CALIDAD
DE
VOZ Y VIDEO

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

Marzo 2015
Versión 5

Contenido

1 Calidad de la voz

1.1 Introducción

La voz puede sufrir diversos tipos de distorsiones y degradaciones a lo largo de un sistema de telecomunicaciones. Estas distorsiones pueden estar generadas por diversos aspectos, entre los que se pueden mencionar las distorsiones en el sistema de transmisión analógico, las introducidas por los codificadores (codecs), las introducidas por ruidos externos al sistema, la demora, el eco y más recientemente, la pérdida de paquetes en redes de datos.

El progreso de la tecnología ha hecho que la calidad percibida de la voz y las conversaciones en los sistemas de telecomunicaciones hayan sufrido cambios a lo largo del tiempo. Originalmente, las distorsiones que se producían en las redes analógicas estaban centradas en las variaciones de las pérdidas y el ruido externo. Con la digitalización de la red, y con los terminales (teléfonos) electrónicos, estos factores su fueron mitigando y minimizando, pero comenzaron a introducirse otros, como las distorsiones de la codificación. La telefonía celular introdujo nuevos factores de distorsión, entre ellos las demoras y el eco. Finalmente, la Voz sobre IP ha introducido nuevos factores de distorsión, dados por la pérdida de paquetes y ha agravado algunos, como las demoras y el eco. Esto ha llevado a que los aspectos de la calidad de la voz y las conversaciones hayan tomado especial relevancia en la actualidad, ya que muchas veces, con los sofisticados sistemas inalámbricos e IP, los usuarios terminan percibiendo una calidad peor de la que se tenía hace décadas, con la telefonía analógica. En la siguiente figura ilustra la evolución de la calidad de la voz a lo largo de los cambios tecnológicos.

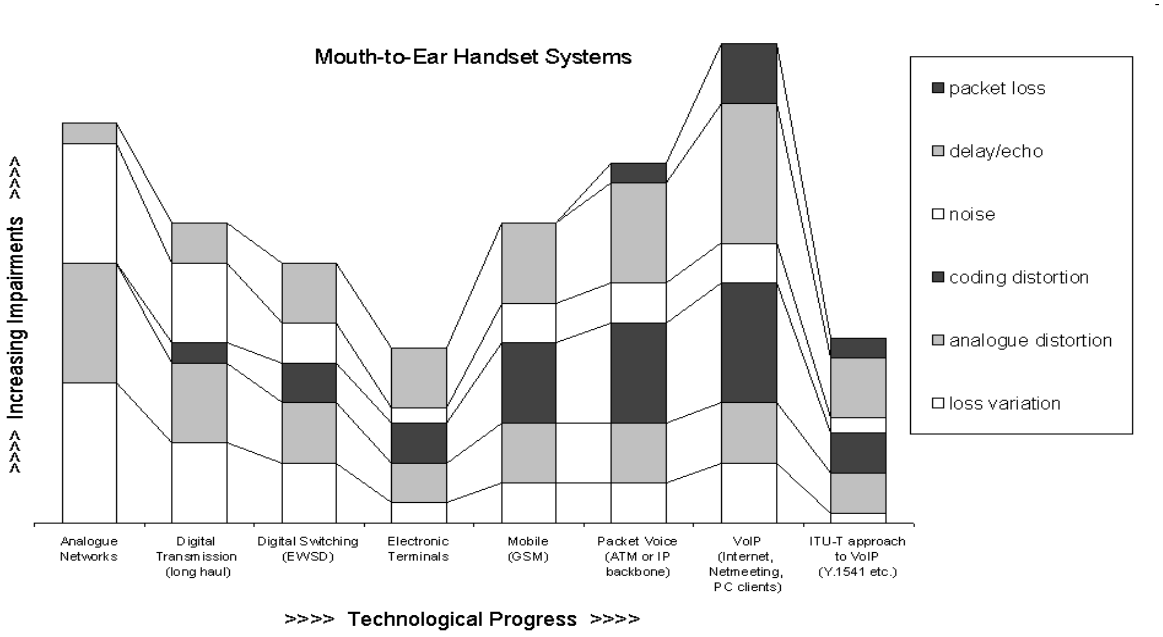


Figura 1.1

1.2 Medida de la calidad de voz

Para evaluar la calidad de voz percibida se han estandarizado métodos. Estos métodos se dividen en *subjetivos* y *objetivos*. Los métodos *subjetivos* de medida de la calidad de voz, se basan en conocer directamente la opinión de los usuarios. Típicamente resultan en un promedio de opiniones (es decir, en un valor de MOS – Mean Opinion Score). Estos métodos se describen en las siguientes secciones.

Los métodos *objetivos* intentan predecir la calidad percibida mediante la aplicación de algoritmos. Estos métodos objetivos, a su vez se subdividen en *intrusivos* (se inyecta una señal de voz conocida en el canal y se estudia su degradación a la salida) y *no intrusivos* (monitorean ciertos parámetros en un punto de la red y en base a estos permite establecer en tiempo real la calidad que percibiría un usuario). Estos métodos se describen en las siguientes secciones.

La terminología utilizada para describir los posibles resultados del promedio de opiniones (MOS) se describe en la Recomendación ITU-T P.800.1 [1], y se puede resumir en la siguiente tabla.

	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

La letra “y” se corresponde con N para pruebas de audio de banda angosta (300 – 3400 Hz) y con W para pruebas de audio de banda ancha (50– 7000 Hz).

MOS-xQS (MOS Quality Subjetive) es el MOS obtenido directamente de pruebas subjetivas.

MOS-xQO (MOS Quality Subjetive) es el MOS obtenido a través de modelos objetivos (como el P.862, descrito más adelante)

MOS-xQE (MOS Quality nEtwork Planning) es el MOS obtenido a través de modelos diseñados para la planificación de redes (como el G.107, descrito más adelante)

MOS-LQx (“Listenting –only”) refieren a pruebas de escucha únicamente.

MOS-CQx (“Conversational”) refieren a pruebas de calidad de la conversación.

MOS-TQx (“Talking”) refieren a pruebas de calidad del retorno del audio de la persona que habla (por ejemplo eco, retorno alto, ruido mientras habla, etc.).

1.3 Métodos Subjetivos de medida de la calidad de voz

La calidad de la voz se establece a través de la opinión del usuario. La calidad de audio puede ser evaluada directamente (ACR = Absolute Category Rating), o en forma comparativa contra un audio de referencia (DCR = Degradation Category Rating). Con evaluaciones directas (del tipo ACR) se califica el audio con valores entre 1 y 5, siendo 5 “Excelente” y 1 “Malo”. El MOS (Mean Opinión Store) es el promedio de los ACR medidos entre un gran número de usuarios.

Si la evaluación es comparativa (del tipo DCR), el audio se califica también 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. El promedio de los valores DCR es conocido como DMOS (Degradation MOS).

La metodología de evaluación subjetiva más ampliamente usada es la del MOS (Mean Opinion Score), estandarizada en la recomendación ITU-T P.800 [2]. Adicionalmente, se puede evaluar la calidad del audio y la calidad de la conversación, las que pueden ser diferentes. La calidad de la conversación implica una comunicación bidireccional, donde, por ejemplo, los retardos juegan un papel muy importante en la calidad percibida. Los valores obtenidos con las técnicas ACR (es decir, el MOS) puede estar sujeto al tipo de experimento realizado. Por ejemplo, si se utilizan varias muestras de buena calidad, una en particular puede ser calificada peor que si esa misma muestra se presenta junto a otras de peor calidad.

Los métodos subjetivos son en general caros y lentos porque requieren un gran panel de usuarios. Son dependientes entre otros factores del país, del idioma, de las experiencias previas de los usuarios.

1.4 Métodos Objetivos de medida de la calidad de voz

Los métodos subjetivos de evaluación de la calidad de la voz son complejos de implementar, ya que requieren de ambientes controlados, y un número apreciable de opiniones de personas, las que deben ser ponderadas y promediadas. No son posibles de implementar en “tiempo real”, por lo que su aplicación se limita a estudios académicos y de laboratorio. Es por esto que existe gran interés en desarrollar métodos objetivos, que traten de predecir las calificaciones subjetivas, en base a parámetros medibles.

1.4.1 ITU-T P.862 (PESQ)

La recomendación ITU-T P.862 [3] presenta un método objetivo para la evaluación de la calidad vocal de extremo a extremo de redes telefónicas de banda angosta y codecs vocales.

Esta recomendación describe un método objetivo para predecir la calidad subjetiva de la voz telefónica utilizando los codecs más comunes de banda angosta. Presenta una descripción de alto nivel del método, explica la forma de utilizar este método y parte de los resultados de referencia obtenidos por la Comisión de Estudio 12 de la ITU-T en el periodo 1999-2000. Proporciona adicionalmente una implementación de referencia escrita en el lenguaje de programación ANSI-C.

El método objetivo descrito se conoce por "evaluación de la calidad vocal por percepción" (PESQ, *perceptual evaluation of speech quality*) y es el resultado de varios años de trabajos de desarrollo.

PESQ 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).

El primer paso de PESQ consiste en una alineación temporal entre las señales iniciales X(t) y degradada Y(t). Para cada intervalo de señal se calcula un punto de arranque y un punto de parada correspondientes.

Una vez alineadas, PESQ compara la señal (entrada) inicial con la salida degradada alineada, utilizando un modelo por percepción, como el representado en la Figura 1.2.

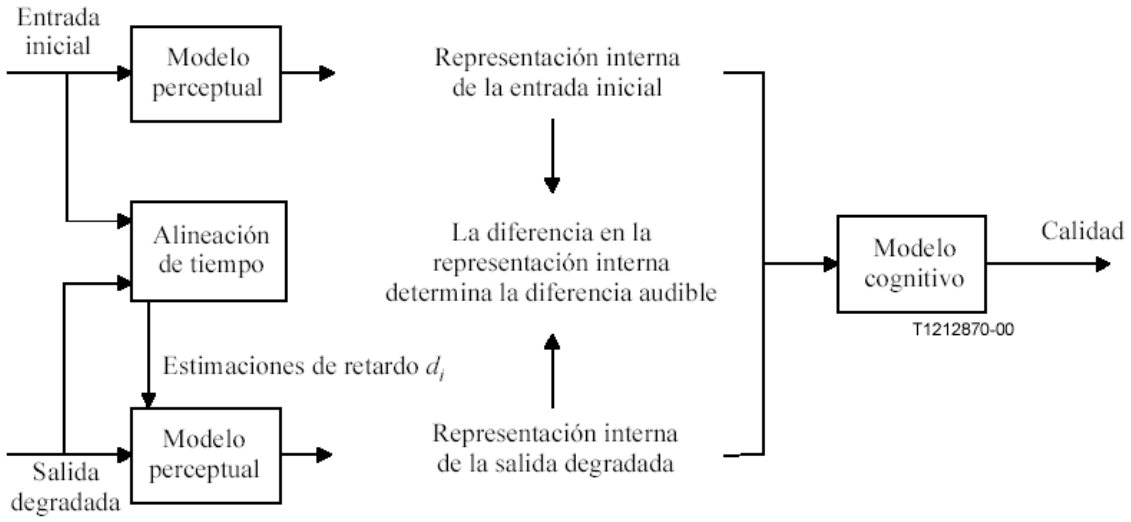


Figura 1.2

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 representación

psicoacústica de señales de audio en el sistema auditivo humano, teniendo en cuenta la frecuencia por percepción (Bark) y la sonoridad (Sone).

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 correspondencia entre la nota PESQ y la estimación del valor de MOS-LQO (Mean Opinion Score of Listening Quality Objective) se establece en la Recomendación ITU-T P.861.1 [4], y está dada por la siguiente relación:

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

Un gráfico de esta relación se presenta en la siguiente figura

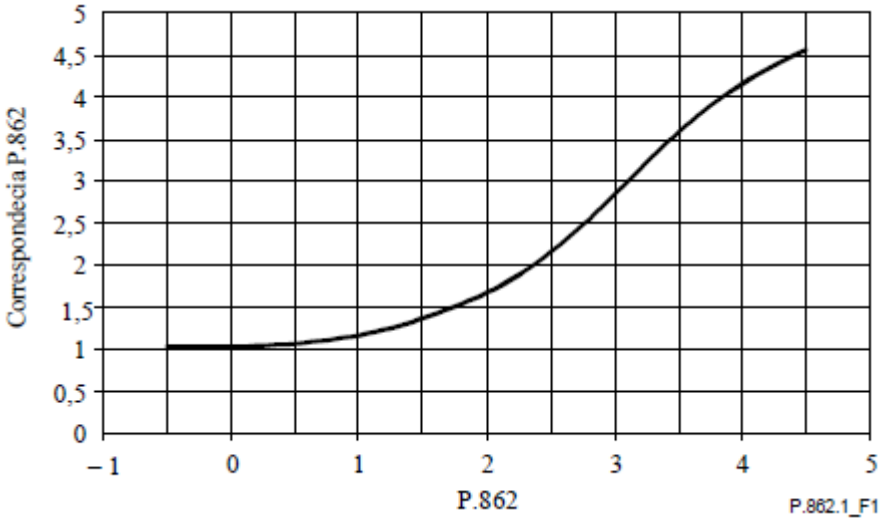


Figura 1.3

Los resultados de PESQ se encuentran en una escala de –0,5 a 4,5, aunque en la mayoría de los casos la gama de las salidas estará entre 1,0 y 4,5, que es la gama normal de valores de MOS que suelen darse en un experimento sobre la calidad de voz. En este rango, la relación de la gráfica se aproxima razonablemente a una recta.

La descripción detallada del algoritmo es compleja, y puede verse en la recomendación referenciada.

El método PESQ es objetivo e intrusivo, ya que requiere del envío de una señal conocida de referencia para evaluar la calidad percibida de la voz. Algunos sistemas lo implementan enviando un par de segundos de audio conocido, lo que basta para poder aplicar el método.

1.4.2 ITU-T P.862.2 (PESQ)

La recomendación ITU-T P.862.2 (aprobada en noviembre de 2007) [5] es una extensión de la recomendación original P.862, extendiendo el área de aplicación a señales de banda ancha (de 50 a 7000 Hz).

