

ESTUDIO COMPARATIVO DE LAS RECOMENDACIONES ITU-T G.107, P.862 Y
P.563 PARA EVALUAR LA CALIDAD DE LA VOZ EN REDES IP.

YONY FERNANDO DIAZ

Anteproyecto de grado presentado
como requisito parcial para optar
al título de Ingeniero Electrónico

UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA
PROGRAMA ACADÉMICO DE INGENIERÍA ELECTRÓNICA
SANTIAGO DE CALI
2006

ESTUDIO COMPARATIVO DE LAS RECOMENDACIONES ITU-T G.107, P.862 Y
P.563 PARA EVALUAR LA CALIDAD DE LA VOZ EN REDES IP.

YONY FERNANDO DIAZ
COD. 0131381

Aprobación del director del proyecto

Ing. Fabio Germán Guerrero M.Sc.

Aprobación del Jefe de área

Ing. Leandro Antonio Villa M.Sc.

UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA
PROGRAMA ACADÉMICO DE INGENIERÍA ELECTRÓNICA
SANTIAGO DE CALI
2006

TABLA DE CONTENIDO

1. INTRODUCCIÓN.....	1
2. ANTECEDENTES.....	1
3. FORMULACIÓN DEL PROBLEMA.....	2
4. JUSTIFICACIÓN.....	2
5. OBJETIVOS.....	3
5.1. GENERAL.....	3
5.2. ESPECÍFICOS.....	3
6. MARCO TEÓRICO.....	3
7. ASPECTOS METODOLÓGICOS.....	12
8. RESULTADOS ESPERADOS.....	13
9. PRESUPUESTO.....	13
9.1. RECURSOS FÍSICOS.....	13
9.2. RECURSOS HUMANOS.....	14
10. CRONOGRAMA.....	14
11. GLOSARIO.....	14
12. REFERENCIAS.....	15
13. BIBLIOGRAFÍA.....	16

1. INTRODUCCIÓN

La Voz sobre IP (Internet protocol) enfrenta problemáticas propias de las redes de datos, que se manifiestan como degradaciones en la calidad del servicio percibida por los usuarios (QoS). Estas degradaciones pueden deberse por ejemplo a retardos en la llegada de los paquetes, jitter (diferencia de retardos) y pérdida de paquetes, entre otros factores. Para que la tecnología de Voz sobre IP pueda ser utilizada comercialmente, 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 diversos 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 (Mean Opinión Store) llamado MOS. Los 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 se establece en tiempo real la calidad que percibiría un usuario). En éste trabajo de grado se busca establecer por medio de una comparación detallada, basada en la implementación de los algoritmos de medida ITU-T G.107 [7], P.563 [8], P.862 [9] y una evaluación experimental, cual de éstos, es el más apropiado para determinar la calidad de la voz en redes LAN corporativas (IEEE 802.11g e IEEE 802.3).

2. ANTECEDENTES.

Desde el año 2001, la recomendación ITU-T P.862 [9], también conocida como algoritmo PESQ ha representado el estado del arte de las técnicas y estándar internacionales para análisis de calidad de voz perceptual. El PESQ es una prueba para calidad de voz intrusiva, aplicable para estimar la calidad de voz extremo a extremo en redes de VoIP. El nuevo estándar ITU-T P.563 [8] conocido como P.SEAM publicado en mayo de 2004, es una prueba avanzada no intrusiva para medir la calidad de la voz en VoIP; a si mismo la ITU-T ha creado un modelo en la recomendación ITU-T G.107 [7] (marzo del 2004), llamado E-Model, 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; estos tres métodos son los principalmente usados para medir ó estimar calidad de voz en IP. Finalmente ya conocidos los antecedentes de [7], [8], [9], se buscaron trabajos previos a este proyecto de grado en Colombia, pero no se encontró ningún trabajo similar publicado, por tal razón para éste trabajo, se tomaran como guías trabajos internacionales como los descritos en las referencias [1], [2], [3], [4], [5], [6] para alcanzar los objetivos esperados.

