

Medida de la calidad de voz en redes IP

José Joskowicz
IIE/FING/UDELAR
Montevideo, Uruguay
josej@fing.edu.uy

Rafael Sotelo
FI/UM
Montevideo, Uruguay
rsotelo@um.edu.uy

Resumen – En este artículo se realiza un estudio de la medición de Calidad de Voz en redes IP. Se presentan los métodos y arquitecturas de medición aceptados, estandarizados y normalizados, así como el “estado del arte” en la materia a través de un estudio de las técnicas propuestas recientemente. Se analizan trabajos que ilustran el estado de los “backbones” de Internet para transmisión de VoIP.

Términos Clave – VoIP, QoS, QoE.

I. INTRODUCCIÓN

La transmisión de voz sobre redes IP (VoIP) tiene un rol cada vez más importante y un uso cada vez más difundido. Los usuarios esperan ver satisfechas sus expectativas de calidad del servicio con independencia de la tecnología utilizada. En este sentido, la “calidad de experiencia” QoE mide que tan bien un servicio de red satisface las expectativas y necesidades vistas por el usuario. Por otro lado, la “calidad de servicio” QoS refiere a la medida del rendimiento de la red desde el punto de vista técnico, y a la posibilidad de gestionarla para cumplir con las prestaciones necesarias para las aplicaciones.

La VoIP enfrenta problemáticas propias de las redes de datos, que se manifiestan como degradaciones en la calidad del servicio percibida por los usuarios (QoE). 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 en forma masiva y comercial, es esencial garantizar una calidad de voz aceptable. Para ello se han desarrollado métodos para medirla. Estos métodos se dividen en *subjetivos* y *objetivos*. Los métodos *subjetivos* de medida de la calidad de servicio, se basan en conocer directamente la opinión de los usuarios. Típicamente resultan en un promedio de opiniones (por ejemplo, en un valor de MOS – Mean Opinión Store). Los métodos *objetivos* miden propiedades físicas de una red para prever o estimar el rendimiento percibido por los usuarios. 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, por ejemplo PESQ) 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).

El presente artículo brinda una breve introducción a la transmisión de voz sobre redes IP, luego describe los métodos y arquitecturas de medición aceptados, estandarizados y normalizados. Presenta la evaluación realizada por varios artículos acerca del estado del backbone de Internet para

soportar VoIP y finalmente presenta los aportes y tendencias recientes en la materia.

II. VOZ SOBRE REDES DE DATOS

A. Paquetización de la Voz

Para poder transmitir las muestras codificadas de voz sobre redes de datos, es necesario armar “paquetes”. Para minimizar la sobrecarga del armado de paquetes y no introducir demoras inaceptables, se toman “ventanas” de 10 a 30 ms. Las muestras de voz de cada una de estas ventanas consecutivas se acumulan y con ellas se forman los “paquetes de voz”.

B. RTP – Real-Time Transport Protocol, RTCP – RTP Control Protocol

El protocolo RTP, basado en el RFC 3550 [1], establece los principios de un protocolo de transporte sobre redes que no garantizan calidad de servicio para datos “de tiempo real”, como por ejemplo voz y video.

Cada paquete RTP consiste en un cabezal y los datos de voz. El cabezal contiene números de secuencia, marcas de tiempo, y monitoreo de entrega. Las aplicaciones típicamente utilizan RTP sobre protocolos de red “no confiables”, como UDP. Los “bytes” obtenidos de cada conjunto de muestras de voz o video son encapsulados en paquetes RTP, y cada paquete RTP es a su vez encapsulado en segmentos UDP.

RTP soporta transferencia de datos a destinos múltiples, usando facilidades de “multicast”, si es provisto por la red.

El RFC 3550 establece, además, un protocolo de control, RTCP, encargado de enviar periódicamente paquetes de control entre los participantes de una sesión.

C. Ancho de Banda

Dado que para el envío de voz sobre redes es necesario armar “paquetes”, el ancho de banda requerido dependerá de la “sobrecarga” (“overhead”) que generen estos paquetes.

