

Unión Internacional de Telecomunicaciones

# UIT-T

SECTOR DE NORMALIZACIÓN  
DE LAS TELECOMUNICACIONES  
DE LA UIT

# P.862.3

(11/2005)

SERIE P: CALIDAD DE TRANSMISIÓN TELEFÓNICA,  
INSTALACIONES TELEFÓNICAS Y REDES LOCALES

Métodos de evaluación objetiva y subjetiva de la calidad

---

**Guía de aplicación para medir la calidad  
objetiva de acuerdo con las  
Recomendaciones P.862, P.862.1 y P.862.2**

Recomendación UIT-T P.862.3

UIT-T



RECOMENDACIONES UIT-T DE LA SERIE P

**CALIDAD DE TRANSMISIÓN TELEFÓNICA, INSTALACIONES TELEFÓNICAS Y REDES LOCALES**

Vocabulario y efectos de los parámetros de transmisión sobre la opinión de los clientes	Series	P.10
Líneas y aparatos de abonado	Series	P.30
		P.300
Patrones de transmisión	Series	P.40
Aparatos para mediciones objetivas	Series	P.50
		P.500
Medidas electroacústicas objetivas	Series	P.60
Medidas relativas a la sonoridad vocal	Series	P.70
<b>Métodos de evaluación objetiva y subjetiva de la calidad</b>	<b>Series</b>	<b>P.80</b>
		<b>P.800</b>
Calidad audiovisual en servicios multimedios	Series	P.900
Aspectos de calidad de transmisión y de calidad de servicio en los puntos extremos de redes de protocolo Internet	Series	P.1000

*Para más información, véase la Lista de Recomendaciones del UIT-T.*

## **Recomendación UIT-T P.862.3**

### **Guía de aplicación para medir la calidad objetiva de acuerdo con las Recomendaciones P.862, P.862.1 y P.862.2**

#### **Resumen**

En esta Recomendación figuran observaciones importantes que conviene tener en cuenta en las evaluaciones objetivas de la calidad vocal o calidad de la voz de acuerdo con las Recs. UIT-T P.862, P.862.1 y P.862.2. Sería conveniente que los usuarios de la Rec. UIT-T P.862 comprendieran y siguieran los consejos que se dan en esta Recomendación.

Esta Recomendación constituye una guía complementaria para usuarios de la Rec. UIT-T P.862, en la que se recomienda un mecanismo para evaluar la calidad vocal en escucha utilizando muestras vocales de referencia y degradadas. El alcance de la Rec. UIT-T P.862 está claramente definido en sí mismo. La presente Recomendación ni amplía ni reduce el alcance de la misma, sino que ofrece información importante y necesaria para obtener en la práctica resultados estables, fiables y significativos con las mediciones objetivas.

#### **Orígenes**

La Recomendación UIT-T P.862.3 fue aprobada el 29 de noviembre de 2005 por la Comisión de Estudio 12 (2005-2008) del UIT-T por el procedimiento de la Recomendación UIT-T A.8. Este texto incluye las aclaraciones aceptadas el 13 de junio de 2006 por la Comisión de Estudio 12.

## PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) es un órgano permanente de la UIT. Este órgano estudia los aspectos técnicos, de explotación y tarifarios y publica Recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

La Asamblea Mundial de Normalización de las Telecomunicaciones (AMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución 1 de la AMNT.

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

## NOTA

En esta Recomendación, la expresión "Administración" se utiliza para designar, en forma abreviada, tanto una administración de telecomunicaciones como una empresa de explotación reconocida de telecomunicaciones.

La observancia de esta Recomendación es voluntaria. Ahora bien, la Recomendación puede contener ciertas disposiciones obligatorias (para asegurar, por ejemplo, la aplicabilidad o la interoperabilidad), por lo que la observancia se consigue con el cumplimiento exacto y puntual de todas las disposiciones obligatorias. La obligatoriedad de un elemento preceptivo o requisito se expresa mediante las frases "tener que, haber de, hay que + infinitivo" o el verbo principal en tiempo futuro simple de mandato, en modo afirmativo o negativo. El hecho de que se utilice esta formulación no entraña que la observancia se imponga a ninguna de las partes.

## PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente Recomendación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT no ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB.