3. FORMULACION DEL PROBLEMA

Dado que el mercado de las telecomunicaciones depende directamente de la calidad del servicio prestado, es de suma importancia para los proveedores, administradores y diseñadores de redes VoIP optimizar el desempeño de sus redes, y así garantizar la satisfacción de sus usuarios. Para esto es necesario poder estimar y monitorear la calidad de la voz en redes VoIP. Pero al existir más de una técnica (recomendaciones ITU-T G.107 [7], P.563 [8] y P.862 [9]), para medir la calidad de la voz, surge la necesidad de determinar cuál de estas, es la más apropiada para la evaluación de QoS (Quality of service) en redes VoIP del tipo LAN corporativas (IEEE 802.11g e IEEE 802.3), que son las redes más populares a nivel comercial.

4. JUSTIFICACION

Hay un aumento en la demanda de comunicación (datos y voz) en el mundo, y la infraestructura de red de conmutación de circuitos existente (PSTN) no esta preparada para llevar tal tráfico. Esto ha generado la creciente tendencia de las compañías de telecomunicaciones a integrar las redes de datos y voz en una misma infraestructura, el concepto de convergencia de las redes en una misma infraestructura se agrupa bajo el concepto de redes de próxima generación (NGN), y está motivada por la disminución en el costo por canal en una llamada IP, además de permitir la integración de tecnologías tales como multimedia y la capacidad de soportar un gran volumen de datos, generando nuevos servicios y beneficios para los usuarios. El servicio de voz sobre este tipo de redes se llama VoIP, éste nuevo servicio hace necesario el contar con herramientas de planificación, diseño y de medida, para poder obtener el mejor desempeño posible en éstas redes. Entonces para poder monitorear, medir ó estimar la calidad de voz, en VoIP, es necesario contar con métodos apropiados para tal fin, debido a esto La ITU (Unión Internacional de Telecomunicaciones) en sus recomendaciones ITU-T G.107 [7], P.563 [8] y P.862 [9] propone tres métodos diferentes para medir ó estimar tal calidad, debido ha esto se plantea el siguiente trabajo de grado con el fin de determinar a partir de un estudio comparativo detallado, cual de éstas recomendaciones, es la más adecuada para evaluar la calidad de la voz, en redes VoIP de tipo LAN corporativas (IEEE802.11g e IEEE 802.3).

5. OBJETIVOS

GENERAL

Realizar un estudio de la calidad de voz sobre IP, en redes LAN corporativas usando los algoritmos de evaluación objetiva propuestos en las recomendaciones ITU-T G.107, P.862 y P.563.

ESPECIFICOS

- Estudiar los algoritmos de evaluación objetiva de calidad de voz especificados en las recomendaciones ITU-T G.107, P.862 y P.563.
- Implementar los algoritmos de evaluación objetiva de calidad de voz especificados en las recomendaciones ITU-T G.107, P.862 y P.563.
- Hacer pruebas de evaluación objetiva de la calidad de la voz sobre una red LAN corporativa.
- Hacer un estudio comparativo de los resultados obtenidos usando las recomendaciones ITU-T G.107, P.862 y P.563 sobre una red LAN corporativa.

6. MARCO TEORICO

6.1. Características y expectativas del servicio VoIP

En general, el servicio VoIP puede dividirse en tres componentes de flujos de datos:

Los paquetes de portador/voz (normalmente cursados como paquetes RTP), señalización/control (éstos pueden incluir H.323, H.248, SIP, SIP-T, BICC), y operaciones y mantenimiento (OAM) (éstos incluyen, entre otros, SNMP, TFTP, COPS).

