

Calidad de voz y video

Curso Multimedia sobre IP 2017



Calidad de voz y video

Introducción





- ¿Qué es la “Calidad de la Experiencia”?

Según la RAE:

CALIDAD

- Propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor
- Superioridad o excelencia
- Adecuación de un producto o servicio a las características especificadas

EXPERIENCIA

- Hecho de haber sentido, conocido o presenciado alguien algo
- Conocimiento de la vida adquirido por las circunstancias o situaciones vividas



- ¿Qué es la “Calidad de la Experiencia”?



Según la ITU:

CALIDAD de la EXPERIENCIA (QoE)

- Es el grado de deleite o molestia del usuario de una aplicación o servicio

NOTA: “Reconociendo la investigación en curso sobre este tema, se trata de una definición de trabajo, que se espera evolucione en el tiempo”

ITU-T P.10/G.100 Amendment 5, julio 2016



- ¿Qué factores influyen en la QoE?
 - Tipo y características de la aplicación o servicio
 - Contexto de uso
 - Expectativas del usuario
 - Antecedentes culturales del usuario
 - Factores socio-económicos
 - Perfiles psicológicos del usuario
 - Estado emocional del usuario
 - Otros factores



ITU-T P.10/G.100 Amendment 5, julio 2016





- Evaluación de QoE

- Es el proceso de **medir** o **estimar** la QoE para un conjunto de usuarios de una aplicación o un servicio con un procedimiento específico y considerando los factores influyentes (posiblemente controlados, medidos o simplemente recopilados y reportados).
- El resultado del proceso puede ser un valor escalar, una representación multidimensional de los resultados, y/o descriptores verbales.
- Todas las evaluaciones de QoE deben ir acompañadas de la descripción de los factores de influencia que se incluyen.
- La evaluación de la QoE puede describirse como exhaustiva cuando incluye muchos de los factores específicos, por ejemplo la mayoría de los factores conocidos.
- Una evaluación limitada de QoE puede incluir sólo uno o un pequeño número de factores.

ITU-T P.10/G.100 Amendment 5, julio 2016

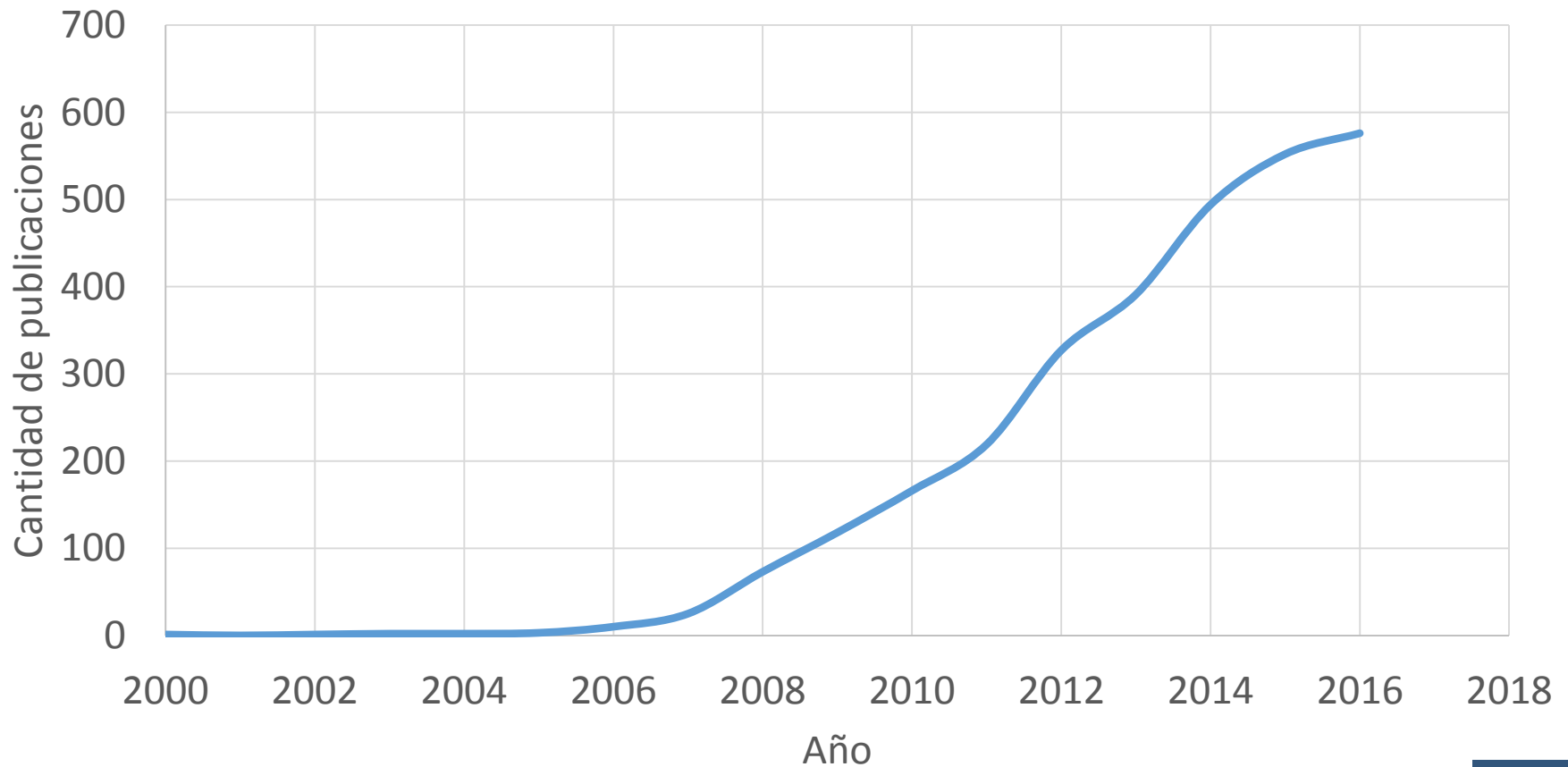


- ¿Cómo se **mide** la QoE?
 - Hay que preguntarle a los usuarios!
 - Se realizan “pruebas subjetivas” y se toman promedios
 - Puede ser complejo y tedioso!

- ¿Cómo se puede **estimar** la QoE?
 - Con el desarrollo de algoritmos que tratan de “emular” o “predecir” lo que opinaría el promedio de los usuarios
 - Se basan en diversos aspectos “medibles”



Cantidad de publicaciones en IEEE con el término "Quality of Experience"

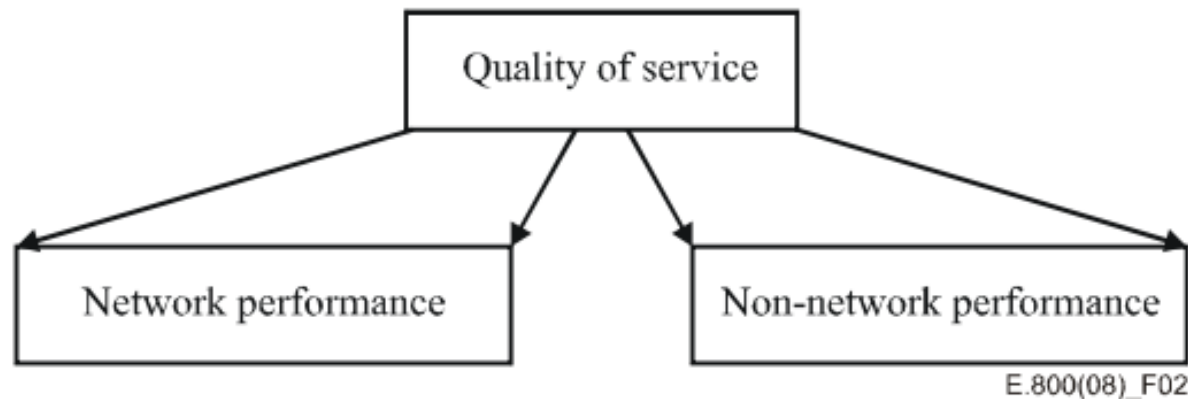


- ¿Qué es la Calidad de Servicio?

Según la ITU:



La totalidad de las características de un servicio de telecomunicaciones que determinan su capacidad para satisfacer las necesidades explícitas e implícitas del usuario del servicio.



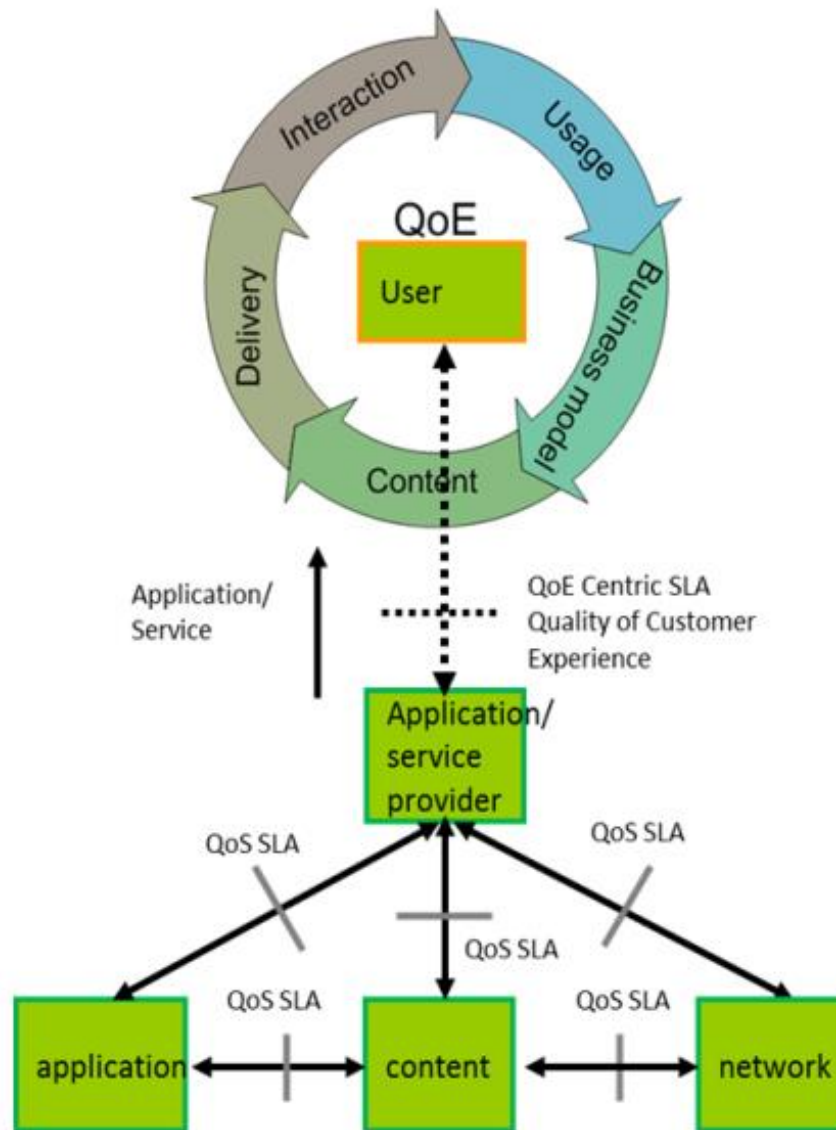
ITU-T E.800, setiembre 2008



- Ejemplo de aspectos de QoS relacionados al desempeño de la red:
 - Bitrate
 - Latencia
 - Tasa de errores
 - ...
- Ejemplo de aspectos de QoS independientes del desempeño de la red:
 - Tiempo de aprovisionamiento del servicio
 - Tiempo de reparación ante fallas
 - Tiempo de resolución de quejas
 - Esquema tarifario
 - ...



QoE vs QoS



Tomado de: Qualinet White Paper on Definitions of Quality of Experience, Output from the fifth Qualinet meeting, Novi Sad, March 12, 2013



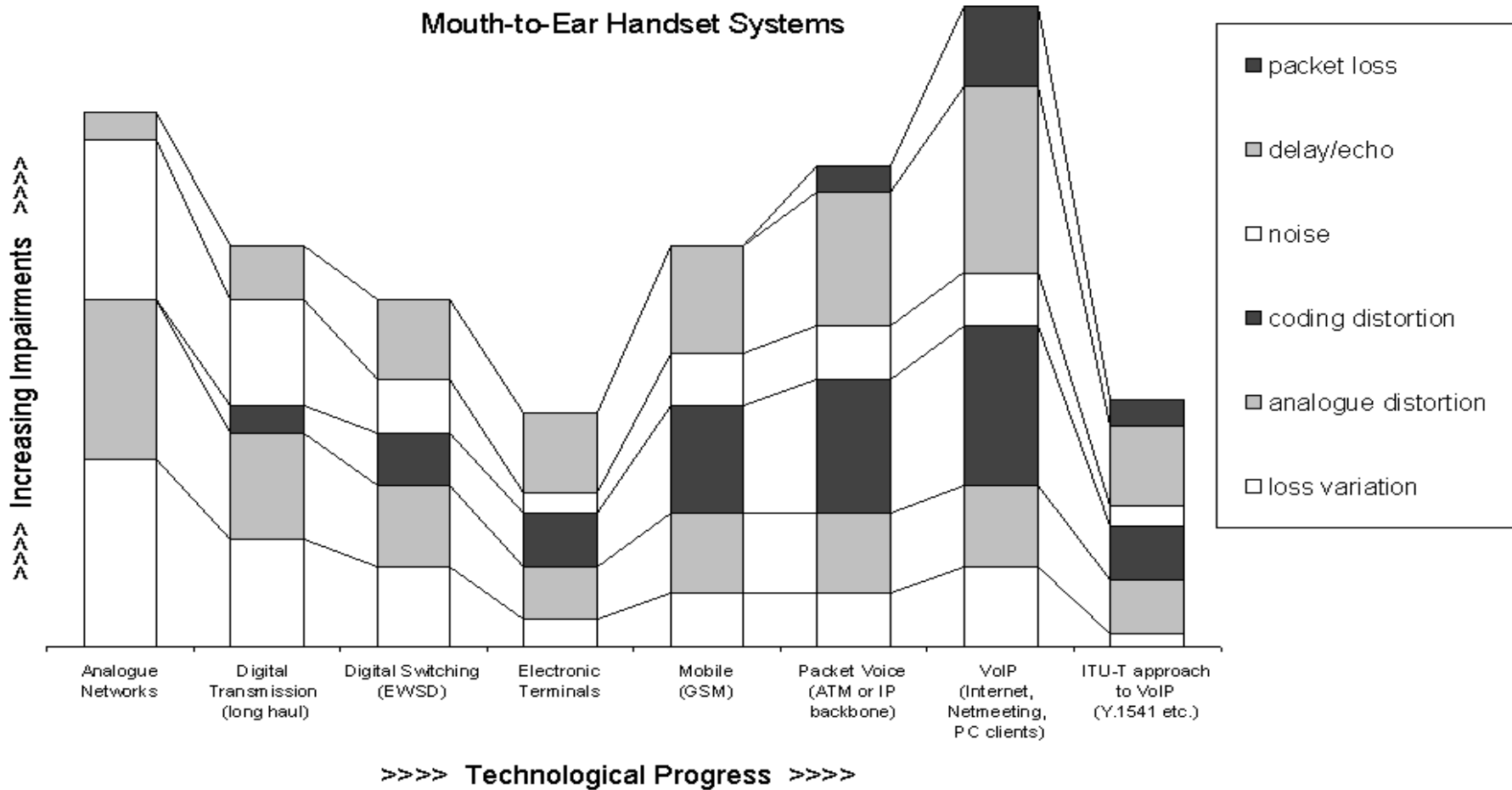
- Los términos calidad de la experiencia (QoE) y Calidad de servicio (QoS), se utiliza ampliamente hoy en día, no sólo en el mundo de las telecomunicaciones, en el que tiene sus raíces, sino cada vez más en diversos tipos de servicios, incluyendo los asociados a multimedia basados en IP.
- En esta presentación, nos enfocamos en algunos aspectos fundamentales de la QoE y QoS aplicada a servicios Multimedia:
 - Calidad de audio
 - Calidad de video
 - Técnicas de QoS sobre redes de paquetes



Calidad de la Voz



Evolución de la calidad percibida de la voz en telecomunicaciones



- Métodos de medida Subjetivos
 - Se basan en conocer directamente la opinión de los usuarios
 - Resultan en un promedio de opiniones (por ejemplo, en un valor de MOS – Mean Opinion Score)
- Métodos de medida Objetivos
 - Se basan en aspectos medibles y algoritmos matemáticos para estimar la calidad percibida por los usuarios
 - Resultan en una “predicción” de la calidad (por ejemplo, una estimación del MOS)



- ***Métodos de determinación subjetiva de la calidad de transmisión***
- Establece los métodos de evaluación subjetiva de la calidad de voz transmitida por sistemas de telecomunicaciones
- Indica como deben realizarse las pruebas, las escalas a utilizar, el acondicionamiento ambiental, guías de instrucciones para los participantes, etc.



- Establece la terminología de las notas medias de opinión (MOS) para audio

	Listening-only	Conversational	Talking
Subjective	MOS-LQSy	MOS-CQSy	MOS-TQSy
Objective	MOS-LQOy	MOS-CQOy	MOS-TQOy
Estimated	MOS-LQEy	MOS-CQEy	MOS-TQEy

NOTE – The letter "y" at the end of above acronyms is a placeholder for the descriptor of the respective audio bandwidth, see the following provisional instructions.

Valores de y:

N = Narrowband

W = Wideband

S = Super-wideband

F = Fullband

(última actualización de la recomendación: Feb. 2016)



- Establece la terminología de las notas medias de opinión (MOS) para video

	Video
Subjective	MOS-VQSz
Objective	MOS-VQOz
Estimated	MOS-VQEz

NOTE – The letter "z" at the end of above acronyms is a placeholder for the descriptor of the respective viewing context, see the following provisional instructions.