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. En este caso, la relación entre la nota PESQ de banda ancha y el MOS-LQO es

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

1.4.3 ITU-T P.862.3 (PESQ)

La recomendación ITU-T P.862.3 (aprobada en noviembre de 2007) [6] es un guía de aplicación de la serie de recomendaciones P.862. Establece lineamientos y áreas de aplicabilidad de estas recomendaciones.

Un esquema básico de aplicación se muestra en la siguiente figura

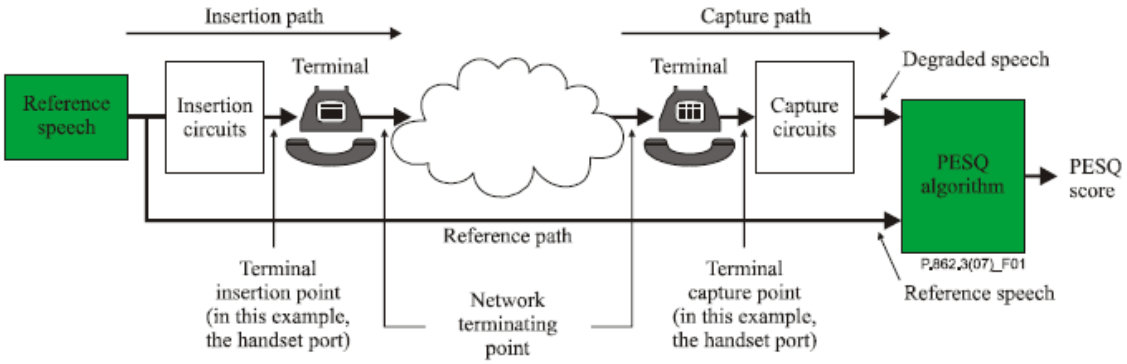


Figura 1.4

Esta recomendación también establece valores de referencia de MOS para diferentes codecs. Para el idioma español, se presenta un resumen en las siguientes tablas. Los valores indican el MOS-LQO estimado por PESQ.

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

1.4.4 ITU-T P.863 (POLQA)

En enero de 2011 ITU-T estandarizó la recomendación P.863 POLQA "Perceptual Objective Listening Quality Assessment", como una evolución de PESQ. Este nuevo estándar de estimación de calidad puede trabajar en la banda "super ancha" (de 50 a 14000 Hz), y es a su vez compatible con las bandas angosta y ancha.

Un cuadro gráfico comparativo de la evolución de la serie P.86x se muestra a continuación

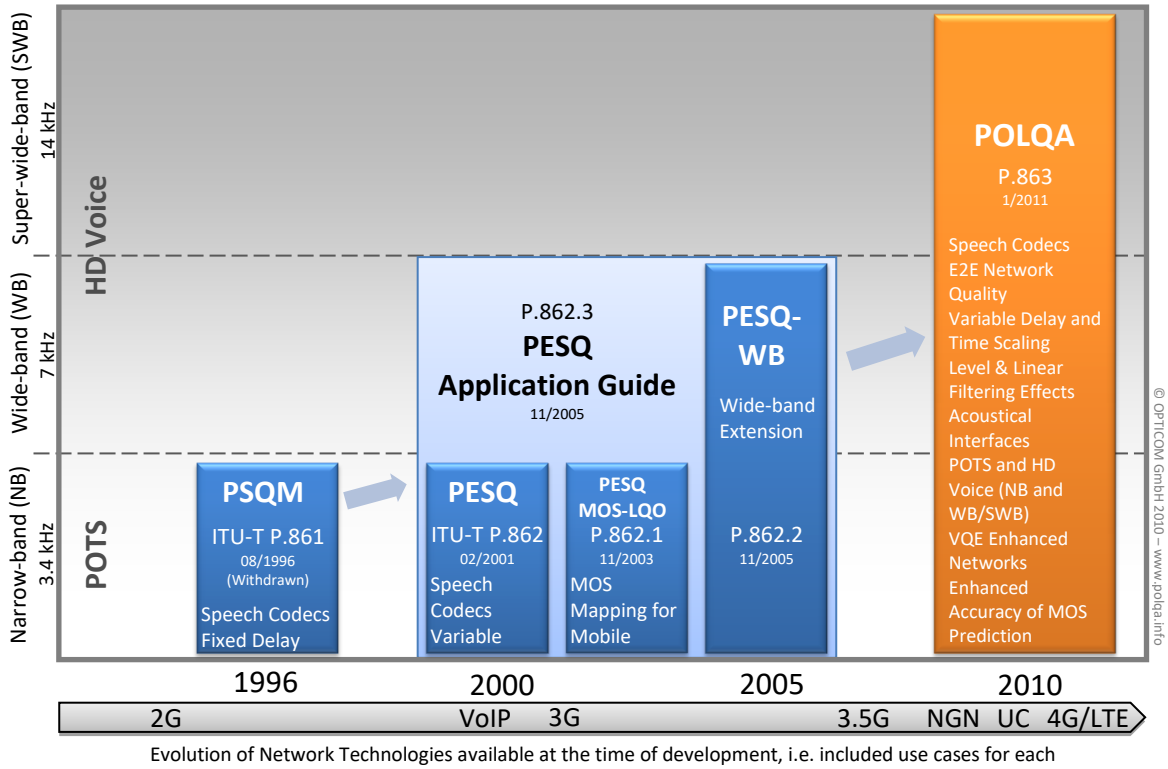


Figura 1.5

1.4.5 ITU-T P.563

El algoritmo P.563 es aplicable para la predicción de la calidad vocal sin una señal de referencia independiente. Por ese motivo, este método se recomienda para la evaluación no intrusiva de la calidad vocal y para la supervisión y evaluación con la red en funcionamiento, empleando en el extremo lejano de una conexión telefónica fuentes de señal vocal desconocidas.

En comparación con la Rec. ITU-T P.862 (que utiliza el método “basado en dos extremos” o “intrusivo”) 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. En la Figura 1.6 se ilustran las diferencias entre ambos métodos.

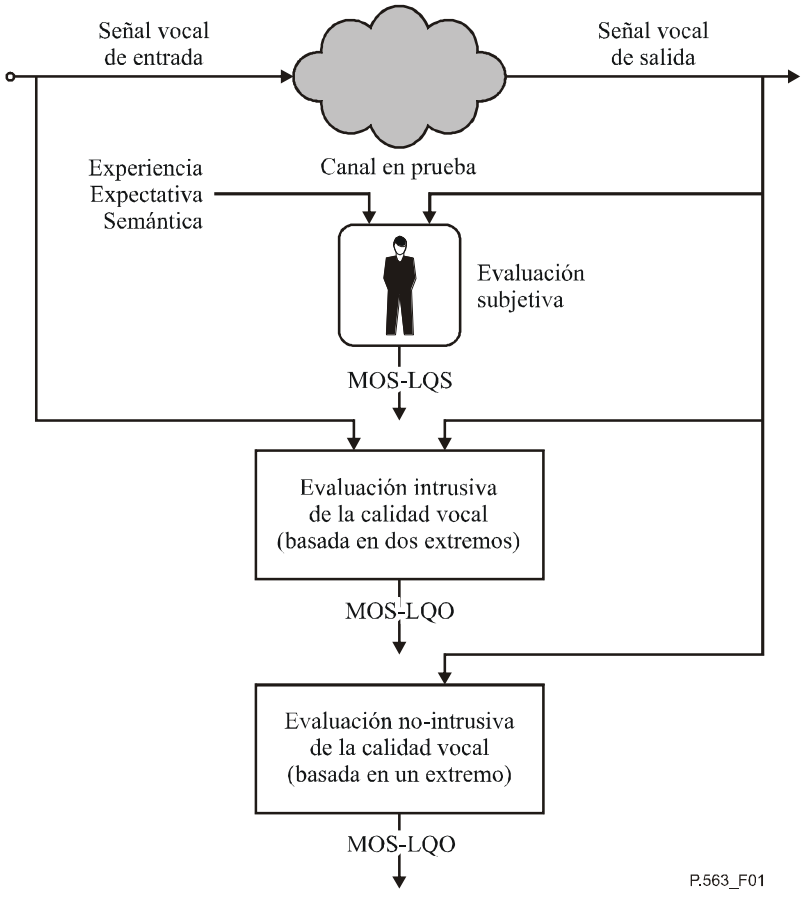


Figura 1.6

El enfoque utilizado en P.563 puede visualizarse como un experto que escucha una llamada real con un dispositivo de prueba, tal como un microteléfono convencional conectado en paralelo a la línea. Esta visualización permite explicar la principal aplicación y permite al usuario clasificar las puntuaciones obtenidas mediante P.563. La puntuación de calidad que se predice mediante P.563 está relacionada con la calidad percibida en extremo receptor.

La señal vocal que debe evaluarse se analiza de varias formas, que detectan un conjunto de **parámetros de señal** característicos. En base a un conjunto restringido de **parámetros clave** se establece la asignación a una **clase de distorsión** principal.

Básicamente, la parametrización de la señal del algoritmo P.563 puede dividirse en tres bloques funcionales independientes que se corresponden con las tres **clases de distorsión** principales:

- Análisis del tracto vocal y desnaturalización de la voz
  - voces masculinas
  - voces femeninas
  - marcada robotización
- Análisis de un ruido adicional intenso
  - SNR estática reducida (nivel básico del ruido de fondo)
  - SNR por segmentos reducida (ruido relacionado con la envolvente de la señal).
- Interrupciones, silenciamientos y recorte temporal

El modelo de calidad vocal de P.563 se compone de tres bloques principales:

1. Decisión sobre la clase de distorsión de que se trata.
2. Evaluación de la calidad vocal intermedia para la correspondiente clase de distorsión.
3. Cálculo global de la calidad vocal.

Cada clase de distorsión utiliza una combinación lineal de varios parámetros para generar la calidad vocal intermedia.
La calidad vocal definitiva se calcula combinando los resultados de calidad vocal intermedia con algunas características adicionales de la señal.

La descripción detallada del algoritmo es compleja, y puede verse en la Recomendación referenciada.

1.5 Calidad de voz en redes IP

La VoIP enfrenta problemáticas propias de las redes de datos, que se manifiestan como degradaciones en la calidad del servicio (QoS) o la calidad de la experiencia (QoE) percibida por los usuarios. Estas degradaciones pueden deberse por ejemplo a retardos, jitter (diferencia de retardos) y pérdida de paquetes, entre otros factores. Para que la tecnología de VoIP pueda ser utilizada tanto a nivel corporativo como a nivel de operadores telefónicos, es esencial garantizar una calidad de voz aceptable. Un análisis acerca de la medida de la calidad de voz se puede ver en [7].

Se analizarán a continuación los factores específicos que afectan la calidad de voz percibida sobre redes de paquetes.

1.5.1 Factores que afectan la calidad de la voz sobre redes de paquetes

Realizaremos una pequeña discusión acerca de los parámetros que influyen en la calidad de la voz transmitida a través de la red de datos:

1.5.1.1 Factor de compresión y codificación

Para poder transmitir la voz a través de una red de datos, es necesario realizar previamente un proceso de digitalización y codificación, el que puede degradar la señal de voz original, debido a la utilización de técnicas de compresión (Ver **¡Error! No se encuentra el origen de la referencia.**).

1.5.1.2 Pérdida de paquetes

A diferencia de las redes telefónicas, donde para cada conversación se establece un vínculo “estable y seguro”, las redes de datos admiten la pérdida de paquetes. Esto está previsto en los protocolos “seguros” de alto nivel, y en caso de que ocurra, los paquetes son reenviados. En los protocolos diseñados para tráfico de tiempo real generalmente no se recibe confirmaciones de recepción de paquetes, ya que si el canal es suficientemente seguro, estas confirmaciones cargan inútilmente al mismo.
En aplicaciones de voz y video, el audio es “encapsulado” en paquetes y enviado, sin confirmación de recepción de cada paquete.
Si el porcentaje de perdida es pequeño, la degradación de la voz también lo es. Los porcentajes de perdida admisibles dependen de otros factores, como por ejemplo la demora de transmisión y el factor de compresión de la voz.
Existen técnicas para hacer menos sensible la degradación de calidad en la voz frente a la pérdida de paquetes. La más sencilla consiste en simplemente repetir el último paquete recibido.
También cuentan como “perdidos” los paquetes que llegan a destiempo o fuera de orden.

Existen métodos para mitigar el efecto de la pérdida de paquetes. Un ejemplo se describe en el Anexo I de la Recomendación ITU-T G.711 [8], donde se detalla un método de cancelación de paquetes perdidos (PLC, Packet Loss Concealment). Este método propone regenerar la forma de onda del paquete perdido en base a información extraída de la señal previa a la pérdida del paquete.

1.5.1.3 Demora

Un factor importante en la percepción de la calidad de la voz es la demora. La demora total está determinada por varios factores, entre los que se encuentran

- Demora debida a los algoritmos de codificación
  En forma genérica, cuanto mayor es la compresión, más demora hay en el proceso (los codecs requieren más tiempo para codificar cada muestra).

Algoritmo de muestreo/compresión	Demora típica introducida
G711 (64 kb/s)	125 µs
G.728 (16 kb/s)	2.5 ms
G.729 (8 kb/s)	10 ms
G.723 (5.3 o 6.4 kb/s)	37.5 ms
RTAudio (8 kb/s)	40 ms

Tabla 1.1

- Demoras de procesamiento
  Es el tiempo involucrado en el procesamiento de la voz para la implementación de los protocolos. Generalmente puede ser despreciado.

- Demoras propias de la red (latencia)
  Las demoras propias de la red están dadas por la velocidad de transmisión de la misma, la congestión, y las demoras de los equipos de red (routers, switches, etc.)