Cuando se trata con la QoS para el servicio de voz, el interés principal tiende a ser en el tren de portadores, ya que esto es lo que generalmente afectará a un abonado (y, más concretamente, su impresión de la calidad de la voz). Los demás componentes son igualmente importantes en lo que toca a la QoS general del servicio. Sin una QoS adecuada para la señalización/control, las llamadas podrán no establecerse o tomar mucho tiempo para hacerlo. Todo esto se reflejaría por último en la impresión que el abonado tenga del servicio ofrecido.

A continuación se muestran algunas reglamentaciones de desempeño que cumplen una función importante, en la medida en que expresan las expectativas públicas últimas o los requisitos formales de los usuarios. Algunas de las que tienen relación con la calidad del servicio vocal, incluidos los aspectos de la señalización, se enumeran más abajo. Otros objetivos pueden deducirse o han sido recomendados por varios organismos normalizadores/reguladores.

Demora del tono para marcar: no más del 1,5% de las llamadas (durante la hora cargada) recibirán una demora del tono para marcar de más de 3 segundos.

Atenuación de adaptación para el eco (línea): más de 20 dB.

Pérdida: 3,0 dB en la línea del abonado (nivel de transmisión de 0 dB).

Ruido: menos de 20 dBmC (nivel de enlace) y menos de 23 dBmC (95% de las líneas).

Demora:

Para comunicaciones nacionales – menos de 150 ms en una dirección, para comunicaciones internacionales con conexiones por satélite – menos de 400 ms en una dirección, para cables submarinos – menos de 170 ms en una dirección.

Demora después de marcar: nominalmente, para llamadas locales – menos de 3 s, para llamadas interurbanas – menos de 5 s, para llamadas internacionales – menos de 8 s.

Pérdida de bloqueo/concordancia: red - 2% durante hora cargada media

Disponibilidad del servicio: 99,999%

Viendo los objetivos anteriores, puede verse que no siempre se identifican los atributos de QoS para cada uno de los componentes. No obstante, pueden deducirse o implicarse. Por ejemplo, la demora después de marcar (el tiempo desde el recibo del último dígito marcado hasta que la parte del extremo lejano es notificada) provee un límite de tiempo por el cual los mensajes de control son procesados y propagados a través de una red para establecer una conexión entre partes. De esa forma, hay un límite implicado a la QoS de demora que los mensajes de control podrán encontrar al atravesar la red IP. Nótese que éste no es un valor absoluto totalmente reflejado en la QoS de la red de transmisión IP, porque también incluye los tiempos de procesamiento en los diversos puntos extremos y nodos a lo largo de la ruta.

Existen interpretaciones similares para aquellos objetivos que afectan a las características del tráfico portador. El E-Model (Recomendación G-107 del UIT-T [7]) se usa para caracterizar las interpretaciones de paquetes portadores de voz. En general, las características de voz (lo que uno escucha en el teléfono) son afectadas por diversos factores cuando hay una red de paquetes en el trayecto del habla. La figura 1 ilustra dichos deterioros en el caso de un ejemplo de red típica simplificada.