Valores de **z**:

M = Mobile (Smartphone, Tablet < 25 cm)

T = PC o TV monitor

(última actualización de la recomendación: Feb. 2016)



- Establece la terminología de las notas medias de opinión (MOS) para contenido audiovisual

	Video
Subjective	MOS-AVQSyz
Objective	MOS-AVQOyz
Estimated	MOS-AVQEyz

(última actualización de la recomendación: Feb. 2016)



- La calidad de la voz se establece a través de la opinión del usuario
- Métodos de “estímulo simple” (Single Stimulus, “SS”)
 - Los “estímulos” se presentan al evaluador de forma aislada.
 - Los estímulos pueden incluir o no la referencia. En caso de incluirla, la referencia se presenta como un estímulo independiente que se califica como cualquier otro.
 - Las muestras se presentan en forma aleatoria para cada observador
- ACR: Absolute Category Rating
 - Se califica el “estímulo” con un valor, entre un mínimo y un máximo. Típicamente, en una escala discreta entre 1 y 5, siendo 5 “Excelente” y 1 “Malo”
 - **MOS** (Mean Opinión Score) es el promedio de los ACR medidos entre un gran número de usuarios



Descripción (inglés)	Descripción (Español)	Puntaje
Excellent	Excelente	5
Good	Buena	4
Fair	Regular	3
Poor	Mediocre	2
Bad	Mala	1

- MOS = Mean Opinion Score = Promedio (media aritmética) de los puntajes



- Métodos de “comparación de estímulos” (Stimulus Comparison, “SC”)
 - También llamados métodos de “doble estímulo”
 - El evaluador debe comparar dos estímulos que le son presentados.
 - Se pueden presentar en simultaneo o uno a continuación del otro
- DCR: Degradation Category Rating
 - También se califica entre un mínimo y un máximo, por ejemplo, entre 1 y 5, siendo 5 cuando no hay diferencias apreciables entre el audio de referencia y el medido y 1 cuando la degradación es muy molesta
 - **DMOS** (Degradation MOS) el promedio de los valores DCR medidos entre un gran número de usuarios



Descripción (inglés)	Descripción (Español)	Puntaje
Degradation is inaudible	Degradación inaudible	5
Degradation is audible but not annoying	Degradación audible, pero no molesta	4
Degradation is slightly annoying	Degradación ligeramente molesta	3
Degradation is annoying	Degradación molesta	2
Degradation is very annoying	Degradación muy molesta	1

- DMOS = Degradation Mean Opinion Score = Promedio (media aritmética) de los puntajes



Métodos subjetivos definidos por ITU-T

- P.805 (2007): Subjective evaluation of conversational quality
 - Se definen métodos de pruebas de conversación para evaluar la calidad de la comunicación en aplicaciones telefónicas. Permite evaluar degradaciones producidas por efectos de delay, echo, voice clipping y dropped packets
- P.830 (1996): Subjective performance assessment of telephone-band and wideband digital codecs
 - Se describen métodos y procedimientos para llevar a cabo evaluaciones subjetivas de la calidad de funcionamiento de códecs de señales vocales digitales.
- P.832 (2000): Subjective performance evaluation of hands-free terminals
 - Se definen métodos y procedimientos para llevar a cabo evaluaciones subjetivas de la calidad de funcionamiento de los terminales en funcionamiento del tipo “manos libres”.



- P.835 (2003): Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm
 - Se describe una metodología para evaluar la calidad subjetiva de la voz en presencia de ruido y, en particular, para la evaluación de los algoritmos de cancelación de ruido.
- P.880 (2004): Continuous evaluation of time-varying speech quality
 - Se define la metodología denominada evaluación continua de la calidad vocal que varía con el tiempo (CETVSQ). Esta metodología es indicada para evaluar los efectos de las fluctuaciones temporales de la calidad vocal.



- P.911 (1998): Subjective audiovisual quality assessment methods for multimedia applications
 - Se definen métodos para evaluar la calidad subjetiva audiovisual no interactiva para aplicaciones multimedia. También aporta indicaciones sobre la relación entre la calidad de audio, de vídeo y audiovisual.
- P.913 (2014): Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment
 - Se definen métodos de evaluación subjetivos de calidad de audio, video y audiovisual



Métodos subjetivos definidos por ITU-T

- P.920 (2000): Interactive test methods for audiovisual communications
 - Se definen métodos de evaluación interactivos. Los métodos propuestos se consideran una extensión de los métodos de la recomendación P.800, y se basan en pruebas de conversación (Conversation test) con hablantes activos.
- P.1302 (2014) Subjective method for simulated conversation tests addressing speech and audiovisual call quality
 - Para evaluar la calidad del habla en aplicaciones telefónicas o de video llamada, se define un método denominado “simulated conversation tests”, que consiste en secuencias de voz o material audiovisual ordenados de forma que sigan una historia definida por el escenario de prueba. Este método se desarrolló para mejorar los métodos no interactivos tradicionales en el que el evaluador solo escucha piezas de audio siguiendo una estructura fija del material juzgado, y limitando la atención del evaluador.



- MUSHRA: MUltiple Stimuli with Hidden Reference and Anchor
 - Es un método utilizado para evaluación de audio de “calidad intermedia”
 - Sirve tanto para sonido mono como estéreo, y se puede realizar tanto con altavoces como con auriculares.
 - Utiliza una escala de calidad continua.
 - Cada evaluador debe calificar sobre una escala graduada de 0 a 100
 - Durante la prueba los evaluadores disponen de todas las muestras (o “estímulos”) entre las que se encuentra la “referencia oculta” y las “anclas”, y puede escuchar cada muestra las veces que quiera y modificar las calificaciones cuantas veces quiera



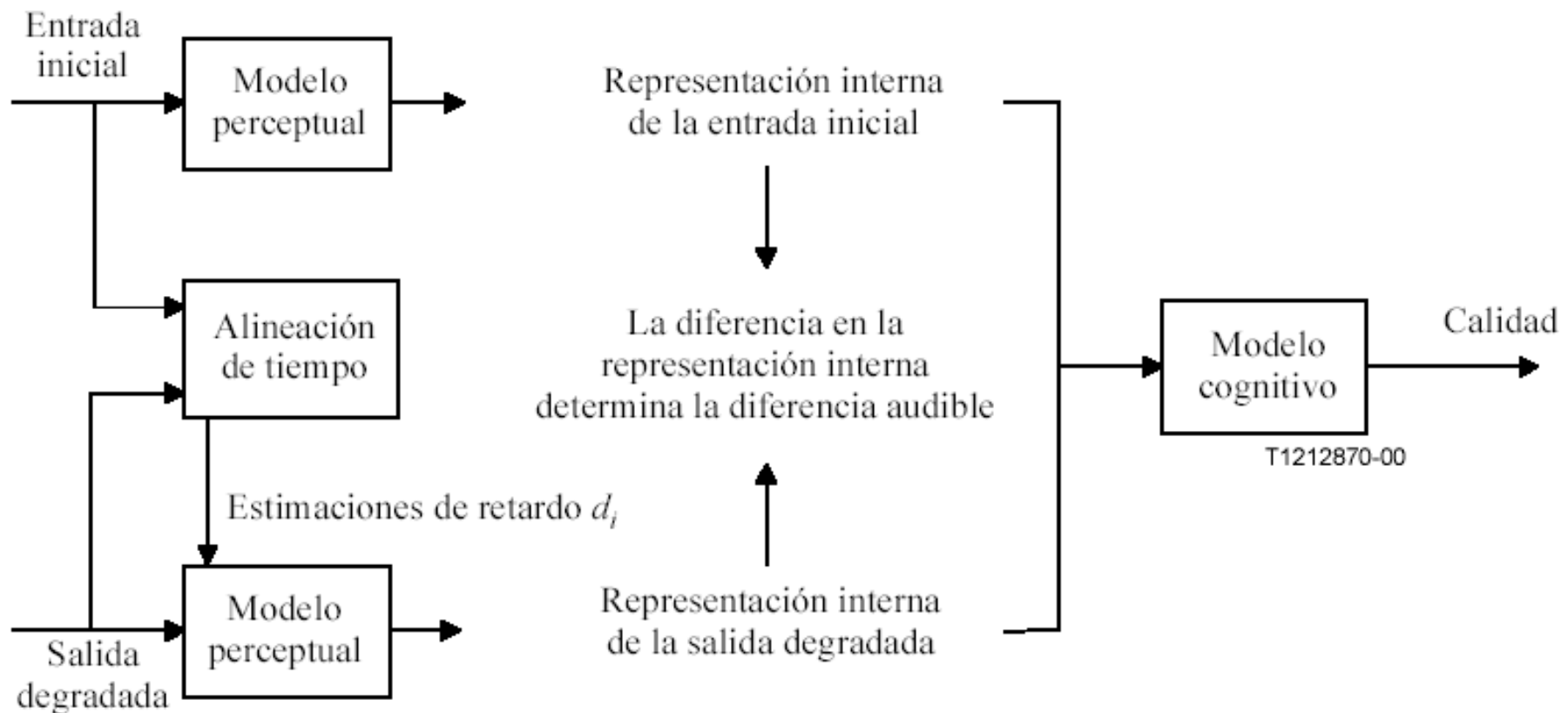
Ejemplo de método subjetivo: MUSHRA



Métodos objetivos: ITU-T P.862

- La recomendación ITU-T P.862 presenta un método objetivo para la evaluación de la calidad vocal de extremo a extremo
- El método objetivo descrito se conoce por "evaluación de la calidad vocal por percepción" (**PESQ**, *perceptual evaluation of speech quality*).
- Se compara una señal inicial $X(t)$ con una señal degradada $Y(t)$ que se obtiene como resultado de la transmisión de $X(t)$ a través de un sistema de comunicaciones (por ejemplo, una red IP).
- La salida de PESQ es una predicción de la calidad percibida por los sujetos en una prueba de escucha subjetiva que sería atribuida a $Y(t)$.



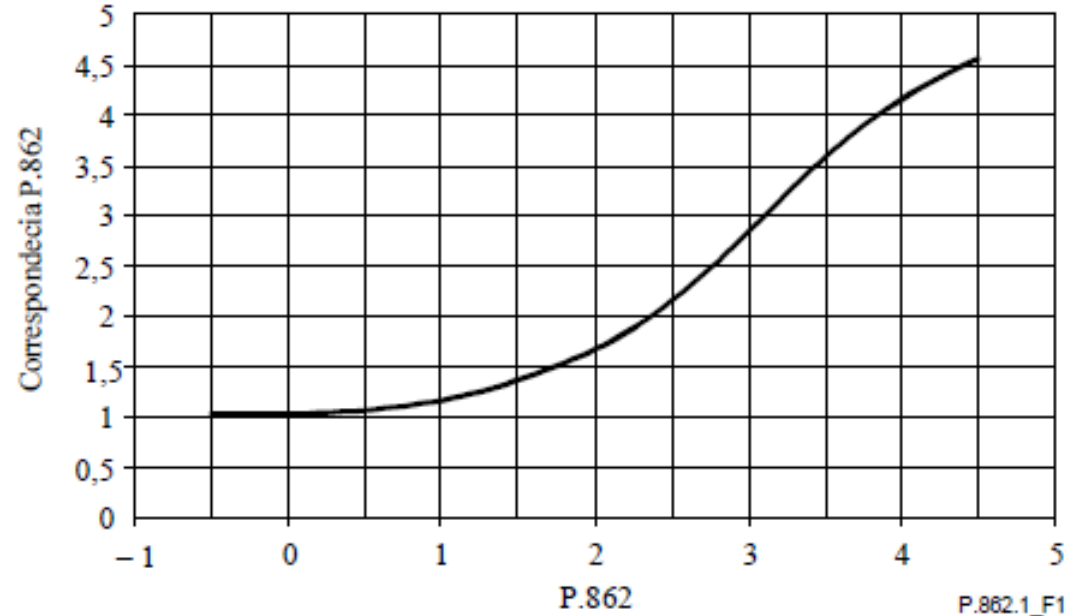


- Lo esencial en este proceso es la transformación de las dos señales, la inicial y la degradada, en una representación interna que intenta reproducir la percepción psicoacústica de señales de audio en el sistema auditivo humano.
- El modelo cognitivo de PESQ termina brindando una distancia entre la señal vocal inicial y la señal vocal degradada (“nota PESQ”), la que corresponde a su vez con una predicción de la MOS subjetiva. La nota PESQ se hace corresponder a una escala “similar” a la de MOS



- Establece la correspondencia entre la nota PESQ y la estimación del valor de MOS-LQO (Mean Opinion Score of Listening Quality Objective)

$$y = 0,999 + \frac{4,999 - 0,999}{1 + e^{-1,4945 \cdot x + 4,6607}}$$



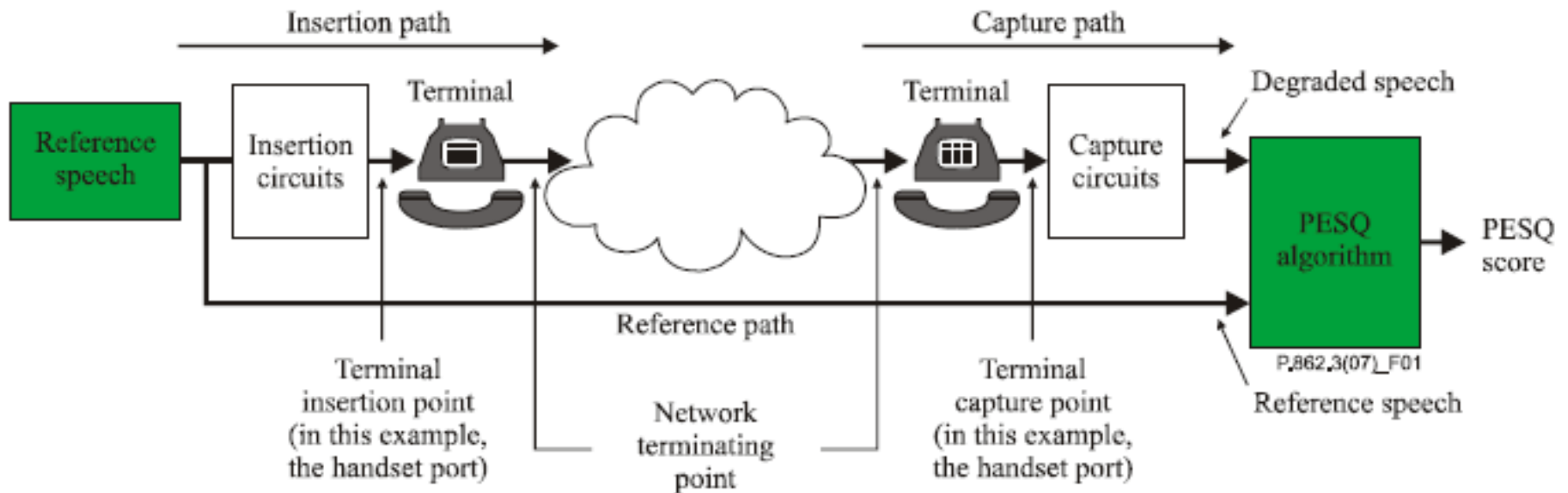
- Extiende el área de aplicación a señales de banda ancha (de 50 a 7000 Hz)

$$y = 0.999 + \frac{4.999 - 0.999}{1 + e^{-1.3669 \times x + 3.8224}}$$

- Los resultados de la predicción del MOS dados por P.862.2 no pueden ser directamente comparados con los valores de predicción del MOS de la recomendación “base” (P.862), ya que las expectativas de los usuarios son diferentes en cada caso



- Es un guía de aplicación de la serie de recomendaciones P.862



- Establece valores de referencia de MOS para diferentes codecs

G.711		G.726				G.728	G.729	G.729 A	G.723.1	
u-Law	A-Law	16 kbps	24 kbps	32 kbps	40 kbps				5.3 kbps	6.3 kbps
4.46	4.36	2.56	3.34	3.93	4.25	3.99	3.80	3.69	3.45	3.60

AMR								FR	HR	EFR
12.2 kbps	10.2 kbps	7.95	7.4	6.7	5.9	5.15	4.75			
3.92	3.73	3.56	3.50	3.40	3.25	3.11	3.06	3.99	3.16	3.12

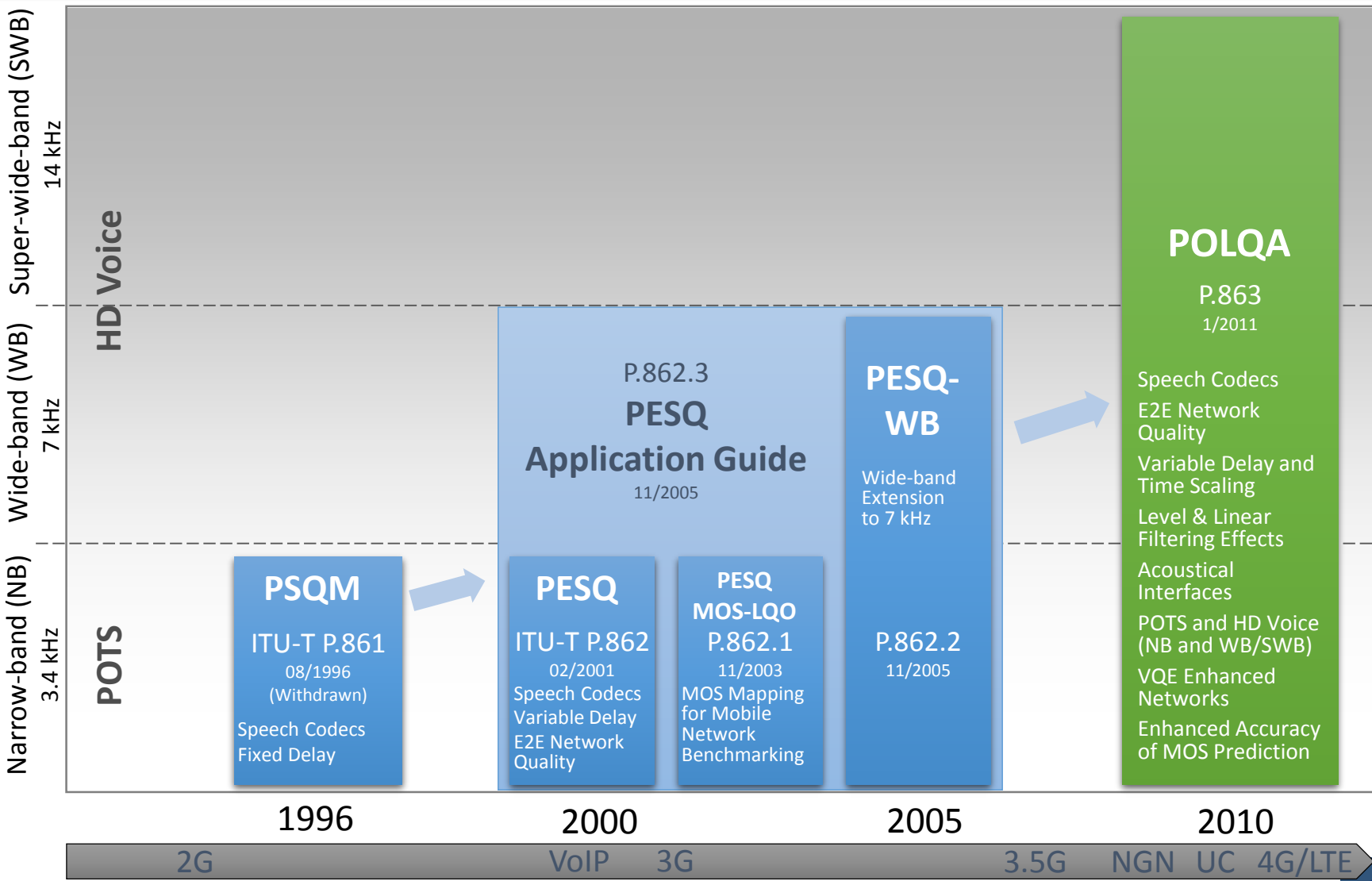


POLQA

- Estandarizada en enero de 2011 (actualizada 2014)
- Evolución de PESQ
- Puede trabajar en la banda “super ancha” (de 50 a 14000 Hz)
- Es a su vez compatible con las bandas angosta y ancha



