3. CALIDAD DE SERVICIO EN VOIP

3.1. Introducción

Es bien sabido que las redes basadas en el protocolo IP ofrecen un único nivel de servicio: Best Effort. Es decir, no se lleva a cabo ningún tipo de reserva de recursos, no se conoce el retraso máximo sufrido por los paquetes, ni garantiza que los paquetes lleguen correctamente a su destino. De tal forma que se ofrece el mejor servicio posible para un instante determinado.

Así, se hace imperativo definir y establecer la calidad de servicio (QoS) en las redes de este tipo para el funcionamiento correcto de las aplicaciones multimedia y servicios síncronos, como la voz y vídeo en tiempo real. De modo que la QoS hace referencia a la capacidad de garantizar que se cumplen los requisitos de tráfico para un determinado flujo de información. Esto se traduce para el usuario final en la adquisición de una conexión y un servicio descrito en términos de rapidez, fiabilidad, rendimiento y disponibilidad.

Si atendemos al funcionamiento de VoIP, el hecho de usar una red de conmutación de paquetes, implica que la información vocal es codificada y fragmentada originando un flujo de paquetes que siguen rutas distintas entre el origen y el destino. Esto conlleva a que los paquetes lleguen al destino, generalmente, desordenados y con distinto retardo acumulado, e incluso puede ocurrir que no lleguen al destino debido a las pérdidas. Debido a este fenómeno, es necesario establecer ciertos requisitos en el tráfico de VoIP generado:

- Ancho de banda ocupado.
- Retardo de los paquetes.
- Pérdidas de paquetes.
- Calidad de la comunicación.

3.2. Ancho de banda

El ancho de banda (BW) establece la tasa máxima de transferencia de datos entre los extremos de la comunicación. Debido a que las comunicaciones en tiempo real son muy sensibles al retraso y a la congestión que pueda sufrir la red, se tiene que garantizar un ancho de banda mínimo para este tipo de paquetes.

Centrándonos en la voz sobre IP, los requerimientos de acho de banda para la transmisión de la señal vocal vienen determinados por el códec de voz usado en cuestión. En la siguiente tabla puede observarse la relación existente entre los códecs de voz analizados en este estudio y el BW consumido.

Información del códec				**Cálculo de BW**			
Códec	**Tasa de bit (Kbps)**	**Tamaño trama (bytes)**	**Tamaño trama (ms)**	**Tamaño datos (bytes)**	**Tamaño datos (ms)**	**Paquetes por segundo**	**Ancho de banda (Kbps)**
G711	64	80	10	160	20	50	87,2
G729	8	10	10	20	20	50	31,2
AMR	12,20	30,5	20	30,5	20	50	35,4
	10,20	25,5	20	25,5	20	50	33,4
	7,95	19,875	20	19,875	20	50	31,15
	7,40	18,5	20	18,5	20	50	30,6
	6,70	16,75	20	16,75	20	50	29,9
	5,90	14,75	20	14,75	20	50	29,1
	5,15	12,875	20	12,875	20	50	28,35
	4,75	11,875	20	11,875	20	50	27,95
iLBC	15,2	38	20	38	20	50	38,4
	13,33	50	30	50	30	33,3	28,8

Tabla 6. ***Ancho de banda usado por cada códec.***

Para realizar el cálculo del ancho de banda necesario por cada algoritmo de codificación de voz hay que tener en cuenta que se han añadido las cabeceras pertinentes al empaquetamiento de los datos de voz.

El desglose de dichas cabeceras queda como sigue: 12 octetos para RTP, 8 para UDP, 20 para IP y 18 bytes de cabecera de nivel de enlace, suponiendo que se trata de Ethernet con 4 octetos de código de redundancia cíclica (CRC).