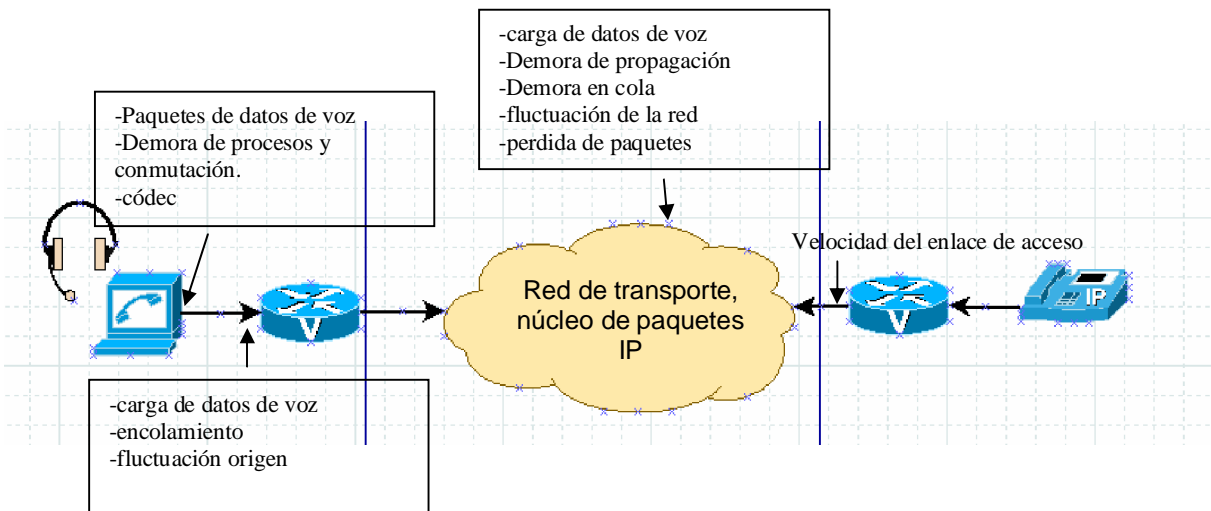


Figura 1: Deterioros de la voz en una red IP

Por la figura anterior, se aprecia que puede ser muy difícil determinar la calidad prevista de la voz de una llamada VoIP mediante la inspección de valores concretos. Además, también pueden influir otros factores fuera del dominio IP. Por ello, el E-Model cumple la función analítica de poder combinar todo lo anterior y producir los resultados esperados de calidad teórica del habla. Cuando se compara con los ejemplos existentes de PSTN, se puede determinar un nivel relativo de calidad. Pero además del modelo E ó E-model, existen otras técnicas que permiten evaluar la calidad de la voz en una comunicación VoIP, estas técnicas son MOS (Mean Opinión Score), estandarizada en la recomendación ITU-T P.800 [10], El algoritmo P.563 definido en ITU-T p.563 [8], ITU-T P.862 (PESQ) [9], estas técnicas serán estudiadas en el punto 6.4 de éste apartado.

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

Factor de compresión

Para poder transmitir la voz a través de una red de datos, es necesario realizar previamente un proceso de digitalización. En telefonía clásica, éste proceso se realiza utilizando CODECs, obteniendo una señal digital de 64 kb/s. Este proceso, se realiza de acuerdo a la recomendación G.711 de la ITU-T. Sin embargo, cuando se dispone de velocidades de red reducidas, es conveniente tratar de minimizar el ancho de banda requerido por las señales de voz. Para ello, se han desarrollado varias recomendaciones, que reducen la velocidad de transmisión requerida, a expensas de degradar la calidad de la voz.

La siguiente tabla resume las recomendaciones de la ITU-T respecto a los algoritmos estandarizados de compresión de voz:

Algoritmo	Descripción
G.711	Audio encoding at 64 k bit/s (μ -law and A-law)
G.722	7 kHz speed at 48, 56 and 64K bit/s (hi-fi voice)
G.723.1	Dual Rate Speed at 6.4 and 5.3 k bit/s
G.728	16 k bit/s speech
G.729 Annex A	8 k bit/s speech (Conjugate structure- algebraic code excited linear prediction or CS-ACELP). Reduce Complexity
G.729 Annex B	8 k bit/s speech (Conjugate structure- algebraic code excited linear prediction or CS-ACELP). Silence Compression
G.729 Annex AB	8 k bit/s speech (Conjugate structure- algebraic code excited linear prediction or CS-ACELP). Reduce Complexity & Silence Compression

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 pérdida es pequeño, la degradación de la voz también lo es. Los porcentajes de pérdida 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.

Demora (retardo ó delay en ingles)

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 compresión:

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
G.711 (64 kb/s)	125 μ s
G.728 (16 kb/s)	2.5 ms
G.729 (8 kb/s)	10 ms
G.723 (5.3 o 6.4 kb/s)	30 ms

Demoras de procesamiento

Es el tiempo involucrado en el procesamiento de la voz para la implementación de los protocolos. Dependen de los procesadores utilizados.