Evolución de P.86x



© OPTICOM GmbH 2010 – www.polqa.info

Evolution of Network Technologies available at the time of development, i.e. included use cases for each Recommendation

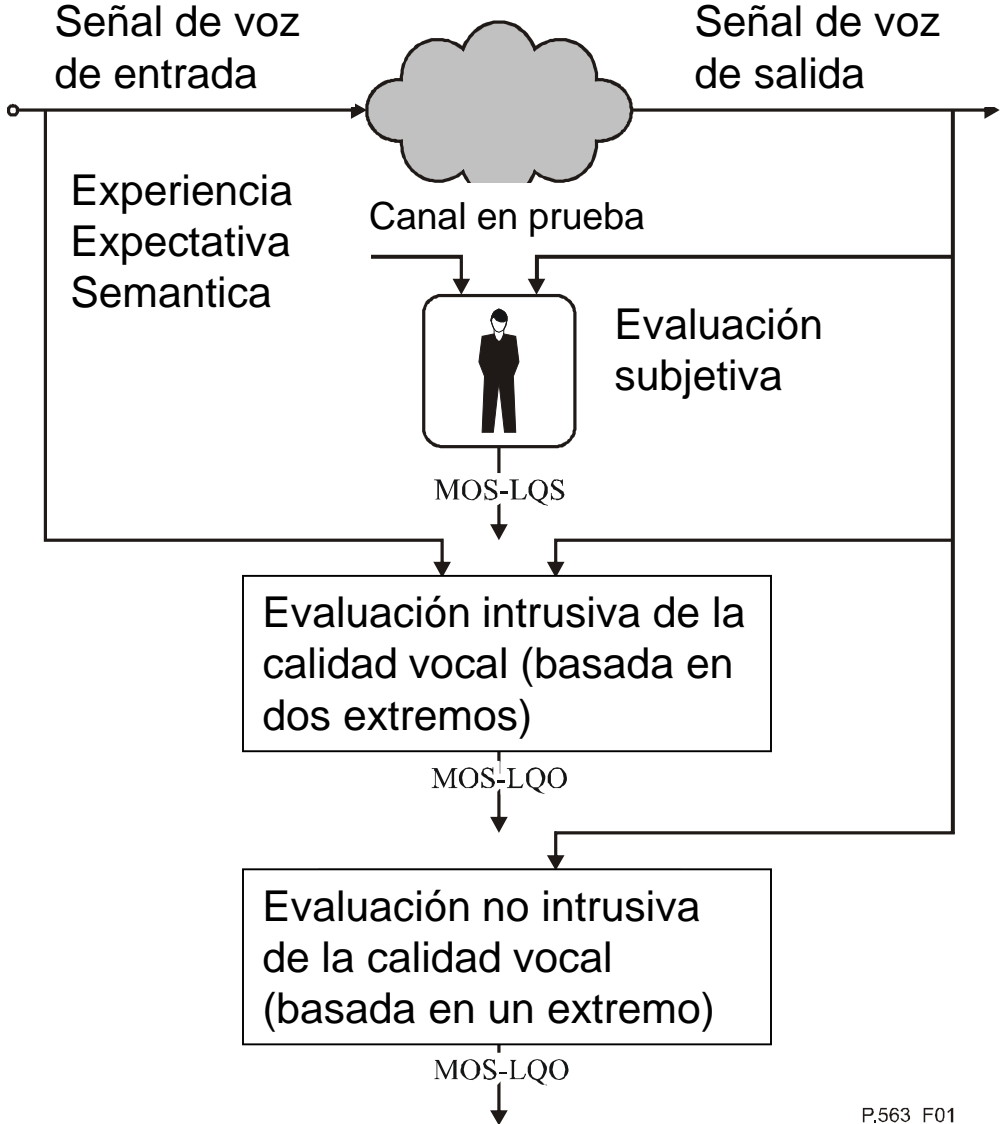


Métodos objetivos: ITU-T P.563

- El algoritmo P.563 es aplicable para la predicción de la calidad vocal sin una señal de referencia independiente
- En comparación con la ITU-T P.862 que compara una señal de referencia de elevada calidad con la señal degradada en base a un modelo perceptual, P.563 predice la calidad de la voz de una señal degradada **sin** una señal vocal de referencia dada
- El enfoque utilizado en P.563 puede visualizarse como un experto que escucha una llamada real con un dispositivo de prueba



ITU-T P.563 vs ITU-T P.862



P.563_F01



- Trata de detectar tres clases de degradación de la señal de voz:
 - Desnaturalización de la voz
 - Análisis del tracto vocal, tratando de identificar si existe una marcada “robotización”
 - Análisis de ruidos adicionales intensos
 - SNR estática (nivel básico del ruido de fondo)
 - SNR por segmentos de voz
 - Interrupciones, silenciamientos y recortes temporales



- Cada clase de distorsión utiliza una combinación lineal de varios parámetros, con lo que se genera una calidad vocal intermedia.
- La calidad vocal definitiva se calcula combinando los resultados de calidad vocal intermedia con algunas características adicionales de la señal.



Calidad de Voz sobre redes IP

Calidad de voz y video



- La calidad de la voz sobre redes de paquetes se ve afectada por varios factores
 - Compresión o códec utilizado
 - Pérdida de paquetes
 - Demora
 - Eco
 - Jitter



Pérdida de paquetes

- A diferencia de las redes telefónicas, donde para cada conversación se establece sobre un vínculo “estable y seguro”, las redes de datos admiten la pérdida de paquetes.
- En aplicaciones de voz y video el audio y video es “encapsulado” en paquetes y enviado, sin confirmación de recepción de cada paquete.
- Puede haber un porcentaje de paquetes que no llegan al destino
- Se escucha como interrupciones en la voz, o cortes de video

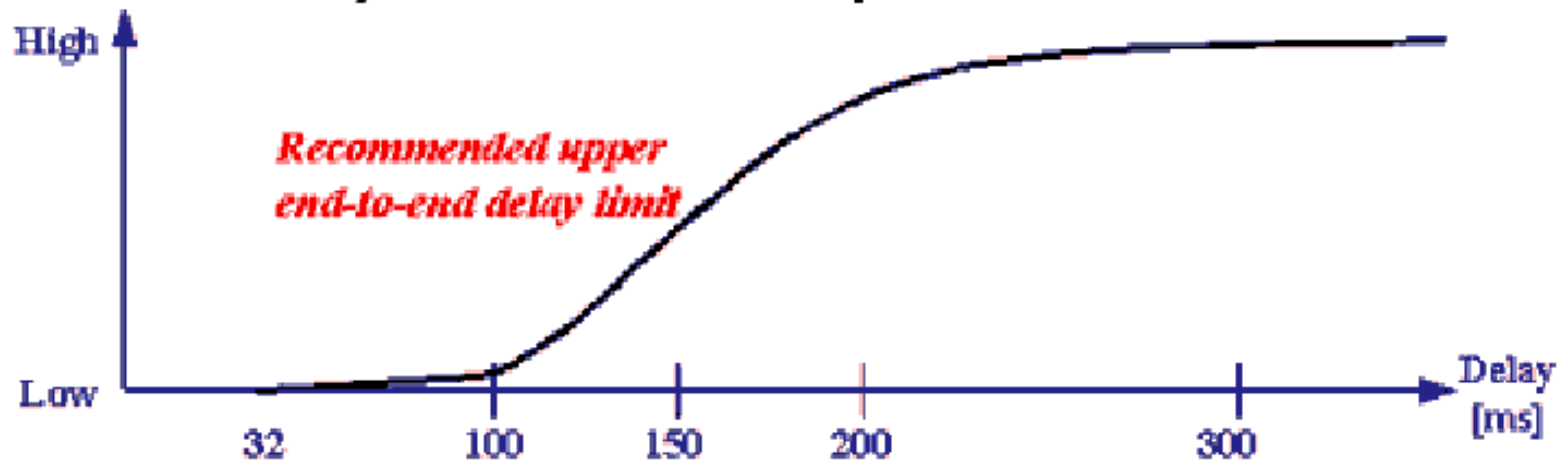


Demora (Delay)

- Se deben a varios factores:
 - Codificación
 - G.711 (64 kb/s) 0,13 – 0,75 ,ms
 - G.728 (16 kb/s) 2.5 ms
 - G.729 (8 kb/s) 10 – 15 ms
 - G.723.1 (5.3 o 6.4 kb/s) 37.5 ms
 - RTAudio < 40 ms
 - Red (latencia)
 - Cantidad de muestras/bytes por paquete
 - Velocidad de transmisión
 - Congestión
 - Demoras de los equipos de red (colas en routers, gateways, etc.)
 - Jitter Buffer



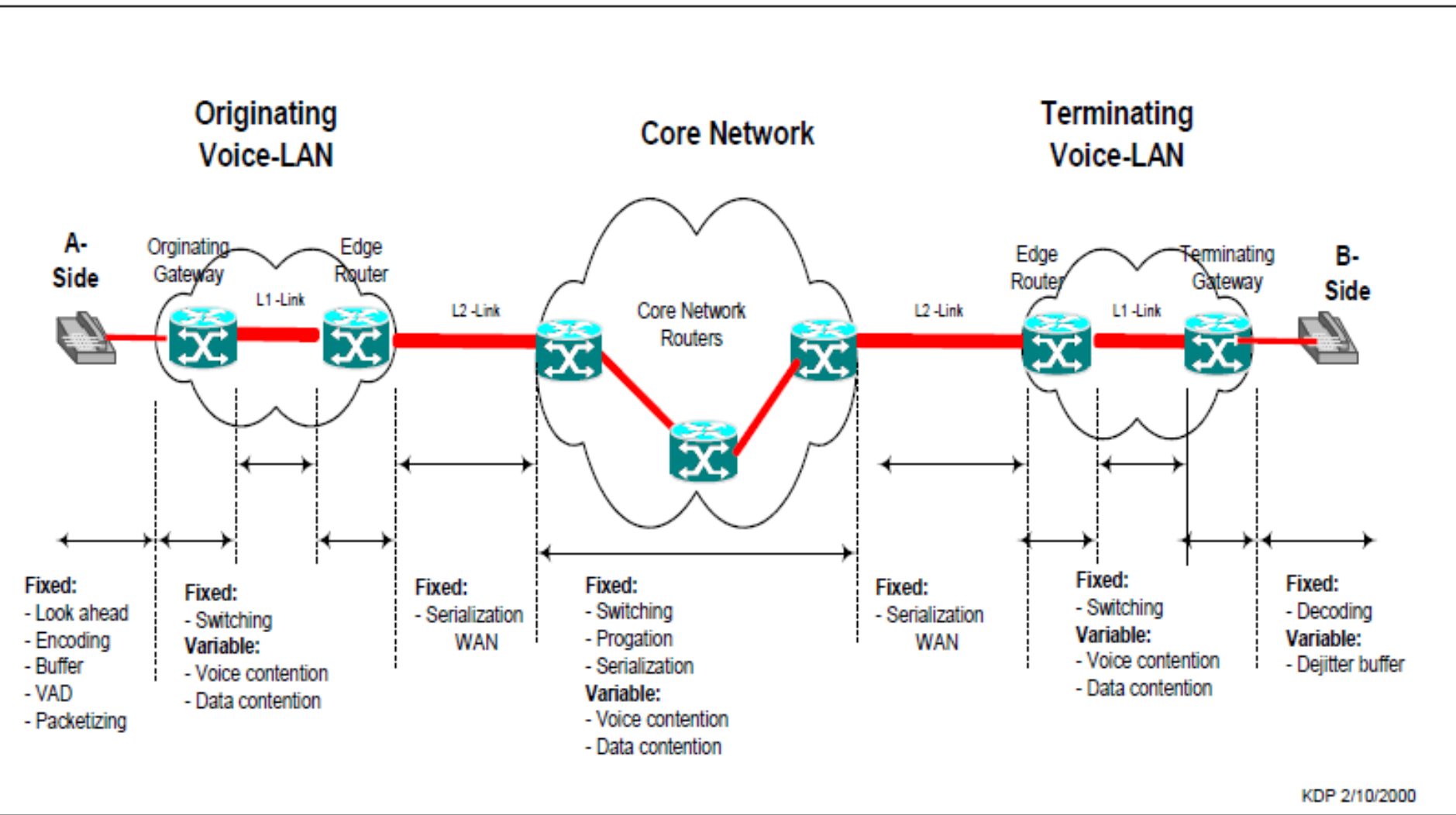
Delay's Effect on User Experience



- Es la variación en las demoras (latencias).
 - Por ejemplo, si dos puntos comunicados reciben un paquete cada 20 ms *en promedio*, pero en determinado momento, un paquete llega a los 30 ms y luego otro a los 10 ms, el sistema tiene un “jitter” de 10 ms.
- El jitter afecta la percepción de la voz, y puede evitarse mediante *buffers*
 - Los *buffers* agregan una demora adicional al sistema, ya que deben “retener” paquetes para poder entregarlos a intervalos constantes. Cuánto más variación de demoras (*jitter*) exista, más grandes deberán ser los *buffers*, y por lo tanto, mayor demora total tendrá el sistema.



