Comportamiento de los Codecs de Voz frente a la Pérdida de Paquetes

Carlos Baladrón[*], Javier M. Aguiar[*], Belén Carro[*], Borja de la Cuesta[*], Julio Tejedor[+]
[*] Universidad de Valladolid, Campus Miguel Delibes 47011, Valladolid (España), E-mail: cbalzor@ribera.tel.uva.es, javagu@tel.uva.es, belcar@tel.uva.es, bcuedie@ribera.tel.uva.es, Fax: +34983423667.
[+] Accenture S.L., Plaza Pablo Ruiz Picasso s/n 28020 Madrid (España), E-mail: j.x.tejedor.pallares@accenture.com, Fax: +34915966695.

Resumen

El rendimiento de cada codec de audio en términos de la calidad de voz percibida por el receptor de la transmisión varía en función de la tasa de pérdida de paquetes a la que se enfrenta. El resultado es que para una tasa de pérdida de paquetes determinada, habrá ciertos codecs que se comporten mejor que otros. El objetivo de este trabajo es precisamente determinar, para diferentes tasas de pérdida de paquetes, qué codec de los considerados es el que ofrece un rendimiento óptimo en el sentido de la calidad de audio, y estudiar si el nuevo codec de código abierto Speex puede ofrecer un rendimiento comparable al de otros estándares. Para ello se han llevado a cabo emulaciones en entornos de red controlados modificando la tasa de pérdida de paquetes extremo a extremo.

1. Introducción

Una de las tendencias actuales más importantes en cuanto a comunicaciones de voz en tiempo real consiste en integrar éstas en las redes de datos de conmutación de paquetes basadas en el protocolo IP. Dicha forma de transmitir voz presenta, además de una gran multitud de ventajas respecto a su integración con toda clase de dispositivos multimedia, algunas dificultades que no aparecían en la clásica conmutación de circuitos.

De entre esas dificultades, la que tiene uno de los mayores impactos negativos en la calidad de la comunicación [1] es el fenómeno conocido como "pérdida de paquetes", que hace referencia a paquetes de red que no llegan a su destino porque son descartados, bien por congestión en la red, errores de transmisión, encaminamiento defectuoso o llegadas demasiado tardías. Hay que recordar que las redes IP no ofrecen (por sí mismas) ningún tipo de garantía de calidad de servicio (QoS, *Quality of Service*) más allá del servicio básico "*best effort*", por lo que los paquetes descartados jamás llegarán a su destino.

La estrategia típica que utilizan las aplicaciones para enfrentarse a esta problemática es solicitar a la fuente de los paquetes la retransmisión de los mismos por medio de protocolos situados en capas inferiores (por ejemplo TCP). Sin embargo en las comunicaciones de voz en tiempo real no habrá tiempo para que los paquetes perdidos sean retransmitidos, puesto que, al introducir la red cierto retardo, llegarían al receptor demasiado tarde como para ser reproducidos a tiempo. Estos sistemas en tiempo real se enfrentan por tanto a los paquetes perdidos empleando otras estrategias, como rellenar los huecos en el discurso introduciendo fragmentos de voz sintéticos o repitiendo fragmentos anteriores.

La señal de voz analógica se transforma en una señal digital de un formato específico utilizando un codec. Los diferentes codecs implementan distintas filosofías de codificación, provocando un comportamiento diferente en función de la pérdida de paquetes en el sistema de transmisión. El resultado es que para una tasa de pérdida de paquetes determinada, habrá ciertos codecs que se comporten mejor que otros (en términos de calidad de audio), pero para otra tasa de pérdidas diferente es posible que los codecs que ofrecen la mejor calidad sean otros distintos. Este estudio trata de determinar y comparar el comportamiento para los codecs Speex, ITU-T G.711 (ley μ) y LPC-10, a través de una serie de pruebas experimentales utilizando un emulador de red. Será especialmente interesante observar en qué condiciones es ventajoso el uso de Speex, ya que su condición de codec gratuito puede hacerlo conveniente en ciertas aplicaciones para abaratar costes.