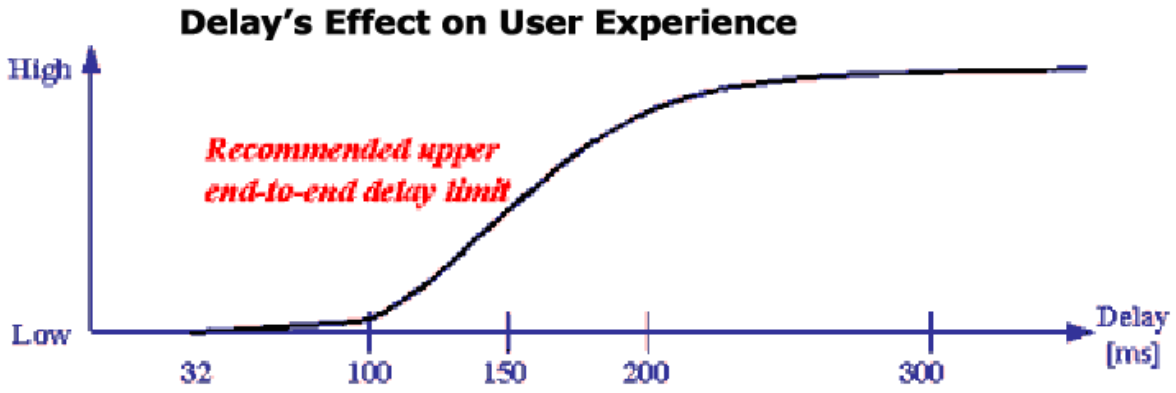


Figura 1.7

Las demoras no afectan directamente la calidad de la voz, sino la calidad de la conversación. Como se puede ver en la Figura 1.7, hasta 100 ms son generalmente tolerados, casi sin percepción de los interlocutores. Entre 100 y 200 ms las demoras son notadas. Al acercarse a los 300 ms de demora, la conversación se vuelve poco natural. Pasando los 300 ms la demora se torna crítica, haciendo dificultosa la conversación.

Un efecto secundario, generado por las demoras elevadas, es el eco. El eco se debe a que parte de la energía de audio enviada es devuelta por el receptor. En los sistemas telefónicos este efecto no tiene mayor importancia, ya que los retardos o demoras son despreciables, y por lo tanto, el "eco" no es percibido como tal.
Cuando la demora de punta a punta comienza a aumentar, el efecto del eco comienza a percibirse.

1.5.1.4 Eco

Si el tiempo transcurrido desde que se habla hasta que se percibe el retorno de la propia voz es menor a 30 ms, el efecto del eco no es percibido. Asimismo, si el nivel del retorno está por debajo de los –25 dB, el efecto del eco tampoco es percibido. En las conversaciones telefónicas habituales, generalmente existe un retorno de la propia voz en niveles audibles (mayores a –25 dB), pero la demora es mínima, por lo que este retorno no es percibido como eco.

El retorno que produce el eco se produce en diferentes elementos de la red, varios de los cuales se esquematizan en la siguiente figura.

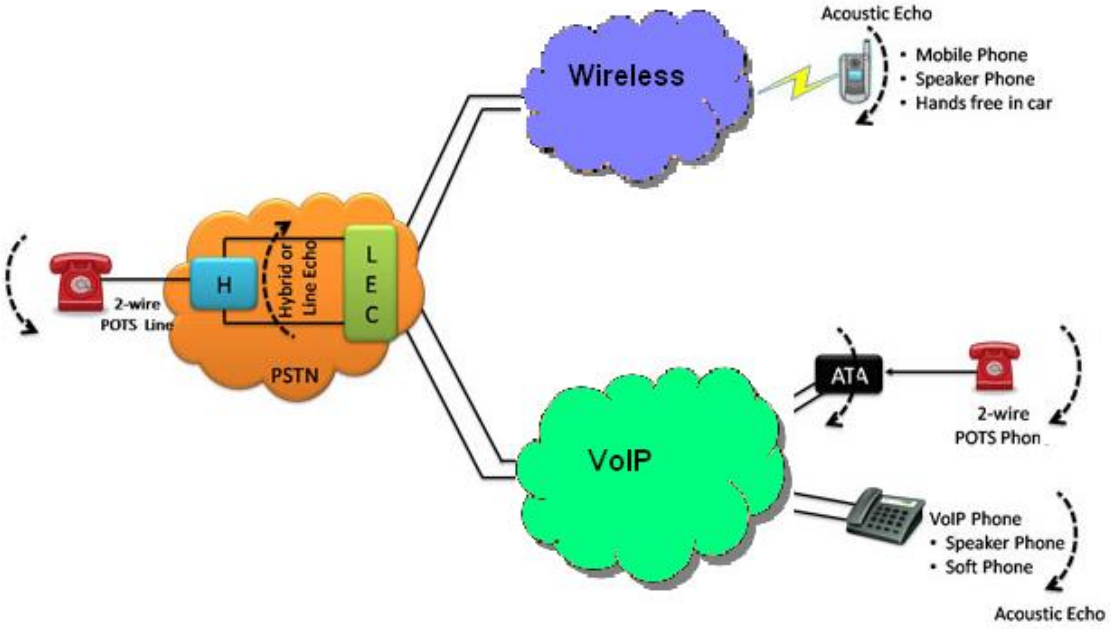


Figura 1.8

Los teléfonos analógicos pueden generar retorno en sus “híbridas”. Las “híbridas” de las tarjetas de abonado también pueden generar retorno. Los teléfonos celulares tienen el micrófono muy cerca del auricular, y pueden generar retorno acústico. Los teléfonos IP de hardware pueden generar retorno acústico, si se utilizan en “manos libres”. Los teléfonos IP de software pueden generar retorno en la tarjeta de sonido del PC, o en las diademas.

Todos estos retornos pueden ser percibidos como eco, si las demoras entre su generación y su escucha es apreciable. Como las redes IP tienen retardos de punta a punta muy superiores a los existentes en las redes TDM, todos estos retornos se pueden percibir como eco, y deben ser evitados o cancelados.

En la Recomendación ITU-T G.168 se describen las características que deben tener los sistemas digitales diseñados para realizar “cancelación de eco”. Estos sistemas funcionan como se muestra en la Figura 1.10. Mediante un procesamiento digital se evalúa si parte de la señal en el camino de “recepción” (Receive Path) se ha introducido en el camino de “Transmisión” (Send Path), con cierto retardo. Si esto es detectado, la señal del canal de “Transmisión” es procesada, restándole la estimación de la señal correspondiente al eco. Luego de este proceso, la señal pasa por un proceso no lineal, que suprime las señales que están por debajo de cierto umbral.

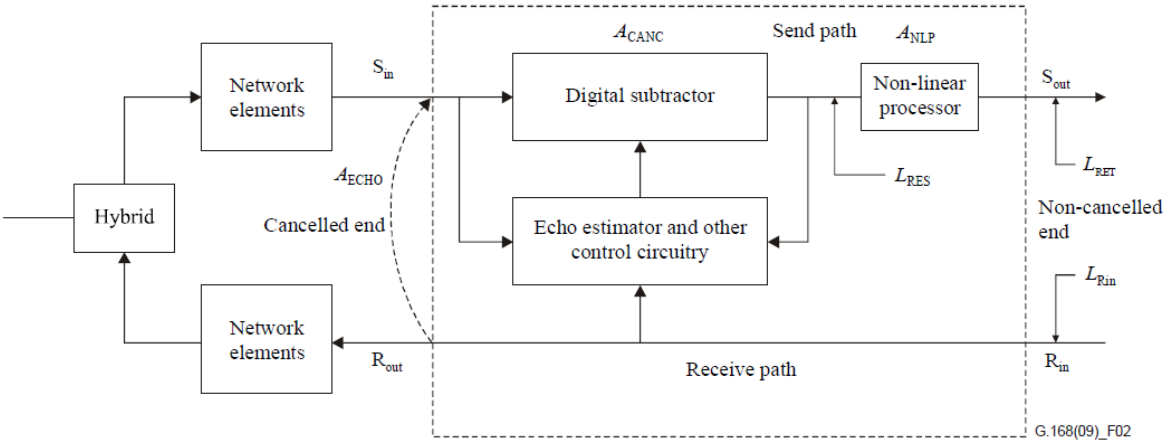


Figura 1.9

La mayoría de los sistemas que utilizan VoIP disponen de canceladores de eco en algún punto del camino de audio.

1.5.1.5 Variaciones en la demora (Jitter)

El “jitter” 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 receptor debe recibir los paquetes a intervalos constantes, para poder regenerar de forma adecuada la señal original. Dado que el “jitter” es inevitable, los receptores disponen de un “buffer” de entrada, con el objetivo de “suavizar” el efecto de la variación de las demoras. Este buffer recibe los paquetes a intervalos variables, y los entrega a intervalos constantes.
Es de hacer notar que este “buffer” agrega una demora adicional al sistema, ya que debe “retener” paquetes para poder entregarlos a intervalos constantes. Cuánto más variación de demoras (“jitter”) exista, más grande deberá ser el buffer, y por lo tanto, mayor demora será introducida al sistema. Típicamente los jitter-buffers introducen una demora de entre 10 ms a 60 ms.

1.5.1.6 Tamaño de los paquetes

El “tamaño” de los paquetes influye en dos aspectos fundamentales en la transmisión de la voz sobre redes de datos: La demora y el “ancho de banda” requerido.

Para poder transmitir las muestras codificadas de voz sobre una red de datos, es necesario armar “paquetes”, según los protocolos de datos utilizados (por ejemplo, IP). Un paquete de datos puede contener varias muestras de voz. Por ello, es necesario esperar a recibir varias muestras para poder armar y enviar el paquete. Esto introduce un retardo o demora en la transmisión. Desde éste punto de vista, parece conveniente armar paquetes con la mínima cantidad de muestras de voz (por ejemplo, un paquete por cada muestra). Sin embargo, hay que tener en cuenta que cada paquete tiene una cantidad mínima de información (bytes) de control (cabezal del paquete, origen, destino, etc.). Esta información (“sobrecarga” u “overhead”), no aporta a la información real que se quiere transmitir, pero afecta al tamaño total del paquete, y por tanto al ancho de banda, como se vio en **¡Error! No se encuentra el origen de la referencia.**.

La duración de las “ventanas” de voz se encuentran entre 10 a 30 ms, valor que aporta a la demora total.

1.5.2 ITU-T G.107 (E-Model)

La industria de las telecomunicaciones ha aceptado una representación numérica de la calidad de la voz, llamada “MOS” (Mean Opinion Score), y estandarizada en la recomendación ITU-T P.800. La calidad de la voz es calificada con un número, entre 1 y 5. El valor numérico de MOS es proporcional a la calidad de la voz. 1 significa muy mala calidad y 5 significa excelente. Los valores son obtenidos mediante el promedio de las opiniones de un gran grupo de usuarios.

La ITU-T ha creado un “modelo” en la recomendación ITU-T G.107, llamado “E-Model” [9], para estimar o predecir la calidad de la voz en redes IP (VoIP) percibida por un usuario típico, en base a parámetros medibles de la red. El resultado del E-Model es un factor escalar, llamado “R” (“Transmission Rating Factor”), que puede tomar valores entre 0 y 100.

El “E-model” toma en cuenta una gran cantidad de factores que pueden deteriorar la calidad de la voz percibida, como por ejemplo, el uso de compresión, los retardos de la red, así como también los factores “típicos” en telefonía como ser pérdida, ruido y eco. Puede ser aplicado para estimar la calidad de la voz en redes de paquetes, tanto fijas como inalámbricas [10].

El E-Model puede ser utilizado para evaluar como se verá afectada la calidad de la voz en una red en base a parámetros mensurables. El modelo parte de un puntaje "perfecto" (100) y resta diversos factores que degradan la calidad, según se puede ver en la ecuación (1.4.1).

$$R = R_o - I_s - I_d - I_{e,eff} + A \qquad (1.4.1)$$

donde

R_o Representa la relación señal/ruido básica (antes de ingresar en la red) que incluye fuentes de ruido, tales como ruido ambiente. El valor inicial puede ser como máximo 100. Las fuentes de ruido independientes del sistema como ser el ruido ambiental, pueden hacer que este valor inicial sea menor a 100.

I_s Es una combinación de todas las degradaciones que aparecen de forma más o menos simultánea con la señal vocal. Por ejemplo, volumen excesivo y distorsión de cuantización.

I_d Representa las degradaciones producidas por el retardo y el eco

$I_{e,eff}$ "Effective equipment impairment factor". Representa las degradaciones producidas por los códecs y por las pérdidas de paquetes de distribución aleatoria.

A Factor de Mejoras de Expectativas. Muchas veces, los usuarios están dispuestos a aceptar peor calidad de voz si saben que se están utilizando tecnologías "no clásicas" (por ejemplo celulares o VoIP). Permite compensar los factores de degradación cuando existen otras ventajas de acceso para el usuario.

Los valores de R varían entre 0 y 100, correspondiendo los valores más altos a mejores calidades de voz.

Los tres tipos de degradaciones (I_s, I_d y $I_{e,\ eff}$) se subdividen, a su vez, en la combinación de otros factores, como se detalla a continuación.

Cálculo de Is

$$I_s = I_{olr} + I_{st} + I_q \qquad (1.4.2)$$

donde

I_{olr} Representa la disminución de calidad producida por valores demasiado bajos de OLR (Overall Loudness Rating). El OLR se calcula, a su vez, como

$$OLR = SLR + RLR \qquad (1.4.3)$$

Siendo

SLR (Send Loudness Rating), es la pérdida entre la boca del emisor y el micrófono del aparato telefónico

RLR (Receive Loudness Rating), es la pérdida entre el parlante del aparato telefónico y el oído del receptor

I_{st} Representa la degradación producida por efectos locales no óptimos, y depende esencialmente del factor STMR (Side Tone Masking Rating). Parte de la señal recibida por el micrófono es transmitida, dentro del mismo teléfono, al parlante, generando un "efecto local" que hace que la persona que habla se escuche por el oído en el que tiene el tubo o microteléfono. La atenuación de la señal que pasa del micrófono al parlante del mismo aparato se conoce como STMR. Si este valor no está dentro de los parámetros adecuados, genera una sensación de "eco", o de "línea muerta", según el caso, bajando la calidad de la comunicación.

I_q Representa la degradación producida por la distorsión de cuantificación. Se calcula en base a "unidades qdu" . 1 qdu se define como el "ruido" de cuantización" que resulta de una codificación y decodificación completas en Ley A o Ley μ

La fórmula de cálculo detallada de los parámetros (I_{olr}, I_{st}, I_q) puede verse en la recomendación G.107 [**¡Error! Marcador no definido.**].

Cálculo de Id

$$I_d = I_{dte} + I_{dle} + I_{dd} \qquad (1.4.4)$$

Donde

I_{dte} Expresa una estimación para las degradaciones debidas al eco para el hablante. Se calcula en base al factor TELR (Talker Echo Loudness Rating) y la demora media T de punta a punta en un sentido. El factor TELR es la medida de la atenuación del eco percibido por el hablante.

