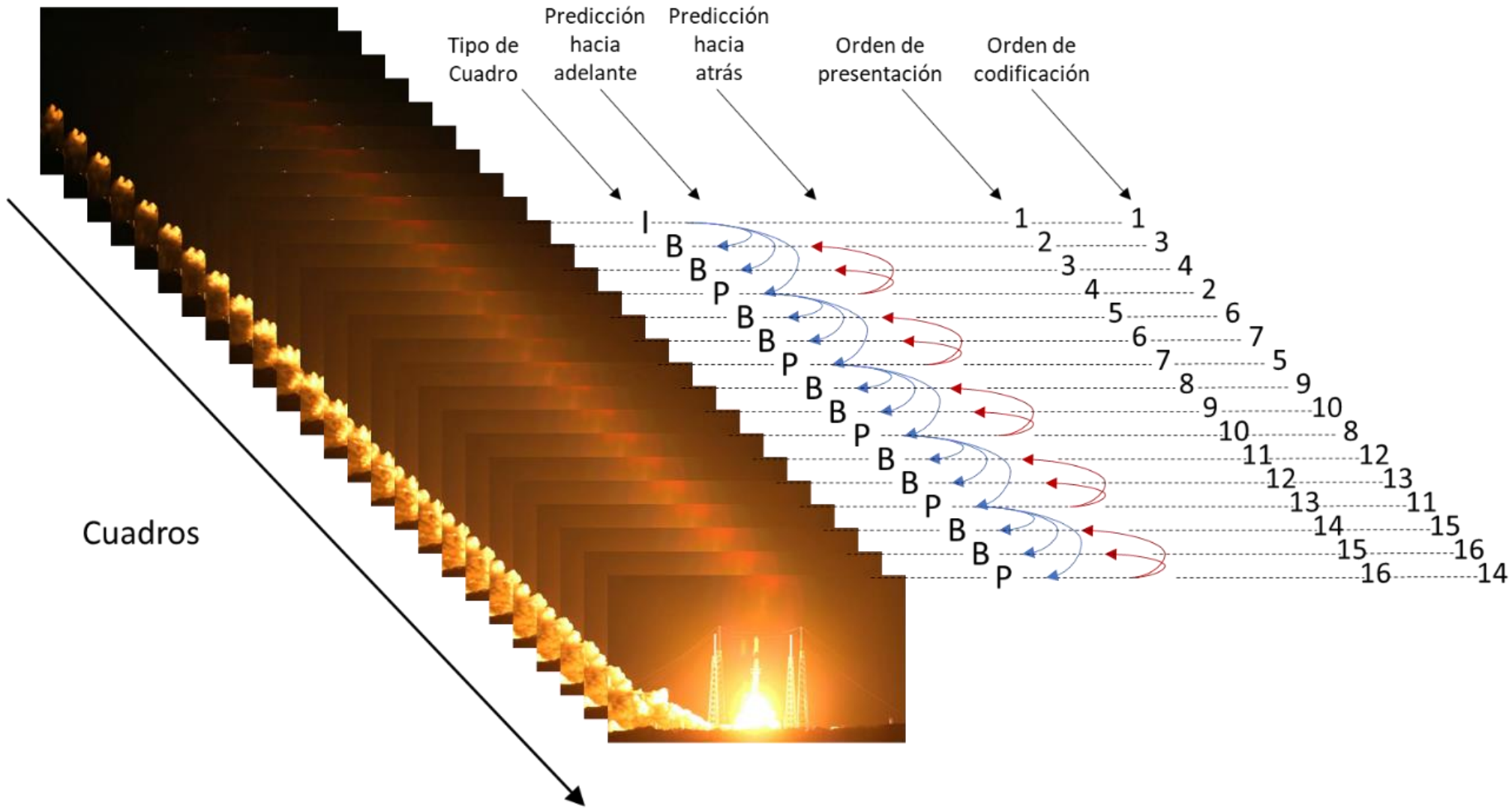
Predicción entre cuadros



Predicción entre cuadros



Group of Pictures (GoP)



MPEG-4 y H.264/AVC

MPEG-4

- Es la evolución de MPEG-1 y 2, y provee la tecnología base para la codificación en base a contenidos, y su almacenamiento, transmisión y manipulación
- Puede codificar múltiples “Objetos de video” (MVO – Multiple Video Objects)

H.264/MPEG-4 Part 10

- JVT/H.26L/AVC (Advanced Video Coding) o H.264/AVC
- Con AVC, para una misma calidad de video, se logran mejoras en el ancho de banda requerido de aproximadamente un 50% respecto estándares anteriores
- Ampliamente utilizado (por ejemplo, el que utiliza la TV Digital abierta)

H.264/SVC y MVC

SVC: “Scalable Video Coding” (Anexo G, 2007)

- Permite la construcción de sub-flujos de datos dentro de un flujo principal.
- El flujo principal o “capa base” (base layer) puede ser decodificado por cualquier equipo que soporte H.264/AVC, aunque no soporte SVC.
- Los flujos adicionales pueden contener información adicional del flujo, brindando mayor definición.