Como se ha visto, para el envío de voz sobre redes de paquetes se utiliza el estándar RTP. Este protocolo a su vez se monta sobre UDP, el que a su vez se monta sobre IP, el que, en la LAN, viaja típicamente sobre Ethernet

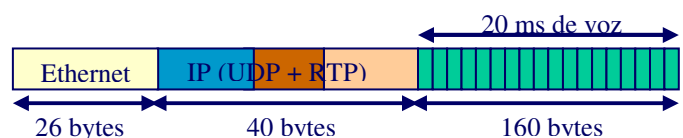


Fig. 1

Esta suma de protocolos hace que el ancho de banda requerido para el tráfico de voz sobre Ethernet sea bastante mayor al ancho de banda del audio. Para una ventana de 20 ms, y con codificación de audio Ley A, se obtienen 160 bytes de voz por trama, lo que equivale a un ancho de banda de 90,4 kb/s (compárese con los 64 kb/s del flujo de audio)

Este cálculo fue hecho para el envío de audio en una dirección. Como las comunicaciones son bidireccionales, el ancho de banda real requerido en la LAN será el doble. Pueden utilizarse técnicas de “supresión de silencio”, en las que no se envían paquetes cuando no hay audio. En este caso, el ancho de banda total es similar al ancho de banda unidireccional.

El ancho de banda de la voz paquetizada en la LAN depende del tamaño de la “ventana” (típicamente 10, 20 o 30 ms) y el CODEC utilizado.

TABLA I

Tipo de Codec	Duración de Trama (ms)	Bytes de voz/Trama	Bytes de paquete IP	Bytes de trama Ethernet	Ancho de Banda en LAN (kbps)
G.711 (64 kb/s)	10	80	120	146	116,8
	20	160	200	226	90,4
	30	240	280	306	81,6
G.729 (8 kb/s)	10	10	50	76	60,8
	20	20	60	86	34,4
	30	30	70	96	25,6
G.723 (6.3 kb/s)	20	16	56	82	32,7
	30	24	64	90	23,9
G.723 (5.3 kb/s)	20	13	53	79	31,7
	30	20	60	86	22,9

La Tabla I muestra los anchos de banda unidireccionales necesarios utilizando redes IP sobre Ethernet. Como se puede ver, el ancho de banda puede variar de 23 kb/s a 117 kb/s, dependiendo del CODEC y la ventana seleccionada.

D. Factores que afectan la calidad de voz sobre redes de paquetes

Varios factores afectan la calidad de voz percibida por los usuarios, cuando se transmite a través de una red de datos. Estos factores incluyen la compresión utilizada (Codec), el porcentaje de pérdida de paquetes, los retardos debidos a diversas causas (algoritmos de compresión, tiempo de procesamiento, latencia de la red, etc.), el eco, las variaciones en la demora (jitter) y el tamaño de los paquetes.

III. MÉTODOS SUBJETIVOS

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 Opinión 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.

El valor obtenido 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. Por esto se ha estandarizado también el “Q-Method”, en la recomendación ITU-T P.810 [3].

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 y de las experiencias previas de los usuarios.

IV. MÉTODOS OBJETIVOS

A. E-Model

Muchos estudios realizaron pruebas para relacionar los varios parámetros de red medibles con el MOS. La ITU ha recomendado el modelo E [4], un modelo informático para utilización en planificación de la transmisión. Este modelo ha sido adoptado también por la ETSI y por la TIA [5] y es el modelo de opinión más ampliamente difundido. El resultado primario del modelo es una cuantificación escalar de la calidad de audio que se estima percibirá un usuario. Una característica fundamental de este modelo es la utilización de factores de degradación de la transmisión que reflejen los efectos de los modernos dispositivos de procesamiento de señales. El E-model calcula, en base a varios parámetros medibles de la red, un parámetro R que puede relacionarse con el MOS de acuerdo a la Tabla II.

TABLA II

R	User Satisfaction	MOS
100		4.5
94.3	Very Satisfied	4.4
90	Satisfied	4.3
80	Some users dissatisfied	4.0
70	Many users dissatisfied	3.6
60	Nearly all users dissatisfied	3.1
50	Not recommended	2.6
0		1

R se obtiene según la Ecuación 1:

$$R = (R_o - I_s) - I_d - I_e\text{-eff} + A \quad (1)$$

El E-model calcula los sumandos R_o , I_s , I_d , I_e -eff a partir de los parámetros de transmisión que se ven en la figura 2.