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, gateways, etc.)

Las demoras no afectan directamente la calidad de la voz, sino la calidad de la conversación. 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 muy 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 extremo a extremo comienza a aumentar, el efecto del eco comienza a percibirse.

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, el eco existe en niveles perceptibles (mayores a -25 dB), pero la demora es mínima, por lo que el eco no es perceptible. Las excepciones son las comunicaciones vía satélite, en las que la demora promedio es del orden de los 150 ms. Para estos casos, las compañías telefónicas disponen generalmente de sofisticados equipos canceladores de eco.

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 se introducirá al sistema.

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 (encabezado del paquete, origen, destino, etc.). Esta información no aporta a la información real que se quiere transmitir, pero afecta al tamaño total del paquete, y por tanto al ancho de banda.

6.3 Medida de la calidad de voz en redes VoIP

La VoIP enfrenta problemáticas propias de las redes de datos, que se manifiestan como degradaciones en la calidad del servicio percibida por los usuarios (QoS). 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 las Empresas, 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, el valor MOS (Mean Opinión Store). Los 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).

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 Opinion Score) 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 [10]. 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) pueden estar sujetos 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.

E-Model (ITU-G.107) [7]

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 [10]. 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, 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 codecs, los retardos de la red, así como también los factores típicos en telefonía como son pérdida, ruido y eco. Éste método puede ser aplicado para estimar la calidad de la voz en redes de paquetes, tanto fijas como inalámbricas. El E-Model puede ser utilizado para evaluar como se verá afectada la calidad de la voz en una red en base a parámetros medibles. El modelo parte de un puntaje perfecto (100) y resta diversos factores que degradan la calidad, según se puede ver en la ecuación (6.3.1).

$R = R_o - I_s - I_d - I_{e,eff} + A$ (6.3.1), 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 son el ruido ambiental, pueden hacer que este valor inicial sea menor a 100.

Is. 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.

Id. Representa las degradaciones producidas por el retardo y el eco.

le,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 como 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.

ITU-T P.862 (PESQ) [9]

La recomendación ITU-T P.862 presenta un método objetivo para la evaluación de la calidad de la voz 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. El método objetivo descrito se conoce por evaluación de la calidad vocal por percepción (PESQ, *perceptual evaluation of 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. 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. 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. 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.

ITU-T P.563 [8]

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 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

voz robotizada

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:

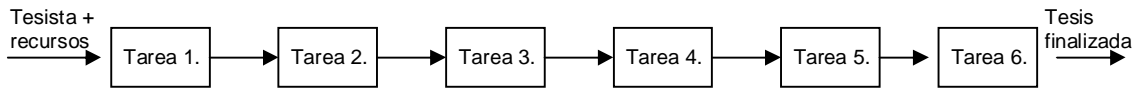
El modelo de calidad vocal de P.564 se compone de tres bloques principales:

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.

7. ASPECTOS METODOLÓGICOS

Las tareas que se realizarán para alcanzar el objetivo de terminar el trabajo de grado, se presentan a continuación, cada tarea será explicada posteriormente:



Tarea 1: para desarrollar éste proyecto el primer paso es realizar un estudio general de VoIP, para conocer sus protocolos, sus estándares, sus requerimientos, su funcionamiento, etc., esto es necesario para contextualizar el desarrollo del trabajo.

Tarea 2: el segundo paso consiste en realizar un estudio general de la QoS en redes VoIP, para identificar los requerimientos, las problemáticas, etc. para garantizar calidad de servicio.

Tarea 3: el tercer paso es uno de los mas importantes y trata de estudiar las recomendaciones de la ITU-T: G.107, P.563, P.862 y la P800 que será usada como referencia para realizar la comparación de las tres primeras recomendaciones.