Este artículo consta de las siguientes secciones. Primero, esta introducción. A continuación, en la Sección 2, se presenta el banco de pruebas utilizado. El apartado 2.1 es un estudio de cómo se lleva a cabo el modelado de la pérdida de paquetes en el emulador para que su comportamiento se asemeje lo más posible a la realidad. En el apartado 2.2 se analizan los métodos de medida de la calidad del audio empleados para determinar la eficiencia de los codecs, y se introduce el algoritmo PESQ como método perceptual de medida objetiva de calidad. En la Sección 3 se presentan los resultados experimentales obtenidos. Y finalmente, en la Sección 4, se exponen las principales conclusiones extraídas del trabajo presentado.

2. Banco de Pruebas

A la hora de implementar el emulador para llevar a cabo las pruebas experimentales, es fundamental que el mismo refleje fielmente la realidad.

Para diseñar el banco de pruebas adecuado en este caso es necesario estudiar dos puntos clave. Primero, el modelo estadístico que seguirán las pérdidas de paquetes. En [2], entre otros, se pone de manifiesto que la pérdida de paquetes en las redes IP presenta unas características concretas, como por ejemplo, el hecho de que las pérdidas ocurren generalmente a ráfagas. El modelo que emplee el banco de pruebas tendrá que ajustarse lo mejor posible a dichas características.

El segundo punto clave será la forma de medir la calidad de audio de una determinada comunicación, lo cual no representa en absoluto un problema trivial, debido a la naturaleza subjetiva de dicha medida: cada sujeto puede percibir la misma comunicación de una forma diferente.

2.1. Modelado de Pérdida de Paquetes

A lo largo de este estudio se han considerado dos modelos diferentes. El modelado independiente o de Bernoulli y el de Gilbert.

El modelo de Bernoulli presenta como principal ventaja su sencillez, si bien no permite incorporar la característica de pérdidas a ráfagas, lo cual constituye su principal problema, puesto que afecta de forma determinante a la calidad percibida [1].

En la Fig. 1 puede verse el diagrama de estados de este modelo. p representa la probabilidad de que el paquete que se está considerando se pierda, mientras que q es la probabilidad de que se reciba correctamente. La suma $p+q$ siempre es igual a 1.

Al utilizar este modelado, cada pérdida se considera un suceso independiente: la probabilidad p de pérdida de un paquete no se ve afectada por lo que haya ocurrido con los paquetes anteriores.

Para poder modelar ráfagas de pérdidas es necesario utilizar un modelo más complicado. Un modelado de cadena de Markov de orden k [3] determina la probabilidad de pérdida de un paquete en función del estado de pérdida de los k paquetes anteriores. De esta forma es sencillo diseñar un modelo de tráfico que incluya ráfagas de pérdidas, e incluso determinar la probabilidad de que dichas ráfagas sean de una longitud u otra.

Un caso concreto de modelo de Markov es aquél en que k=1, y es conocido como modelo de 2 estados de Gilbert, cuyo diagrama de estados puede verse en la Fig. 2. El estado 0 corresponde a una correcta recepción del paquete, y el estado 1, por el contrario,

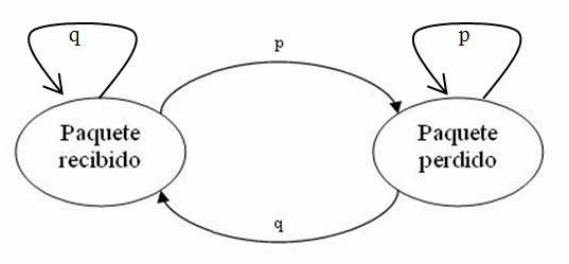


Fig. 1: Diagrama de estados del modelo de Bernoulli, donde p+q=1.

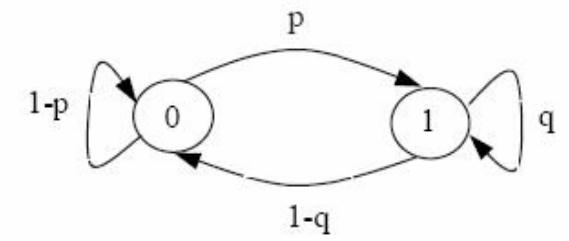


Fig. 2: Diagrama de estados del modelo de Gilbert

a una pérdida de dicho paquete. La probabilidad de que un paquete se pierda cuando el anterior se ha recibido correctamente (pasar del estado 0 al estado 1) es p. La probabilidad de que un paquete se pierda cuando el paquete anterior se ha perdido también (mantenerse en el estado 1), es q. A partir de aquí es sencillo calcular la probabilidad de que un paquete se reciba cuando el anterior se ha recibido correctamente (mantenerse en el estado 0) y de que un paquete se reciba cuando el anterior se ha perdido (pasar del estado 1 al 0), que son, respectivamente, 1-p y 1-q.