© UIT 2006

Reservados todos los derechos. Ninguna parte de esta publicación puede reproducirse por ningún procedimiento sin previa autorización escrita por parte de la UIT.

## ÍNDICE

	<b>Página</b>
1 Alcance .....	1
2 Referencias .....	1
3 Definiciones.....	2
4 Abreviaturas, siglas o acrónimos .....	3
5 Convenios .....	3
6 Observaciones generales.....	3
6.1 Factores de las pruebas .....	4
6.2 Aplicaciones .....	4
7 Características de las señales de referencia .....	4
7.1 Longitud de la señal.....	5
7.2 Voz activa.....	5
7.3 Estructura temporal .....	5
7.4 Nivel vocal activo (nivel con voz activa) .....	5
7.5 Aplicación de voz artificial .....	6
7.6 Requisitos de las grabaciones de voz .....	6
7.7 Variación debida al hablante y al contenido de la voz .....	6
7.8 Silencios al principio y al final.....	6
7.9 Filtrado previo .....	7
7.10 Fondo de ruido.....	7
7.11 Temas de implementación.....	8
8 Características de la señal degradada que han de evaluarse .....	8
8.1 Diferencia en la duración de la voz activa entre la señal de voz de referencia y la de voz degradada .....	8
8.2 Nivel vocal activo.....	9
8.3 Diferencia en la duración del silencio al principio y al final entre la voz de referencia y la degradada.....	9
9 Características de los trayectos de inserción y adquisición de señales.....	9
9.1 Influencia de los circuitos de medición y la configuración de pruebas en el trayecto de inserción.....	9
9.2 Influencia de los circuitos de medición y la configuración de pruebas en el trayecto de adquisición .....	10
10 Análisis de los resultados .....	11
10.1 Promedio de los resultados de medición .....	11
10.2 Fiabilidad de los resultados de las mediciones de la PESQ .....	11
10.3 Precisión de los valores de las mediciones PESQ.....	12
10.4 Interpretación de los resultados de precisión.....	13
11 Informe de los resultados.....	13

**Página**

12	Orientaciones sobre la utilización de la ampliación a banda ancha (P.862.2) a la P.862 .....	14
	Apéndice I – Valores de referencia para determinar la calidad objetiva mediante la Rec. UIT-T P.862 para códecs normalizados UIT-T/GSM.....	14
I.1	Los valores de referencia de la Rec. UIT-T P.862.1 se calcularon para las siguientes condiciones códecs/MNRU a partir de la base de datos de voz del anexo B/P.501:.....	14
I.2	Preprocesamiento de la voz fuente .....	16
I.3	Procesamiento de la G.711 .....	17
I.4	Procesamiento de la G.726 .....	17
I.5	Procesamiento de G.728, G.729, anexo A/G.729 y G.723.1.....	17
I.6	Procesamiento de MNRU .....	17
	Apéndice II – Bases de datos de pruebas para P.862/P.862.1 .....	21
	Apéndice III – Informe sobre las mediciones P.862/P.862.1.....	22
III.1	Informe e interpretación de los resultados PESQ promediados .....	22
III.2	Informe e interpretación de los resultados de las mediciones individuales de la PESQ.....	22
	Apéndice IV – Método de calibración de las interfaces de marca.....	24
IV.1	Calibración del nivel en emisión (extremo próximo) del equipo de pruebas.....	24
IV.2	Calibración del nivel en recepción (extremo distante) del equipo de pruebas.....	24
	BIBLIOGRAFÍA .....	25

## Recomendación UIT-T P.862.3

### Guía de aplicación para medir la calidad objetiva de acuerdo con las Recomendaciones P.862, P.862.1 y P.862.2

#### 1 Alcance

En la presente Recomendación figuran observaciones importantes que conviene tener en cuenta en las evaluaciones objetivas de la calidad de la voz de acuerdo con las Recs. UIT-T P.862, P.862.1 y P.862.2. Sería conveniente que los usuarios de la Rec. UIT-T P.862 comprendieran y siguieran los consejos que se dan en esta Recomendación.