I_{dle} Representa degradaciones debidas al eco para el oyente. Se calcula en base al factor WEPL (Weighted Echo Path Loss) y la demora media Tr de ida y vuela. El factor WEPL es la medida de la atenuación entre la señal "directa" recibida por el oyente, la señal retardada recibida como eco.

I_{dd} Representa la degradación producida por retardos absolutos demasiado largos T_a, que se producen incluso con compensación perfecta del eco. Si T_a < 100 ms, el factor I_{dd} es 0.

La fórmula de cálculo detallada de los parámetros (I_{dte}, I_{dle}, I_{dd}) puede verse en la recomendación G.107.

El efecto de la demora en el valor de R se grafica en la Figura 1.10, asumiendo todos los otros factores ideales [11].

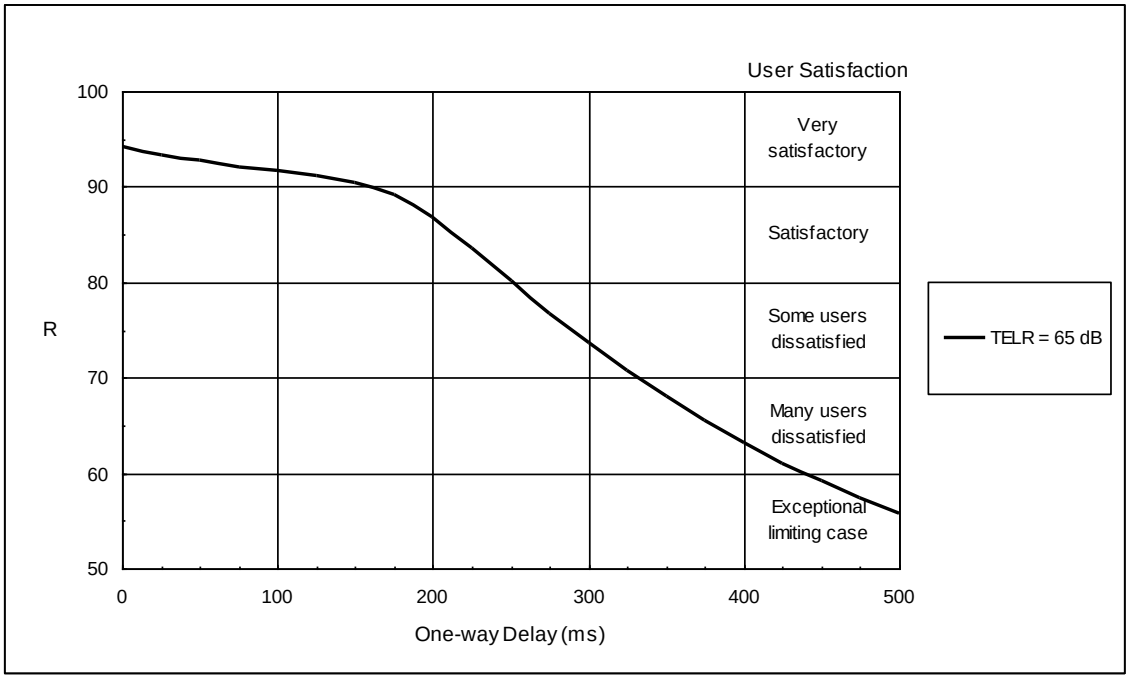


Figura 1.10

Puede verse como hasta 175 ms el valore de R es mayor que 90, y se encuentra en la zona de “Muy satisfechos”. Sin embargo, luego de los 175 ms, el efecto de las demoras degrada fuertemente la comunicación, haciéndola poco natural.
Si a la gráfica anterior se le suma el efecto del eco, el modelo E predice las curvas presentadas en la Figura 1.11.

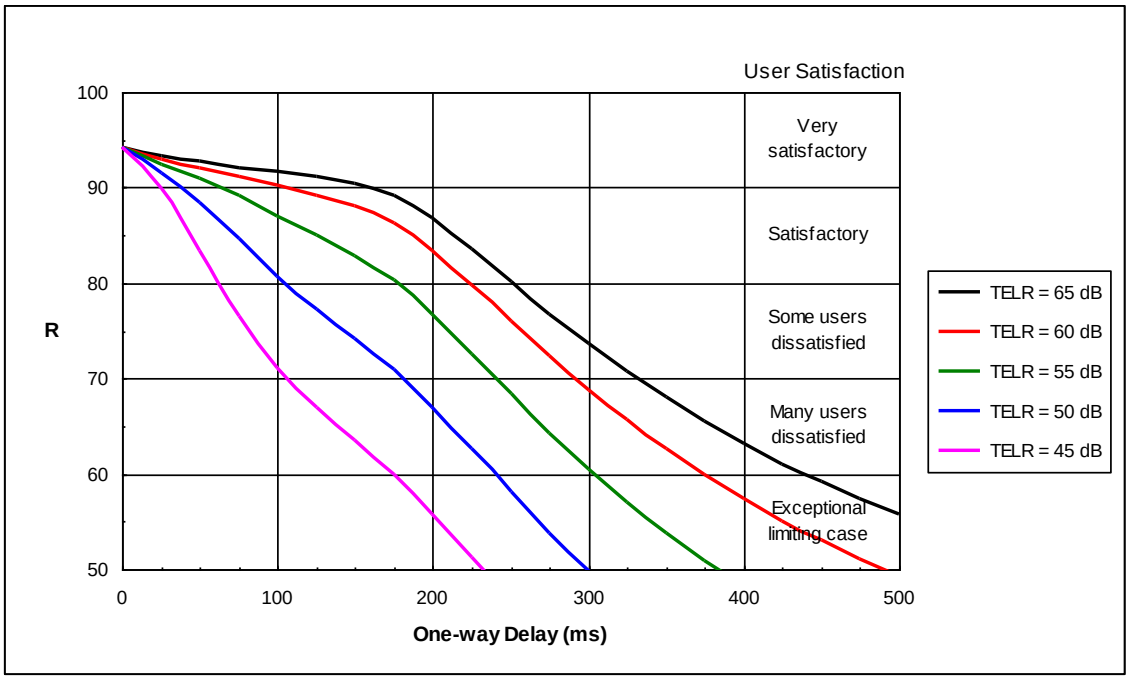


Figura 1.11

Es de hacer notar que el valor TELR es la medida de la atenuación del eco percibido por el hablante. Cuanto más atenuado el eco percibido (mayor valor en db de TELR), menor efecto tiene el eco sobre la degradación. En la medida que aumenta el eco, el valor de R decrece rápidamente con el retardo.

Cálculo de Ie-eff

$I_{e\text{-}eff}$ representa las degradaciones producidas por los códecs y por las pérdidas de paquetes, según la siguiente fórmula:

$$Ie\text{-}eff = Ie + (95 - Ie) \cdot \frac{Ppl}{\frac{Ppl}{BurstR} + Bpl} \qquad (1.4.5)$$

Donde

I_e Es un valor que depende del codec utilizado, y representa la degradación percibida producida por los diferentes algoritmos de compresión.

P_{pl} Representa la probabilidad de pérdida de paquetes

B_{pl} Se define como el "factor de robustez" contra pérdida de paquetes, y es un valor preestablecido para cada codec

BurstR Es la “Relación de ráfaga”, y se define como

$$BurstR = \frac{\text{Longitud media de las ráfagas observadas en una secuencia de llegada}}{\text{Longitud media de las ráfagas previstas en la red en condiciones de pérdida "arbitraria"}}$$

Si no existen pérdida de paquetes (P_{pl}=0), el factor $I_{e\text{-}eff}$ depende únicamente del tipo de codec utilizado

Los valores de I_e para los diferentes codecs se detallan en la Tabla 1.2:

Codec Type	Reference	Operating Rate kbit/s	Ie Value
Waveform Codecs			
PCM	G.711	64	0
ADPCM	G.726, G.727	40	2
	G.721, G.726, G.727	32	7
	G.726, G.727	24	25
	G.726, G.727	16	50
Speech Compression Codecs			
LD-CELP	G.728	16	7
		12.8	20
CS-ACELP	G.729	8	10
	G.729-A + VAD	8	11
VSELP	IS-54	8	20
ACELP	IS-641	7.4	10
QCELP	IS-96a	8	21
RCELP	IS-127	8	6
VSELP	Japanese PDC	6.7	24
RPE-LTP	GSM 06.10, Full-rate	13	20
VSELP	GSM 06.20, Half-rate	5.6	23
ACELP	GSM 06.60, EFR	12.2	5
ACELP	G.723.1	5.3	19
MP-MLQ	G.723.1	6.3	15

Tabla 1.2

En una red sin pérdida de paquetes y sin eco, el valor de R dependerá de la demora y de los codecs utilizados, según se muestra en la Figura 1.12, para G.711, G.729A y G.723.1 (notar que la gráfica “negra” coincide con las gráficas anteriores)

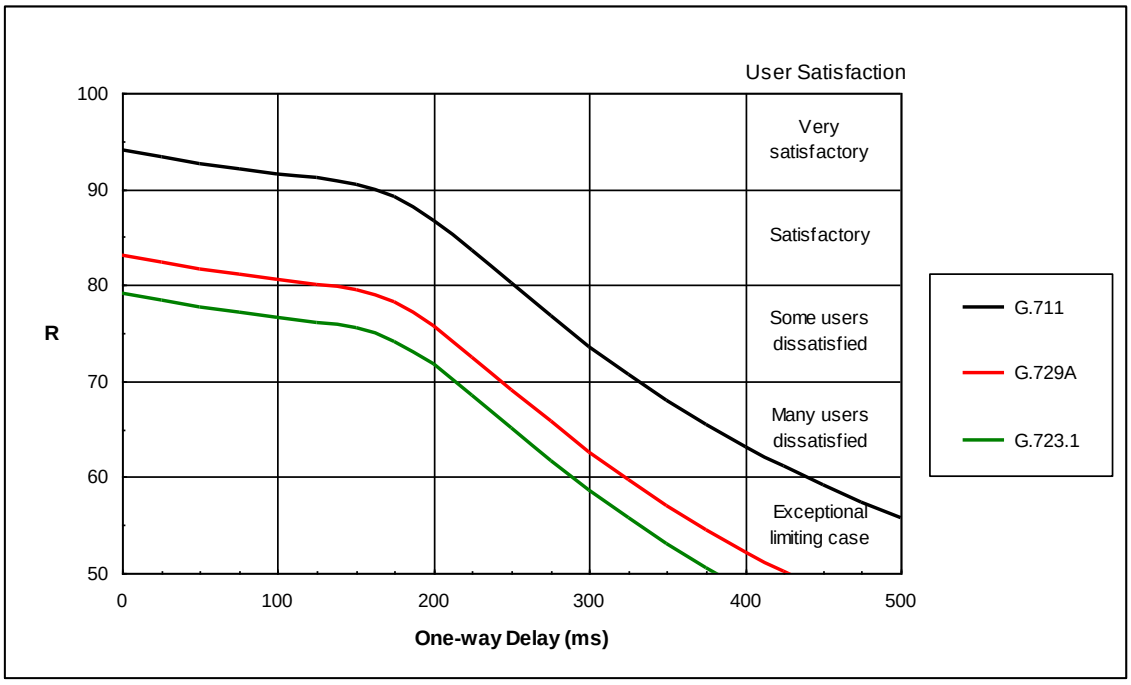
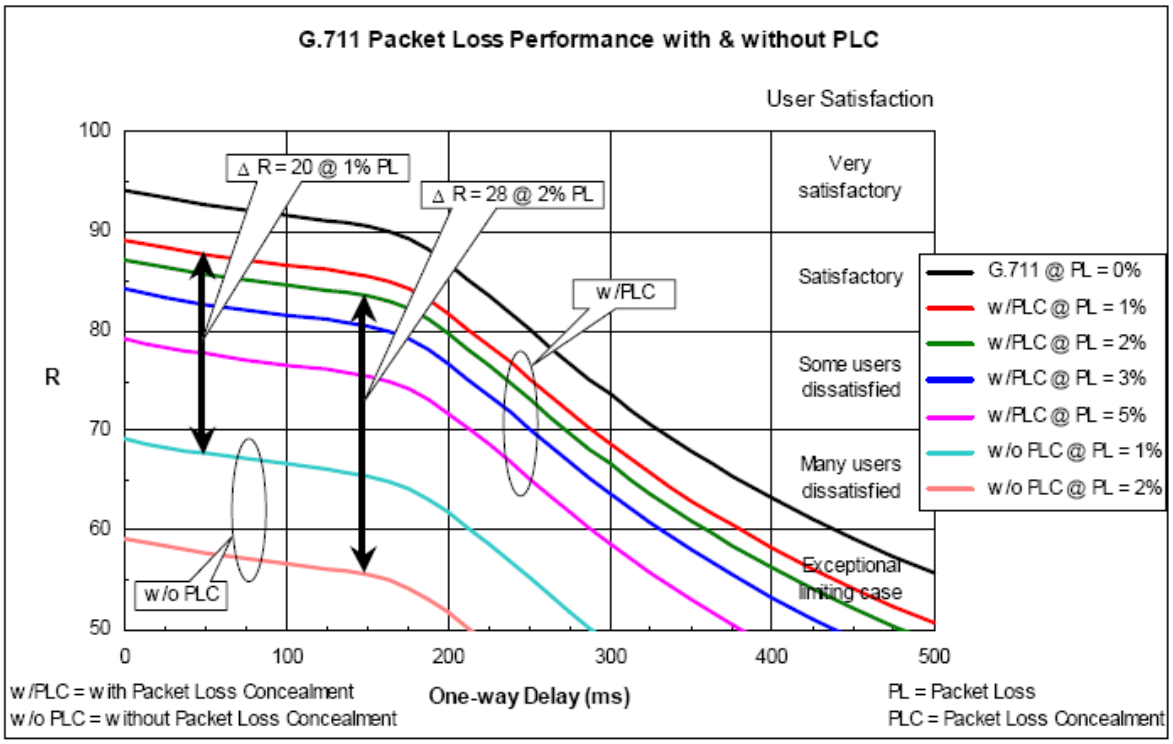


Figura 1.12

En una red con pérdida de paquetes, el valor de R depende de la utilización o no de técnicas de PLC (Packet Loss Compensation). A modo de ejemplo, para los codecs G.711 y G.729, las gráficas de la Figura 1.13 muestran el efecto conjunto de la demora y la pérdida de paquetes.