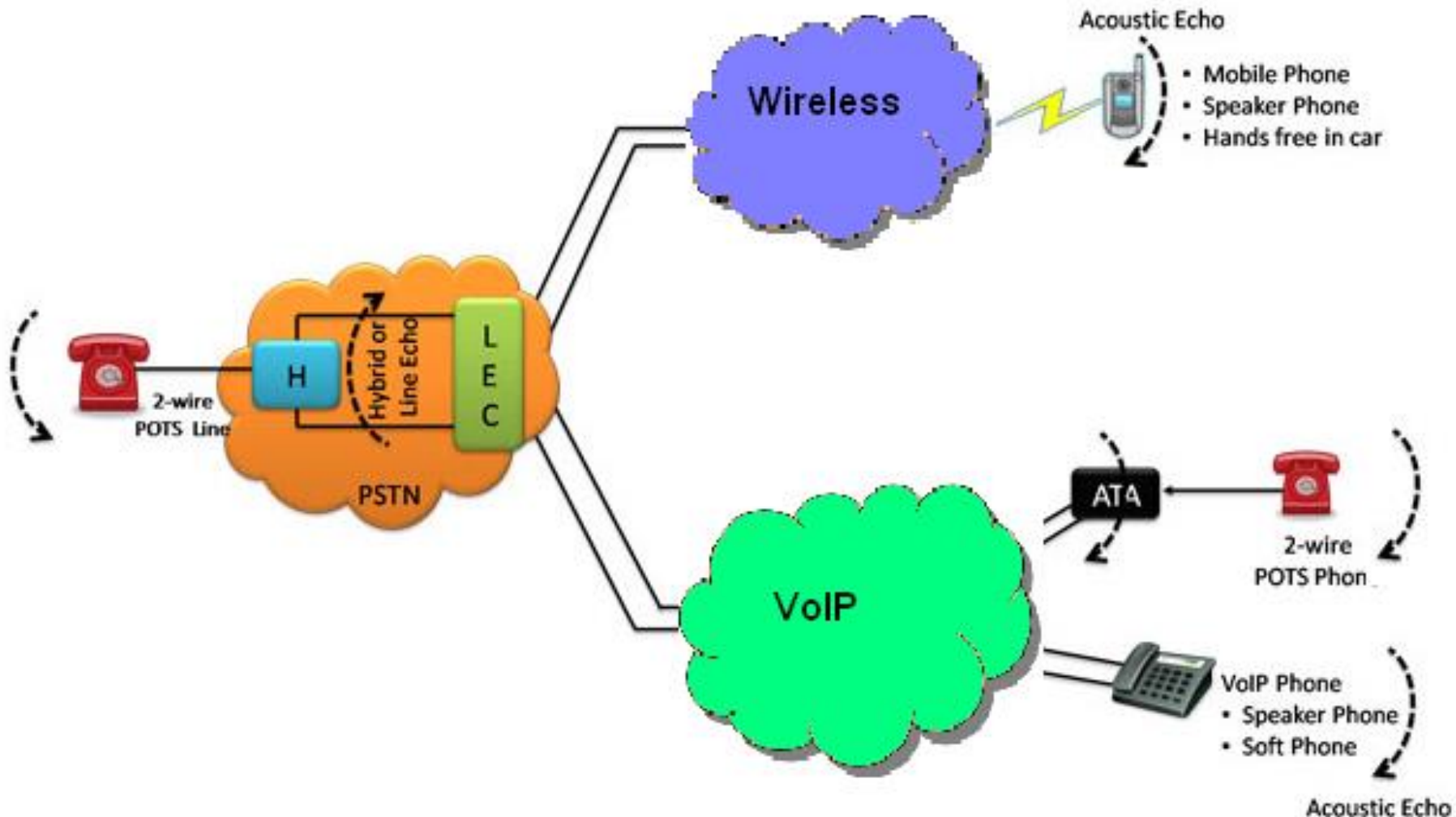
Demora y Jitter



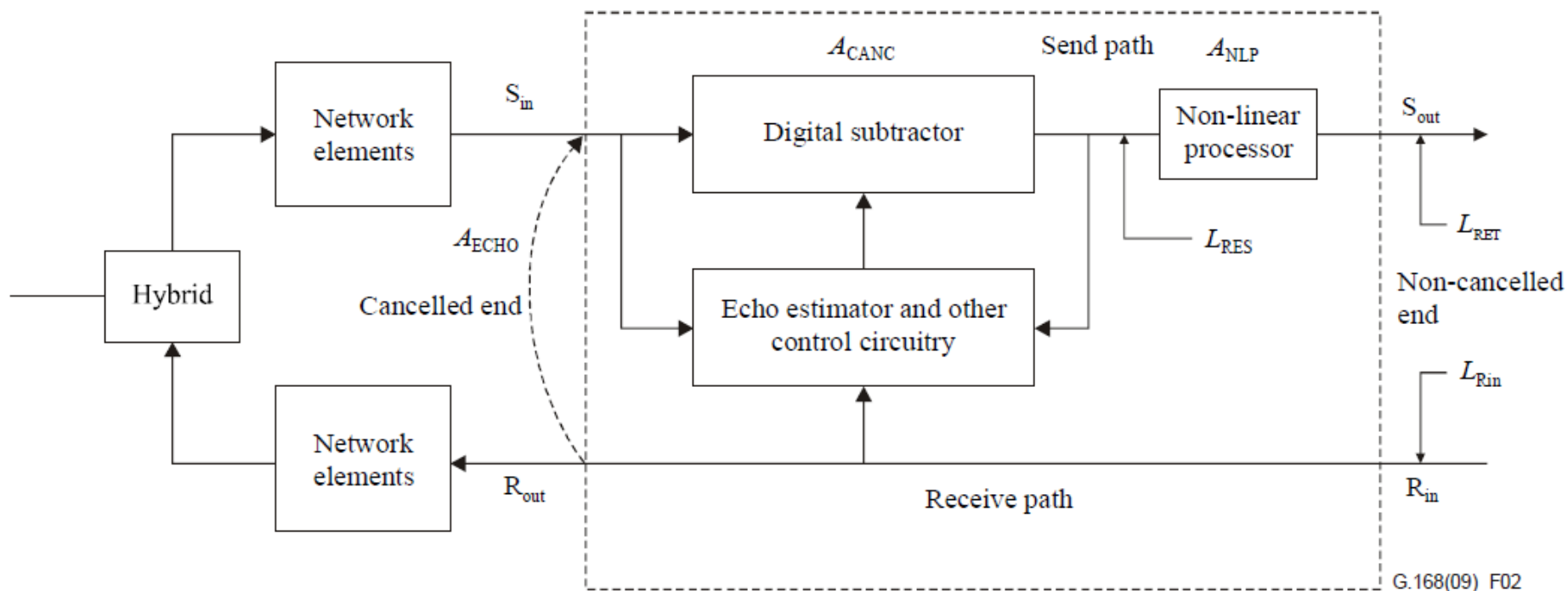
KDP 2/10/2000

- Tiempo transcurrido desde que se habla hasta que se percibe el retorno de la propia voz
- Si la demora de retorno es menor a 30 ms, o el nivel del retorno está por debajo de los -25 dB, el efecto del eco no es percibido.
- Dado que las demoras de voz sobre redes de datos son altas, puede existir eco



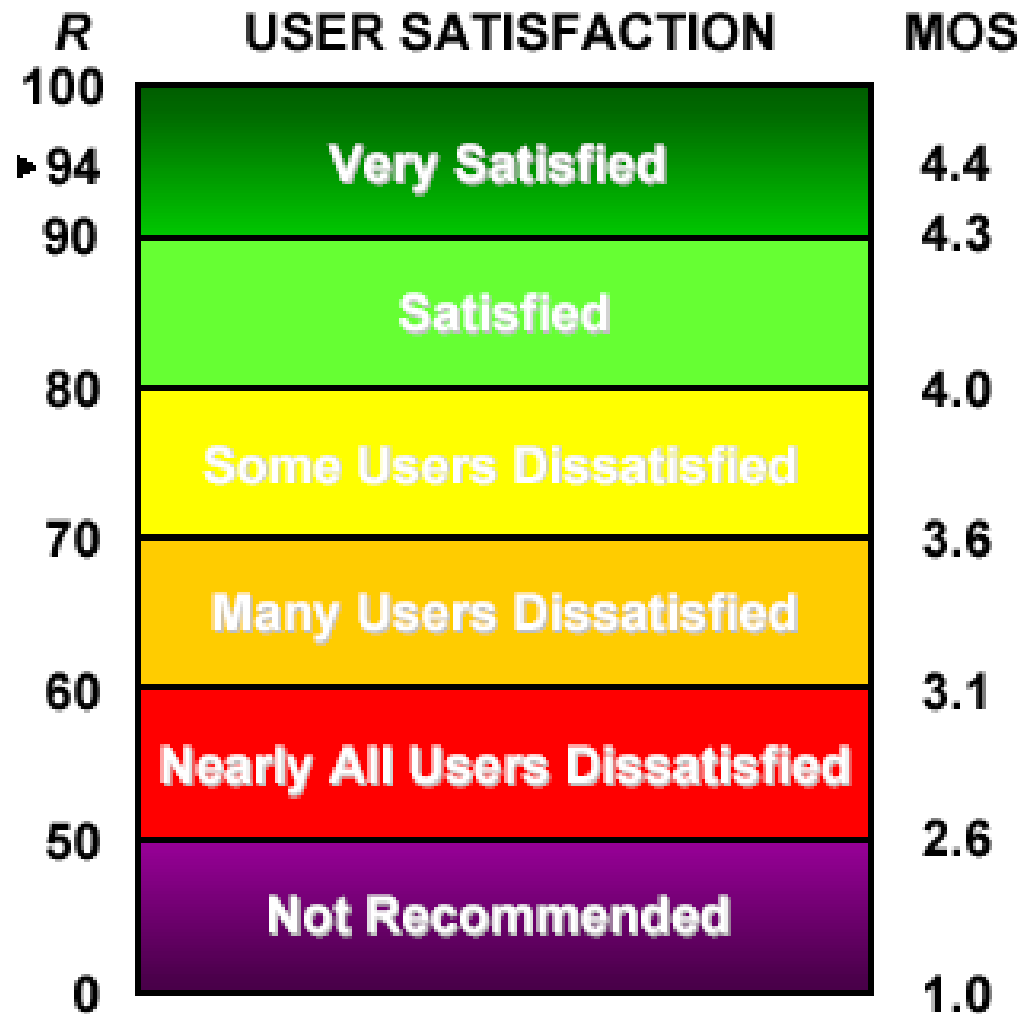


- ITU-T G.168: Digital Network Echo Cancellers



- La ITU ha definido un modelo, llamado “E-Model” (ITU-T G.107), para estimar la calidad de la voz sobre redes de paquetes, teniendo en cuenta factores medibles de la red
 - El resultado del “E-Model” es un valor escalar llamado **R**, que puede ser directamente relacionado con el **MOS** (ITU-T P.800)
 - Inicialmente el modelo aplicaba a comunicaciones de “banda angosta”. En 2009 fue extendido a comunicaciones de banda ancha.



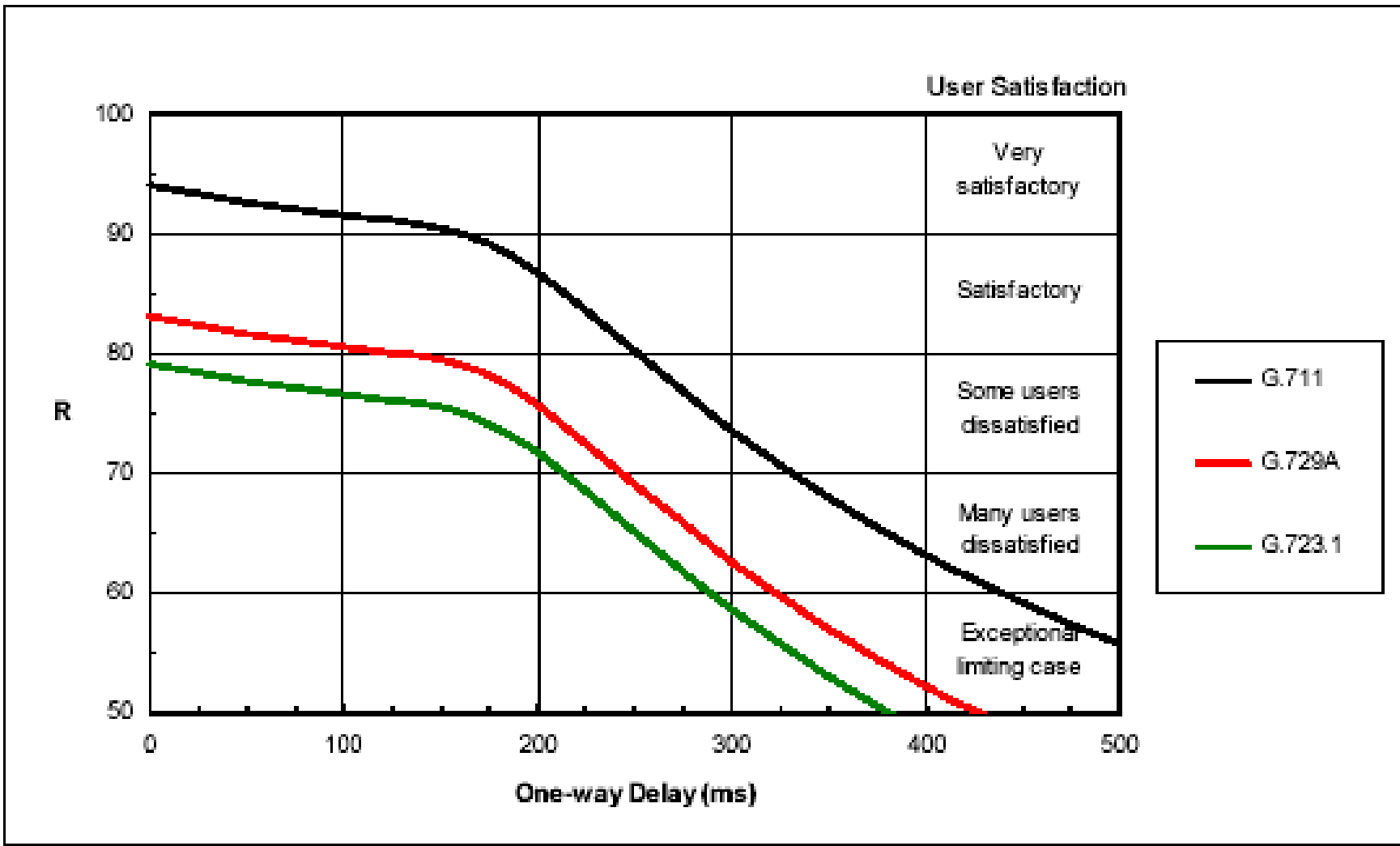


Definición de “R”

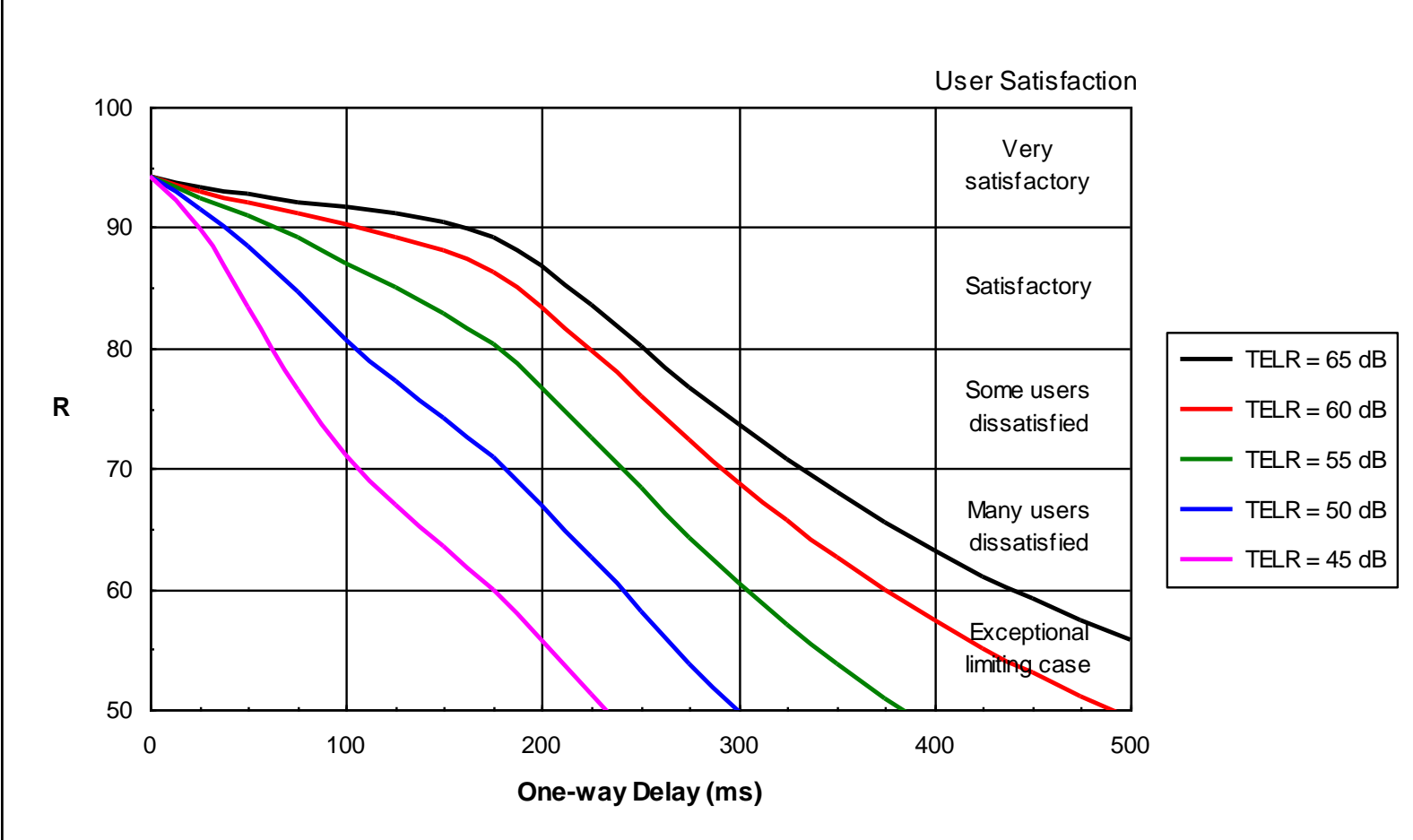
- $R = R_o - I_s - I_d - I_e + A$
 - R_o = Fuentes de ruido independientes del sistema
 - Ruido ambiental, tanto en el origen como en el destino
 - El máximo teórico es 100
 - I_s = Deterioro simultáneo a la generación de la señal digital
 - Volumen excesivo, distorsión de cuantización
 - I_d = Deterioro casusado por las demoras
 - Demoras, Jitter, Eco
 - I_e = Deterioro causado por equipos especiales
 - Codec, pérdidas de paquetes
 - A = Factor de Mejoras de Expectativas



Efectos del Codec y la Demora



Efectos del Eco y la demora

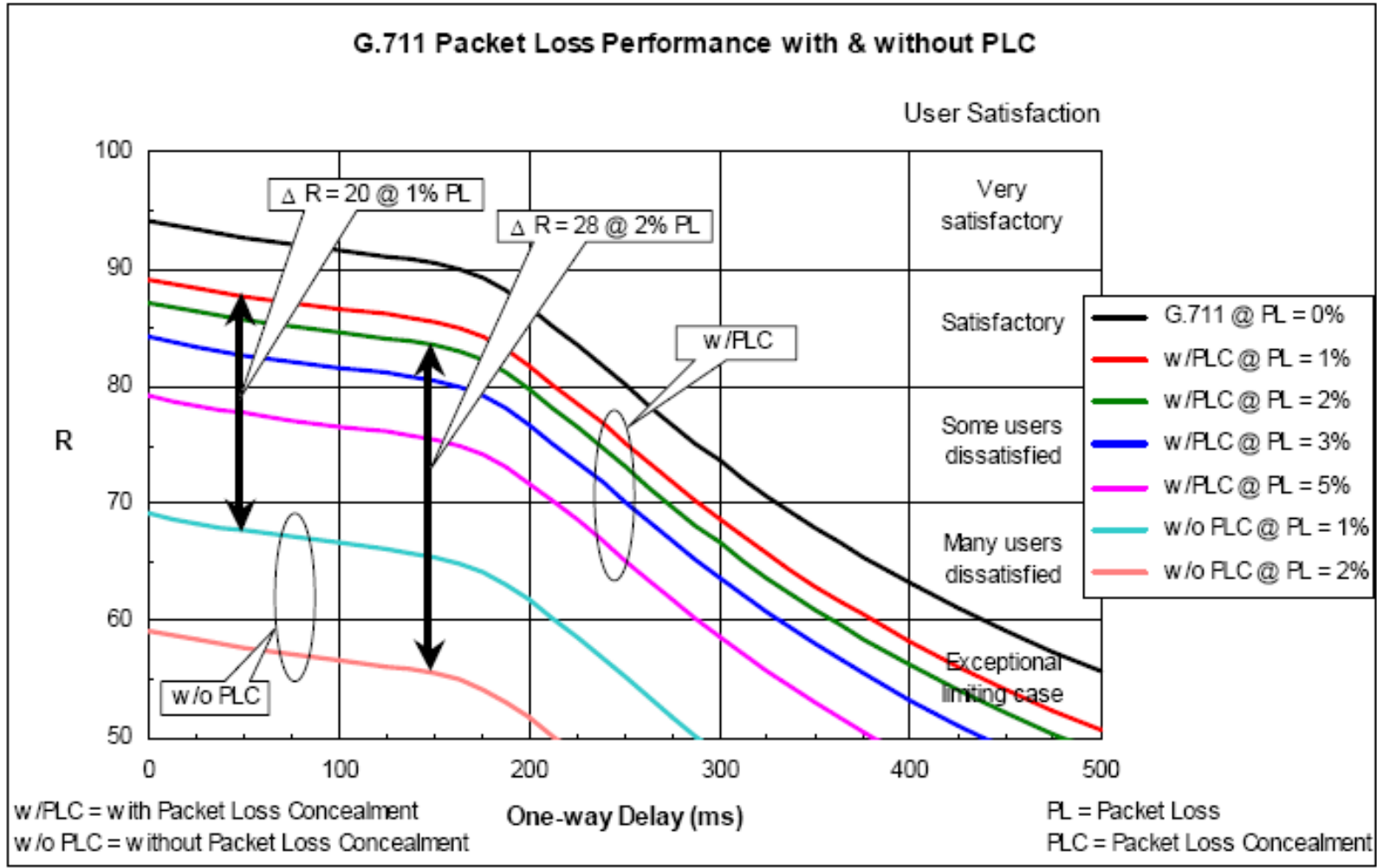


TEL R = Talker Echo Loudness Rating.