Debido a que consigue, con relativa simplicidad, modelar de una forma bastante fiable el comportamiento de una red real, este modelo es el más utilizado a la hora de estudiar comunicaciones de voz sobre redes IP, como se puede ver por ejemplo en [3] y [4].

A lo largo de este estudio se emplean ambos modelos para comparar sus resultados.

2.2. Métodos de Medida de Calidad en Comunicaciones de Voz

La satisfacción del usuario determina directamente la calidad de una comunicación. Sin embargo, al tratarse de una percepción subjetiva, es extremadamente difícil de medir. Existen para ello varias opciones, tal y como puede observarse en la clasificación que aparece en la Fig. 3.

Primeramente se puede considerar la realización de tests subjetivos con varios sujetos de prueba que otorguen una nota de calidad (MOS - *Mean Opinion Score*) a la comunicación de voz que se trata de medir. La recomendación ITU-T P.800 señala cómo llevar a cabo estas medidas de forma adecuada. Sin embargo este método es extremadamente costoso en

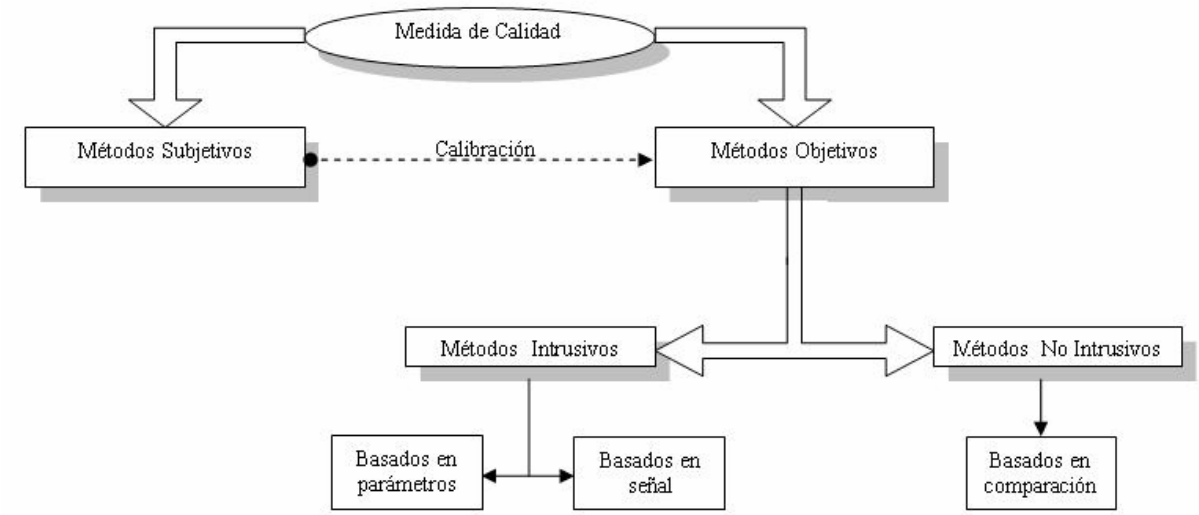


Fig. 3: Métodos de medida de calidad en comunicaciones de voz.

tiempo y recursos, lo que no lo hace adecuado para una experimentación exhaustiva.

La otra opción es realizar medidas objetivas de parámetros que conduzcan a una estimación de la calidad de voz. Tal y como se puede ver en la Fig. 3, existen dos tipos: los intrusivos se basan en una comparación entre la señal original y la señal degradada (una vez ha sido recibida). Las medidas resultantes son precisas, pero como resulta obvio, no es posible llevar a cabo las mediciones en tiempo real. Por este motivo, serán métodos muy adecuados para llevar a cabo medidas experimentales en entornos controlados, pero no tanto para monitorizar comunicaciones reales.