Además, otro parámetro relevante en el cálculo realizado es el número de tramas por paquete (Nfpp). De tal manera que para los dos primeros códec se ha supuesto dos tramas por paquetes, mientras que en el resto se considera una única trama por paquete. No obstante, esto es sólo una ilustración pues este parámetro es objeto de estudio y por tanto posteriormente se presentarán los resultados para un rango comprendido entre una y seis tramas por paquete.

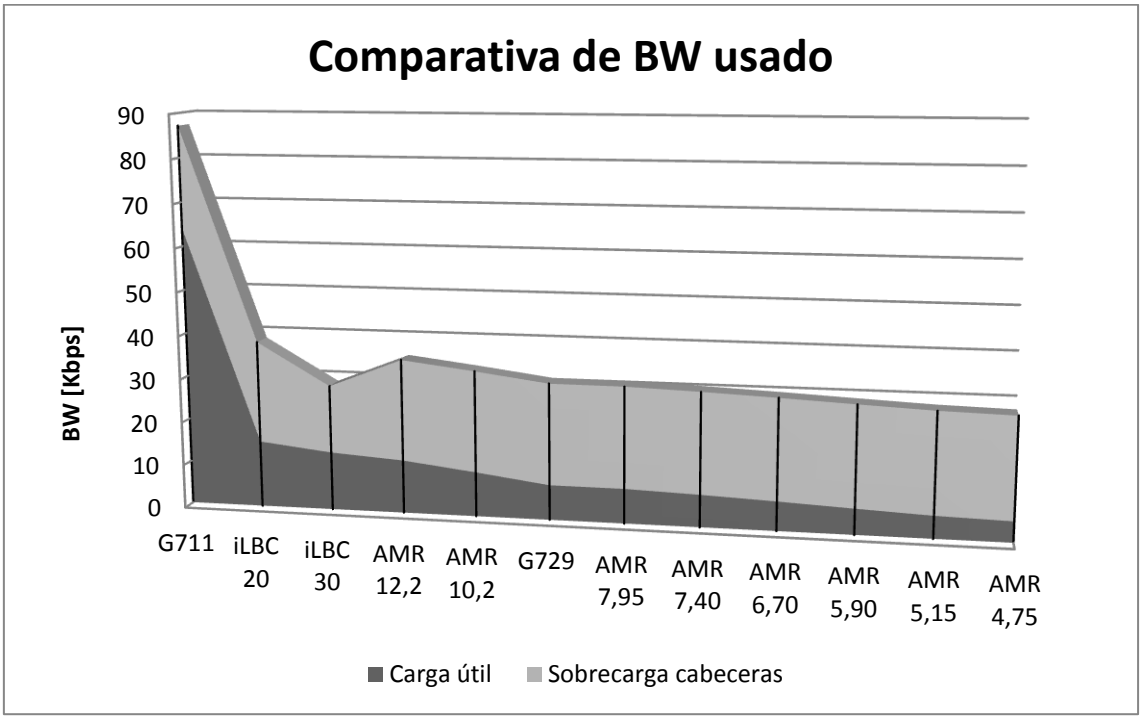

Figura 9. ***Comparativa de ancho de banda para cada códec.***

En la figura anterior puede apreciarse que para el códec iLBC 30 se consigue una menor sobrecarga de cabeceras que para el resto de los códecs. Este hecho se debe a que la duración del conjunto de tramas contenidas en un paquete es de 30 ms a diferencia de los demás que se han considerado de 20 ms. No obstante, se representa una ejemplificación pues,

aunque el tamaño de trama está fijado por el códec usado, la cantidad de tramas que puede transportar un mismo paquete está definida por el parámetro Nfpp, anteriormente comentado.

Por último, comentar que el códec no sólo determina la manera de muestrear y digitalizar la información sino que también se encarga de comprimir la secuencia de datos. Esta compresión de la forma de onda posee bastante relevancia en cuanto al ahorro del ancho de banda. Esto es especialmente importante en los enlaces de baja capacidad y permite realizar un mayor número de conexiones simultáneamente. Otra manera de preservar el ancho de banda consiste en el uso de las tramas SID, supresión de silencios. Proceso por el cual se genera una trama de ruido de confort, en lugar de una ráfaga de tramas activas, que sustituye a los silencios en la conversación.