MVC: “Multiview Video Coding” (Anexo H, 2009)

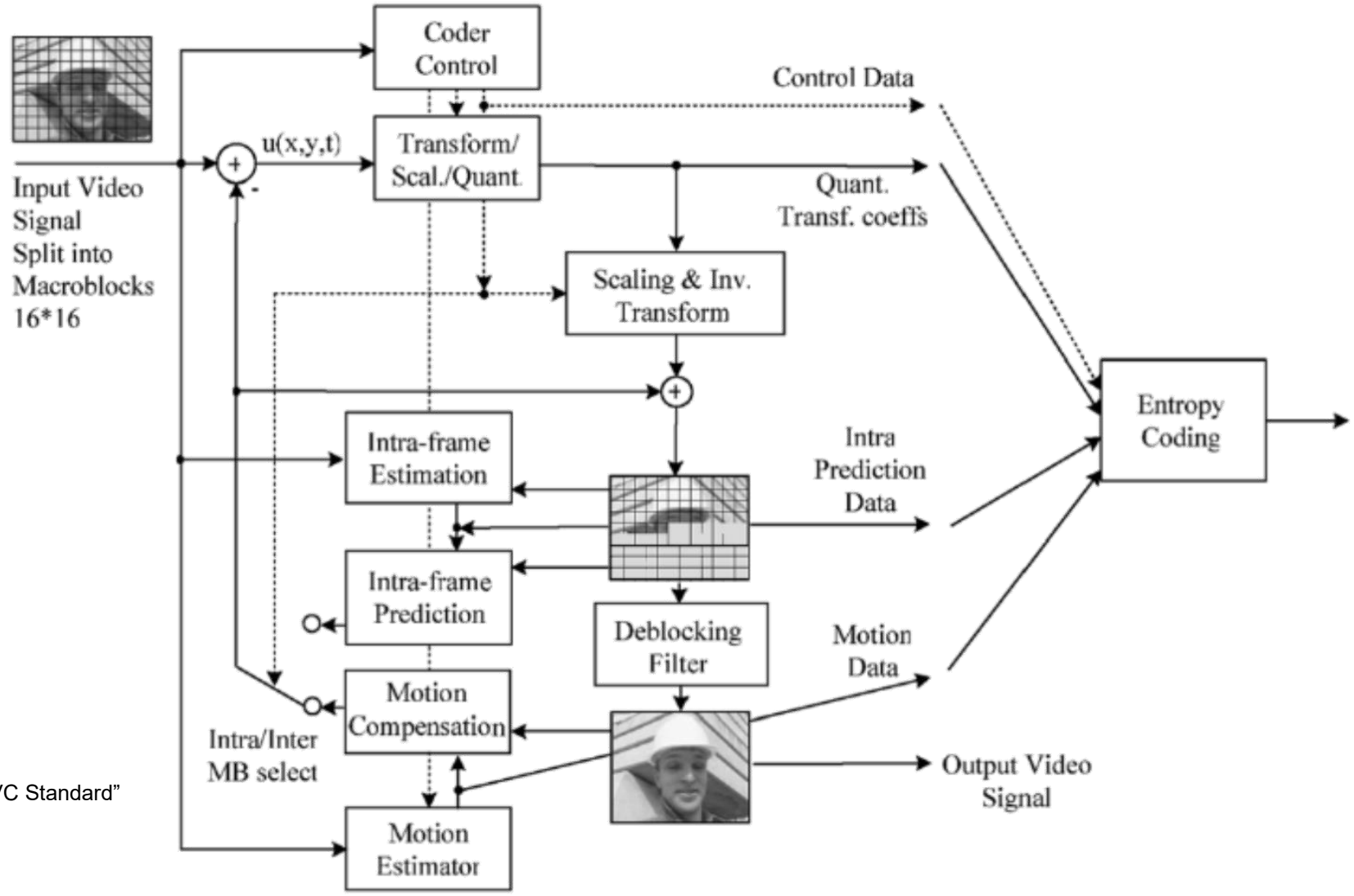
- Diferentes flujos representan diferentes visiones de la misma escena (por ejemplo, para 3D)

Perfiles y niveles

En H.264 se establecen “Perfiles” y “Niveles”

- Baseline Profile (BP)
- Main Profile (MP)
- High Profile (HiP)
- Otros (en total hay 17 perfiles!)