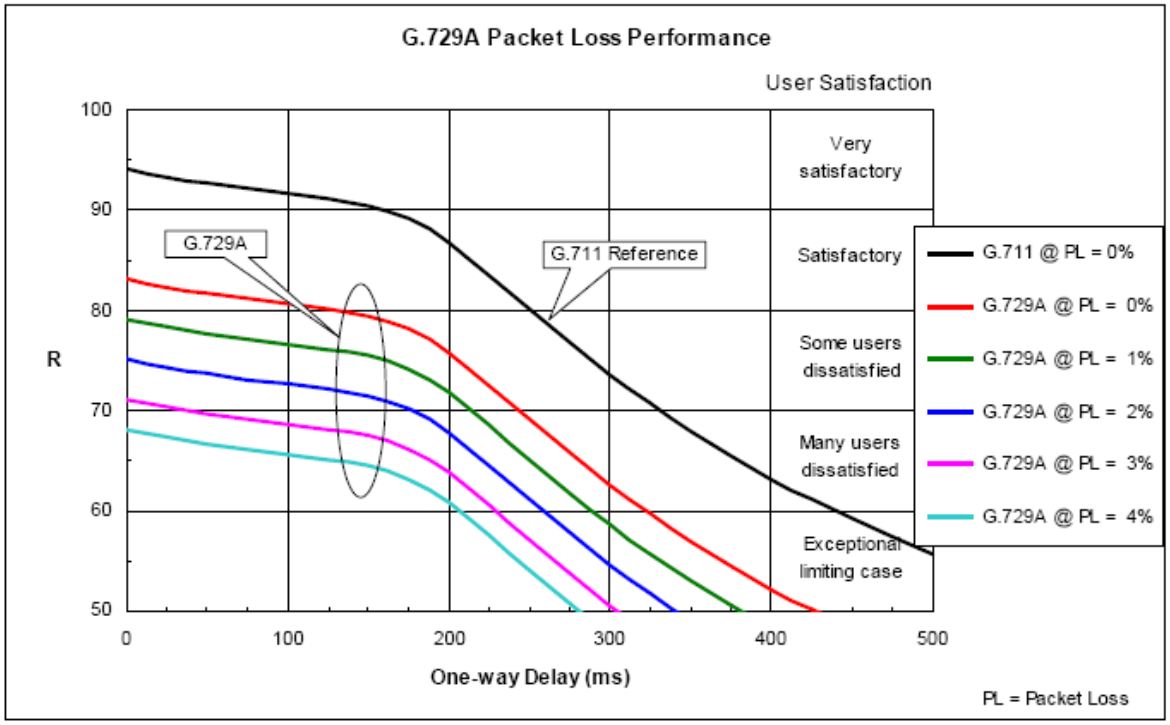


Figura 1.13

Cálculo de A

A representa un "Factor de Mejoras de Expectativas". Muchas veces, los usuarios están dispuestos a aceptar peor calidad de voz si saben que se están utilizando tecnologías "no clásicas" (por ejemplo celulares o VoIP). No existe, por consiguiente, ninguna relación entre A y los demás parámetros de transmisión.

El cuadro siguiente presenta los valores típicos de A para diferentes tecnologías, según la recomendación ITU-T G-113 [12].

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

Tabla 1.3

Relación de R y MOS

El modelo relaciona el valor de "R" con el "MOS", con un gran nivel de aproximación, según la siguiente ecuación:

Para $R < 6.5$: $\text{MOS}_{\text{CQE}} = 1$

Para $6.5 < R < 100$: $\text{MOS}_{\text{CQE}} = 1 + 0{,}035R + R(R-60)(100-R)7 \cdot 10^{-6}$ (1.4.6)

Para $R > 100$: $\text{MOS}_{\text{CQE}} = 4{,}5$

La Figura 1.14y la Figura 1.15 muestran la relación entre R y MOS, según la fórmula anterior:

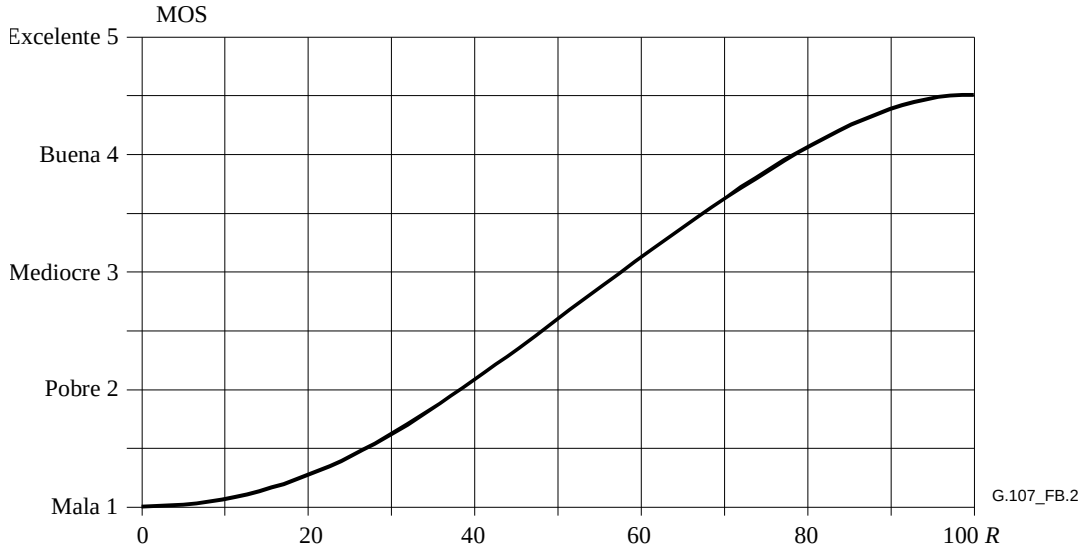


Figura 1.14

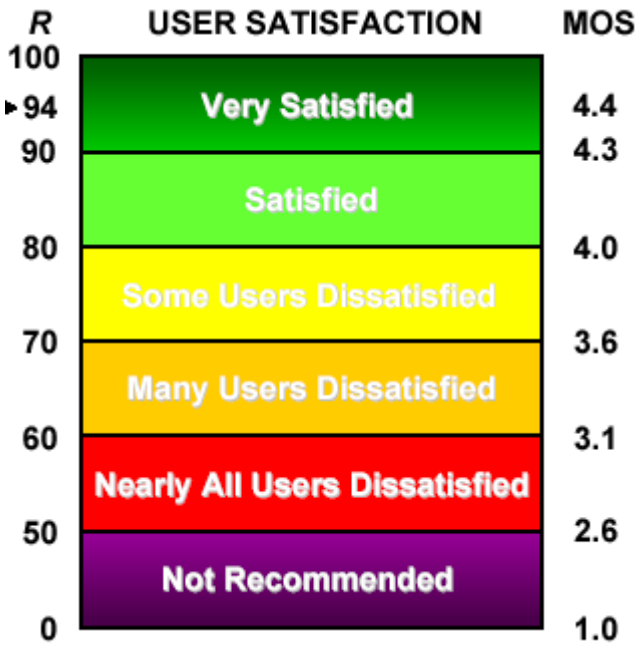


Figura 1.15

Aplicación del E-model

El RFC 3611 [13] define campos de “reportes extendidos” (XR, Extended Reports) en el protocolo RTCP que permiten intercambiar información acerca de la calidad de la comunicación. En este RFC se incluye la posibilidad de intercambiar información del valor de “R” entre fuentes y destinos, así como los valores percibidos de MOS-LQ (MOS listening quality) y MOS-CQ (MOS conversational quality)

En la Figura 1.16 se muestra un ejemplo de un paquete RTCP según el RFC 3611. Se puede ver el intercambio de información donde se incluye el valor de R, MOS-LQ y MOS-CQ.

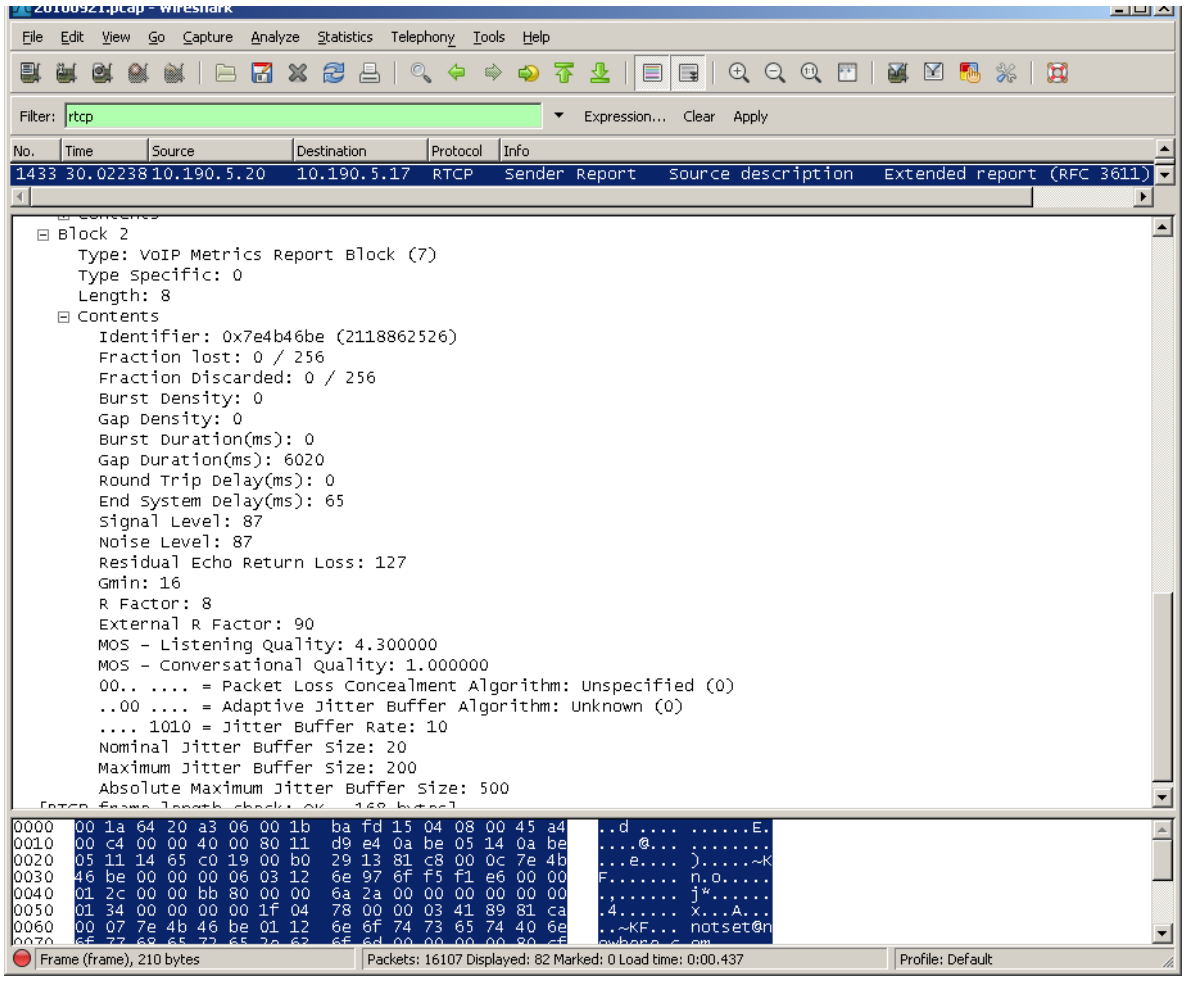


Figura 1.16

Más recientemente, en 2009, se incluyó el Anexo II a la Recomendación, adaptando el modelo a comunicaciones de banda ancha. 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. El E-Model, aplicado a banda ancha, amplía la escala del factor R, extendiéndola de 100 a 129.

2 Calidad de Video

El video digital, distribuido a través de redes de comunicaciones, sufre varios tipos de distorsión durante el proceso de adquisición, compresión, procesamiento, transmisión y reproducción. Por ejemplo, las técnicas utilizadas habitualmente en la codificación digital de video introducen pérdida de información, para reducir el ancho de banda necesario para su transmisión, lo que genera distorsiones. Por otro lado, las redes de paquetes sobre las que se transporta el video (por ejemplo Internet, aunque también redes de área local LAN, las redes extendidas WAN y las redes inalámbricas), pueden introducir distorsiones adicionales, debido a las demoras, los errores y las pérdidas de paquetes, entre otros factores.

El video es utilizado en diversos tipos de aplicaciones, las que a su vez, tienen diversos requerimientos. La TV es, quizás, la aplicación de video más conocida. Sin embargo, existen en forma cada vez más difundida un nuevo conjunto de aplicaciones de video, entre las que se encuentran la video telefonía, los servicios de video conferencia, la distribución de video a demanda a través de Internet y la IP-TV, por mencionar los más relevantes. Cada una de estas aplicaciones tiene sus características propias en lo que respecta a requerimientos de calidad, velocidades, etc.

La videotelefonía es una aplicación típicamente punto a punto, con imágenes del tipo “cabeza y hombros”, y generalmente poco movimiento. Sin embargo, es una aplicación altamente interactiva, dónde los retardos punta a punta juegan un rol fundamental en la calidad conversacional percibida.

Las aplicaciones de video conferencias son típicamente punto a multi-punto. Al igual que la video telefonía, generalmente tienen poco movimiento. Además de la difusión del audio y el video es deseable en estas aplicaciones poder compartir imágenes y documentos. La interactividad también es típicamente un requisito, aunque podrían admitirse retardos punta a punta un poco mayores que en la video telefonía, ya que los participantes generalmente están dispuestos a “solicitar la palabra” en este tipo de comunicaciones.

La distribución de televisión digital, y en particular la IP-TV generan otro tipo de requerimientos. En estas aplicaciones se debe soportar todo tipo de imágenes (desde las estáticas hasta las de mayor movimiento), y la calidad percibida de la imagen juega un rol fundamental. Los usuarios de estos servicios esperan recibir la calidad por lo cual están pagando. Además de esto, otros efectos, como las demoras entre el cambio de canales (“zapping”) juega un papel importante en la experiencia del usuario. En la TV analógica, los usuarios están acostumbrados a que estos cambios de canal son prácticamente instantáneos (lo que es difícil de lograr cuando se debe, por ejemplo, tener un jittter-buffer de algunos segundos).