3.3. Retardo

Como se ha comentado anteriormente, el hecho de que las comunicaciones a través de VoIP sean en tiempo real hace que el tráfico intercambiado sea muy sensible al retardo. Los efectos producidos en la conversación que se derivan del retardo son:

- **Eco**. Es consecuencia de las reflexiones de la señal transmitida. De manera que cuando el retardo supera el umbral de los 25 ms, el transmisor percibe una señal molesta y retrasada de sus propias palabras. Si el retraso alcanzara niveles muy elevados, mantener la conversación sería imposible.
- **Talker overlap**. Este fenómeno ocurre cuando uno de los abonados se superpone a la voz del otro debido a un gran retardo de los paquetes no permitiendo la interactividad entre los locutores. Esto comienza a ser apreciable a partir de los 150 ms aproximadamente.

Por regla general, la voz comienza a degradarse a partir de los 150 ms, aunque, en condiciones extraordinarias, se puede aceptar hasta 400 ms. No obstante, el retardo es un factor crítico en este tipo de comunicaciones.

Para solventarlo es relevante conocer las fuentes que lo provocan. Éstas pueden agruparse en dos grupos:

- **Retardo en la pasarela**.
  - Retardo algorítmico. Es introducido por el códec usado y es inherente al algoritmo de codificación.

  - Retardo de empaquetamiento. Hace referencia al tiempo que se requiere para rellenar un paquete de información, es decir la carga útil del paquete sin cabeceras, con los datos ya codificados y comprimidos. Depende directamente del tamaño de trama definido por cada códec y el número de tramas por paquetes.

Códec	Tamaño muestreo (ms)	Retardo de conversión (ms)
G711	0,125	5
G729	10	15
AMR	20	20

Tabla 7. ***Retardo algorítmico y de empaquetamiento.***

  - Retardo de serialización. Está relacionado con la tasa del reloj de transmisión. Se define como el tiempo requerido para inyectar una unidad de información en la interfaz de red.

  - Retardo de supresión de jitter. Este retardo se debe al almacenamiento temporal del flujo de paquetes en un buffer del extremo receptor. Esto se hace así para eliminar la variabilidad del retardo transformando los retrasos variables a retrasos fijos.

- **Retardo en la red**.
  - Retardo de propagación. Es el tiempo requerido por el paquete para llegar desde su origen al destino. Depende tanto del medio de propagación como del estado de la red.

- Retardo de encolado. Tiempo que esperan los paquetes almacenados en una cola de salida antes de ser transmitidos por la red.

Todos estos tipos de retardos son considerablemente importantes para un correcto funcionamiento de una conversación a través de la red de paquetes. Sin embargo, en el caso de estudio sólo podemos considerar los retardos originados en el extremo transmisor. Aun más, dentro de este grupo es posible considerar ciertos retrasos como despreciables frente al componente fundamental. Así, nos referimos al retardo de encolado que existe en el multiplexor como el predominante respecto al retardo algorítmico, de empaquetado y de serialización.

3.4. Pérdida de paquetes

La pérdida de paquetes es un fenómeno que ocurre con mucha frecuencia en las redes de paquetes como consecuencia de una congestión de red. Los paquetes al atravesarla pueden encontrarse la cola de salida de algún elemento de red llena de tal manera que no se aceptan paquetes nuevos y, consecuentemente, se desechan. Otro factor que puede provocar esta impertinencia es que los paquetes lleguen con demasiado retraso al extremo receptor y tengan que ser descartados.

En aplicaciones que no funcionan en tiempo real es posible solventar en cierta medida el problema de la pérdida de paquetes haciendo uso del protocolo de transporte TCP el cual permite retransmitir los paquetes que no lleguen al destino correctamente. Sin embargo, esto no es procedente para el tipo de tráfico que nos concierne puesto que cualquier retransmisión produce un retardo adicional no tolerable generalmente. Ahí reside la razón por la cual se hace uso del protocolo de transporte UDP.