Los no intrusivos están basados en parámetros o en las señales degradadas recibidas. A partir de aquéllos, el método de medida llevará a cabo una serie de cálculos y devolverá una estimación de la calidad. Al no requerir de la señal original como referencia, estos métodos son los adecuados para monitorizar redes reales.

Como el objetivo de este trabajo es determinar con precisión los diferentes comportamientos de los codecs de audio en función de las condiciones de pérdida de paquetes a las que se enfrentan, es necesario contar con el método de medida más preciso que sea posible. El número de experimentos a realizar será grande, por lo que resulta inviable llevar a cabo un estudio subjetivo. Como además dichos experimentos se van a llevar a cabo en un entorno controlado de laboratorio y no hay necesidad de obtener las medidas en tiempo real, el método de medida empleado será intrusivo.

De entre los muchos disponibles (como PSQM [5], PSQM+ [6], PAMS [7], MNB [8, 9], etc.), el escogido será el algoritmo PESQ (*Perceptual Evaluation of Speech Quality*) [10, 11], ya que aúna muchas de las ventajas de los anteriores y es especialmente adecuado para su utilización en la medición de la calidad de voz en entornos de redes IP. Además se trata del más moderno estándar ITU para la medición de la calidad en conversaciones de voz.

El algoritmo PESQ toma como entradas dos señales, una original y otra degradada, y devuelve una predicción de la puntuación que obtendría la señal degradada de ser sometida a un test subjetivo de evaluación de calidad. Las puntuaciones PESQ pueden ser traducidas de una manera muy sencilla a puntuaciones MOS. Para obtener dicha predicción, PESQ se basa en un modelo perceptual del oído humano que trata de simular el comportamiento del sistema auditivo frente a la señal que se está evaluando.

Las medidas que devuelve PESQ se ajustan adecuadamente a los resultados de los tests subjetivos, según [11], cuando:

- Se utilizan codecs de forma de onda.
- No se utilizan codecs de forma de onda, pero con tasa binaria de al menos 4Kbps.
- Se producen errores en el canal.
- Se produce pérdida de paquetes.

El hecho de que el algoritmo PESQ devuelva resultados adecuados en presencia de pérdida de paquetes es determinante para considerar éste el método de medida adecuado para los experimentos que se van a realizar. Sin embargo, PESQ presenta algunas limitaciones, como por ejemplo que no es capaz de determinar adecuadamente el efecto de:

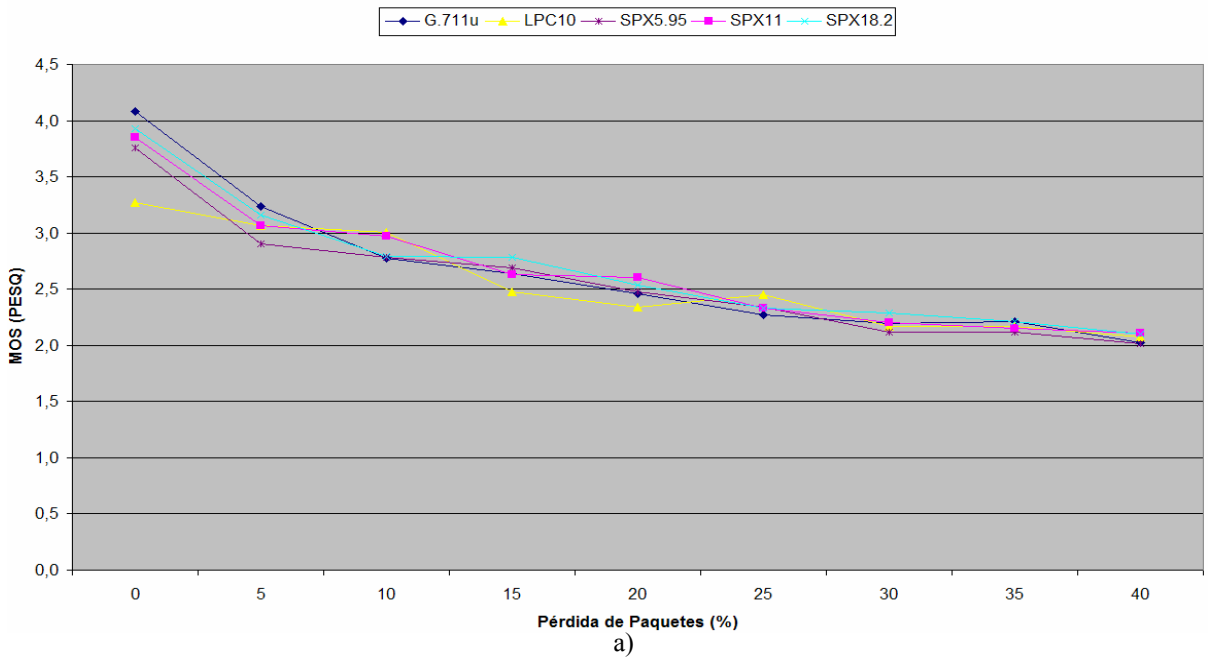
- Retardo.