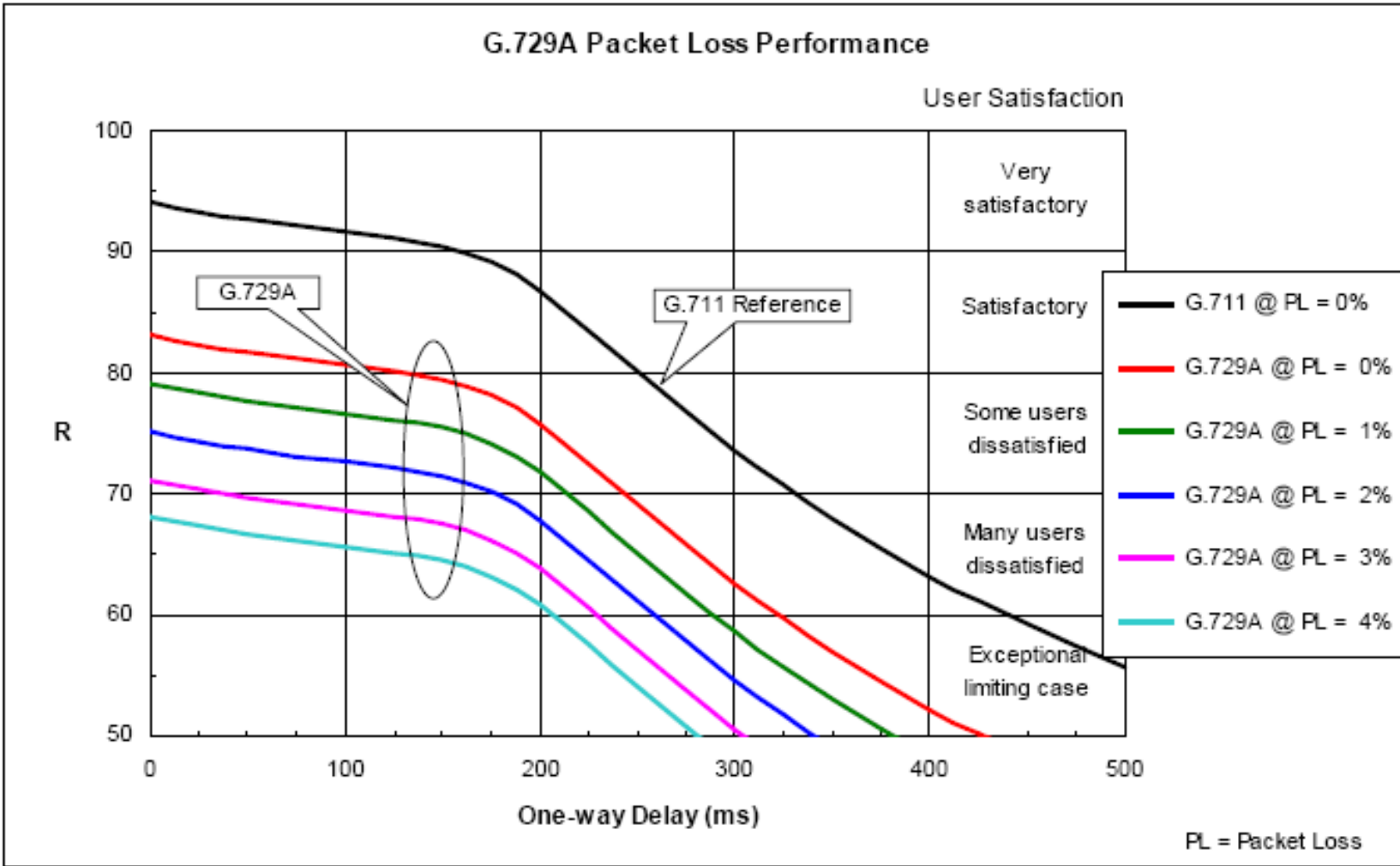
Cuanto más atenuado el eco percibido (mayor valor en db de TELR), menor efecto tiene el eco sobre la degradación



Efectos de la Demora y la Pérdida de paquetes



Efectos de la Demora y la Pérdida de paquetes



Ejemplo de sistema de comunicación	Valor máximo de A
Convencional (alámbrico)	0
Movilidad mediante redes celulares en un edificio	5
Movilidad en una zona geográfica o en un vehículo en movimiento	10
Conexión con lugares de difícil acceso, por ejemplo, mediante conexiones de múltiples saltos por satélite	20



Factor R en paquetes RTCP

The screenshot shows a Wireshark interface with a filter set to 'rtcp'. The selected packet is a 'Sender Report' (No. 1433) from 10.190.5.20 to 10.190.5.17. The packet details pane shows a 'Block 2' of type 'VoIP Metrics Report Block (7)'. Under 'Contents', the 'R Factor' is listed as 8, which is circled in red. Other metrics include Gmin: 16, External R Factor: 90, and MOS - Listening Quality: 4.300000. The packet bytes pane at the bottom shows the raw data in hexadecimal and ASCII.

No.	Time	Source	Destination	Protocol	Info
1433	30.02238	10.190.5.20	10.190.5.17	RTCP	Sender Report source description Extended report (RFC 3611)

Block 2
Type: VoIP Metrics Report Block (7)
Type Specific: 0
Length: 8

Contents
Identifier: 0x7e4b46be (2118862526)
Fraction lost: 0 / 256
Fraction Discarded: 0 / 256
Burst Density: 0
Gap Density: 0
Burst Duration(ms): 0
Gap Duration(ms): 6020
Round Trip Delay(ms): 0
End System Delay(ms): 65
Signal Level: 87
Noise Level: 87
Residual Echo Return Loss: 127
Gmin: 16
R Factor: 8
External R Factor: 90
MOS - Listening Quality: 4.300000
MOS - Conversational Quality: 1.000000
00.. = Packet Loss Concealment Algorithm: Unspecified (0)
..00 = Adaptive Jitter Buffer Algorithm: Unknown (0)
.... 1010 = Jitter Buffer Rate: 10
Nominal Jitter Buffer Size: 20
Maximum Jitter Buffer Size: 200
Absolute Maximum Jitter Buffer size: 500
RTCP frame length check: OK - 168 bytes

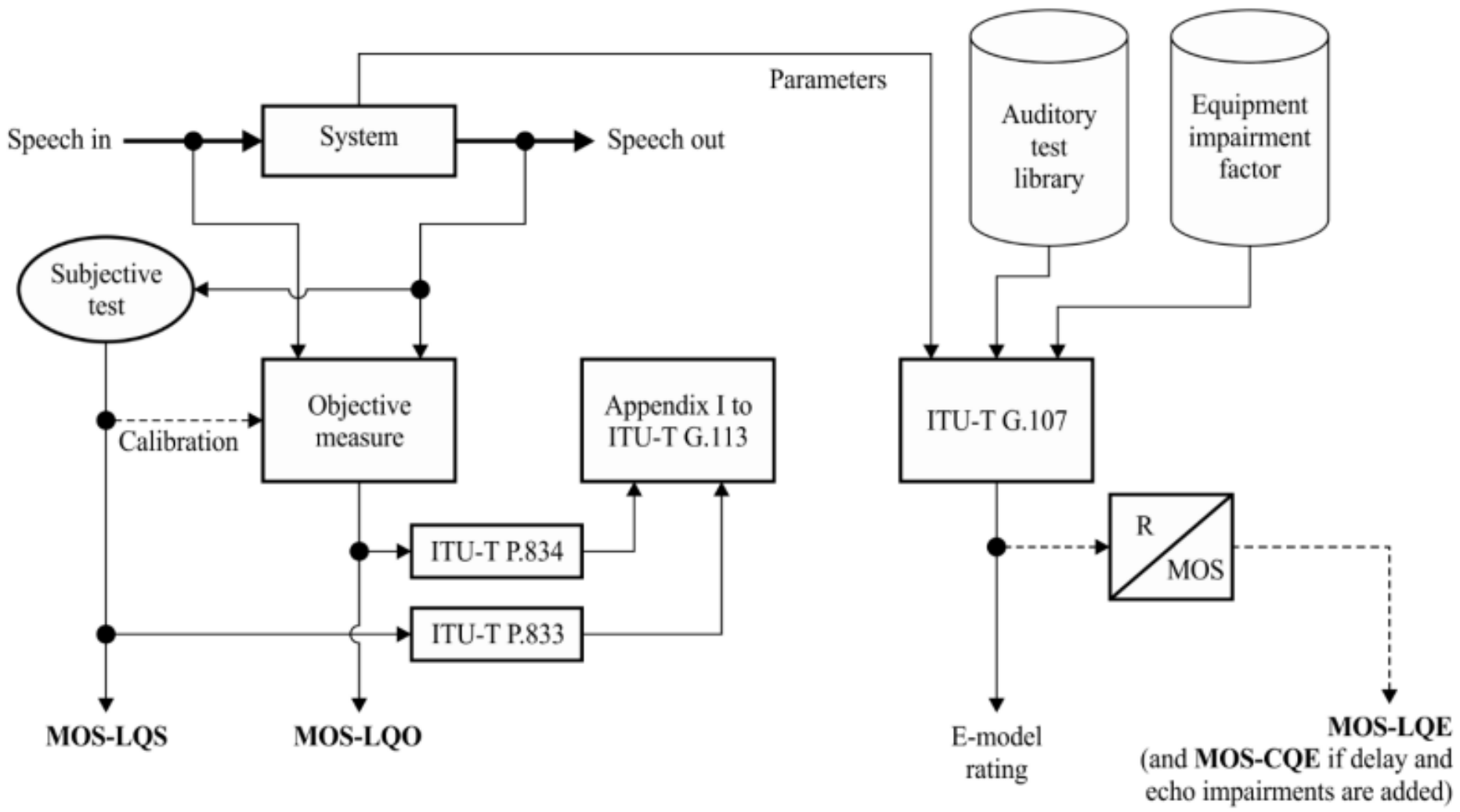
```
0000 00 1a 64 20 a3 06 00 1b ba fd 15 04 08 00 45 a4 ..d... ..E.  
0010 00 c4 00 00 40 00 80 11 d9 e4 0a be 05 14 0a be ..@.....  
0020 05 11 14 65 c0 19 00 b0 29 13 81 c8 00 0c 7e 4b ...e....).....~K  
0030 46 be 00 00 00 06 03 12 6e 97 6f f5 f1 e6 00 00 F.....n.o....  
0040 01 2c 00 00 bb 80 00 00 6a 2a 00 00 00 00 00 00 .....j*.....  
0050 01 34 00 00 00 00 1f 04 78 00 00 03 41 89 81 ca ..4.....X..A..  
0060 00 07 7e 4b 46 be 01 12 6e 6f 74 73 65 74 40 6e ..~KF...notset@n  
0070 65 77 62 65 72 65 70 62 65 6d 00 00 00 00 80 c5 ..w... ..
```



- Las expectativas de los usuarios, y las calificaciones de calidad del audio, son diferentes en comunicaciones de banda ancha respecto de las comunicaciones de banda angosta.
- En 2009 el modelo es extendido para llegar a un $R_{\max} = 129$



Resumen



Tomado de ITU-T P.801

P.800.1(16)_F01

Calidad de Video



- Varios tipos de degradaciones suelen presentarse en las señales de video transmitidas sobre redes de paquetes
- A su vez, varios tipos de degradaciones obedecen al método de codificación utilizado
- El estudio en esta área es todavía un tema de investigación



Degradaciones en video digital

- Efecto de bloques (blocking)
- Efecto de imagen de base (basis image)
- Borrosidad o falta de definición (Blurring)
- Color bleeding (Corrimiento del color)
- Efecto escalera y Ringing
- Patrones de mosaicos (Mosaic Patterns)
- Contornos y bordes falsos
- Errores de Compensaciones de Movimiento (MC mismatch)
- Efecto mosquito
- Fluctuaciones en áreas estacionarias
- Errores de crominancia



- La manera más confiable de medir la calidad de una imagen o un video es la evaluación subjetiva, realizada por un conjunto de personas que opinan acerca de su percepción
 - El “MOS” (Mean Opinion Score) es la métrica generalmente aceptada como medida de la calidad



- ITU-R BT.500-13 (Enero 2012): Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión
- Detalla varios tipos de métodos de evaluación:
 - SS (Single Stimulus)
 - DSIS (Double Stimulus Impairment Scale)
 - DSCQS (Double Stimulus Continuous Quality Scale)
 - SSCQE (Single Stimulus Continuous Quality Evaluation)
 - SDSCE (Simultaneous Double Stimulus for Continuous Evaluation).

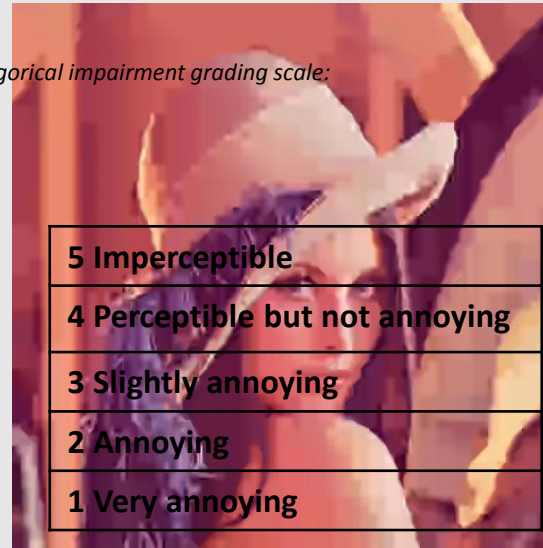


- ITU-T P.910 (1999): Métodos de evaluación subjetiva de la calidad de vídeo para aplicaciones multimedia
- Detalla los métodos
 - ACR (Absolute Category Rating)
 - DCR (Degradation Category Rating)
 - PR (Pair Comparison)



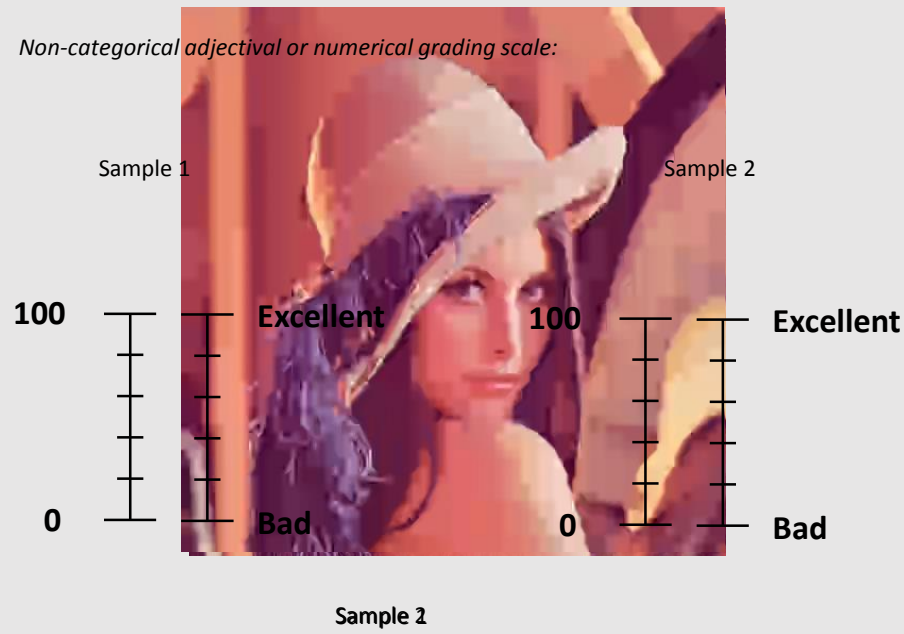
Double Stimulus Impairment Scale (DSIS)

- *Categorical impairment grading scale:*



Double Stimulus Continuous Quality Scale (DSCQS)

- *Non-categorical adjectival or numerical grading scale:*



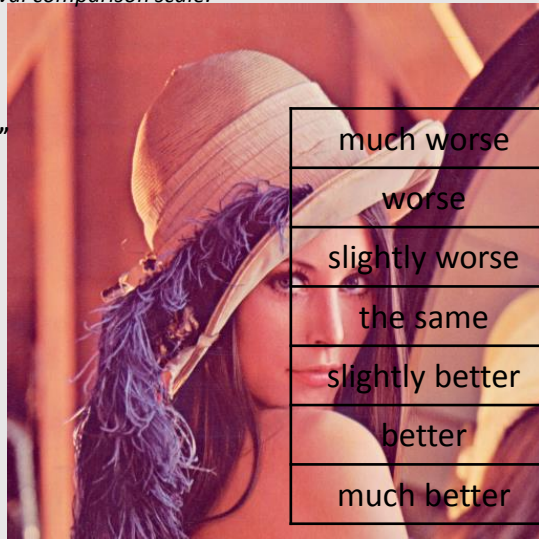
Tomado de: Quality of Multimedia Experience, Past, Present and Future, Touradj Ebrahimi, ACM Multimedia, 2009



Stimulus Comparison (SC)

- *Categorical adjectival comparison scale:*

“same or different”



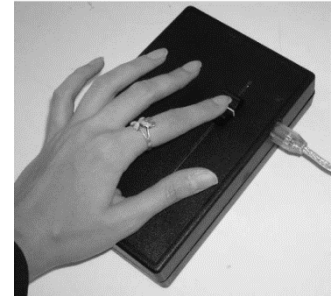
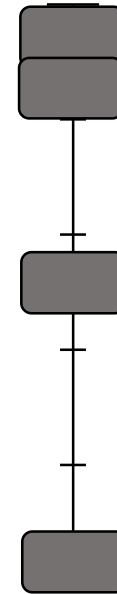
- *Non-categorical judgement:*