La consecuencia primordial de los paquetes perdidos es un deterioro de la señal de voz. Esto se debe a que cada trama contiene, de forma aproximada pues depende del tipo de códec, un fonema de voz. En consecuencia, cuando se pierde un paquete se pierde el fonema correspondiente o, en su defecto, varios de ellos pues depende del número de tramas contenidas en cada uno de los paquetes. Cuando la tasa de

pérdidas es pequeña, aún provocando una disminución en la calidad de servicio, no presenta un gran problema pues los decodificadores incluyen mecanismos de interpolación y el cerebro humano es capaz de reconstruir el intervalo de conversación no recibida. Sin embargo, probabilísticamente, la pérdida de un paquete suele traducirse en la pérdida de varios de ellos, lo que sí degrada severamente la calidad de la comunicación.

De nuevo, en el escenario de simulación propuesto, el fenómeno de la pérdida de paquetes será estudiado en el conmutador del extremo transmisor.

3.4. Calidad de la comunicación

La calidad de servicio en VoIP está íntimamente ligada con la percepción que tienen los usuarios finales de la conversación mantenida. Existen medidas subjetivas que implican a personas físicas y métodos de medición que emulan la percepción del oído humano y permiten obtener una puntuación equivalente.

3.4.1. Métodos subjetivos

En los métodos subjetivos se distinguen dos métodos de evaluación de la calidad del audio: por evaluación directa (ACR) o por comparación contra un audio de referencia (DCR). Con evaluaciones directas se califica la calidad en función de una escala del uno al cinco, siendo cinco "excelente" y uno "malo", tal como muestra la tabla 8. El MOS (Mean Opinion Score) es la escala más utilizada basada en someter a conversación a un gran número de usuarios y promediar la nota otorgada por cada uno de ellos. Esta metodología de evaluación se halla estandarizada en la recomendación ITU-T P.800 [27].

Puntuación	Calidad
5	Excelente
4	Buena
3	Aceptable
2	Pobre
1	Mala

Tabla 8. ***Escala MOS para evaluación directa.***

En cambio, si la evaluación es comparativa, el rango de calificación también oscila entre uno y cinco, pero con la salvedad de que en este caso se califica en función del esfuerzo empleado por los usuarios para entender la conversación debido a las diferencias entre el audio de referencia y el medido. Así, se define una escala equivalente conocida como DMOS (Degradation MOS), tabla 9.

En general, los métodos subjetivos son caros, lentos y dependientes de factores como el país, el idioma, la actitud de los participantes, etc.

Puntuación	Calidad
5	Relajación completa: ningún esfuerzo
4	Necesidad de atención: esfuerzo aceptable
3	Esfuerzo moderado
2	Esfuerzo considerado
1	Imposible de entender

Tabla 9. ***Escala DMOS para evaluación comparativa.***

3.4.2. Métodos objetivos

En cuanto a los métodos objetivos predominan tres recomendaciones de la ITU-T: el modelo E, la recomendación P.862 y la P.563. Todos ellos tienen en común que se basan en mediciones de propiedades físicas de una red para estimar la evaluación de los usuarios. A su vez, se diferencian entre métodos intrusivos, basados en medir la degradación a la salida

sufrida por una señal conocida previamente inyectada en la red; y no intrusiva, consistente en monitorizar ciertos parámetros para determinar la calidad en tiempo real.

3.4.2.1. Modelo E

El modelo E [28] es el modelo de opinión más ampliamente difundido. Se basa en una cuantificación escalar de la calidad del sonido que se estima que percibirá un usuario. Este modelo suele usarse como una herramienta informática para la planificación de la transmisión. Una característica fundamental es que incluye factores de degradación de la señal transmitida que reflejen los efectos de los dispositivos de red que intervienen en la generación, transmisión y conmutación de la señal de audio. De manera que dicho modelo proporciona, en base a diversos parámetros medibles en la red, el factor R, el cual es posible traducirlo a una escala MOS tal como se refleja en la siguiente tabla.