Codificador H.264



Tomada de: “Video Compression – From Concepts to the H.264/AVC Standard”
Gary J. Sullivan, Thomas Wiegand
Proceedings of the IEEEIssue 1, pp. 18 - 31, Jan 2005

H.265

Estandarizado por ITU en 2013:

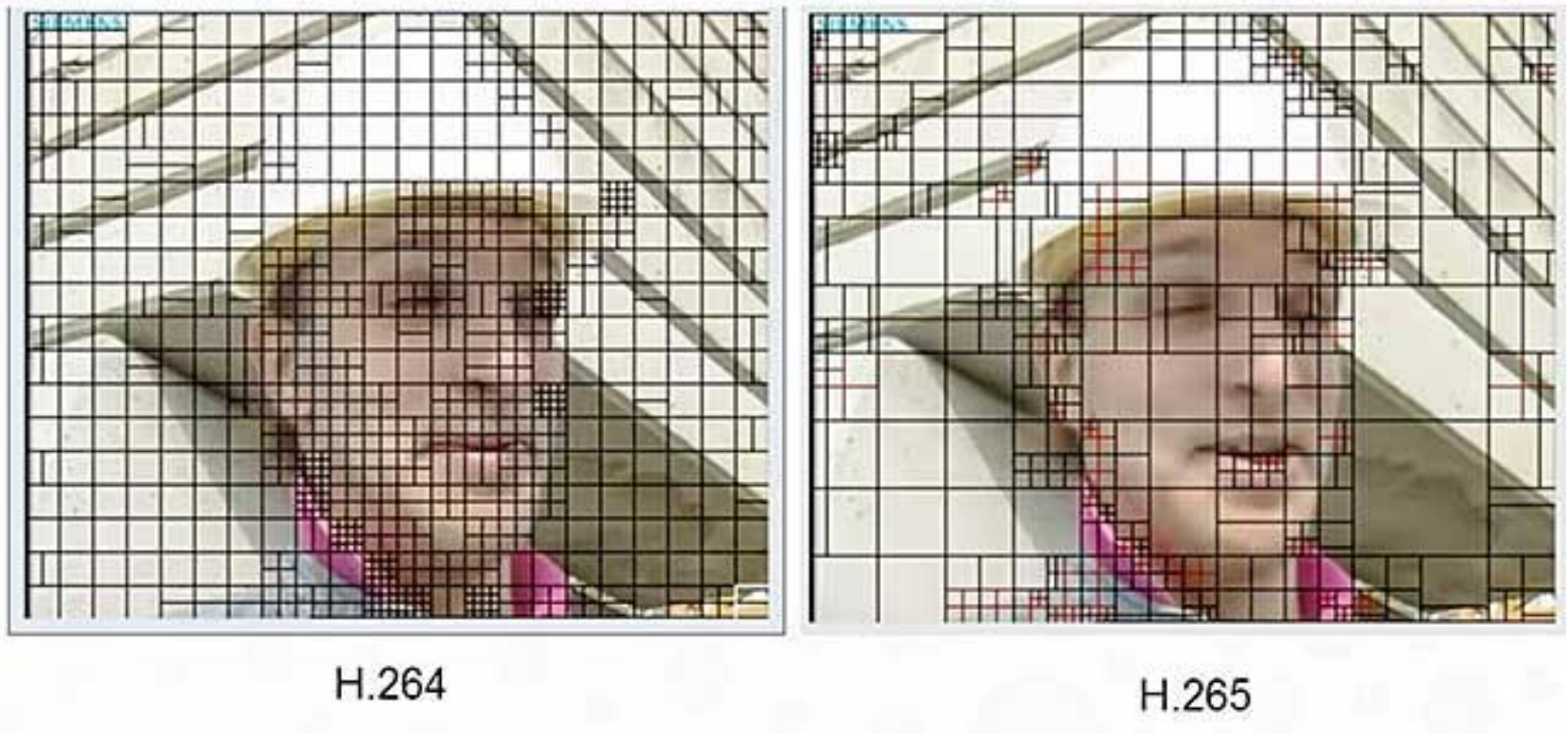
H.265 o **MPEG-H Parte2** o **High Efficiency Video Coding** (**HEVC)**