Single Stimulus Continuous Quality Evaluation (SSCQE)



(Very annoying)

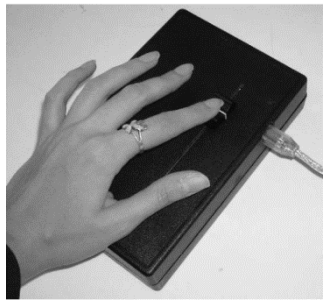
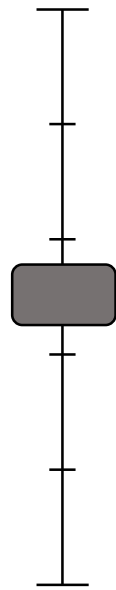


(Imperceptible)

Simultaneous Double Stimulus for Continuous Evaluation (SDSCE)



(Much better)



(Much worse)



- PSNR (Peak Signal to Noise Ratio)

$$MSE = \frac{1}{TMN} \sum_{n=1}^N \sum_{m=1}^M \sum_{t=1}^T [x(m,n,t) - y(m,n,t)]^2$$

$$RMSE = \sqrt{MSE}$$

$$PSNR = 10 \log_{10} \left(\frac{L^2}{MSE} \right)$$

Evaluaciones de imágenes comprimidas con JPEG

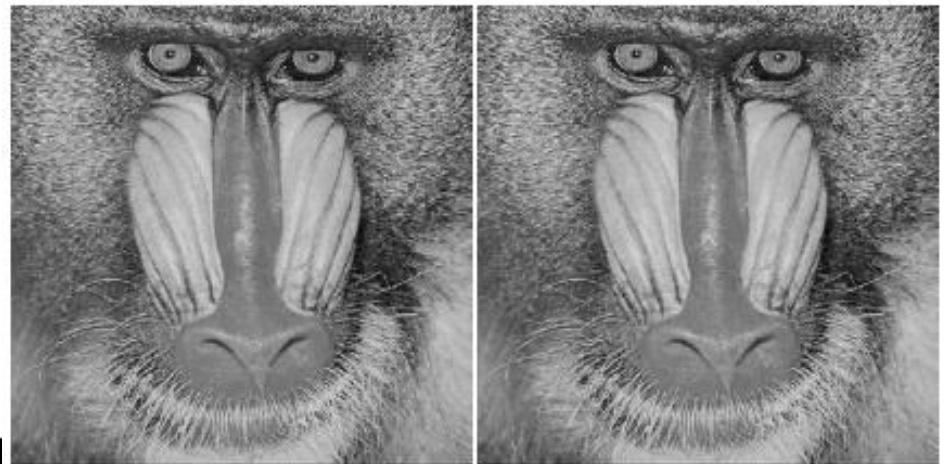
“Tiffany” original y comprimida
MSE=165



“Lago” original y comprimida
MSE=167



“Mandrill” original y comprimida
MSE=163



Evaluación de imágenes comprimidas



ORIGINAL

COMPRESIÓN A BAJA TASA DE BIT

COMPRESIÓN A ALTA TASA DE BIT



- Métodos objetivos
 - FR - Full Reference
 - Se basan en la disponibilidad completa de la señal original
 - RR – Reduced Reference
 - Se trata de enviar, junto con el video codificado, algunos parámetros que caractericen a la señal, y que sirvan de referencia en el receptor para poder estimar la calidad percibida
 - NR - No Reference
 - Intentan estimar la calidad percibida basándose únicamente en el análisis de la señal recibida



- El VQEG (Video Quality Expert Group) está llevando a cabo un gran trabajo sistemático y objetivo de comparación de modelos
- El objetivo del VQEG es proporcionar evidencia para los organismos internacionales de estandarización acerca del desempeño de diversos modelos propuestos, a los efectos de definir una métrica estándar y objetiva de calidad percibida de video digital (VQM – Video Quality Metric)



- Proyectos terminados
 - FR-TV (Full Reference TV)
 - RRNR-TV (Reduced Reference, No Reference TV)
 - Multimedia
 - HDTV
 - Hybrid Perceptual / Bitstream
- Algunos proyectos en curso (2017)
 - Audivisual HD (AVHD)
 - High Dynamic Range Video / Wide Color Gamut
 - Monitoring of Audio Visual Quality by Key Indicators (MOAVI)
 - Ultra HD
 - Quality Assessment for Recognition and Task-based multimedia applications (QART)



- ITU ha estandarizado en junio de 2004 a los cuatro mejores algoritmos de VQEG FRTV en las recomendaciones ITU-T J.144 y ITU-R BT.1683
 - British Telecom BTFR (Reino Unido)
 - Yonsei University/Radio Research Laboratory/SK Telecom (Corea)
 - Centro de Investigación y Desarrollo en Telecomunicaciones (CPqD) (Brasil)
 - National Telecommunications and Information Administration (NTIA)/Institute for Telecommunication Sciences (ITS) (Estados Unidos)



- ITU ha estandarizado en agosto de 2008 a los mejores algoritmos de VQEG MM en las recomendaciones ITU-T J.246 y ITU-T J.247
- Se trata de modelos del tipo FR, RR para aplicaciones multimedia, en formatos de pantalla VGA, CIF y QCIF
- Los modelos NR no presentaron resultados lo suficientemente satisfactorios
- En total se presentaron 5 modelos, propuestos por NTT, OPTICOM, Psytechnics, SwissQual y Yonsei University



- Se han evaluado modelos del tipo FR y RR, para la predicción de la calidad de video percibida en aplicaciones de televisión digital de alta resolución (HDTV).
- Las pruebas se limitaron a codecs MPEG-2 y H.264, incluyendo eventuales errores de transmisión.
- La resolución fue 1080i @ 50/60 Hz y 1080p @ 25/30 fps
- El reporte final de VQEG fue aprobado en junio de 2010
 - Se presentaron modelos propuestos por NTT, OPTICOM, SwissQual, Tektronix y Yonsei University.
 - El modelo del tipo FR que tuvo mejor desempeño fue el propuesto por SwissQual, seguido del de Tektronix
- Se estandarizó en la Recomendación ITU-T J.341 en enero de 2011



- Modelos objetivos de calidad de vídeo que utilizan tanto la secuencia de vídeo procesada como la información de flujo de bits (“bitstream”).
- Videos WVGA / VGA / HDTV.
- Modelos NR, RR, FR
- Se estandarizó en 2014 en ITU-T J.343 (varias recomendaciones, para NR, RR y FR)



Calidad de Video sobre redes IP

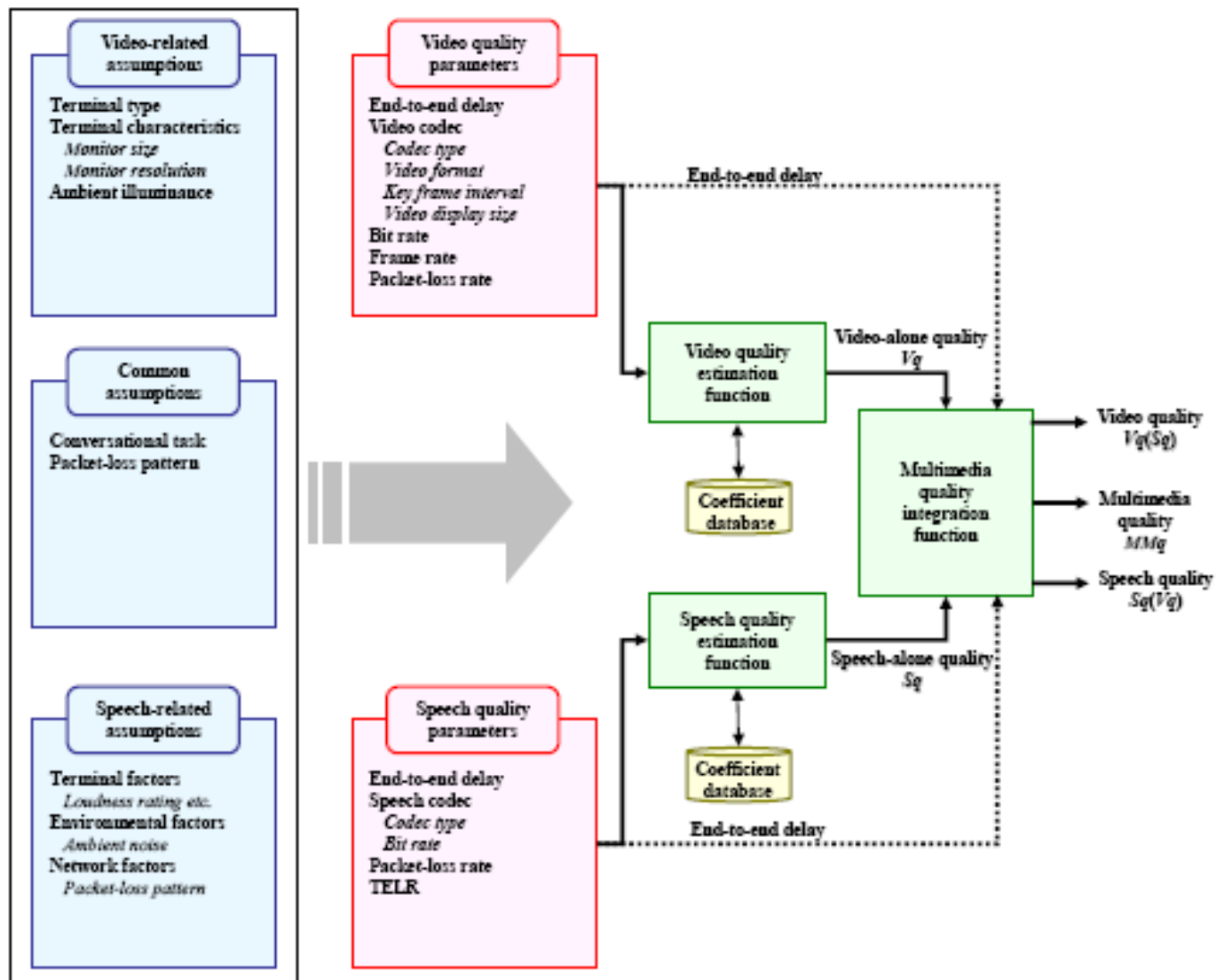
Calidad de voz y video



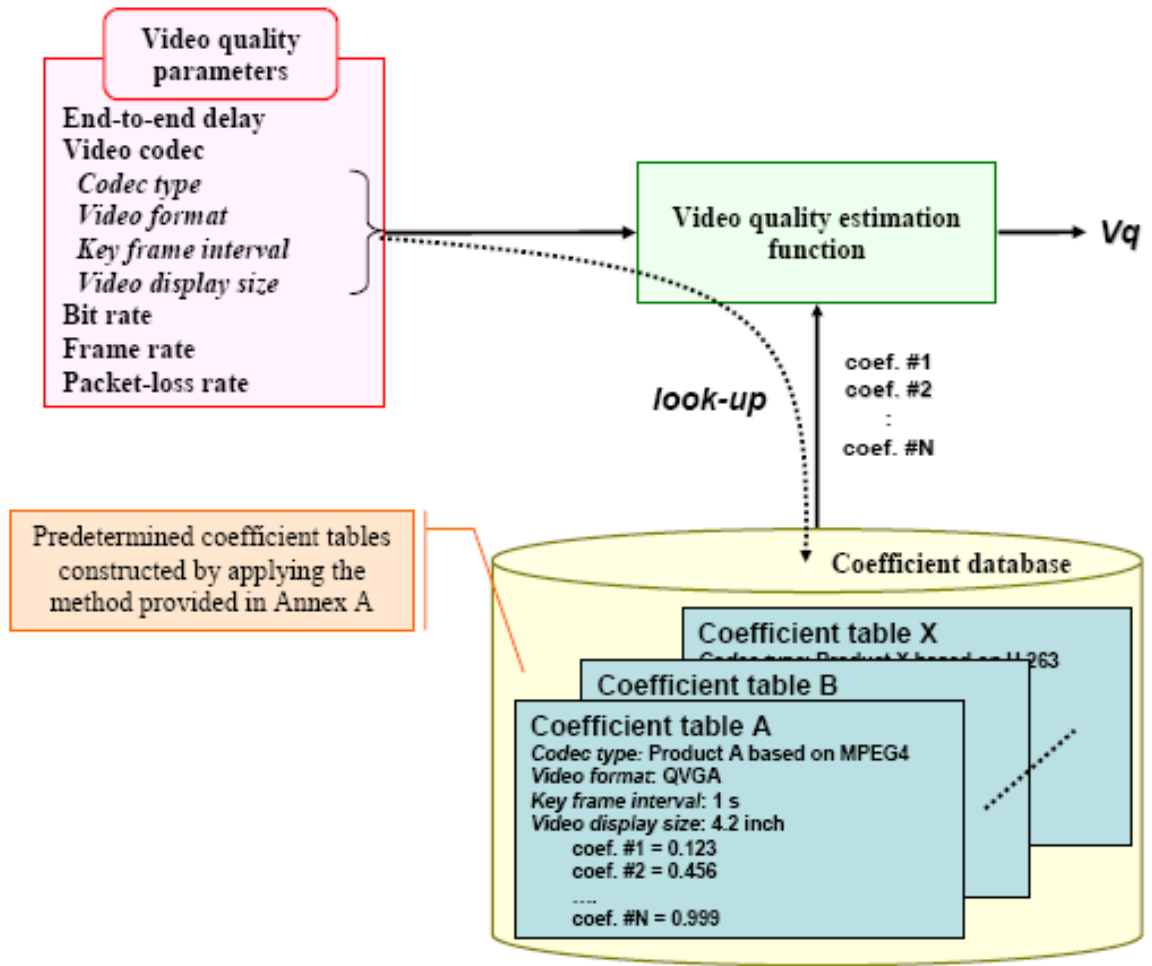
- Aprobada por ITU-T en abril 2007, sobre la base de propuestas de NTT. Actualizada en 2012.
- Propone un algoritmo de estimación de la calidad de video teléfonos en ambientes de redes de datos
- Para ser utilizada como herramienta de diseño o planificación
- Estima tres parámetros de calidad
 - Sq Speech Quality
 - Vq Video Quality
 - MMq Multimedia Quality



ITU-T G.1070 Framework



ITU-T G.1070: Coeficientes para cada Codec



- Básicamente se reduce al E-Model, simplificado

$$Q = R_o - I_d - I_{e,eff}$$

$$S_q = f(Q), \text{ similar al E-Model}$$

- Los efectos de la demora se incluyen en el MM_q , y por lo tanto, se excluyen del S_q



- Se propone

$$V_q = 1 + I_c e^{-\frac{P_{plv}}{D_{Pplv}}}$$

$I_c = f(\text{codec, bitrate, frame rate})$

$D_{Pplv} = f(\text{codec, bitrate, frame rate})$



- Se propone

$$MM_q = m_1 MM_{SV} + m_2 MM_T + m_3 MM_{SV} MM_T + m_4$$

Calidad AudioVisual

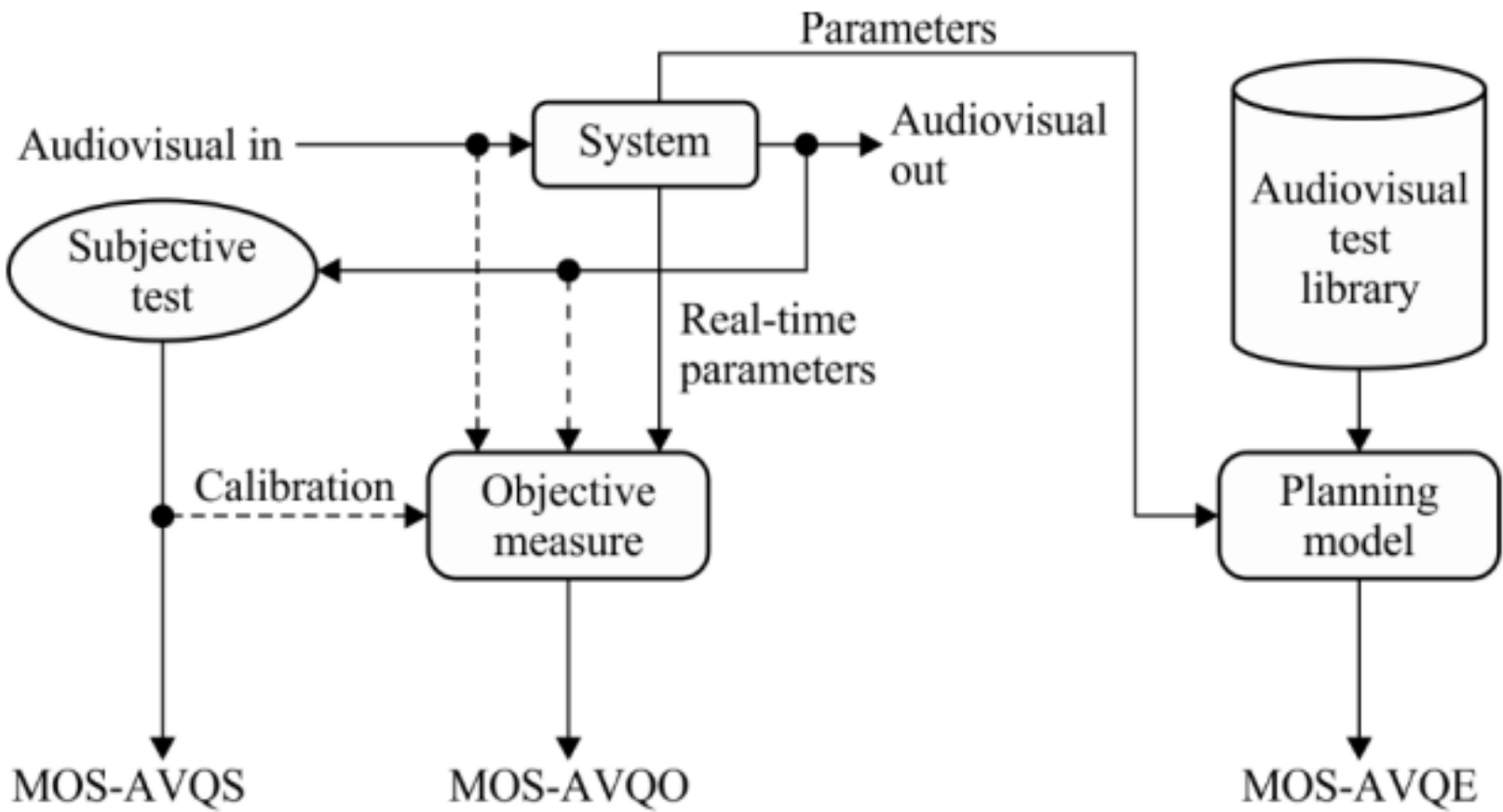
$$MM_{SV} = f(Vq, Sq)$$

Efectos de las demoras

$$MM_T = f(\text{Speech Delay}, \text{Video Delay})$$



Resumen (Calidad Audiovisual)