R	Satisfacción usuarios	MOS	
100	Muy satisfechos	4,5	Deseable
94,3		4,4	
90	Satisfechos	4,3	
80	Algunos insatisfechos	4.0	Aceptable
70	Muchos insatisfechos	3,6	
60	Casi todos insatisfechos	3,1	No aceptable
50	No recomendado	2,6	
0		1,0	

Tabla 10. ***Traducción factor R a escala MOS.***

El factor R viene dado por la siguiente expresión:

$$R = Ro - Is - Id - Ie,eff + A \tag{3.1}$$

Cada uno de los componentes de la ecuación anterior se calcula a partir de los parámetros de transmisión que se muestran en la figura contenida en la recomendación de la ITU-T [28] y que se presenta en este documento, figura 10.

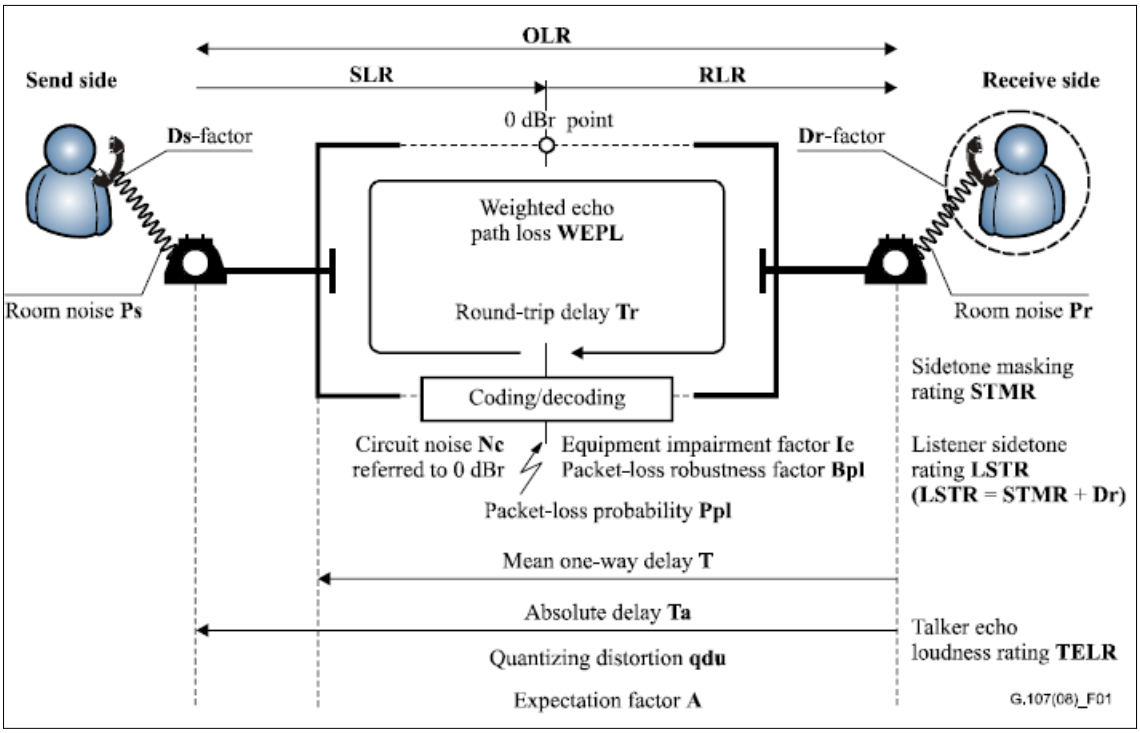

Figura 10. ***Conexión de referencia para el modelo E.***

El componente Ro representa la relación señal a ruido, es decir, describe el efecto del ruido en la comunicación incluyendo fuentes de ruido tales como el ruido de circuito y el ruido ambiente. El factor Is es la suma de todos los impedimentos que ocurren simultáneamente con la transmisión de la voz y se divide en factores más específicos que contemplan el volumen de la conexión y la distorsión de la cuantificación.