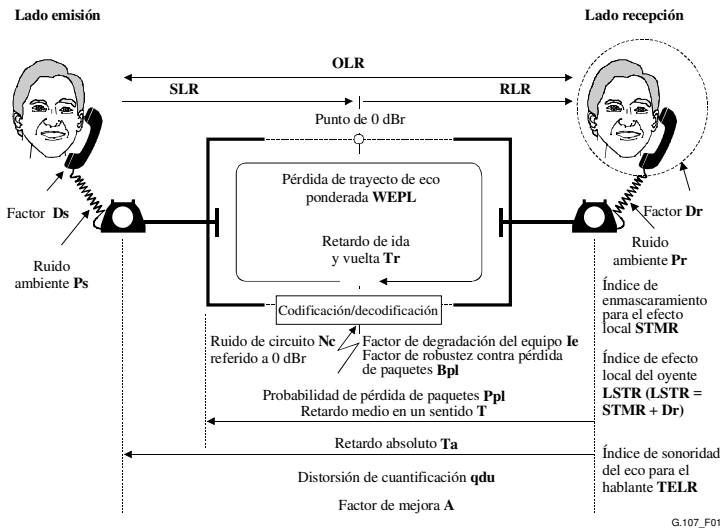


Fig. 2

R_o (efecto del ruido) e I_s (relacionado con el volumen de la conexión y con la cuantificación) son intrínsecos a la señal de voz en la entrada de la red y no dependen de ésta. Son irrelevantes al comparar VoIP a llamadas PSTN.

I_d modela las degradaciones producidas por los retardos y el eco, mientras que I_e -eff representa las degradaciones producidas por los códecs y por las pérdidas de paquetes de distribución aleatoria.

A representa el factor de ventaja que significa que el usuario aceptaría una degradación en la calidad a cambio de facilidad de acceso (por ejemplo, en telefonía móvil o satelital).

Posteriormente se propuso tener en cuenta el "recency effect". Para establecer la calidad de voz de una llamada, se la divide en períodos de tiempo con pérdidas uniformes. Al variar las pérdidas, variará el I_e . Sin embargo el usuario percibe los cambios más tarde, según una curva exponencial, que será más rápida cuando empeoran las pérdidas que cuando mejoran.

El factor I_e del E-Model corresponde a las degradaciones producidas por los codecs y por las pérdidas de paquetes. Sin embargo, en una conversación telefónica realizada a través de VoIP, la tasa de pérdida de paquetes no puede considerarse constante. Por el contrario, es generalmente típico que la tasa de pérdida de paquetes sea baja durante ciertos períodos y crezca considerablemente durante "ráfagas". Por ello, la calidad de la voz tampoco puede considerarse constante durante la conversación, sino que podrá variar según el momento. Asumiendo que existen dos "estados" relativos a la pérdida de paquetes (un estado de baja pérdida y otro de alta pérdida), también existirán dos valores instantáneos de I_e , los que se pueden denominar I_{eg} ("gap") durante el período de baja tasa de pérdida de paquetes y I_{eb} ("burst") durante el período de

ráfaga de pérdida de paquetes. Sin embargo, la percepción humana no varía en forma instantánea. La recomendación ETSI 102 024-5 [6] propone un modelo exponencial para modelar la evolución de la percepción en función de los cambios instantáneos de I_e , según se puede ver en la figura 3.

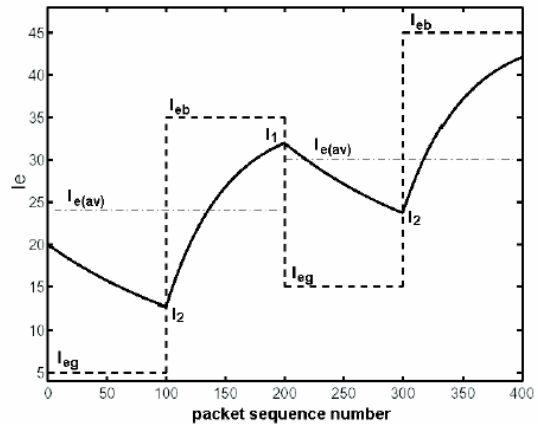


Fig. 3

En la revisión efectuada en el año 2000, se proporcionó una versión mejorada del modelo E con el fin de tener en cuenta mejor los efectos del ruido ambiente en el lado emisor y la distorsión de cuantificación. En la revisión de 2002 se ha incluido la degradación debida a la pérdida aleatoria de paquetes en una manera paramétrica para diferentes códecs. En relación con la versión de 2003, puede verse un modelado mejorado de la calidad para niveles bajos del efecto local del hablante. La revisión 2005 permite realizar predicciones más precisas sobre la calidad de los códecs en condiciones de pérdida de paquetes dependiente (a corto plazo).