- Versión 1 (Abril 2013)
- Versión 2 (Octubre 2014), agrega un gran número de “perfiles”
- Versión 3 (Abril 2015), agrega un perfil para 3D
- Versión 4 (Diciembre 2016), agrega “Screen Content Coding (SCC)”
- Versión 5 (Febrero 2018)

Reduce aproximadamente a la mitad el bitrate de H.264 para obtener la misma calidad

H.265

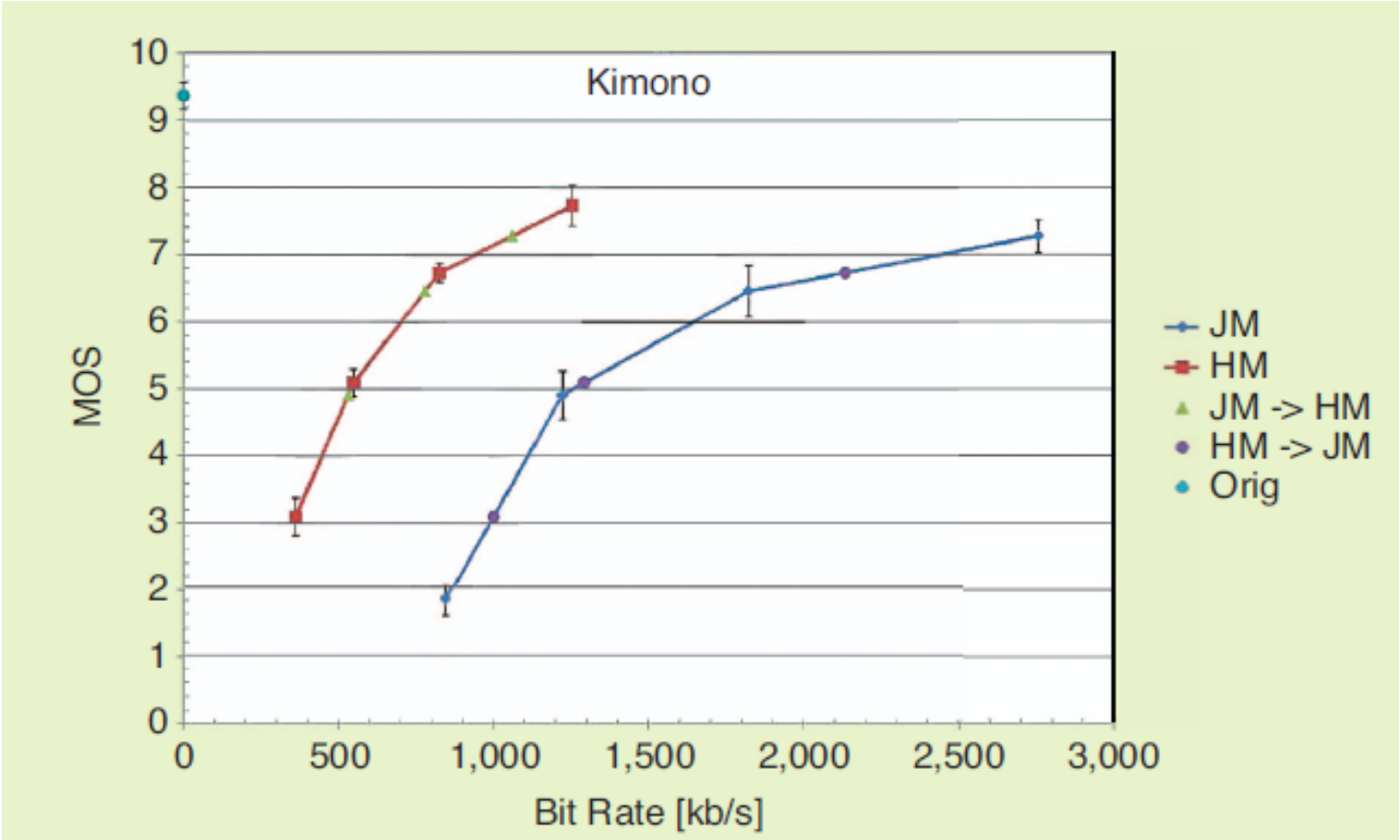
H.264 Macroblocks y H.265 Coding Tree Units – CTU