Esta Recomendación constituye una guía complementaria para los usuarios de la Rec. UIT-T P.862, en la que se recomienda un mecanismo para evaluar la calidad vocal en escucha utilizando muestras de voz de referencia y degradadas. No puede emplearse para evaluar la calidad del habla sólo o con interacción. Se parte del supuesto de que el algoritmo de evaluación objetiva de la calidad es totalmente conforme con la Rec. UIT-T P.862, lo cual puede comprobarse mediante la prueba de conformidad incluida en un anexo a la Rec. UIT-T P.862.

El alcance de la Rec. UIT-T P.862 está claramente definido en sí mismo. La presente Recomendación ni amplía ni reduce el alcance de la misma, sino que ofrece información importante y necesaria para obtener en la práctica resultados estables, fiables y significativos con las mediciones objetivas.

En la cláusula 12 se analizan las aplicaciones y limitaciones relativas a la ampliación a banda ancha de la Recomendación P.862 que se definen en la Rec. UIT-T P.862.2.

#### 2 Referencias

Las siguientes Recomendaciones del UIT-T y otras referencias contienen disposiciones que, mediante su referencia en este texto, constituyen disposiciones de la presente Recomendación. Al efectuar esta publicación, estaban en vigor las ediciones indicadas. Todas las Recomendaciones y otras referencias son objeto de revisiones por lo que se preconiza que los usuarios de esta Recomendación investiguen la posibilidad de aplicar las ediciones más recientes de las Recomendaciones y otras referencias citadas a continuación. Se publica periódicamente una lista de las Recomendaciones UIT-T actualmente vigentes. En esta Recomendación, la referencia a un documento, en tanto que autónomo, no le otorga el rango de una Recomendación.

- [1] Recomendación UIT-T P.501 (2002), *Señales de prueba para utilización en telefonometría, anexo B – Ficheros de voz y secuencias de ruido.*
- [2] Recomendación UIT-T P.56 (1993), *Medición objetiva del nivel vocal activo.*
- [3] Recomendaciones UIT-T de la serie P – Suplemento 23 (1998), *Bases de datos del UIT-T de voces codificadas.*
- [4] Recomendación UIT-T P.50 (1999), *Voces artificiales.*
- [5] Recomendación UIT-T P.800 (1996), *Métodos de determinación subjetiva de la calidad de transmisión.*
- [6] Recomendación UIT-T P.830 (1996), *Evaluación subjetiva de la calidad de funcionamiento de los códecs digitales de banda telefónica y de banda ancha.*
- [7] Recomendación UIT-T P.862 (2001), *Evaluación de la calidad vocal por percepción: Un método objetivo para la evaluación de la calidad vocal de extremo a extremo de las redes telefónicas de banda estrecha y códecs vocales.*

- [8] Recomendación UIT-T P.862.1 (2003), *Función de correspondencia para convertir los resultados brutos de la prueba P.862 en nota media de opinión de la calidad de escucha objetiva*.

### 3 Definiciones

En esta Recomendación se definen los términos siguientes.

**3.1 voz/señal fuente:** Es la señal de voz original sin degradación alguna. Debe grabarse y almacenarse de conformidad con la Rec. UIT-T P.830. No tiene por qué ser la misma que la voz de referencia que se define a continuación.

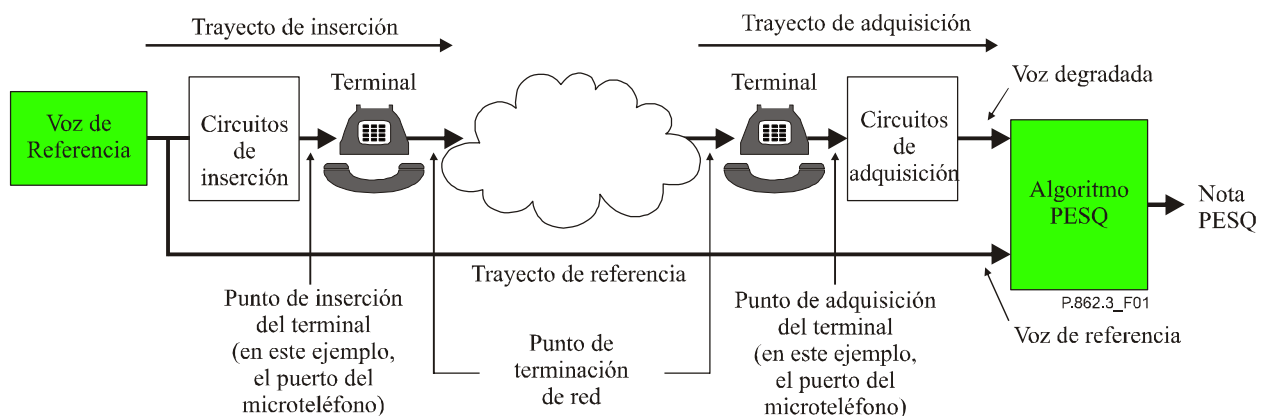
**3.2 voz/señal de referencia:** Es la señal vocal que se utiliza como referencia en el algoritmo de la Rec. UIT-T P.862 para poner de manifiesto los efectos del sistema sometido a prueba.