- Características del hablante (idioma, género, etc.).
- Niveles de ganancia del sistema.
- Varios hablantes.
- Distintas tasas binarias entre codificador y decodificador.

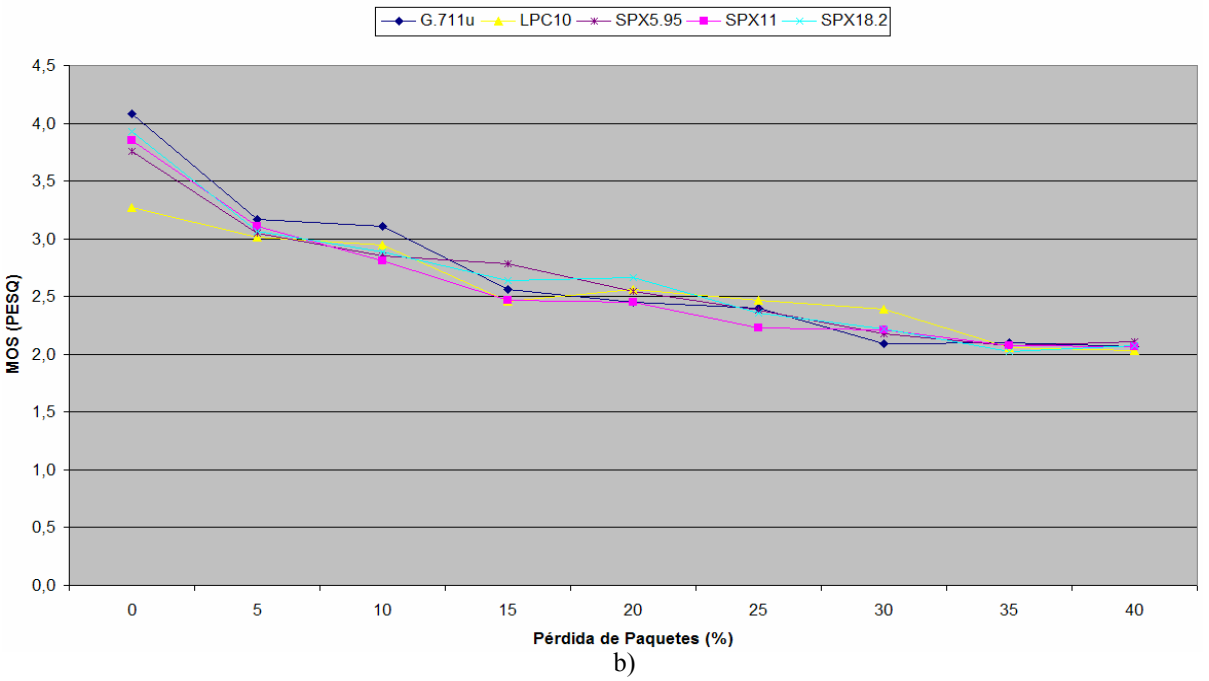
El banco de pruebas constará de dos clientes de comunicación (un emisor y un receptor) y un emulador de red a través del cual se transmite la comunicación y que introduce pérdidas siguiendo los modelos discutidos en la Sección 2.1. La señal original del cliente emisor y la degradada a través de la red, obtenida en el cliente receptor, se introducen en un evaluador de calidad basado en el algoritmo PESQ, que devuelve finalmente la nota MOS.