Id modela las degradaciones producidas por los retardos y el eco, el cual a su vez se divide en factores más específicos que distinguen entre el eco del transmisor y del receptor. El sumando A representa el factor de

ventaja que significa que se admite una degradación de la calidad de la comunicación por parte del usuario a cambio de otras facilidades en el acceso al servicio.

El factor Ie,eff es el factor de degradación por equipo efectivo y se corresponde a las degradaciones producidas por los códec y por las pérdidas de paquetes. Para cuantificar esta degradación se definen cuatro parámetros: factor de degradación por equipo (Ie), su valor se halla tabulado para distintos códecs [29]; robustez ante pérdida de paquetes (Bpl), tabulado igualmente; probabilidad de pérdida de paquetes (Ppl) y ratio de ráfagas ($BurstR$). Cuando la conversación se realiza a través de VoIP, este último parámetro juega un rol muy importante ya que las pérdidas, por lo general, se producen a ráfagas haciendo que tome un valor superior a uno.

En la revisión efectuada en el año 2000, se proporcionó una versión mejorada de este modelo con la finalidad de considerar los efectos de ruido ambiente en extremo emisor y la distorsión de cuantificación de una forma más acertada. En el año 2002 se incluye la degradación causada por la pérdida aleatoria de paquetes en una manera paramétrica para diferentes códecs. Desde la versión del año 2003, se proporciona un modelo mejorado para evaluar la calidad en situaciones para bajos niveles de volumen en el lado del emisor. Hasta el año 2005 no se tiene en cuenta la pérdida de paquetes dependiente que es la que atañe a las comunicaciones mediante VoIP. Por último, en 2009, se incluyen apéndices para considerar transmisiones de voz de banda ancha.

Algunas críticas realizadas a este modelo es que no tiene en cuenta degradaciones debidas a adaptaciones dinámicas [30], como por ejemplo cambio de códecs o variación del tamaño del buffer de reproducción.

3.4.2.2. PESQ

La recomendación ITU-T P.862 [31] describe un método para la evaluación de la calidad de la señal de audio en una comunicación a través

de redes telefónicas de banda estrecha. Este método se conoce por sus siglas en inglés PESQ (Perceptual Evaluation of Speech Quality).

Es un método intrusivo que necesita inyectar en la red una señal de audio conocida para evaluar la calidad de la comunicación en el otro extremo. Así, PESQ compara una señal inicial $X(t)$ conocida con una señal $Y(t)$ obtenida al otro lado de la red como consecuencia de transmitir $X(t)$ por la misma. El primer paso consiste en realizar una correlación temporal de ambas señales. Posteriormente, una vez alineadas, se comparan utilizando un modelo de percepción basado en la transformación se sendas señales en una representación que intenta reproducir la percepción psico-acústica del sistema auditivo humano. Esto se representa de forma gráfica en la figura 11.