**3.3 voz/señal de entrada:** Es la señal que se aplica al sistema sometido a prueba en el punto de inserción de señales. Se obtiene a partir de la señal de voz de referencia. Puede ser idéntica a la señal de referencia o haberse aplicado algún tipo de procesamiento a la misma, por ejemplo adición de ruido. Para más información véase 7.10.

**3.4 voz/señal degradada:** Es la voz de referencia después de haber pasado a través del sistema sometido a prueba.

**3.5 trayecto de inserción de señales:** Consta del trayecto de conexión (cableado, dispositivos electrónicos, etc.) entre la señal de referencia para el algoritmo de la Rec. UIT-T P.862 y la interfaz de entrada, denominada punto de inserción.

En la figura 1, el circuito de pruebas de ejemplo se divide en trayecto de inserción, sistema sometido a prueba y trayecto de adquisición, y se muestran los posibles puntos de inserción y adquisición *para el caso de mediciones hardware*. Los puntos de inserción y adquisición concretos dependerán del sistema específico sometido a prueba y de la configuración de pruebas.



NOTA – En función del modo de inserción, adquisición y almacenamiento de las señales, así como de si los cálculos de las mediciones según la Rec. UIT-T P.862 se realizan en tiempo real o en diferido, estos trayectos pueden ser trayectos eléctricos físicos continuos o trayectos lógicos, como el caso de una muestra de salida que se almacena para posteriores análisis.

**Figura 1/P.862.3 – Ejemplo de configuración de medición y terminología**

**3.6 trayecto de adquisición de señales:** Consta del trayecto de conexión entre el punto de adquisición (la interfaz de salida de la red sometida a prueba) y el punto donde se aplica el algoritmo de la Rec. UIT-T P.862 (véase la figura 1).



**3.7 dBov:** Valor en dB relativo al punto de sobrecarga de un sistema digital. Según la Rec. UIT-T G.711, 0 dBm0 en representación analógica equivale a -6,15 dBov y -6,18 dBov para los códecs de ley A y ley  $\mu$ , respectivamente.

#### 4 Abreviaturas, siglas o acrónimos

En esta Recomendación se utilizan las siguientes abreviaturas, siglas o acrónimos.

IRS	Sistema intermedio de referencia ( <i>intermediate reference system</i> )
MOS	Nota media de opinión ( <i>mean opinion score</i> )
MOS-LQO	Nota media de opinión – Calidad de audición objetiva (estimación de la calidad en escucha subjetiva a partir de una técnica de medición objetiva)
MOS-LQS	Nota media de opinión – Calidad de audición subjetiva (medición directa de la calidad en escucha a partir de índices subjetivos de muestras)
PESQ	Evaluación de la calidad vocal por percepción ( <i>perceptual evaluation of speech quality</i> )
RMS	Valor cuadrático medio ( <i>root mean square</i> )

#### 5 Convenios

Se recomienda convertir los resultados en bruto de las mediciones definidas en la Rec. UIT-T P.862 a MOS-LQO (definida en la Rec. UIT-T P.800.1) utilizando para ello la relación especificada en la Rec. UIT-T P.862.1<sup>1</sup>, lo que evitará confusiones al comparar e interpretar los resultados, que podrían producirse debido a la aparente similitud de las dos escalas.