Finalmente es de hacer notar que en casi todas las aplicaciones de video, el audio también está presente, y juega un papel muy importante en la calidad perceptual

general. La percepción del usuario respecto al video no sólo se ve condicionada por la calidad del audio, sino también por la sincronización existente entre el audio y el video. Pequeños tiempos de defasaje entre ambas señales son muy notorios (por ejemplo, al ver una persona hablando), lo que produce sensaciones molestas, y afecta notoriamente a la calidad percibida, aún cuando la calidad de las señales de audio y de video que se estén recibiendo sean excelentes.

2.1 Medida de la Calidad de video

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. La opinión media, obtenida mediante el "MOS" (Mean Opinion Score) es la métrica generalmente aceptada como medida de la calidad. Para ello, los experimentos subjetivos controlados continúan siendo actualmente los métodos de medida reconocidos en la estimación perceptual de la calidad del video.

Por otra parte, existe un gran esfuerzo en generar métodos objetivos, que estimen la calidad de video percibida por los usuarios. Si bien éste continúa siendo un teman en investigación y desarrollo, existen ya algunos avances y recomendaciones al respecto.

2.2 Métodos Subjetivos de medida de la calidad del video

Diversos métodos subjetivos de evaluación de video son reconocidos, y están estandarizados en las recomendaciones ITU-R BT.500-13 [14], especialmente desarrollada para aplicaciones de televisión y ITU-T P.910 [15], para aplicaciones multimedia. En todos los métodos propuestos, los evaluadores son individuos que juzgan la calidad en base a su propia percepción y experiencia previa. Estos métodos tienen en común la dificultad y lo costoso de su implementación.

La recomendación ITU-R BT.500 detalla los métodos DSIS, DSCQS, SSCQE y SDSCE. La ITU-T P.910 los métodos ACR, DCR y PR. Todos ellos se describen brevemente a continuación.

Métodos propuestos en ITU-R BT.500-13

2.2.1 DSIS – Double Stimulus Impairment Scale

El método DSIS (Double Stimulus Impairment Scale o Escala de degradación con doble estímulo) consiste en la comparación de dos estímulos, uno dado por la señal original (no degradada) y el otro por la señal degradada. En forma genérica, las señales pueden ser una imagen, una secuencia de imágenes o un video. Las condiciones de visualización, iluminación del ambiente, disposición de las

personas respecto al monitor o televisor, etc. están controladas y detalladas en la recomendación.
Los participantes deben seleccionar una de entre cinco opciones, como se muestra en la siguiente tabla:

La degradación es imperceptible	5
La degradación es perceptible, pero no molesta	4
La degradación es ligeramente molesta	3
La degradación es molesta	2
La degradación es muy molesta	1

La estructura de la prueba consiste en mostrar la señal de referencia por unos segundos, y luego la señal degradada, por la misma cantidad de segundos. La misma secuencia se repite dos veces para cada prueba. Se solicita a los participantes que esperen hasta el final de cada secuencia de prueba para realizar su calificación (es decir, deben realizar la calificación luego de ver las dos secuencias de señal original / señal degradada).

2.2.2 DSCQS – Double Stimulus Continuos Quality Scale

Al igual que en el método anterior, en el método DSCQS (Double Stimulus Continuos Quality Scale o Escala de calidad continua de doble estímulo) se presentan dos señales. Sin embargo, en este método se solicita a los participantes que califiquen la calidad de ambas señales, en lugar de la degradación. Es decir, se debe calificar tanto la señal de referencia (señal “A”) como la señal procesada o degradada (señal “B”). La calificación se realiza en base a una escala continua, utilizando una plantilla impresa de 10 cm de largo, donde se presentan las calificaciones indicadas en la siguiente figura:

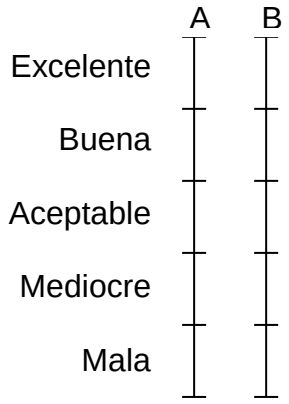


Figura 2.1

La escala continua permite medir diferencias más precisas entre ambas señales, e incluso permite categorizar a la segunda señal con mejor calidad que la primera. Esto último podría ser posible en el caso que se utilice este método para evaluar

algoritmos de realce de video, o algoritmos que intenten mejorar la calidad (por ejemplo, compensar el “efecto de bloques”, que se detallará más adelante).

2.2.3 SSCQE– Single Stimulus Continuous Quality Evaluation

A diferencia de los métodos anteriores, en el método SSCQE (Single Stimulus Continuous Quality Evaluation o Evaluación de calidad continua de estímulo único), se presenta una única secuencia de video a ser evaluada. Este video puede o no tener degradaciones.
También a diferencia de los otros métodos, se propone aquí una evaluación continua de la calidad del video, y no sólo una calificación global que integra las degradaciones de varios segundos. Para ello se utiliza un cursor móvil, conectado a una computadora, que permite registrar en forma continua las calificaciones.
Se utiliza la misma escala presentada en la
. Dado que se toman muestras de la evaluación en forma continua, se puede asignar a cada instante del video su correspondiente calificación, lo que permite tener en forma mucho más detallada el efecto perceptual de cada una de las degradaciones.

2.2.4 SDSCE - Simultaneous Double Stimulus for Continuous Evaluation

Cuando hay que evaluar la fidelidad, es necesario compara una señal contra su referencia. El método SDSCE (Simultaneous Double Stimulus for Continuous Evaluation o Método de doble estímulo simultáneo para evaluación continua) ha sido elaborado a partir del SSCQE, con ligeras diferencias en cuanto a la manera de presentar las imágenes a los sujetos y con respecto a la escala de apreciación. El método fue propuesto a MPEG para evaluar el comportamiento frente a errores a velocidades de transmisión muy bajas, pero puede ser aplicado adecuadamente a todos los casos en los que hay que evaluar la fidelidad de la información visual afectada por la degradación que varía en función del tiempo.
Con este método, el grupo de personas observa dos secuencias al mismo tiempo: una es la referencia, la otra es la señal degradada a evaluar. Ambas pueden ser presentadas dentro de un mismo monitor, o en dos monitores alineados. Se pide a los sujetos que comprueben las diferencias entre las dos secuencias y juzguen la fidelidad de la señal a calificar moviendo el cursor de un dispositivo de voto manual. Cuando la fidelidad es perfecta, el cursor debe estar en la parte superior de la escala (codificada con el valor100), cuando la fidelidad es nula, el cursor debe estar en la parte inferior de la escala (codificada con el valor 0).

Métodos propuestos en ITU-T P.910

2.2.5 ACR - Absolute Category Rating

El método ACR (Absolute Category Rating o Índices por categorías absolutas) es un juicio de categorías en el que las secuencias de prueba se presentan una por

vez y se califican independientemente en una escala de categorías. Se utiliza una escala de 5 niveles, como se presenta en la tabla siguiente

Excelente	5
Bueno	4
Aceptable	3
Mediocre	2
Mala	1

2.2.6 DCR - Degradation Category Rating

El DCR (Degradation Category Rating o Índices por categorías de degradación) es un método de doble estímulo, donde las secuencias se presentan por pares: el primer estímulo presentado en cada par es siempre la señal de referencia, mientras que el segundo estímulo es la señal degradada. Se pueden presentar las señales de referencia y la degradada en forma serial, una a continuación de la otra, o en forma conjunta, en el mismo monitor. La evaluación se realiza con la escala de 5 valores presentada en la de la sección anterio

2.2.7 PC - Pair Comparison

El método PC (Pair Comparison o Método de comparación por pares) se utiliza cuando se desean comparar degradaciones producidas por dos sistemas diferentes, sobre una misma señal de referencia. En el método PC se presenta una señal luego de pasar por un sistema, y a continuación la misma señal luego de pasar por el otro sistema. Estos sistemas pueden ser simplemente codificadores, medios de transmisión, etc. Después de ver cada par de secuencias, se hace una apreciación sobre qué señal sufrió “menos degradaciones”, en el contexto del escenario de prueba.

2.3 Métodos Objetivos de medida de la calidad del video

Los métodos subjetivos, presentados en la sección anterior, son costosos, difíciles de realizar, e impracticables en aplicaciones de tiempo real.
Por esto se hace necesario el uso de métodos objetivos y automáticos, que puedan predecir con fiabilidad la calidad percibida, en base a medidas objetivas tomadas en algún punto del sistema.

Las primeras medidas objetivas de la calidad del video están basadas en obtener las diferencias, píxel a píxel, entre las imágenes originales (previo a la compresión y transmisión) y las imágenes presentadas (luego de la recepción y reconstrucción). Dado que los sistemas de video utilizan técnicas de compresión con pérdida de información, y que los medios de transmisión a su vez pueden introducir factores distorsionantes (retardos, pérdida de paquetes, etc.), las imágenes presentadas serán diferentes a las originales.

Las medidas más simples son las de error cuadrático medio (MSE - Mean Square Error) y su raíz cuadrada (RMSE = Root Mean Square Error) y la relación señal a ruido de pico (PSNR – Peak Signal to Noise Ratio), definidas en las ecuaciones (2.1) a (2.3) más adelante. Estas métricas requieren de la referencia completa de la señal original para poder ser calculadas.

$$MSE=\frac{1}{TMN}\sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{t=1}^{T}\left[x(m,n,t)\ -y(m,n,t\ )\right]^2 \qquad (2.1)$$

$$RMSE=\sqrt{MSE} \qquad (2.2)$$

$$PSNR=10\log_{10}\left(\frac{L^2}{MSE}\right) \qquad (2.3)$$

donde la imagen tiene N x M píxeles y T cuadros, x, y son los píxeles de la imagen original y la distorsionada respectivamente. L es el rango dinámico que pueden tomar los valores de x o y, y toma el valor 255 para 8 bits por píxel.

Estas métricas son fáciles de calcular, y tienen un claro significado. Por estas razones, han sido ampliamente usadas como métricas en la estimación de la calidad de video. Hay que poner especial énfasis en la alineación espacial y temporal de las imágenes a comparar, ya que la referencia y la imagen degradada pueden estar desfasadas en el tiempo o en el espacio.

Sin embargo, también han sido ampliamente criticadas por no tener correlación directa con la calidad percibida. Por ejemplo, en la Figura 2.2, tomada de [16], se muestran tres ejemplos de imágenes comprimidas, donde se puede ver claramente que con similares valores de MSE, la calidad percibida puede ser esencialmente diferente (comparar, por ejemplo, “Tiffany” con “Mandril”, sobre el lado derecho de la figura), lo que pone en duda la utilidad de este tipo de métrica como indicador de calidad. En la Figura 2.3, presentada en [17], se puede ver como la misma imagen, con el mismo valor de PSNR, puede tener diferente calidad percibida, dependiendo del lugar en el que se presenten las degradaciones. En la figura (b), se nota claramente la degradación en el cielo (parte superior), mientras que en la figura (c), una degradación similar en la parte inferior prácticamente no es perceptible. Este fenómeno se conoce como “enmascaramiento”. En zonas texturadas o con gran “actividad espacial”, las degradaciones quedan “enmascaradas” y son menos percibidas por el sistema visual humano. El enmascaramiento también puede darse en video, donde cambios rápidos temporales pueden enmascarar cierta pérdida de calidad en cada cuadro.

Debido a la baja correlación entre el MSE, RMSE y PSNR con la calidad percibida, en los últimos tiempos, se ha realizado un gran esfuerzo para desarrollar nuevos modelos que tengan en cuenta las características de percepción del sistema visual humano y que permitan calcular métricas objetivas que lo simulen. Sin embargo, no se existen, al momento de escribir el presente trabajo, métodos objetivos de la medida ***perceptual*** de calidad de video estandarizados, que apliquen a todos los casos y con buenos resultados respecto a las medidas subjetivas. Sin embargo, existen varias propuestas de métricas de medida, con diversa complejidad y precisión de sus resultados. Las métricas basadas en analizar una imagen degradada y compararla píxel a píxel con su imagen de referencia (MSE, PSNR), no son suficientes para estimar la calidad ***percibida*** de la misma. El sistema visual humano es extremadamente complejo, y puede detectar fácilmente algún tipo de distorsión, mientras que puede pasar por alto otras, dependiendo de diversos factores. Estos factores pueden incluir el tipo de aplicación (TV, video conferencia, video telefonía, etc.), el lugar de la imagen en donde se produce la degradación (generalmente las degradaciones son menos visibles en regiones con muchos detalles o "actividad espacial", o con gran movimiento, y son más visibles en imágenes estacionarias, o en fondos poco texturados). Incluso la calidad percibida puede depender del tipo de dispositivo utilizado y del tamaño del monitor [18]. En general, el sistema de visión humano juega un rol fundamental, y la ciencia no tiene aún una compresión total del mismo.

En forma genérica, los métodos objetivos de medida de calidad pueden clasificarse según la disponibilidad total, parcial o nula de la señal original, como se detalla a continuación.

2.3.1 FR - Full Reference

Estos métodos se basan en la disponibilidad de la señal original, la que puede ser contrastada con la señal degradada, cuadro a cuadro. Esto presupone una severa restricción al uso práctico de este tipo de métodos, ya que en varias aplicaciones reales esto no es posible. Los métodos que utilizan métricas del tipo FR (Full Reference) pueden ser utilizados para categorizar en forma objetiva un sistema de transmisión, un codec, el efecto de un reducido ancho de banda, o de diversos factores que degraden una señal, en ambientes controlados. Sin embargo, no son adecuados para aplicaciones de tiempo real (TV, video conferencias, etc.), ya que no es posible tener las señales originales.

2.3.2 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. Puede pensarse en la reserva de un pequeño ancho de banda (comparado con el del video) para el envío de este tipo de información adicional.

2.3.3 NR - No Reference

Las personas no necesitan señales de referencia, ni información adicional para juzgar la calidad de una señal de video. Se basan en sus experiencias previas, y en las expectativas que tengan respecto al sistema. De igual manera, estos métodos intentan estimar la calidad percibida basándose únicamente en el análisis de la señal recibida. Son los métodos más complejos de implementar, pero no requieren de otra información que la propia señal de video.