P.800.1(16)_F03



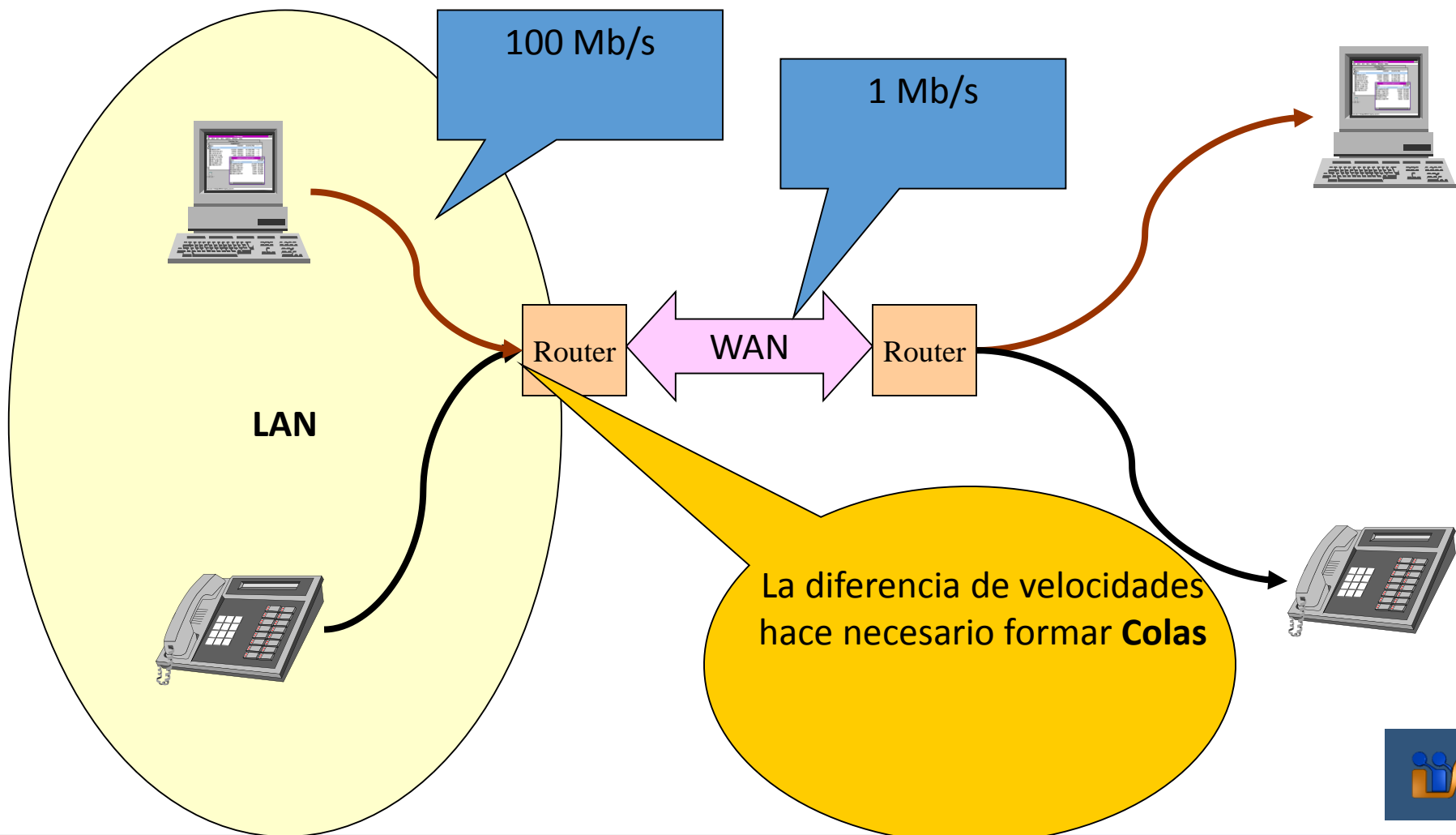
QoS en Redes de Datos



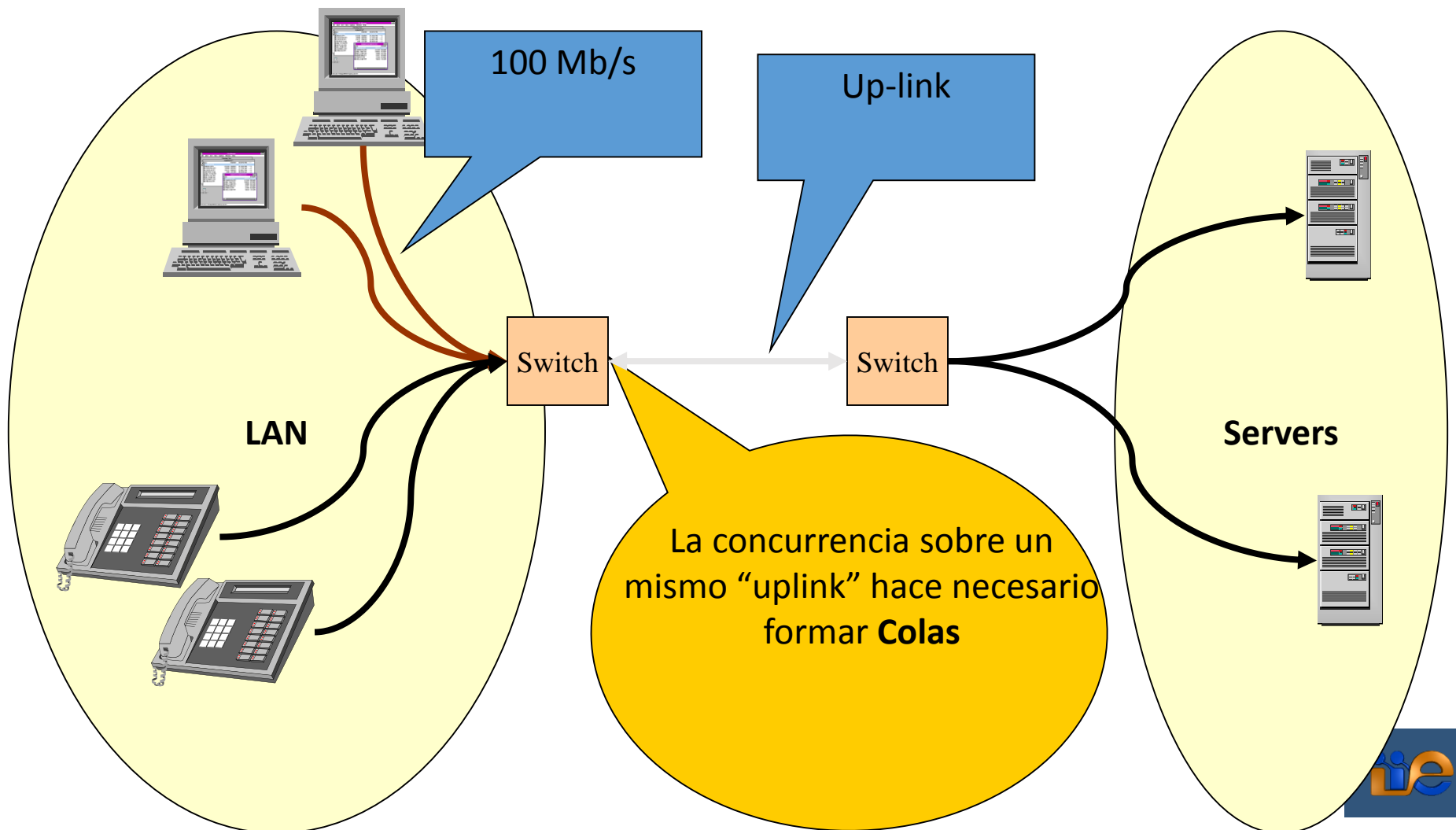
- Algunas de las técnicas utilizadas
 - Priorización
 - A nivel de capa 2, capa 3, capa 4, etc.
 - Fragmentación
 - Necesaria en enlaces de baja velocidad
 - Control de los retardos máximos
 - Fundamental para la calidad conversacional



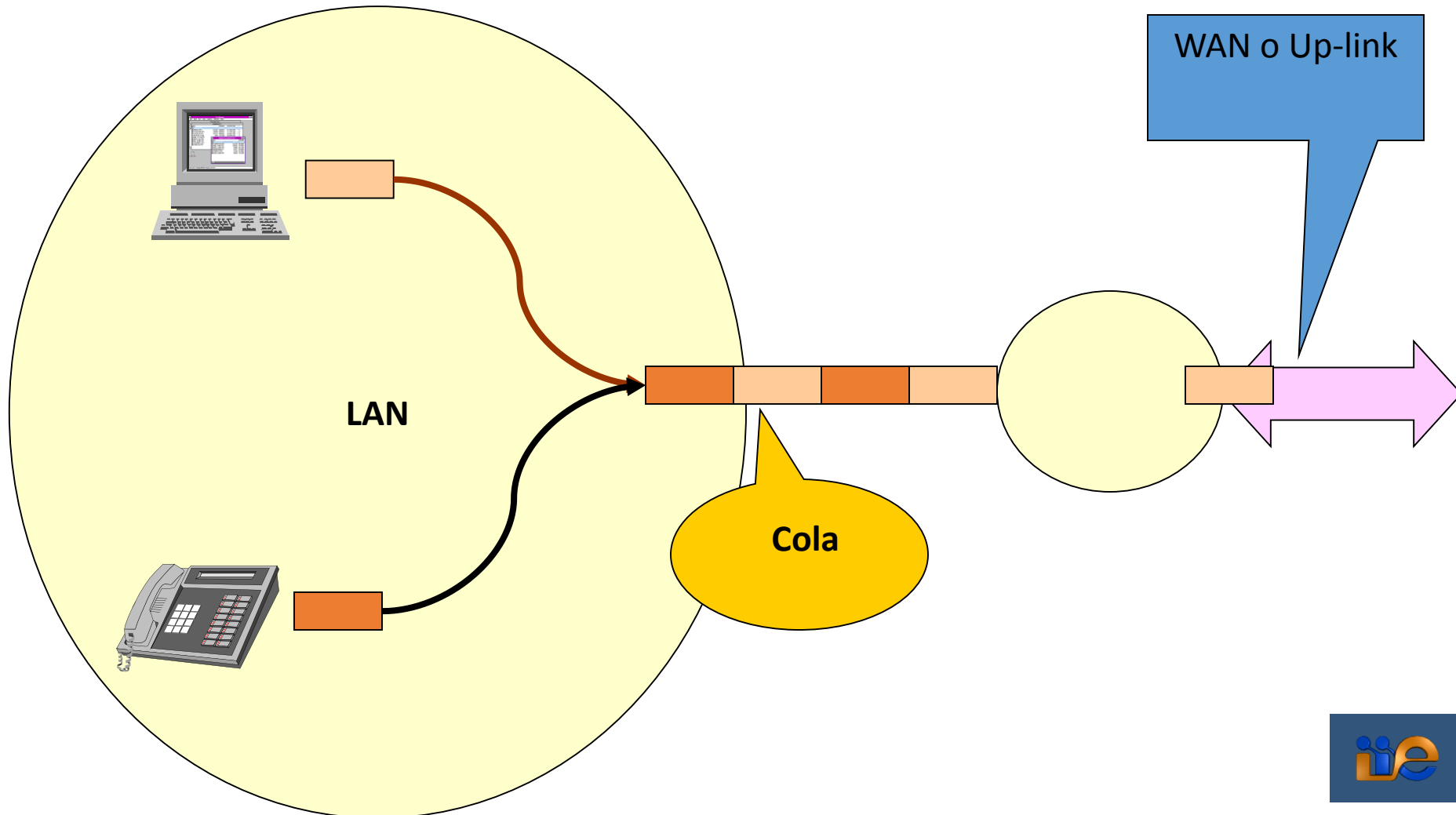
Problemas en enlaces de baja velocidad



Problemas en enlaces compartidos



Problemas en enlaces compartidos



Priorización

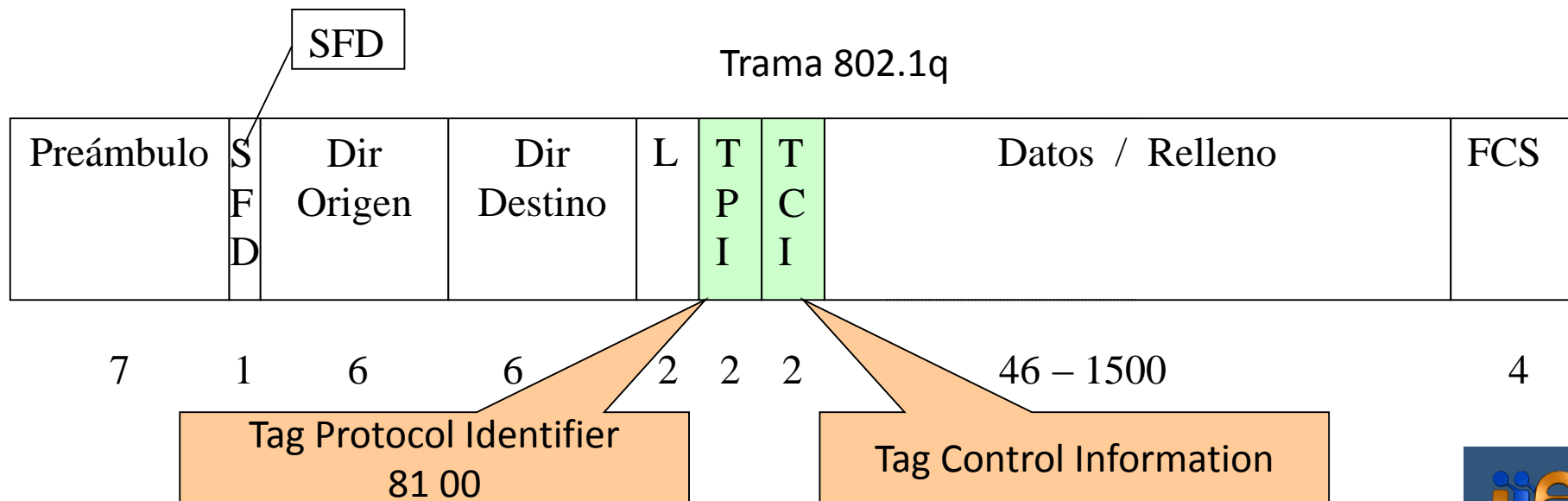
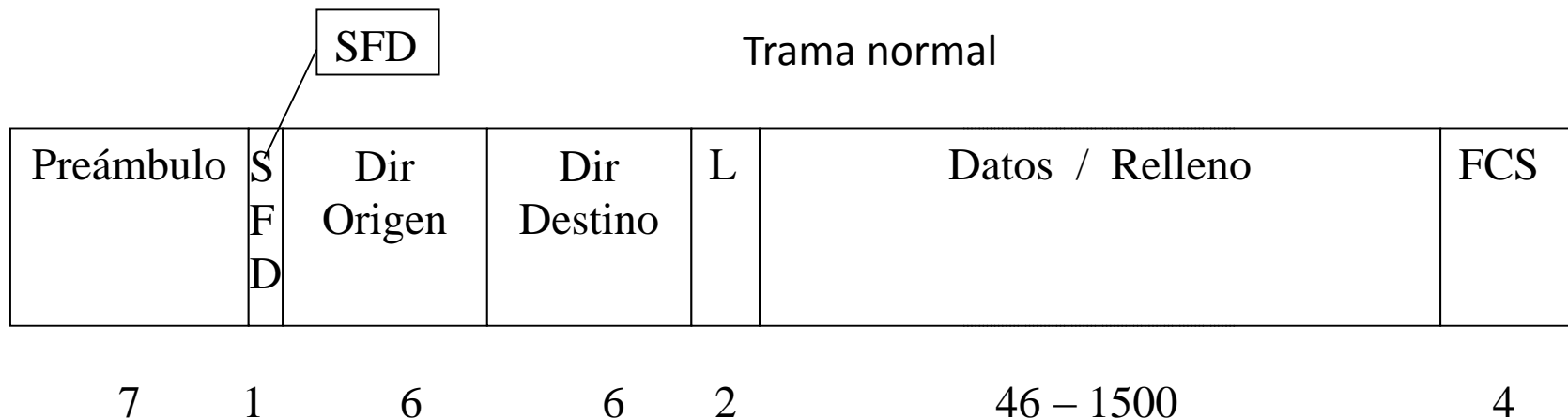
- Permite “marcar” tramas, paquete o cierto tipo de tráfico con diferentes prioridades
- En los switches o routers se pueden formar varias “colas”, según las prioridades de los paquetes



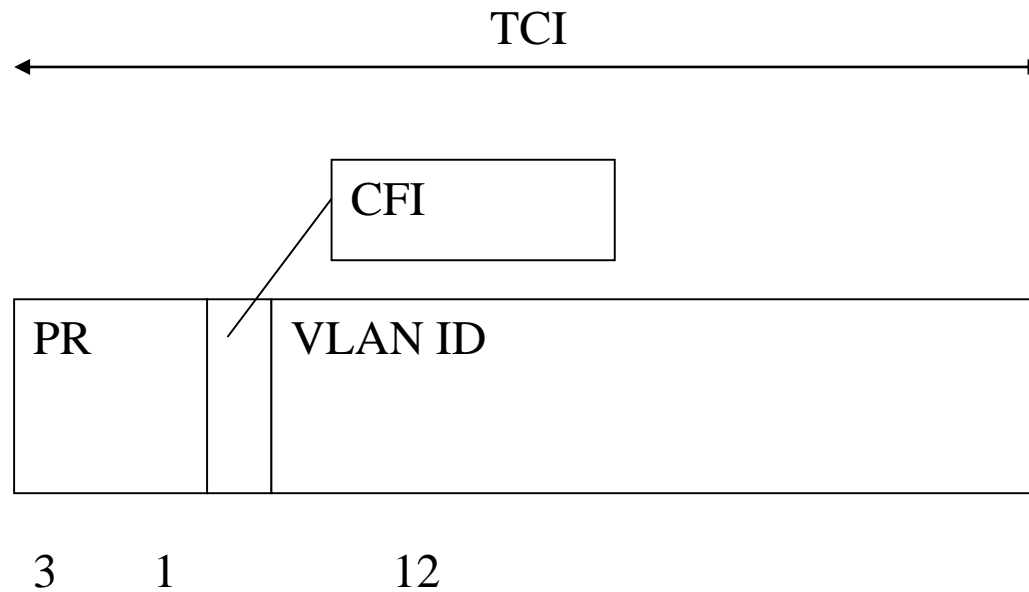
- Las recomendaciones IEEE 802.1q y IEEE 802.1p incorporan 4 bytes adicionales a las tramas Ethernet, donde se puede incluir información acerca de VLANs y etiquetas que identifican la “prioridad” de la trama.