Tarea 4: el tercer paso consiste en realizar la comparación de los tres métodos dados en las recomendaciones de la ITU-T: G.107, P.563, P.862 para evaluar la calidad de la voz en redes VoIP, en este paso se espera hacer simulaciones, implementaciones software, e implementaciones de redes VoIP del tipo LAN (IEEE 802.11g e IEEE 802.3) que permitan determinar cual de las tres técnicas es la más apropiada para medir la calidad de la voz en éstas redes.

Tarea 5: el quinto paso se espera estandarizar los procedimientos realizados, que permitieron determinar cual de las tres técnicas es la más apropiada para medir la calidad en las redes LAN ya mencionadas, además de la recopilación de toda información, conclusiones y resultados obtenidos en los pasos previos, con el fin de tener toda la información disponible para formar el cuerpo del documento de grado.

Tarea 6: el sexto paso consiste en formar el documento de tesis a partir de toda la los resultado obtenidos en los pasos anteriores.

8. RESULTADOS ESPERADOS

1. Conjunto de pruebas obtenidas con las recomendaciones ITU-T G.107, P.563 y P.862 al evaluar la calidad de la voz en redes VoIP, del tipo LAN corporativas (IEEE 802.11g e IEEE 802.3)
2. Estudio comparativo y evaluación de meritos de cada una de las recomendaciones ITU-T G.107, P.563, P.862, para determinar la calidad de la voz en redes IP (VoIP) de tipo LAN corporativa (IEEE 802.11g e IEEE 802.3)
3. Implementación software en Matlab, de los algoritmos dados en las recomendaciones ITU-T G.107, P.563, P.862, para determinar la calidad de voz en redes IP (VoIP).
4. Publicación de los resultados obtenidos en un artículo de categoría nacional.

9. PRESUPUESTO

Para la realización del presente proyecto se presenta la siguiente lista de elementos necesarios, costos y disponibilidad.

9.1. Recursos Físicos

Recurso	Fuente	Costos
Computador	Univalle/Estudiante	\$ 2.000.000
Fotocopias, impresiones y papelería	Estudiante	\$ 800.000
Visitas, seminarios, congresos, simposios	Estudiante	\$ 3.500.000
Espacio de trabajo con punto de red	Univalle	\$ 1.000.000
Libros	Estudiante	\$ 600.000
Internet	Univalle/Estudiante	\$ 800.000
Llamadas LD, Fax	Estudiante	\$ 50.000
		Total \$ 8.750.000

9.2. Recursos Humanos

Recurso	Fuente	Función	Disponibilidad
Ing. Fabio Guerrero	Univalle	Director	2h. Semana
Yony Fernando Díaz	Tesista	encargado	30h. Semana

10. CRONOGRAMA

La siguiente tabla ubica cada una de las tareas a desarrollar y el tiempo disponible para su ejecución. Los pasos que se mencionan en la tabla corresponden a los pasos expuestos en los aspectos metodológicos.

TAREA	Ene	Feb.	Marzo	Abril	Mayo	Junio	Julio	Agos	Sept	Octub
Paso 1	X	X								
Paso 2			X	X						
Paso 3					X	X				
Paso 4							X	X	X	
Paso 5									X	
Paso 6										X

11. GLOSARIO

Voz sobre IP (VoIP): Conjunto de aplicaciones que permiten la transmisión de voz en vivo a través de Internet utilizando los protocolos TCP/IP.

QoS: Quality of Service (Calidad del Servicio) es una característica de una red de telecomunicaciones que permite garantizar al cliente una calidad pactada por cada servicio contratado.

MOS: Mean Opinión Store es un promedio de opiniones de la calidad del servicio en VoIP, formado por usuarios de este servicio.

LAN: Local Area Network (Red de Área Local) es una red de dispositivos conectados (como son PCs, impresoras, servidores y concentradores) que cubren un área geográfica relativamente pequeña (generalmente no más grande de 5 kilómetros)

IEEE 802.3: red de área local (LAN) cableada.