#### 6 Observaciones generales

En esta cláusula figuran observaciones complementarias sobre el alcance de la Rec. UIT-T P.862. Dicho alcance se resume claramente en la propia Rec. UIT-T P.862.

La fiabilidad y coherencia de los resultados dependen de varios factores, por ejemplo:

- Número de llamadas.
- Número de mediciones.
- Longitud de las muestras de voz.
- Tipo de voz empleada, es decir, natural o artificial.

Estos factores, y las consideraciones que figuran a continuación, afectan a la estructura y la complejidad de las pruebas:

- Objeto de las mediciones (por ejemplo, el establecimiento de las características de las conexiones, la supervisión de rutina o el diagnóstico de fallos).
- Características del canal de transmisión (por ejemplo ¿varían estas características en el tiempo como en el caso de una conexión móvil o ciertos tipos de conexión VoIP?).
- Tiempo disponible para efectuar las mediciones (lo que puede repercutir en la amplitud de las pruebas).

Cuando se sospeche que ciertos tipos de conexión pueden verse afectadas por las condiciones de "horas punta", también puede ser importante llevar a cabo varias series de mediciones a diferentes horas del día.

---

<sup>1</sup> En la cláusula 10 puede verse el procedimiento detallado para obtener la MOS-LQO.

La estructura de pruebas anterior debe citarse siempre junto con los valores procesados de MOS-LQO.

## **6.1 Factores de las pruebas**

La Rec. UIT-T P.862 ha demostrado ser válida para evaluar los factores de prueba, tecnologías de codificación y aplicaciones que se indican en el cuadro 1/P.862. Concretamente, deben tomarse las precauciones oportunas cuando se realicen pruebas de una red con tráfico real, puesto que algunos equipos pudieran causar degradaciones que no contempla la Rec. UIT-T P.862, por ejemplo, fenómenos indeseados que pudieran introducir los sistemas de reducción de ruido que se encuentren entre el punto de inserción de señales y el punto de adquisición de señales. Por otra parte, es bien sabido que la PESQ subestima las grandes distorsiones en la respuesta en frecuencia lineal. Esto se aplica especialmente, por ejemplo, a las limitaciones de la anchura de banda inferiores a la de 300 Hz a 3,4 kHz.

No se recomienda utilizar la Rec. UIT-T P.862 en sistemas que disponen de algoritmos de supresión de ruido entre el punto de inserción de señales y el punto de adquisición de señales.

## **6.2 Aplicaciones**

La Rec. UIT-T P.862 puede emplearse para realizar pruebas en una red con tráfico real, lo que permite evaluar el sistema en condiciones reales en lugar de realizar simulaciones por computador o emplear una determinada configuración de pruebas en un entorno de laboratorio.

Las pruebas de campo con tráfico real no generan resultados reproducibles debido a que los canales de transmisión varían en el tiempo de manera incontrolada. Las alternativas son las simulaciones controladas de la red que generan resultados exactamente reproducibles. En este último caso deben promediarse los resultados.

Las pruebas de la red con tráfico real, tales como las pruebas con móviles en vehículos en movimiento, afectan a la estructura y contenido de las señales de voz de referencia. En las pruebas con vehículos en movimiento esto se debe a la necesidad de evaluar una calidad muy variable con el tiempo a fin de obtener información precisa sobre la calidad geográfica.

Las pruebas de la red con tráfico real también exigen evaluar la calidad muestra por muestra, dado que no es posible sacar promedios por condiciones, con condiciones de la red con tráfico real, variables en el tiempo.

Las dos condiciones anteriores afectarán a la estabilidad, y posiblemente a la precisión de los resultados de la Rec. UIT-T P.862. Por esa razón, los usuarios de dicha Recomendación deberán verificar minuciosamente los resultados de las pruebas de la red con tráfico real a fin de comprobar la estabilidad de los mismos. En 10.2 se muestran resultados de calidad de funcionamiento.

Si el sistema sometido a prueba dispone de un terminal de banda ancha (como es el caso de ciertos cascos telefónicos manos libres o teléfonos IP de banda ancha), la PESQ dará como resultado una predicción de calidad similar a la que se hubiera obtenido con un filtrado en recepción de tipo IRS.