B. Perceptual Evaluation of Speech Quality (PESQ)

La recomendación ITU-T P.862 [7] presenta un método objetivo para la evaluación de la calidad vocal de extremo a extremo de redes telefónicas de banda estrecha y códecs vocales.

Esta Recomendación describe un método objetivo para predecir la calidad subjetiva de la voz telefónica utilizando los códecs más comunes. 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. Es un método "intrusivo" ya que es necesario enviar un audio conocido para poder evaluar la QoE.

El método objetivo descrito se conoce por "evaluación de la calidad vocal por percepción" (PESQ, *perceptual evaluation of evaluation of speech quality*).

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 inicial $X(t)$ 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 4.

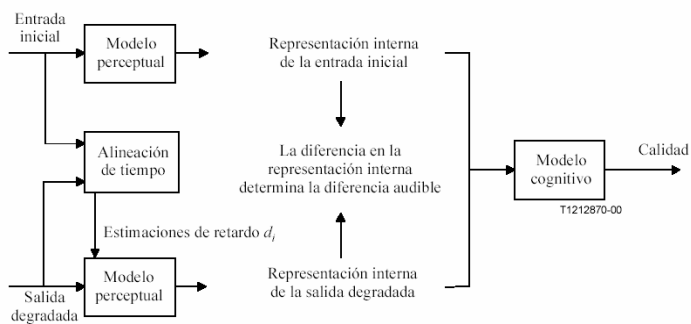


Fig. 4

Lo esencial en este proceso es la transformación de la señal 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 nota PESQ se hace corresponder a una escala similar a la de MOS, un número único 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.

C. ITU-T P.563

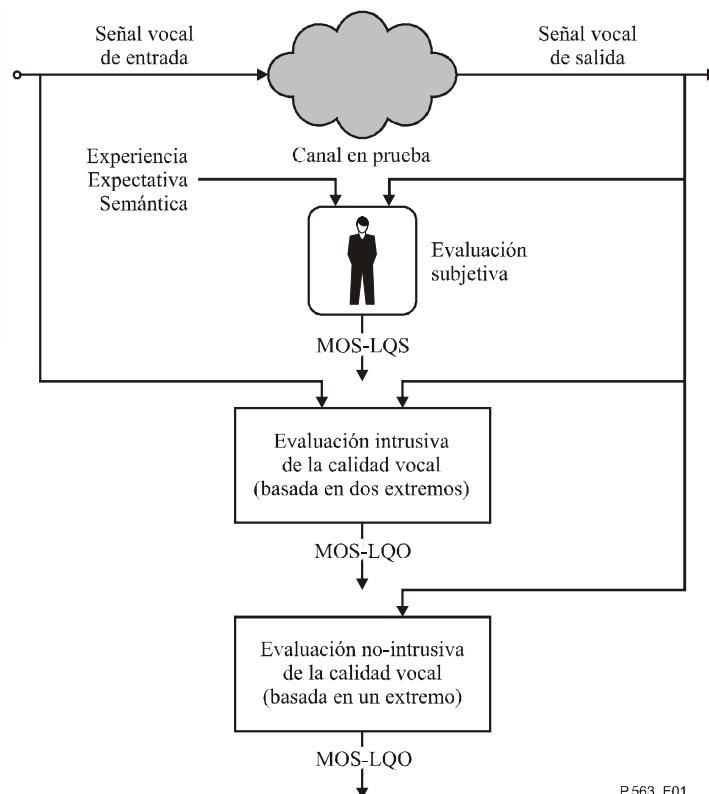
El algoritmo P.563 [8] es aplicable a la predicción de la calidad vocal sin una señal de referencia independiente. Por ese motivo, este método se recomienda para la evaluación no intrusiva de la calidad vocal y para la supervisión y evaluación con la red en funcionamiento, empleando en el extremo lejano de una conexión telefónica fuentes de señal vocal desconocidas.