Tramas 802.1q



Campo TCI: Tag Control Information



- PR = Prioridad
- CFI = Canonical Format Indication
- VLAN ID = Identificador de VLAN (LAN Virtual)



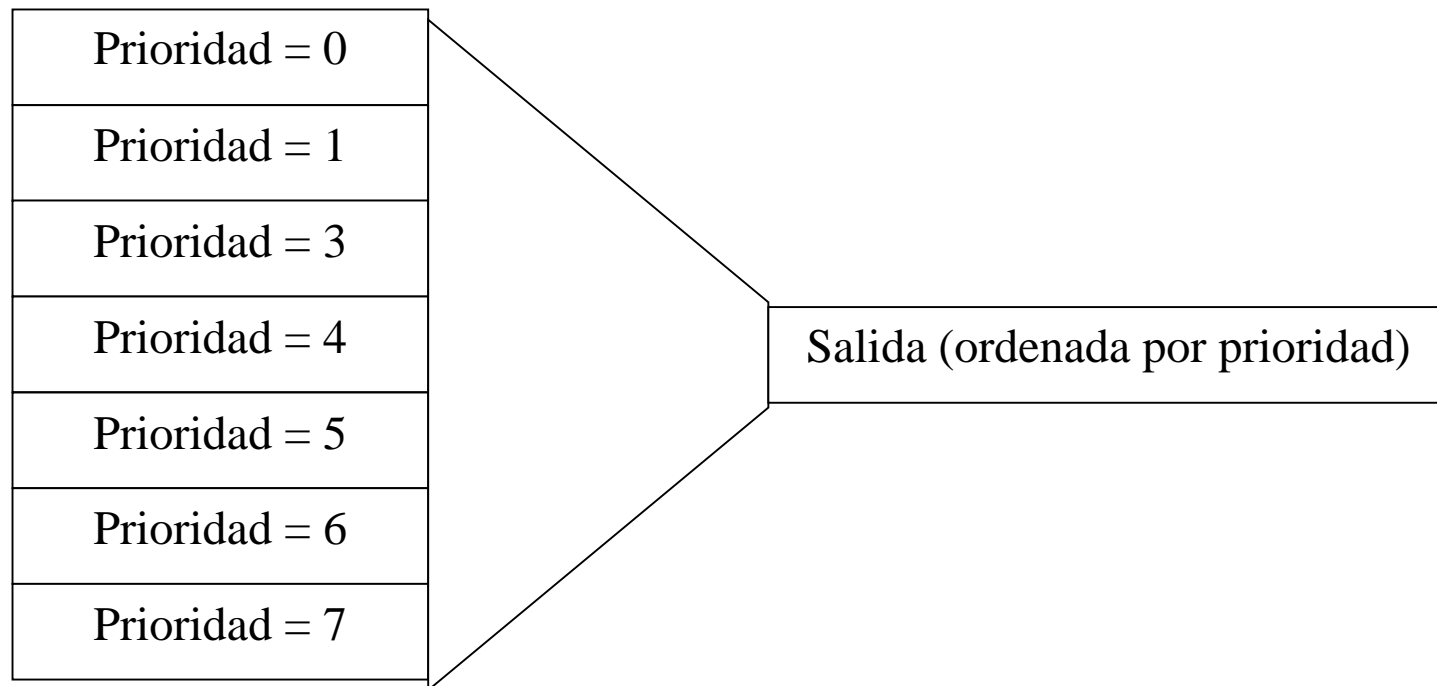
Prioridad en 802.1p

- Permite 8 prioridades: 0-7

user_priority	Acronym	Traffic type
1	BK	Background
2	—	Spare
0 (Default)	BE	Best Effort
3	EE	Excellent Effort
4	CL	Controlled Load
5	VI	“Video,” < 100 ms latency and jitter
6	VO	“Voice,” < 10 ms latency and jitter
7	NC	Network Control



- Colas



Estrategias de encolamiento y priorización

- **FIFO (First In, First Out)**
 - El primer paquete que haya ingresado en una cola, es el primero en salir.
- **PQ (Priority Queuing)**
 - La salida de los paquetes se realiza según el orden estricto de prioridad, y dentro de cada prioridad, según el orden de llegada. Este tipo de encolamiento puede hacer que, si existe siempre tráfico de alta prioridad, el tráfico de baja prioridad nunca sea enviado.
- **FQ (Fair Queuing)**
 - Es un esquema en el que cada cola se accede en forma circular, asegurando una distribución uniforme de ancho de banda entre todas las colas.



Estrategias de encolamiento y priorización

- **WRR (Weighted Round Robin)**
 - Permite asignar diferentes anchos de banda a cada cola.
- **WFQ (Weighted Fair Queuing)**
 - Es una combinación de PQ y FQ, garantizando que aplicaciones de alto tráfico no monopolicen el enlace.



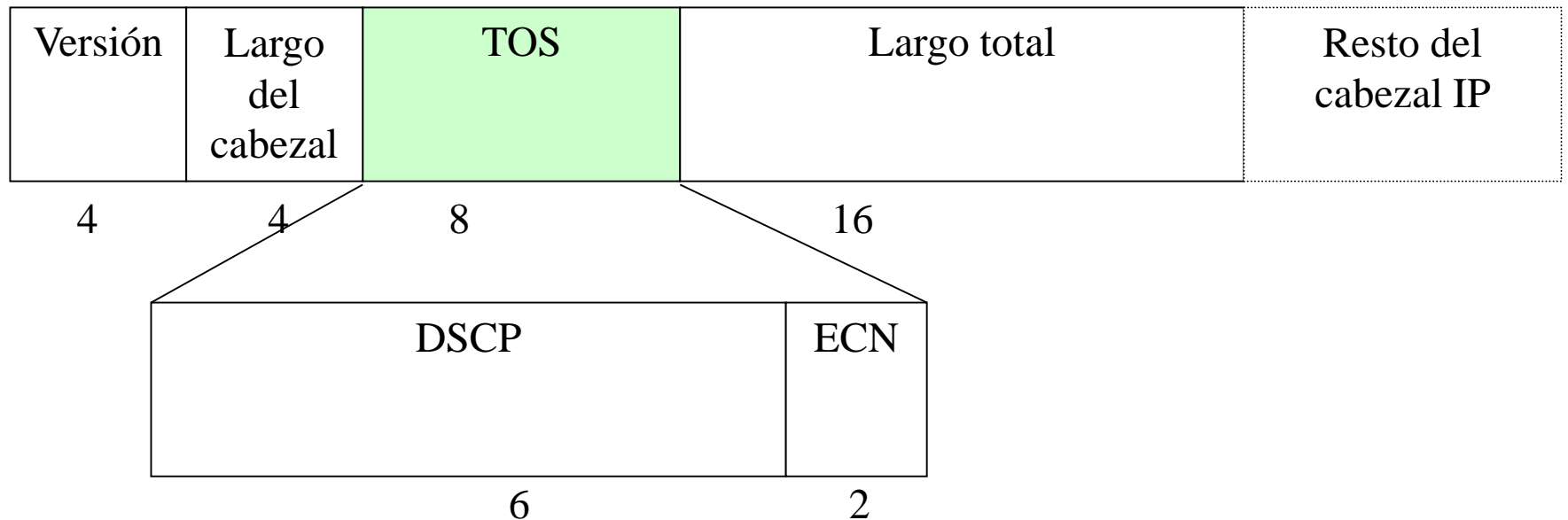
- Muchos switches de datos permiten implementar cierta priorización del tráfico basado en VLANs
- De esta forma, se puede poner a todos los dispositivos de VoIP en la misma VLAN, y darle prioridad frente al tráfico de otras VLANs, dedicadas a aplicaciones de datos
- Adicionalmente, en este caso el tráfico de voz no se ve afectado por el de datos



- **DiffServ** (Differentiated Services) es comúnmente utilizado para gestionar prioridad en los paquetes
- La información de priorización se encuentra en el cabezal del paquete IP, en un campo llamado **TOS** (Type Of Service)



Cabezal IP con TOS



- DSCP = Differentiated Services Code Point
- ECN = Explicit Congestion Notification



DSCP: Differentiated Services Code Point

- Es posible codificar hasta $2^6 = 64$ posibles prioridades.
- De éstas, 32 están reservadas para usos experimentales
- 32 pueden ser utilizadas
 - 21 están estandarizadas por el IETF
 - Las prioridades estandarizadas se dividen en 3 grupos



- **DE (Default)**

- Se asume el comportamiento por defecto, utilizando por tanto técnicas de encolamiento de “mejor esfuerzo”. El valor típico de DSCP para este tipo de tráfico es 000000.

- **AF (Assured Forwarding)**

- Estandarizado en el RFC 2597, donde se definen 4 clases de prioridades dentro de este tipo de priorización.

- **EF (Expedited Forwarding)**

- Estandarizado en el RFC 3246, establece las máximas prioridades para el tráfico marcado con este identificador. El valor típico de DSCP utilizado es 101110 (46 decimal).



ECN: Explicit Congestion Notification

- Permite conocer el estado de congestión del destino.
- Es utilizado para que el destino pueda indicarle a la fuente, aún antes de perder paquetes, que existe cierto estado de congestión, de manera que la fuente pueda tomar los recaudos apropiados, por ejemplo, disminuyendo el ancho de banda utilizado.
 - ECN = 11 indica que existe congestión
 - ECN = 10 o 01 indican que no existe congestión.
 - ECN = 00 indica que el extremo distante no soporta la función de notificación de congestión.



Otros mecanismos de priorización en Capa 3

- RSVP (Resource Reservation Protocol)
 - Establece los mecanismos para reservar cierto ancho de banda en la comunicación entre dispositivos que pasen a través de routers.
- El tráfico también puede ser priorizado en base a la dirección IP de origen o destino.
 - Esto puede ser implementado cuando se utilizan direcciones IP estáticas.

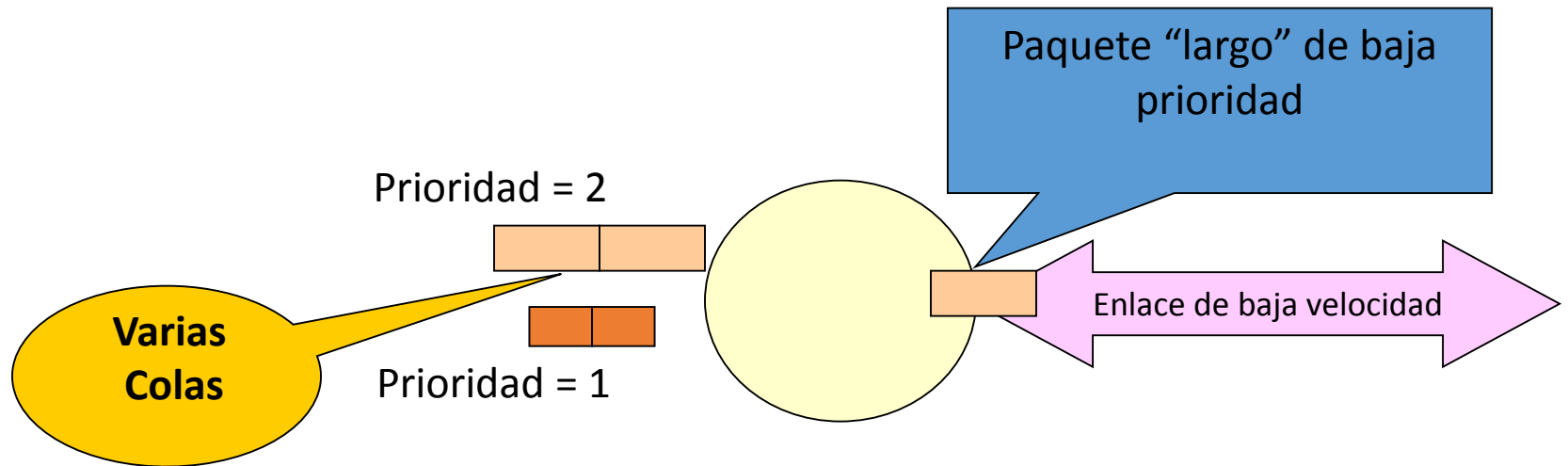


QoS en Capa 4 y superiores

- Los paquetes de datos pueden ser priorizados en base a los puertos TCP o UDP.
 - Sin embargo, diferentes aplicaciones podrían utilizar los mismos puertos, por lo que este tipo de priorizaciones debe ser evaluada en cada caso.
- Es posible también tener prioridades según el protocolo de capas superiores.
 - Por ejemplo, puede ser priorizado el tráfico RTP respecto a otros, y asignarlo a las colas de alta prioridad.

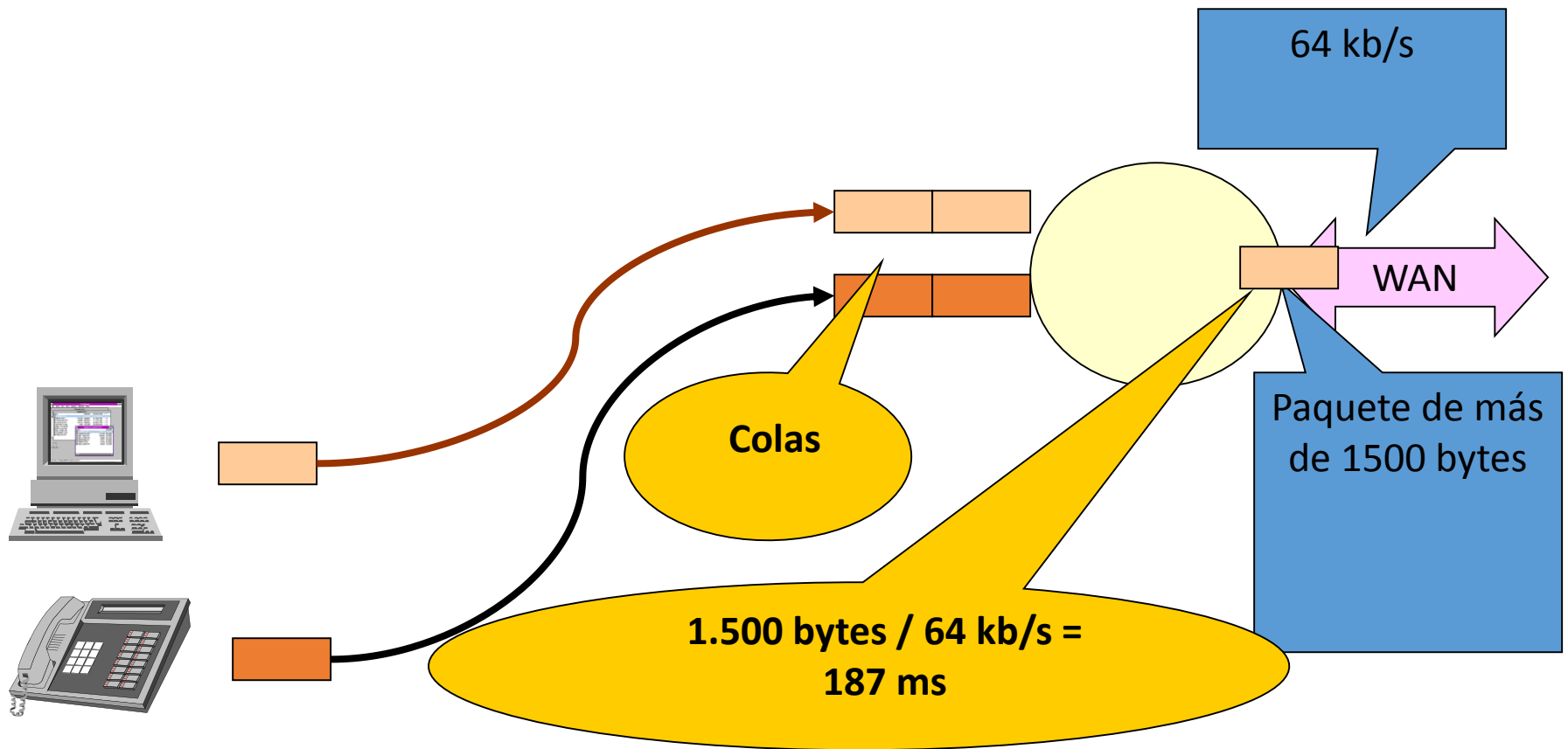


Fragmentación



- Las colas y prioridades no resuelven el problema de “paquetes largos sobre enlaces de baja velocidad”
- Es necesario “Fragmentar”

Fragmentación



Fragmentación

	Trama [Bytes]	
	Ethernet MTU 1500	Paq. de Voz 32
Velocidad [Kbps]	Tiempo [ms]	Tiempo [ms]
64	187,50	4,00
128	93,75	2,00
256	46,88	1,00
512	23,44	0,50
1024	11,72	0,25
2048	5,86	0,13
34000	0,35	0,01
155000	0,08	0,00
622000	0,02	0,00

Requiere fragmentación en las tramas de datos



Bibliografía y material de referencia

- VQEG (<https://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx>)
- Pruebas subjetivas de Calidad de Experiencia en aplicaciones multimedia, Revisión y Tutorial, Marcos Juayek, Enero 2016
- Multimedia Quality of Experience (QoE), Current Status and Future Requirements, Chang Wen Chen et al, Wiley, 2016
- Qualinet White Paper on Definitions of Quality of Experience, Output from the fifth Qualinet meeting, Novi Sad, March 12, 2013



Calidad de voz y video

¡Muchas Gracias!