3. Resultados Experimentales

Los resultados de la experimentación aparecen en las Figs. 4 y 5. La Fig. 4 presenta la calidad media (en unidades MOS) obtenida para un codec determinado frente a una tasa de pérdidas dada, en un intervalo de pérdidas de 0%-40% (que se correspondería con un entorno de comunicación poco fiable como una red *wireless*). La Fig. 4 a) presenta los resultados obtenidos utilizando pérdidas que siguen el modelo de Bernouilli, mientras que b) presenta los resultados obtenidos con pérdidas de Gilbert. La Fig. 5 presenta con detalle el intervalo de pérdidas de 0% a 5% (valores habituales en entornos fiables como redes

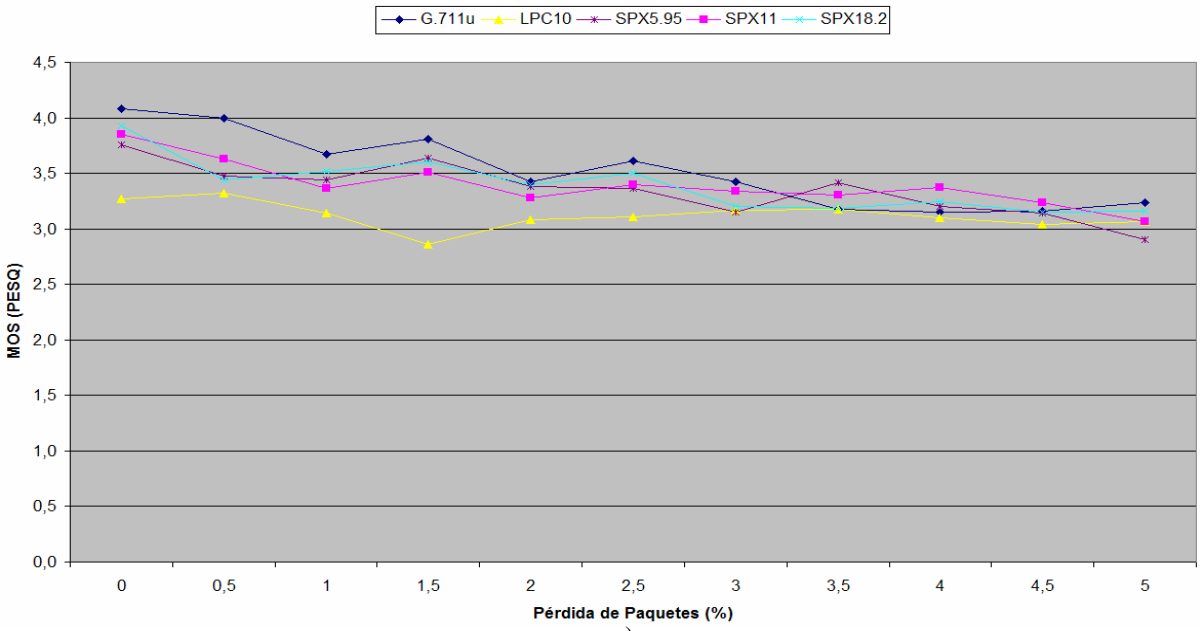


a)