En comparación con la Rec. ITU-T P.862, P.563 predice la calidad de la voz de una señal degradada sin una señal vocal de referencia dada. En la figura 5 se ilustran las diferencias entre ambos métodos.

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**.



P.563_F01

Fig. 5

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, las que se corresponden con el análisis del tracto vocal y desnaturalización de la voz, el análisis de un ruido adicional intenso y las 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.

V - ESTADO DE INTERNET PARA VOIP

Algunos artículos plantean el estado de Internet (particularmente de los backbones) para transmisión de VoIP.

Por ejemplo, en [9] se realizan estudios a partir de trazas obtenidas de cuarenta y tres enlaces de backbone de Internet de siete proveedores distintos de EEUU. Constatan que los backbones estudiados pueden brindar buena calidad para VoIP si tienen ancho de banda disponible. Esto aplica para enlaces cortos o largos, ya que estos últimos (costa a costa de EEUU) tienen retardos no muy mayores a 150 ms. Sin embargo para enlaces costa a costa muy cargados, no se tendrá buena calidad y sería necesario implementar un sistema de QoS para privilegiar a los paquetes de voz. Los autores encuentran, como es lógico, que existe un compromiso entre retardo y pérdidas en el buffer de reproducción y que existe un máximo $MOS(loss, delay)$ correspondiente al mejor rendimiento posible del enlace. Es sumamente importante la elección del buffer de reproducción. Una elección conservadora de un valor fijo alto, previene pérdidas y lleva a un buen rendimiento en enlaces con bajo retardo (menor a 150 ms). También consideraron un buffer de reproducción adaptativo [10] y resultó bueno en los enlaces largos con mucha carga, pero no en los enlaces de otro tipo.

En [11] encontramos buenos reportes de funcionamiento para VoIP de backbones de Internet. En este caso en Holanda. Se experimentó en base a 12 puntos del país en un mismo ISP. Se comprobó que es muy poco frecuente que ocurra reordenamiento de paquetes, que los paquetes experimentan en general bajos retardos y bajas pérdidas, y que las redes pueden soportar calidad de VoIP satisfactoria.

[12] realiza mediciones en la Red Nacional de Educación e Investigación de Brasil (RNP) para probar una herramienta propia creada a partir de una modificación del E-Model. Se realizan transmisiones de voz en un solo sentido entre dos universidades. En este caso se llegan a tener valores MOS muy bajos, que dependen del sentido de la transmisión y del número de llamadas. Se menciona la presencia de otras aplicaciones compartiendo el enlace. Sin embargo no se aportan datos sobre el ancho de banda utilizado por las mismas. No hay QoS implementado. Se constata un comportamiento asimétrico del *backbone*.

Por otro lado, los autores de [13] han evaluado la disponibilidad del servicio VoIP en Internet basándose en medidas realizadas en 14 nodos en EEUU, Suiza, Finlandia, Japón y China. Han encontrado que la pérdida de paquetes no es un evento raro, y en general es peor en enlaces internacionales. En segundo lugar, constatan faltas en la red. En algunos casos, estas faltas son largas, llevando a que los tiempos de restablecimiento sean largos. En tercer lugar, a pesar que Internet2 tiene menores retardos y pérdidas, el efecto de faltas en la red es similar a la Internet pública.

VI - APORTES RECIENTES A LA MEDICIÓN DE CALIDAD DE VOZ EN REDES VOIP

El ITU-T está estandarizando una metodología de medición de calidad de voz basado solamente en información de los paquetes IP, asumiendo un “payload” de voz genérico. Será el estándar P.VTQ. Es una herramienta **no intrusiva** ya que no ocupa el canal. Servirá para monitoreo en tiempo real de la calidad de transmisión. Permitirá a un proveedor conocer la calidad de voz que está ofreciendo en un momento concreto a un cliente para realizar una *gestión dinámica de la red*.

Una de las herramientas candidatas para el estándar es PsyVoIP [14] que está en uso en la industria. Opera monitoreando en tiempo real llamadas VoIP de clientes para determinar la calidad de voz brindada por una red VoIP.