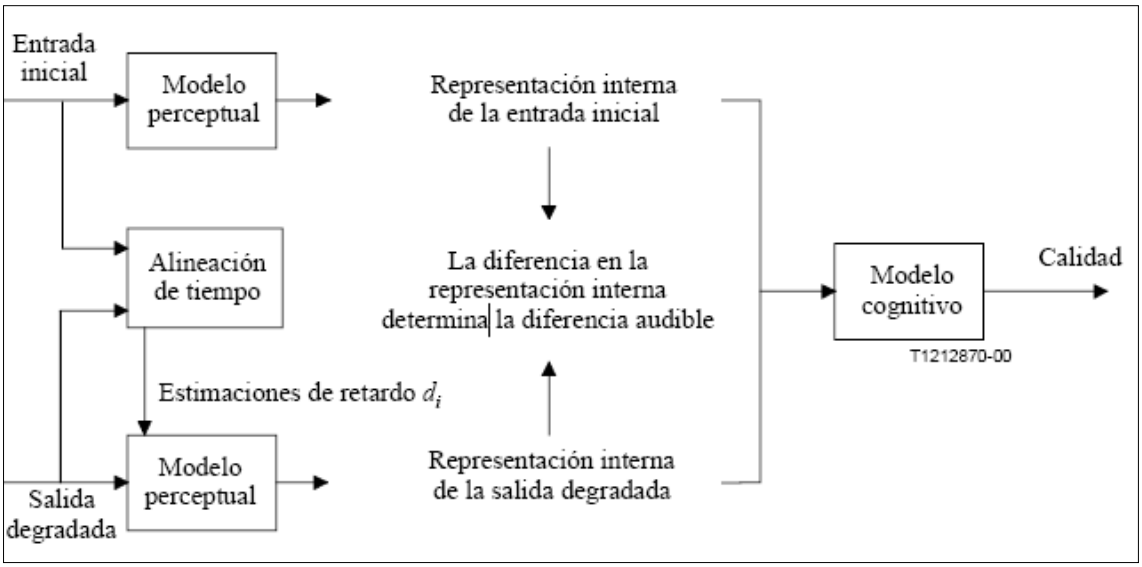

Figura 11. ***Obtención de la calidad mediante el modelo PESQ.***

Así, se obtiene una nota PESQ, equivalente a la puntuación MOS, correspondiente a un escalar en el rango -0,5 a 4,5, aunque en la mayoría de los casos suele estar entre 1,0 y 4,5 como suele suceder de forma general en experimentos sobre la calidad vocal.

3.4.2.3. ITU-T P.563

El algoritmo contenido en la recomendación P.563 [32] propone un método no intrusivo para la evaluación de la calidad vocal, ya que no necesita de una señal de referencia. Así, suele ser indicado este método para la supervisión y evaluación con la red en funcionamiento mediante el empleo de señales de audio desconocidas. La calidad que se obtiene al aplicar P.563 está relacionada con la calidad percibida en el extremo receptor.

El procedimiento en el que se basa este método es evaluar la señal de varias formas para detectar un conjunto de parámetros útiles. A partir de este conjunto de parámetros claves, se determina cuál es la principal clase de distorsión presente. De tal manera que el algoritmo P.563 se divide en tres bloques funcionales que se corresponden con las tres clases de distorsión fundamentales: relativa al tracto vocal y desnaturalización de la voz; debida a un nivel de ruido adicional intenso y las interrupciones, y relacionada con los silenciamientos y recorte temporal. Cada clase de distorsión hace uso de una combinación lineal de diversos parámetros para generar la calidad vocal intermedia.

La calidad vocal definitiva se calcula combinando los resultados de la calidad vocal intermedia con ciertas características adicionales de la señal de audio.

3.4.2.4. Nuevos métodos

El organismo internacional de las telecomunicaciones ITU-T está implicado en la consecución de una nueva metodología de obtención de la calidad vocal basada exclusivamente en la información contenida en los paquetes IP, teniendo en cuenta una carga útil de voz genérica. Este nuevo método se definirá como el estándar P.VQT, aunque actualmente el primer paso es la recomendación P.564 [33]. Este nuevo modelo presenta una herramienta no intrusiva para la monitorización en tiempo real de la calidad de la transmisión de la comunicación vocal, lo que permite realizar una gestión dinámica de la red.

Otra alternativa para la medición de la calidad de manera no intrusiva se presenta en [34]. Se propone extraer las cabeceras de los paquetes recibidos y sustituir la carga útil con datos conocidos. De forma que se utiliza una señal conocida con la misma distorsión que la recibida. Una vez hecho esto se puede aplicar el método de evaluación objetivo PESQ y obtener la puntuación MOS pertinente.