## **7 Características de las señales de referencia**

Las señales de referencia se definen y utilizan como señales de entrada al sistema sometido a prueba y como entrada de referencia de la Rec. UIT-T P.862. Las características del trayecto de inserción de señales se describen en el cláusula 9. Si el idioma en que se realiza la evaluación está incluido en la base de datos de señales vocales que figura en el anexo B/P.501, se recomienda utilizar dichas señales como señales de prueba a fin de mejorar la compatibilidad entre diferentes mediciones y evitar así la utilización de señales de referencia diferentes.

## 7.1 Longitud de la señal

El UIT-T ha verificado la validez de la Rec. UIT-T P.862 con señales de una duración de 8 a 12 s en su mayoría. No obstante, se sabe que la Rec. UIT-T P.862 puede aplicarse a señales vocales de hasta 30 s de duración [B.1]. Por consiguiente, se recomienda que la duración de las muestras de las señales vocales sea de 8 a 30 s, incluidos los silencios antes, después y entre enunciados<sup>2</sup>.

En los casos de pruebas de campo con tráfico real, pueden emplearse señales de referencia más cortas, aunque ello no sirva para evaluar exhaustivamente el sistema. Estas frases más cortas deberían tener como mínimo una voz de 3,2 s de duración, de acuerdo con 7.2.

Debe señalarse que, como el algoritmo definido en la Rec. UIT-T P.862 no es lineal, el resultado obtenido utilizando señales concatenadas no corresponde a la media aritmética simple de los resultados de cada una de las muestras.

## 7.2 Voz activa

La actividad vocal en la voz de referencia, que puede medirse como se describe en la Rec. UIT-T P.56<sup>3</sup>, debería estar entre el 40% y 80%. La duración mínima de voz activa en la señal de referencia debe ser de 3,2 s. Esta condición, junto con la longitud recomendada del fichero de señales, permite garantizar que la duración de la voz es suficiente para obtener una predicción precisa al aplicar la Rec. UIT-T P.862 y la voz debe contener silencios para que intervengan elementos importantes en la red.

## 7.3 Estructura temporal

La voz de referencia debe constar de enunciados separados por periodos de silencio representativos de las pausas naturales cuando se habla. La mayoría de los experimentos que se utilizan para calibrar y validar la Rec. UIT-T P.862 contienen pares de frases separadas por un silencio. Como ejemplos pueden citarse los materiales vocales incluidos en el Suplemento 23 a las Recomendaciones de la serie P<sup>4</sup>, cuya duración es de 8 s y que incluyen dos frases cortas separadas por un periodo de silencio de al menos 1 s, así como el anexo B/P.501 citado anteriormente. Se recomienda que la voz de referencia incluya unos cuantos enunciados continuos en lugar de muchos enunciados cortos de voz, tales como sería el contar rápidamente<sup>5</sup>.

## 7.4 Nivel vocal activo (nivel con voz activa)

El nivel vocal activo al que se refiere esta Recomendación es el nivel equivalente de la señal de referencia digital almacenada, medida de acuerdo con la Rec. UIT-T P.56. El nivel vocal activo que se aplica al trayecto de inserción de señales del sistema de medición se describe por separado en la cláusula 9. Se recomienda que todos los ficheros de voz de referencia se almacenen con un nivel

---

<sup>2</sup> El programa informático de referencia que figura en el anexo A/P.862 tiene la siguiente limitación de la longitud de las señales, aunque esta limitación ya cae fuera de la gama determinada en esta Recomendación: debido a la precisión de la aritmética en coma flotante disponible en la Rec. UIT-T P.862, una vez que las señales procesadas alcanzan una cierta longitud, comienzan a introducirse errores en el cálculo de energía de la señal. Del análisis se desprende que las señales de más de un millón de muestras aproximadamente comienzan a causar problemas. Sesenta segundos de una señal monoaural de 16 kHz contiene 960 000 muestras, valor que podría emplearse como umbral determinante para mostrar un aviso.

<sup>3</sup> La Rec. UIT-T G.191 contiene un software en el fichero sv56demo.c, que sirve para medir la proporción y el nivel vocal activo (nivel con voz activa) de conformidad con la Rec. UIT-T P.56.

<sup>4</sup> Obsérvese que la propiedad intelectual del Suplemento 23 no permite la utilización de las señales en aplicaciones comerciales.