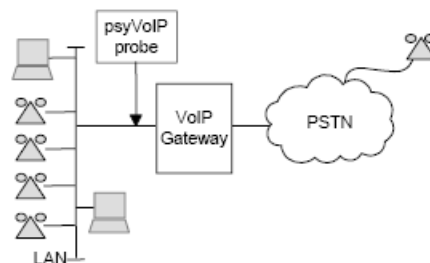


Fig. 6

La punta de prueba psyVoIP captura los datos de una llamada específica, como se muestra en la Figura 6. PsyVoIP clasifica y caracteriza los distintos gateways y teléfonos IP ya que se comprobó que dan distinta calidad ante las mismas condiciones de red.

Otra propuesta de medición no intrusiva se presenta en [15]. Se propone tomar los paquetes recibidos, extraer su encabezamiento y sustituir el “payload” de voz con una señal conocida. De esa forma se obtiene una señal sustituta conocida, con la misma distorsión que la recibida. Luego se puede aplicar cualquiera de las herramientas objetivas que precisan de la señal original (PESQ, PQSM) y obtener un MOS (Ver figura 7).

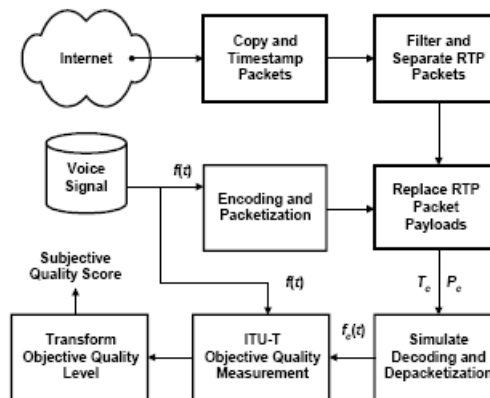


Fig. 7

Existen métodos objetivos de medición de voz a partir de *Data Mining* [16]. Una gran cantidad de características de distorsión en la percepción se extraen de la señal de voz. Se utilizan “classification and regresión trees” (CART) y “multivariate adaptive regression splines” (MARS) para seleccionar las características más salientes del conjunto, y para construir buenos estimadores de la calidad de escucha subjetiva basado en las características seleccionadas. El método del artículo resulta particularmente interesante ya que el algoritmo puede ser implementado en tiempo real y es escalable con más datos para aprender.

Otros autores [17] sugieren utilizar redes neuronales aleatorias (RNN) para realizar mediciones no intrusivas en tiempo real, y muestran buenos resultados en varios idiomas (español, árabe y francés).

Algunas críticas se han realizado al E-model. Varias de ellas cuestionan la linealidad de las degradaciones asumidas en el modelo, y proponen alternativas de cálculo donde los parámetros de las degradaciones se combinan en fórmulas no lineales [18] [3]. Otras, indican que requiere de pruebas subjetivas para caracterizar nuevos Codecs, y proponen para ello mecanismos objetivos basados en PESQ [19].

En [20] se indica que el E-Model no tiene en cuenta degradaciones debidas a adaptaciones dinámicas, por ejemplo cambio de Codecs o cambio del tamaño del buffer de reproducción.

VII - ESTUDIOS FUTUROS

La “calidad de la experiencia” (QoE) en sistemas de voz ha sido estudiada durante muchos años, y los requerimientos son bien conocidos. Sin embargo, estos conceptos se deben extender para una cantidad de nuevas aplicaciones no vocales, por ejemplo, video en sistemas móviles, video en IPTV, “channel zapping”, video conferencias, etc.

También es necesario extender algunos de los estándares actuales para nuevas aplicaciones de audio. El actual E-Model fue pensado para aplicaciones vocales de banda estrecha (300 a 3.400 Hz) y para el uso de microteléfonos. Estudios futuros serán necesarios para extenderlo a banda ancha (150 a 7.000 Hz), aplicaciones “manos libres”, aplicaciones multipartitas y multimodales.

VIII - CONCLUSIONES

Se ha realizado un resumen de los conceptos de digitalización de voz y un estudio de las recomendaciones de la medición de la calidad de voz así como del “estado del arte” en la materia. Complementariamente se relevaron trabajos que han evaluado la capacidad actual de Internet para aplicaciones de VoIP.