El “Video Quality Experts Group” (VQEG) [19] está realizando un interesante y exhaustivo trabajo en el estudio y comparación de desempeño de métricas objetivas, separando el estudio por áreas de aplicación, como se verá a continuación.

Figura 2.2
Evaluaciones de imágenes comprimidas con JPEG.
Arriba: Imagen “Tiffany” original y comprimida, MSE=165; Medio: Imagen “Lago” original y comprimida, MSE=167; Abajo: Imagen “Mandril” original y comprimida, MSE=163

Figura 2.3

2.3.4 El trabajo del VQEG

El VQEG (Video Quality Expert Group) [19] 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).

VQEG ha desarrollado varios proyectos, entre los que se pueden mencionar

- FR-TV (Full Reference TV)
- RRNR-TV (Reduced Reference, No Reference TV)
- Multimedia
- HDTV
- Hybrid Perceptual / Bitstream

Cada aplicación, por sus propias características, requiere de pruebas y modelos diferentes.

2.3.4.1 FR-TV

El primer proyecto, ya terminado, es el correspondiente a FR—TV. Este proyecto fue llevado a cabo en dos fases, como se detalla a continuación.

En la fase I de FR-TV, llevada a cabo entre 1997 y 2000, se evaluaron 9 propuestas, además de la PSNR. Las evaluaciones fueron realizadas con diversos tipos de material de video, incluyendo 20 tipos diferentes de contenidos (entre los que hay deportes, animaciones, escenas de interiores y exteriores, etc.) y con velocidades de 768 kb/s hasta 50 Mb/s. Se utilizó el método DSCQS y las pruebas se realizaron sobre una base de más de 26.000 opiniones subjetivas, en 8 laboratorios independientes en diferentes partes del mundo.

Como se mencionó, el objetivo es contrastar el resultado presentado por cada uno de los modelos propuestos, contra el resultado subjetivo obtenido para el mismo video, utilizando el método DSCQS. El desempeño de cada una de los modelos propuestos fue evaluado respecto a tres aspectos de su capacidad de estimar la calidad subjetiva:

- Precisión de la predicción: La capacidad de predecir la calidad subjetiva con mínimo error
- Monotonicidad: Mide si los incrementos (decrementos) en una variable están asociados a incrementos (decrementos) en la otra. Típicamente se mide con el coeficiente de correlación de Spearman
- Consistencia: Se corresponde con el grado en que el modelo mantiene la precisión a lo largo de las secuencias de pruebas. Se puede evaluar midiendo la cantidad de puntos para los que el error de la predicción es mayor a cierto umbral, por ejemplo, el doble de la desviación estándar.

Como resultado de la fase I de FR-TV, dependiendo de la métrica de comparación utilizada, siete u ocho de los modelos propuestos resultaron estadísticamente equivalentes entre sí, y a su vez, equivalentes a los resultados obtenidos con el PSNR [20]. Este resultado fue realmente desalentador, ya que indica que no existen diferencias apreciables entre el sencillo cálculo del PSNR y los sofisticados métodos preceptuales propuestos. En base a estos resultados, el VQEG ha realizado una segunda fase de pruebas, llamando nuevamente a interesados en contrastar sus modelos. La denominada “fase II” para FR-TV fue realizada entre los años 2001 y 2003 y los resultados finales fueron publicados en agosto de 2003 [21]. El objetivo de esta segunda fase era obtener resultados más discriminatorios que los obtenidos en la fase I. Al igual que en la fase I, las evaluaciones fueron realizadas con diversos tipos de material de video, y en formato de 525 y 625 líneas por cuadro. Se evaluaron velocidades entre 768 kb/s y 5 Mb/s. Las pruebas fueron realizadas en 3 laboratorios independientes, en Canadá, Estados Unidos e Italia. Sobre la base de los resultados obtenidos, la ITU ha estandarizado, en las recomendaciones ITU-T J.144 [22] y ITU-R BT.1683 [23] de 2004, a los cuatro mejores algoritmos de predicción de la calidad percibida, definidos en el marco de esta recomendación:

- British Telecom BTFR (Reino Unido) [24]
- Yonsei University/Radio Research Laboratory/SK Telecom (Corea) [25]
- Centro de Investigación y Desarrollo en Telecomunicaciones (CPqD) (Brasil) [26]

- National Telecommunications and Information Administration (NTIA)/Institute for Telecommunication Sciences (ITS) (Estados Unidos) [27]

Es importante destacar que los modelos propuestos en esta Recomendación pueden utilizarse para evaluar un codec (combinación de codificador/decodificador) o una combinación de varios métodos de compresión y dispositivos de almacenamiento en memoria. Aunque en el cálculo de los estimadores de la calidad objetiva descrito en la Recomendación se considera la degradación provocada por errores (por ejemplo, errores en los bits, paquetes rechazados), no se dispone aún de resultados de pruebas independientes para validar la utilización de estimadores en los sistemas con degradación por errores. El material de pruebas utilizado por el VEQG no contenía errores de canal.

2.3.4.2 Multimedia

El VQEG ha evaluado modelos del tipo FR, RR y NR para aplicaciones multimedia, en formatos de pantalla VGA, CIF y QCIF. Se utilizaron 346 secuencias originales de video y 5420 videos procesados basados en dichas secuencias. Se llevaron a cabo 41 experimentos de pruebas subjetivas, con un total de 984 sujetos evaluadores, contras las que se contrastaron los resultados predichos por diversos modelos propuestos [28]. En total se presentaron 5 modelos, propuestos por NTT (Nippon Telegraph and Telephone), OPTICOM, Psytechnics, SwissQual y Yonsei University.

Sobre la base de los resultados, se han publicado las recomendaciones ITU-T J.246 [29] para los modelos del tipo RR y ITU-T J.247 [30] para los modelos FR. Los modelos NR (sin referencia) no han presentado resultados lo suficientemente satisfactorios como para ser incluidos en recomendaciones de ITU-T

2.3.4.3 HDTV

El objetivo del grupo de HDTV fue evaluar modelos del tipo FR, RR y NR, para la predicción de la calidad de video percibida en aplicaciones de televisión digital de alta resolución (HDTV). Las pruebas se limitan a codecs MPEG-2 y H.264, incluyendo eventuales errores de transmisión. La resolución de pantalla a evaluar es 1080i @ 50 / 60 Hz y 1080p @ 25 / 30 fps.

El reporte final de VQEG fue aprobado en junio de 2010 [31]. Se presentaron modelos propuestos por NTT (Nippon Telegraph and Telephone), OPTICOM, SwissQual, Tektronix y Yonsei University. El modelo del tipo FR que tuvo mejor desempeño fue el propuesto por SwissQual, seguido del de Tektronix. VQEG ha propuesto estandarizar por lo menos uno de estos modelos, y ITU-T lo realizó en la Recomendación ITU-T J.341 en enero de 2011 [32]. El único proponente que presentó modelos del tipo RR fue Yonsei con resultados aceptables. VQEG ha

indicado que estos modelos podrían ser estandarizados. Finalmente no fueron presentados modelos del tipo NR.

2.4 Calidad de video en redes IP

La transmisión de video y multimedia sobre redes de datos enfrenta problemáticas específicas en lo que respecta a la calidad percibida por los usuarios. Varios tipos de degradaciones suelen presentarse en las señales de video transmitidas sobre redes de paquetes. El estudio en esta área es todavía un tema de investigación [33]. Se analizarán a continuación los factores que pueden afectar la calidad de video percibida, y luego los métodos aceptados para medirla en forma subjetiva y objetiva.

2.4.1 Factores que afectan la calidad del video sobre redes de paquetes

La transmisión de video sobre redes de paquetes, y en particular, sobre la Internet, presenta características esencialmente diferentes a la de la difusión de TV por las vías clásicas (radiofrecuencia, TV-cable). Se utilizan rangos de anchos de banda variables, hay congestión y pérdida de paquetes, y típicamente, en las aplicaciones corporativas, la observación se realiza desde distancias más cortas, y generalmente en pantallas más pequeñas.

En esta sección se describirán conceptualmente los factores específicos de las redes IP que afectan a la calidad de video.

2.4.1.1 Factor de compresión

El proceso de digitalización de video utiliza técnicas que transforman una secuencia de píxeles al dominio de la frecuencia espacial (DCT), cuantificando valores, descartando eventualmente componentes de alta frecuencia, y haciendo uso de técnicas de predicción y compensación de movimientos. Esto genera “ruido de cuantificación”, el que puede degradar la imagen original a niveles perceptibles. Es este “ruido de cuantificación” [34], el que genera las clásicas degradaciones que pueden verse en imágenes y videos con alta compresión, entre ellos, el conocido “efecto de bloques”, que hace ver a la imagen como un conjunto de bloques pequeños.

Los algoritmos de compresión utilizados actualmente en la codificación digital de video introducen varios tipos de degradaciones, las que se pueden clasificar según sus características principales [35]. Esta clasificación es útil para poder comprender las causas de las degradaciones y el impacto que tiene en la calidad percibida.

Debido al gran ancho de banda requerido en video para enviar la señal sin comprimir, el uso de codecs con compresión es sumamente habitual en video.

Esto pone un límite a la calidad de imagen recibida, el que es independiente del medio de transmisión sobre el que viaje la señal.

Las principales degradaciones introducidas por el uso de codecs con compresión son las siguientes:

- 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

2.4.1.2 Pérdida de paquetes

La pérdida de paquetes en las redes IP afecta a la calidad percibida de video. En la Figura 0.1, tomada de [36], se muestra como la pérdida de un paquete puede propagarse, afectando no solo a la información de video contenida en dicho paquete, sino a otras partes del mismo o diferentes cuadros. Típicamente, y dado que la codificación se realiza en forma diferencial, la pérdida de un paquete afectará a todos los bloques siguientes en la misma fila (“slice”). Si el paquete perdido corresponde además a un cuadro de referencia (I), también se verán afectados los cuadros predictivos, posteriores o anteriores, propagándose el error en el tiempo.

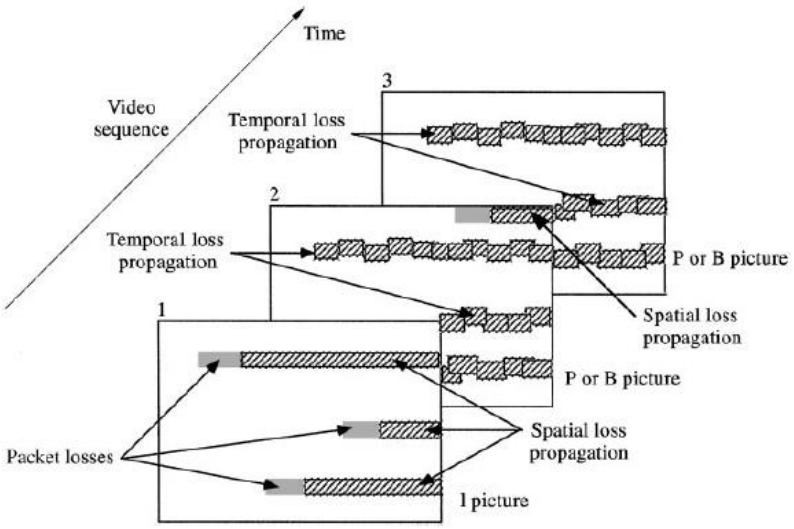


Figura 0.1

Existen técnicas de cancelación de paquetes perdidos, las que tratan de reconstruir la información perdida en base a información disponible. Por ejemplo, reemplazando los píxeles perdidos por los mismos valores de cuadros anteriores.

Varios trabajos se han realizado, estudiando la manera en que la calidad de video se ve afectada por la pérdida de paquetes. Algunos [37] proponen estimar el MSE del video, en base a diferentes técnicas, que tienen en cuenta la pérdida de paquetes. Sin embargo, estos son estimadores no se corresponden con la calidad "perceptual", al basarse en la predicción del MSE o PSNR, los que no tienen relación directa con el MOS.

En [38] se muestra que no basta con conocer el porcentaje de pérdida de paquetes para estimar como se ve afectada la calidad de video percibida. Dependiendo de diversos factores, la pérdida de un paquete determinado de video puede o no afectar la calidad percibida. Por ejemplo, en imágenes casi estáticas, el video perdido puede ser reconstruido en base a imágenes anteriores, casi sin pérdida de calidad, lo que lleva a que la pérdida de un paquete sea prácticamente imperceptible. Algo similar sucede cuando la pérdida solo afecta a un cuadro. De esta manera, se propone un "clasificador de paquetes perdidos", que, en base a un algoritmo, decide si el paquete perdido afectará o no a la calidad percibida del video. El algoritmo toma en cuenta cuantos cuadros se verán afectados por la pérdida del paquete, la movilidad de la imagen y su varianza y el error introducido medido con MSE, entre otros factores. Se define el parámetro VPLR (Visible Packet Loss Rate), en lugar del clásico PLR (Packet Loss Rate), para ser utilizado como entrada a los algoritmos de predicción de calidad en base a la pérdida de paquetes.

La idea de detectar como afecta la posible pérdida de cada paquete en la calidad perceptual es explorada en [39], donde se propone un método que garantice una calidad perceptual constante en la recepción de un flujo de video. La idea en este

caso es marcar solo ciertos paquetes específicos con mayor prioridad (asumiendo una red con soporte para Diff Serv), de manera de "garantizar" su llegada a destino, sin inundar a la red con todos los paquetes de video marcados como prioritarios, sino solo con los paquetes cuya pérdida afecten especialmente la calidad percibida. El algoritmo se basa únicamente en la calidad deseada (PSNR) y la tasa de pérdida de paquetes (PLR).

Una idea similar es presentada en [40], donde se presenta una métrica para priorizar ciertos paquetes de video, en base a la estimación de la distorsión percibida en caso de la pérdida o llegada fuera de tiempo de cada paquete.

Dado que las pérdidas de paquetes se dan generalmente en ráfagas, la calidad puede verse fuertemente degradada por la pérdida de varios paquetes consecutivos, correspondientes a cuadros consecutivos. En [41] se propone una técnica que consiste en reagrupar los cuadros que se envían, generando un buffer en el codificador de 3 GOPs, y reagrupando el orden en el que se envía la información. Con esto se logra difundir los paquetes perdidos entre varios cuadros separados en el tiempo, y de esta manera, según los autores, mejorar la calidad percibida.