Tomada de:
What Is HEVC (H.265)?, Jan Ozer, February 2013
http://www.streamingmedia.com/Articles/Editorial/What-Is-.../What-Is-HEVC-(H.265)-87765.aspx

H.265

Bitrate vs Calidad, para H.264/AVC y H.265

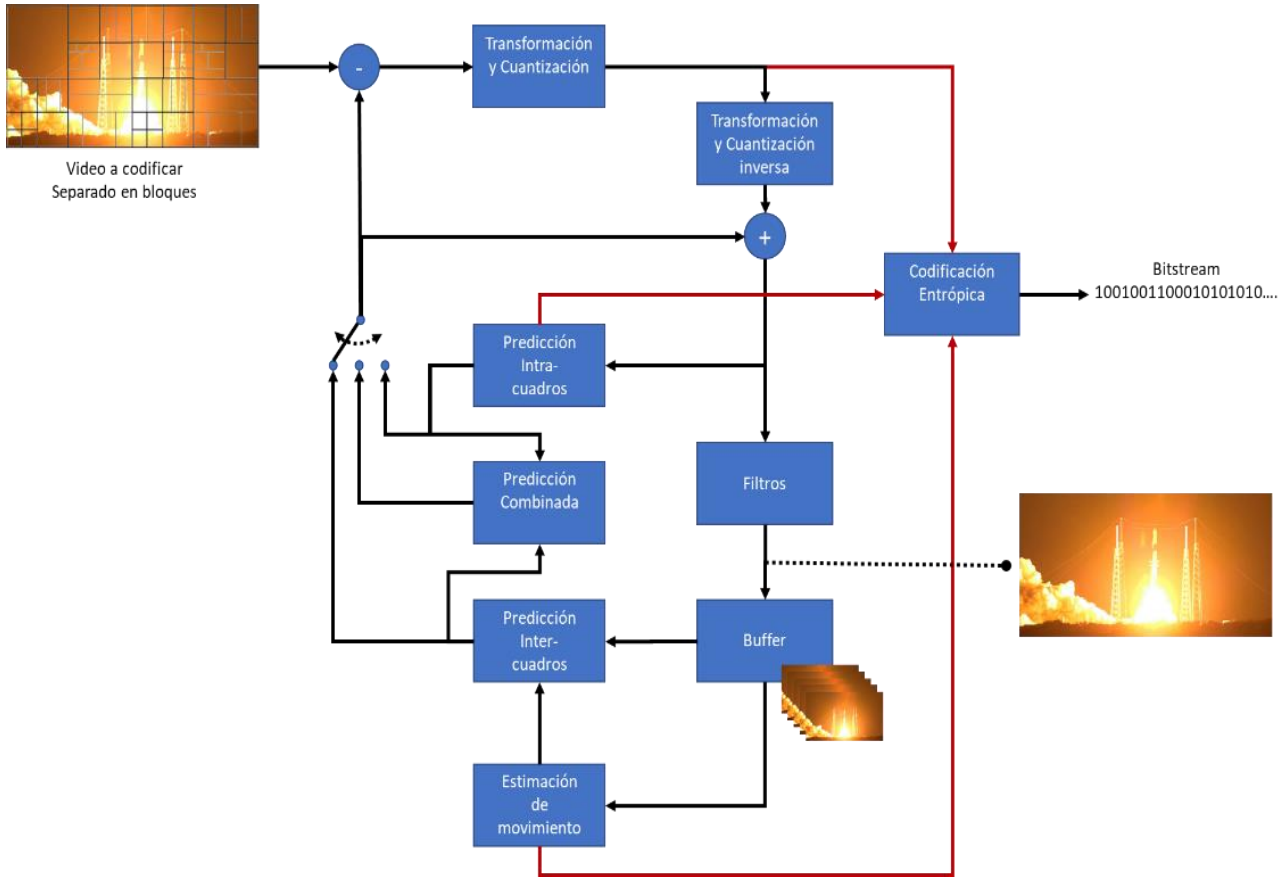


Tomado de: “High Efficiency Video Coding: The Next Frontier in Video Compression”, Jens-Rainer Ohm and Gary J. Sullivan, IEEE SIGNAL PROCESSING MAGAZINE, Jan 2013

H.266

Estandarizado por ITU en 2020: **Versatile Video Coding (VVC):**

Las técnicas utilizadas en el nuevo codificador VVC son básicamente las mismas que sus predecesores, pero *refinadas*.



Muchas Gracias!

CODIFICACIÓN DE VOZ Y VIDEO