Si bien existen recomendaciones de organismos internacionales como ITU-T, ETSI y TIA, hay varias críticas, propuestas de mejoras a las mismas, y propuestas de nuevos métodos de evaluación objetiva de la QoE.

Adicionalmente el advenimiento de nuevas aplicaciones multimedia y de banda ancha hace necesarias nuevas recomendaciones, las que todavía están en una fase temprana de desarrollo.

La evaluación del Backbone de Internet es otro aspecto fundamental, que requerirá de mayores estudios y esfuerzos para establecer políticas de QoS.

REFERENCIAS

- [1] Request for Comments 3550: “RTP: A Transport Protocol for Real-Time Applications”, H. Schulzrinne et al (July 2003)
- [2] ITU-T Rec. P.800. *Methods For Subjective Determination of Transmission Quality*. 1996.
- [3] Akira Takahashi and Hideaki Yoshino, “Perceptual QoS Assessment Technologies for VoIP” IEEE Communications Magazine, July 2004, pp 28-34
- [4] “El modelo E, un modelo informático para utilización en planificación de la transmisión”, Recomendación UIT-T G.107, 03/05
- [5] TIA, “Telecommunications IP Telephony Equipment Voice Quality Recommendations for IP Telephony” – PN-4689 (TIA/EIA/TSB116) TIA Engineering Committee TR-41.1.
- [6] ETSI TS 102 024-5 v4.1.1. “Quality of Service (QoS) Measurement Methodologies.” 2003.
- [7] 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
- [8] ITU-T Rec. P.563, “Single ended method for objective speech quality assessment in narrow-band telephony applications”, International Telecommunication Union, Geneva, Switzerland, May 2004.
- [9] Athina P. Markopoulou, Fouad A. Tobagi, Mansour J. Karam, “Assessment of VoIP Quality over Internet Backbones”, IEEE INFOCOM 2002
- [10] R. Ramachandran, J. Kurose, D. Towsley, H. Schulzrinne, “Adaptive playout mechanisms for packetized audio applications in wide-area networks”, Proc. of Infocom 1994.
- [11] X. Zhou, F. Muller, R. E. Kooij, and P. Van Mieghem, “Estimation of Voice over IP Quality in the Netherlands”, 4th International Workshop on Internet Performance, Simulation, Monitoring and Measurement (IPS-MoMe 2006), February 27-28, Salzburg, Austria
- [12] Leandro Carvalho, Edjair Mota, Regeane Aguiar, Ana F. Lima, José Neuman de Souza, Anderson Barreto, “An E-Model Implementation for Speech Quality Evaluation in VoIP Systems”, Computers and Communications, 2005. ISCC 2005. Proceedings. 10th IEEE Symposium on
- [13] W Jiang, H Schulzrinne, “Assessment of VoIP Service Availability in the Current Internet” PAM 2003 Workshop
- [14] Simon Broom, Mike Hollier, “Speech Quality Measurement Tools for Dynamic Network Management”, MESAQIN 2003
- [15] Adrian E. Conway, “A Passive Method for Monitoring Voice-over-IP Call Quality with ITU-T Objective Speech Quality Measurement Methods”, ICC 2002 - IEEE International Conference on Communications, no. 1, April 2002, pp. 2583 - 2586
- [16] Wei Zha, Wai-Yip Chan, “Objective Speech Quality Measurement Using Statistical Data Mining”, EURASIP Journal on Applied Signal Processing
- [17] S Mohamed, G Rubino, M Varela, “A Method for Quantitative Evaluation of Audio Quality over Packet Networks and its Comparison with Existing Techniques”, Proceedings of MESAQIN’04, June 2004
- [18] A. Takahashi. “Opinion Model for Estimating Conversational Quality of VoIP.” Proc. IEEE ICASSP ’04. vol. III, May 2004, pp. 1072-75.
- [19] L Sun, E Ifeachor, “New Models for Perceived Voice Quality Prediction and their Applications in Playout Buffer Optimization for VoIP Networks”, IEEE International Conference on Communications, 2004
- [20] Hoene C., Karl H., Wolisz A., “A Perceptual Quality Model for Adaptive VoIP Applications”, Proceedings of International SPECTS’04, Julio 2004