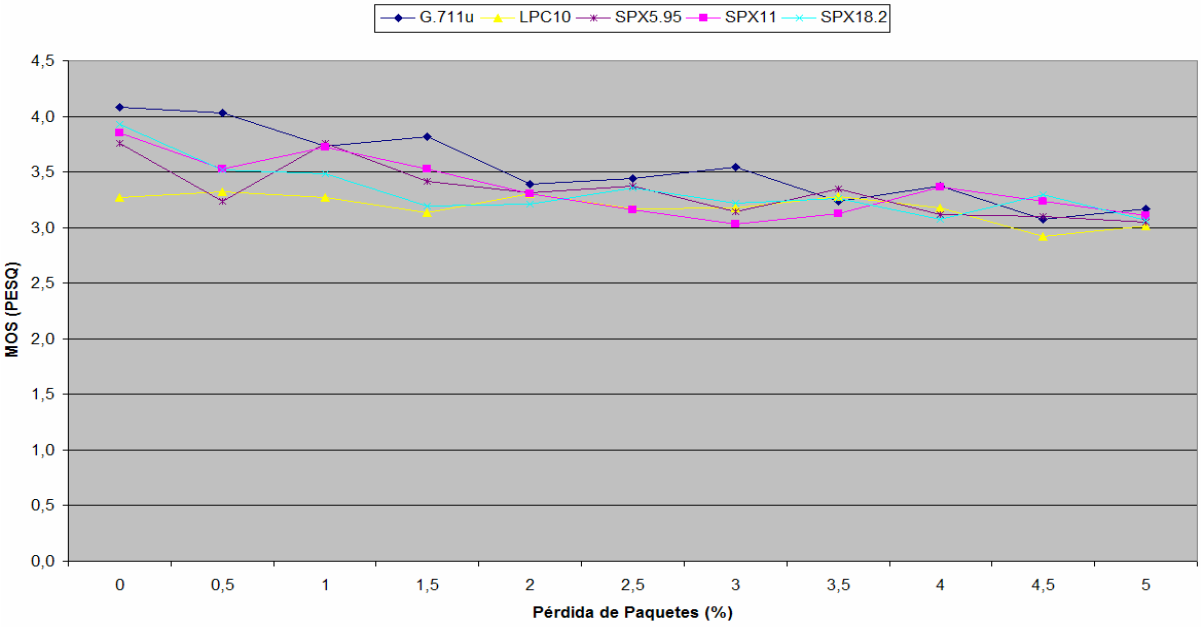


b)

Fig. 4: Resultados experimentales para pérdidas de 0% a 40%. a) con modelado de pérdidas de Bernoulli. b) con modelado de pérdidas de Gilbert.



a)



b)

Fig. 5: Resultados experimentales para pérdidas de 0% a 5%. a) con modelado de pérdidas de Bernoulli. b) con modelado de pérdidas de Gilbert.

locales con hilos).

Los codecs seleccionados para la experimentación presentan características muy diferentes entre sí. G.711 es un codec que ofrece una calidad de audio muy elevada a costa de consumir un gran ancho de banda (64 Kbps). LPC-10 en cambio ofrece una calidad muy baja, pero emplea una tasa de bits muy pequeña, 2.7 Kbps. El tercero de los codecs utilizados es Speex [12], un codec gratuito de código abierto que puede emplear diferentes tasas de bits. La comparación de Speex frente a otros codecs propietarios puede resultar muy útil para determinar si se trata de una opción viable para abaratar costes en aplicaciones que requieran de comunicación de voz. En este estudio se emplea Speex con tasas de bits de 5.98, 11 y 18.2 Kbps.

Los ficheros de audio utilizados durante las pruebas siguen las recomendaciones del estándar ITU-T P.830, que señala que es conveniente utilizar frases sencillas y breves, y que además no existan conexiones entre ellas

4. Conclusiones

A la luz de los datos obtenidos durante la experimentación presentados en las gráficas de las Figs. 4 y 5, se pueden extraer una serie de conclusiones.

LPC-10 ofrece a bajas tasas de pérdidas una calidad peor que los otros codecs, lo cual es lógico debido a que emplea una tasa de bits mucho menor. Su uso debería restringirse a los casos en los que el ancho de banda fuera muy limitante. A tasas de pérdidas altas, cuando la información perdida es tanta que las técnicas de cancelación ya no pueden hacer frente a las mismas, LPC-10 acaba ofreciendo puntuaciones MOS comparables al resto de los codecs. En estos casos, puede resultar ventajoso el uso de LPC-10 debido al bajo consumo de ancho de banda.

En ausencia de pérdidas, G.711 es el codec que ofrece mejor calidad, como es lógico, debido a la alta tasa de bits que emplea. Esta tendencia se mantiene en general cuando las pérdidas son bajas. Sin embargo, al alcanzar éstas valores más elevados, se puede observar que los codecs Speex ofrecen un mejor resultado, empleando además menor ancho de banda.

La familia de codecs Speex presenta buenos resultados de calidad, cercanos a los de G.711, pero con tasas de bits muy inferiores, lo que lo convierte en una buena opción para entornos en los que se cuente con ancho de banda limitado. Además, Speex cuenta con la ventaja adicional de que es un codec gratuito.

Se puede ver que en ausencia de pérdida de paquetes, al aumentar la tasa de bits, los codecs Speex también aumentan su calidad. Sin embargo, en presencia de pérdidas esto no siempre es así, puesto que para algunas tasas, el Speex de 5.98 Kbps ofrece mejor calidad que el Speex de 18.2 Kbps.