Si bien varios trabajos se han realizado acerca de la degradación del video debida a la pérdida de paquetes, el tema está aún abierto, y no hay aún estándares ni trabajos sistemáticos de comparación de diferentes modelos.

2.4.1.3 Demora / Jitter

El receptor debe recibir los paquetes a decodificar a intervalos constantes, para poder regenerar de forma adecuada la señal original. Dado que el jitter es inevitable en las redes de paquetes, los receptores disponen de un buffer de entrada, con el objetivo de "suavizar" el efecto de la variación de las demoras. Este buffer recibe los paquetes a intervalos variables, y los entrega a intervalos constantes.
Es de hacer notar que este buffer agrega una demora adicional al sistema, ya que debe retener paquetes para poder entregarlos a intervalos constantes. Cuánto más variación de demoras (jitter) exista, más grande deberá ser el buffer, y por lo tanto, mayor demora será introducida al sistema. Las demoras son indeseables, y tienen impacto directo en la experiencia del usuario, sobre todo en contenidos de tiempo real (por ejemplo, distribución de eventos deportivos en línea) y en aplicaciones conversacionales (por ejemplo video telefonía, o video conferencias).
Se hace necesario disponer de mecanismos que minimicen el tamaño de los jitter-buffers, pero que a su vez, no comprometan la calidad debida a la perdida de paquetes que no han llegado a tiempo.

En [42] se propone un método dinámico al que llaman AMP (Adaptive Media Playout), en el que se cambia la velocidad de reproducción del medio (video / audio) dependiendo de la condición del canal de transmisión, y logrando de esta

manera reducir el tamaño del jitter-buffer. Se indica que aumentar o disminuir la velocidad de ciertas partes del contenido hasta en un 25% es subjetivamente mejor que aumentar las demoras totales o tener interrupciones.

En el proyecto Multimedia del VQEG se proponen realizar pruebas con demoras entre 2 milisegundos y 5 segundos, y pérdidas de paquetes impulsivas en el rango de 0 a 50%.

2.4.2 ITU-T G.1070

ITU-T ha publicado un modelo de predicción de la calidad de video, para aplicaciones de video telefonía, en base a parámetros medibles de una red IP. Este modelo es similar al E-Model visto anteriormente para audio. Ha sido desarrollador por NTT (de Japón) y estandarizado en la recomendación ITU-T G.1070 “Opinion model for video-telephony applications” [43].

La Recomendación G.1070 propone un modelo que estima la calidad percibida en el uso de aplicaciones de “video telefonía”, que puede ser utilizado al momento de planificar una red de datos que transmita este tipo de servicios a través de IP

El modelo consiste en tres funciones, una para la estimación de la calidad del video (V_q), otra para la estimación de la calidad del audio (S_q), y finalmente una para la estimación de la calidad multimedia (MM_q). En la Figura 2.4 se esquematiza el funcionamiento general del modelo.

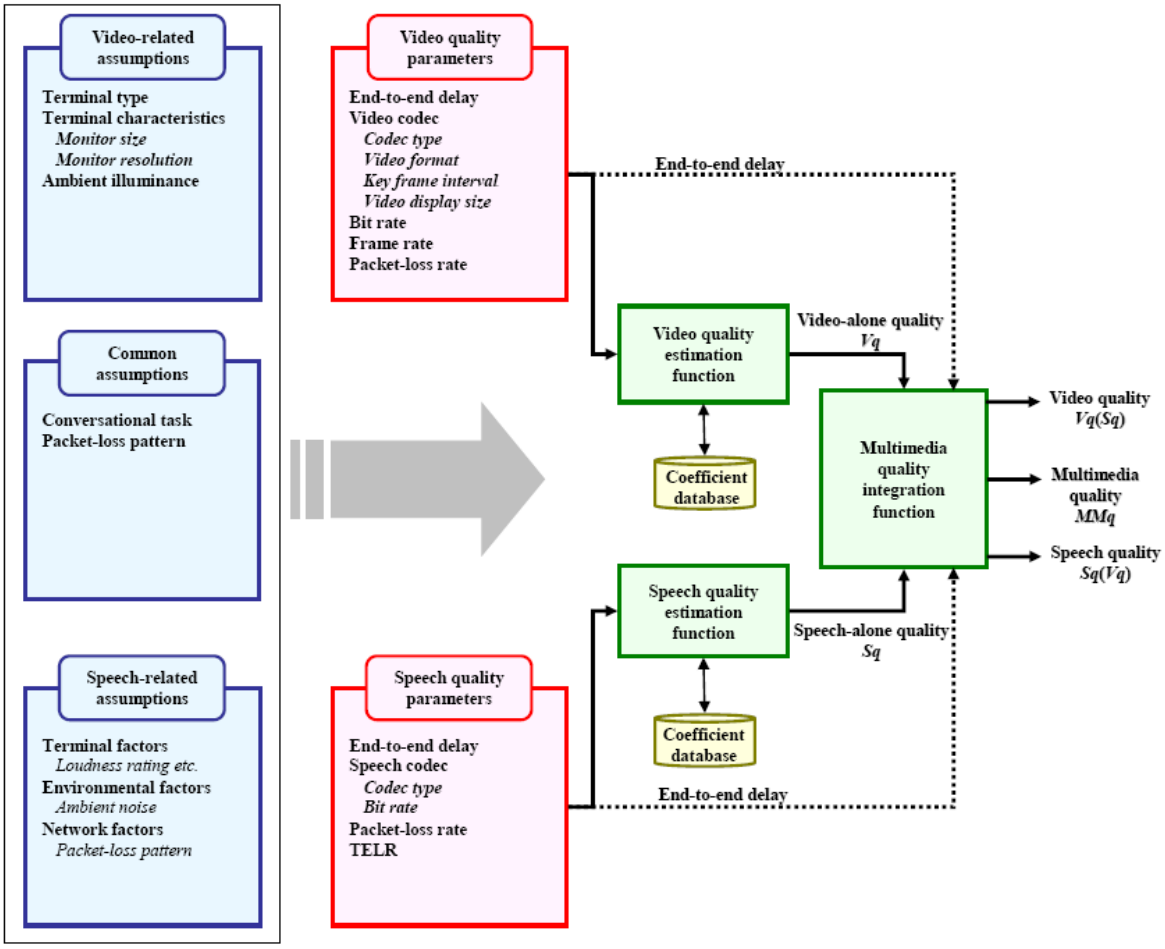


Figura 2.4

La estimación de la calidad de la voz, básicamente, se reduce el E-Model, simplificado.

$$Q = R_o - I_{dte} - I_{e,eff} \qquad (2.3.5.1)$$

donde Q es el equivalente a R en el E-Model. Los parámetros I_{dte} y $I_{e,eff}$ mantienen sus definiciones del E-Model. Las degradaciones introducidas por las demoras son quitadas de la estimación de la calidad de voz, e incluidas en la calidad "multimedia".

La relación entre Q y S_q es similar a la relación entre MOS y R en el E-Model:

Para $Q < 0$: $S_q = 1$

Para $0 < Q < 100$: $S_q = 1 + 0{,}035Q + Q(Q-60)(100-Q)7 \cdot 10^{-6}$ (2.3.5.2)

Para $Q > 100$: $S_q = 4{,}5$

La estimación de la calidad del video se basa en la siguiente fórmula

$$V_q = 1 + I_c e^{-\frac{P_{plv}}{D_{Pplv}}} \qquad (2.3.5.3)$$

Donde I_c representa la calidad del video dada únicamente por las condiciones de codificación, P_{plv} es el porcentaje de pérdida de paquetes y D_{Pplv} representa el grado de robustez respecto a la pérdida de paquetes. Tanto I_c como D_{Pplv} dependen del codec utilizado, el bit rate y el frame rate.

La estimación de la calidad multimedia se expresa en la siguiente fórmula

$$MM_q = m_1 MM_{SV} + m_2 MM_T + m_3 MM_{SV} MM_T + m_4 \qquad (2.3.5.4)$$

Donde MM_{SV} representa la calidad audiovisual, y es función de (V_q, S_q) y MM_T contiene los factores de calidad asociados a las demoras, tanto del audio como del video, y tiene en cuenta las degradaciones producidas por la falta de sincronismo entre ambos medios.

Referencias

[1] Recommendation ITU-T P-800.1 Mean Opinion Score (MOS) Terminology, July 2006

[2] Recommendation ITU-T P-800 Mean Opinion Score (MOS) (Métodos de determinación subjetiva de la calidad de transmisión), Aug 1996, http://www.itu.int/rec/T-REC-P.800/es

[3] Recommendation ITU-T P.862: Evaluación de la calidad vocal por percepción: Un método objetivo para la evaluación de la calidad vocal de extremo a extremo de redes telefónicas de banda estrecha y códecs vocales, Febrero 2001

[4] Recommendation ITU-T P.862.1: Función de correspondencia para convertir los resultados brutos de la prueba P.862 en nota media de opinión de la calidad de escucha objetiva, noviembre 2003

[5] Recommendation ITU-T P.862.2: Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs, Noviembre 2007

[6] Recommendation ITU-T P.862.2: Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs, Noviembre 2007

[7] Calidad de Voz y Video, José Joskowicz (Marzo 2011)

[8] ITU-T G.711 Appendix 1, A high quality low-complexity algorithm for packet loss concealment with G.711, 9/1999.

[9] ITU-T G.107 The E-model, a computational model for use in transmission planning, March 2005, http://www.itu.int/rec/T-REC-G.107/e

[10] E-model tutorial
http://www.itu.int/ITU-T/studygroups/com12/emodelv1/introduction.htm

[11] TIA/TSB 116-A Telecommunications - IP Telephony Equipment – Voice Quality Recommendations for IP Telephony, Mar 1, 2006

[12] “Degradaciones de la transmisión debido al tratamiento de las señales vocales”, Recomendación ITU-T G.113 (2001)

[13] RFC 3611: “RTP Control Protocol Extended Reports (RTCP XR)”, T. Friedman et al (November 2003)

[14] Recommendation ITU-R BT.500-13
Methodology for the subjective assessment of the quality of television pictures
01/2012

[15] Recommendation ITU-T P.910
Subjective video quality assessment methods for multimedia applications
09/1999

[16] The handbook of Video Databases: Dessign and Applications, Chapter 41
B. Furth and O, Marqure
September 2003

[17] Digital Video Quality, Vision Models and Metrics
Stefan Winkler
John Wiley & Sons Ltd, 2005

[18] Effect of Monitor Size on User-Level QoS of Audio-Video Transmission over IP Networks in Ubiquitous Environments
Y. Ito and S. Tasaka
IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2005, September 2005, Vol 3, pp 1806-1812

[19] Video Quality Experts Group
http://www.its.bldrdoc.gov/vqeg/

[20] FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF VIDEO QUALITY ASSESSMENT
June, 2000

[21] FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF VIDEO QUALITY ASSESSMENT, PHASE II ©2003 VQEG
August 25, 2003

[22] Recommendation ITU-T J.144 Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference, February 2004

[23] Recommendation ITU-R BT.1683 Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference, January 2004

[24] Alexandre J Bourret, David S Hands, Damien Bayart, Andrew G Davies: Method and System for Video Quality Assessment, US Patent No. 2006/0152585 A1, July 13, 2006

[25] Sungdeuk Cho, Jihwan Choe, Taeuk Jeong, Wonseok Ahn, and Eunjae Lee: Objective video quality assessment, Optical Engineering Vol. 45 (1), January 2006

[26] Alengar Lotufo, R Da Silva, W D F Falcao, A X Pessoa: Morphological image segmentation applied to video quality assessment, IEEE Proceedings in Computer Graphics, Image Processing and Vision, SIGGRAPI Proceedings, pp 468-475, October 1998

[27] Margaret H Pinson and Stephen Wolf: A New Standardized Method for Objectively Measuring Video Quality, IEEE Transactions on Broadcasting, Volume 50, Issue 3, September 2004, pp. 312-322

[28] FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF MULTIMEDIA QUALITY ASSESSMENT, PHASE I ©2008 VQEG (Version 2.6 September 12, 2008)

[29] Perceptual audiovisual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference (Agosto 2008)

[30] Objective perceptual multimedia video quality measurement in the presence of a full reference (Agosto 2008)

[31] Report on the Validation of Video Quality Models for High Definition Video Content
VQEG Version 2.0 (June 30, 2010)

[32] Recommendation ITU-T J.341, “Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference”, January 2011

[33] Estudio de la Medida de la Calidad Perceptual de Video, José Joskowicz, Universidad de Vigo (Marzo de 2008)

[34] DCT Quantization Noise in Compressed Images, Mark A. Robertson and Robert L. Stevenson, IEEE Transactions on Circuits and Systems For Video Technology, Vol. 15, No. 1, January 2005

[35] Digital Video Image Quality and Perceptual Coding, H.R. Wu and K.R Rao 2006, CRC Press

[36] User-Oriented QoS Analysis in MPEG-2 Video Delivery, O Verscheure, P Frossard, M Hamdi, Real Time Imaging 5, 1999, pp 305-314

[37] Quality Monitoring of Video Over a Packet Network, Amy R. Reibman, Vinay A. Vaishampayan and Yegnaswamy Sermadevi, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 6, NO. 2, APRIL 2004

[38] Visibility of individual packet losses in MPEG-2 video
Amy R. Reibman, Sandeep Kanumuri, Vinay Vaishampayan and Pamela C. Cosman, IEEE International Conference on Image Procecssing) 2004, ICIP’04, Vol 1, pp 171-174

[39] Delivery of MPEG Video Streams with constant perceptual quality of service
Quaglia, D. De Martin, J.C. , IEEE, Proceedings International Conference on Multimedia, 2002

[40] Estimation of packet loss effects on video quality, Bouazizi, I., IEEE, First International Symposium on Control, Communications and Signal Processing, 2004, pp 91-94.

[41] Packet Loss Resilience for MPEG-4 Video Stream over the Internet, Jae-Young Pyun, Jae-Han Jung, Jae-Jeong Shim, IEEE Transactions on consumer electronics, Vol 48, issue 3, August 2002

[42] Adaptive Media Playout of Low Delay Video Streaming Over Error Prone Channels , Mark Kalman, Eckehard Steinbach, Bernd Girod,
IEEE Transactions on Circuits and Systems for Video Technology, Vol 14, no 6, June 2004

[43] Recomendación ITU-T G.1070: “Opinion model for video-telephony applications” (Abril 2007)