IEEE 802.11g: red de área local inalámbrica, conocida como Wi-fi.

NGN: red de próxima generación, red que integra tecnologías como datos y voz en una sola infraestructura de red.

PSTN: Public Switched Telephony Network (red telefónica conmutada) es una red de teléfono diseñada primordialmente para la transmisión de voz, aunque pueda también

transportar datos, por ejemplo, en el caso del fax o de la conexión a Internet a través de un módem acústico.

RTP: Real Time Protocol. Protocolo de Tiempo Real. Protocolo utilizado para la transmisión de información en tiempo real como por ejemplo audio y video en una videoconferencia.

H.323: es un Standard aprobado por la International Telecommunication Union (ITU) que define cómo se transmiten los datos en conferencias audiovisuales a lo largo de una red. En teoría, el H.323 debería permitirle participar a los usuarios en la misma conferencia aún si están usando distintas aplicaciones de videoconferencia.

H.248: es un protocolo de señalización que habilita la conmutación de voz, fax y llamadas multimedia entre la PSTN y las redes de próxima generación ó cualquier red IP.

SIP: es un protocolo de señalización para telefonía IP y multimedia.

SNMP: protocolo simple de gestión de redes, es aquel que permite la gestión remota de dispositivos de red, tales como switches, routers y servidores.

TFTP: son las siglas de Trivial File Transfer Protocol (Protocolo de transferencia de archivos trivial).

CODECS: Codec es una abreviatura de Compresor-Decompresor. Describe una especificación implementada en software, hardware o una combinación de ambos, capaz de transformar un archivo con un flujo de datos (stream) o una señal.

12. REFERENCIAS

[1] Calidad de servicio percibida en servicios de voz y video sobre IP, Pedro Casas Hernandez, Diego Guerra Vidal, Ignacio Irigaray Bayarres, Universidad de la Republica, Proyecto de fin de carrera, Ingeniería Eléctrica Plan 97, telecomunicaciones, 30 de Agosto de 2005.

[2] Transmisión de voz sobre IP con protocolo basado en el algoritmo lms, Carlos Alberto Busso Recabarren, tesis para optar al grado de magister en ciencias de la ingeniería mención eléctrica, santiago de chile JULIO 2003.

[3] An E-Model Implementation for Speech Quality Evaluation in VoIP Systems, Leandro Carvalho, Edjair Mota, Regeane Aguiar, Ana F. Lima, José Neuman de Souza, Anderson Barreto, *Federal University of Amazonas (UFAM)*, *Federal University of Ceará (UFC)*, *Nokia Institute of Technology (INdT)*.

[4] Estudio sobre la percepción de QoS en un demostrador UMTS, Laura Bernabé Miguel, treball de fi de carrera, 30 de junio del 2005.

[5] Simulación of voice over IP for the Battlespace Communications System(land), Ian Grivell and Raymee Chau, department of defense of Australia, 2004.

[6] AdaMOS: Algoritmo MOS-Adaptativo para fonts VoIP, Bruno de Azevedo Vianna, Nilmax Teones Moura, Célio Vinicius Neves de Albuquerque, Vinod E. F. Rebello e Cristina Boeres, Instituto de Computação – Universidade Federal Fluminense(IC/UFF)

- [7] 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>
- [8] ITU-T P.563 Single-ended method for objective speech quality assessment in narrow-band telephony applications, mayo 2004.
- [9] 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.
- [10] ITU-T P-801 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>

13. BIBLIOGRAFÍA

José Joskowicz, Redes de Voz, Versión 05, Agosto 2006.

José Joskowicz, Redes de Datos, Versión 03, Agosto 2006.

E-model tutorial

<http://www.itu.int/ITU-T/studygroups/com12/emodelv1/introduction.htm>.

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

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