Por ello será muy conveniente que al elegir la tasa a la que se va a transmitir utilizando Speex, se estudien a fondo las condiciones de red sobre las que se va a trabajar, para no acabar desperdiciando ancho de banda y obtener, además, menor calidad.

Según aumentan las pérdidas, la calidad ofrecida por los codecs Speex se va acercando cada vez más a la de G.711. En la Fig. 5 se puede observar que a partir de 3.5% de pérdidas los codecs Speex superan a G.711. Esto ilustra que, en ocasiones, será más ventajoso utilizar Speex, que ofrecerá mejor calidad empleando un menor ancho de banda.

Finalmente es importante remarcar dos conclusiones: Primero, que a pesar de que se observan tendencias generales, las líneas de calidad están constantemente cruzándose entre sí en las Figs. 4 y 5, lo que obligará a realizar un análisis preciso de las condiciones de pérdidas de la red a la hora de determinar el codec que ofrezca un comportamiento óptimo. Y segundo, que el codec Speex, siendo gratuito y de código abierto, ofrece un rendimiento bastante aceptable frente a otros estándares, presentando en ocasiones mejores puntuaciones de calidad que G.711 a pesar de emplear tasas de bits sensiblemente menores. Esto lo convierte en un serio competidor cuya utilización merece la pena considerar en la mayoría de los casos.

Agradecimientos

Los autores desean mostrar su agradecimiento al Ministerio de Educación y Ciencia, como entidad financiadora, cofinanciado con FEDER, del proyecto de investigación CICYT TIC2003-07074 “Servicios Multimedia en Tiempo Real sobre Redes IP Heterogéneas” dentro del Plan Nacional de Tecnologías de la Información y las Comunicaciones.

Referencias

[1] W. Jiang, H. Schulzrinne, “Modelling of Packet Loss and Delay and their Effect on Real-Time Multimedia Service Quality”, *Proc. NOSSDAV*, 2000.

[2] V. S. Frost, B. Melamed, "Traffic modeling for telecommunications networks", *Communications Magazine, IEEE*, vol. 32, no. 3, Mar 1994, pp.70-81.

[3] W. Jiang, H. Schulzrinne, “QoS measurement of internet real-time multimedia services,” *Proc. NOSSDAV*, Junio 2000.

[4] M. Yajnik, S. B. Moon, J. F. Kurose, and D. F. Towsley, “Measurement and modeling of the temporal dependence in packet loss”, *INFOCOM* (1), 1999, pp. 345–352.

[5] J. G. Beerends, J. A. Stemerdink. “A Perceptual Speech Quality Measure Based on a Psychoacoustic Sound Representation”. *J. Audio Eng. Soc.*, 42(3), Marzo 1994, pp. 115–123.

[6] ITU-T Contribution: “Improvement of the P.861 Perceptual Speech Quality Measure”, Diciembre 1997.

[7] John Anderson, Agilent Technologies. “Methods for Measuring Perceptual Speech Quality White Paper”. Disponible en: http://cp.literature.agilent.com/litweb/pdf/5988-2352EN.pdf.

[8] S. Voran. “Objective Estimation of Perceived Speech Quality Part I: Development of the Measuring Normalizing Block Technique”, *IEEE Trans. on Speech and Audio Processing*, Vol. 7, No. 4, Julio 1999, pp. 371–382.

[9] S. Voran. “Objective Estimation of Perceived Speech Quality Part 2: Evaluation of the Measuring Normalizing Block Technique”. *IEEE Trans. On Speech and Audio Processing*, Julio 1999, Vol. 7, No. 4, Julio 1999, pp. 373–390.

[10] Recommendation ITU-T P.862: “Perceptual Evaluation of Speech Quality (PESQ), An Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs”, Febrero 2001.

[11] A. W. Rix, J. G. Beerends, M. P. Hollier, A. P. Hekstra. “Perceptual Evaluation of Speech Quality (PESQ). A New Method for Speech Quality Assessment of Telephone Networks and Codecs”,

Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on, Vol. 2, Mayo 2001, pp. 749-752.
[12] "Speex: a free codec for free speech", Disponible en http://www.speex.org.