<sup>5</sup> El número máximo de enunciados que acepta el software de referencia que figura en el anexo A/P.862 está limitado a 50. Si se utilizan señales de referencia con muchos enunciados, es preciso comprobar que la implementación concreta de la Rec. UIT-T P.862 utilizada en las pruebas admite ese número de enunciados.

de -30 dBov para evitar los recortes de crestas. Obsérvese que este nivel es el de la voz fuente almacenada en formato digital y que el nivel de entrada al sistema sometido a prueba se determina por separado de acuerdo con el objetivo de la medición (véase la Rec. UIT-T P.830)<sup>6</sup>.

### **7.5 Aplicación de voz artificial**

Es necesario estudiar más detenidamente la aplicación de señales de voz artificial, desde los puntos de vista del idioma y de la estructura temporal de la potencia de las señales y posiblemente habida cuenta de otros factores [B.2].

### **7.6 Requisitos de las grabaciones de voz**

En las Recs. UIT-T P.800 y P.830 se dan orientaciones sobre cómo realizar las grabaciones de material vocal. En esta Recomendación se parte del supuesto de que la voz fuente está grabada siguiendo dichas pautas. Obsérvese que la voz de referencia puede ser idéntica a esta voz fuente o puede haberse añadido al mismo un fondo de ruido de bajo nivel y/o habersele aplicado una conformación de frecuencia (véanse 7.9 y 7.10).

### **7.7 Variación debida al hablante y al contenido de la voz**

La variación debida al hablante y al contenido de la voz puede controlarse utilizando un conjunto fijo de muestras para todos los casos de prueba que vayan a compararse. Por consiguiente, las bases de datos que figuran en el anexo B/P.501 sirven de gran ayuda para facilitar la comparación e interpretación ulterior de los resultados obtenidos en diferentes laboratorios.

En los casos de simulación de red se recomienda que la señal de voz de referencia conste al menos de dos voces femeninas y dos masculinas, y que cada una diga frases distintas. Las notas P.862.1 que se obtengan de estas muestras distintas deben promediarse ulteriormente para evaluar las diferencias en función de las condiciones.

En los casos de pruebas de campo con tráfico real, puede emplearse un menor número de hablantes, aunque de este modo no se compruebe exhaustivamente el sistema. Si se necesita este formato, pueden incluirse la voz de varios hablantes en estas cortas señales de referencia. En caso de que se desee realizar una evaluación "por muestra", los resultados de las muestras que contengan la voz de más de un hablante mostrarán una menor dependencia con respecto a la muestra.

Al comprobar la validez de la Rec. UIT-T P.862 se disponía de muy pocos datos de voces de niños y de voces con características especiales (por ejemplo, trastornos de la voz o el habla, etc.). Con los pocos datos disponibles, no se observaron problemas con las voces de niños. No debe utilizarse música al aplicar la Rec. UIT-T P.862.

Por otra parte, se recomienda también emplear varias muestras de voces diferentes (de 4 a 10 frases) por hablante para tener en cuenta las variaciones fonéticas.

### **7.8 Silencios al principio y al final**

Para alinear los niveles de las señales de referencia y de las degradadas, en la Rec. UIT-T P.862 se utilizan los valores RMS de los niveles de las mismas. Si se incluyen silencios largos al principio y al final de la señal de referencia, la alineación de niveles puede verse comprometida.

---

<sup>6</sup> Un valor nominal típico del nivel vocal activo es -20 dBm0, que corresponde a -26 dBov aproximadamente. Al realizar pruebas de un sistema específico, el nivel vocal activo medio de dicho sistema sometido a prueba puede diferir considerablemente del valor nominal de -20 dBm0. En tales casos, puede utilizarse como nivel vocal activo de entrada el valor medio medido. Cuando se evalúe la respuesta del sistema con respecto al nivel de entrada, conviene utilizar varios valores con voz activa, por ejemplo, -14, -26 y -38 dBov (lo que equivale, aproximadamente a -8, -20, y -32 dBm0) según se recomienda en la Rec. UIT-T P.830.

Se recomienda un silencio al principio y uno al final de 0,5 s como mínimo, siempre que el equipo de medición tenga tiempo de sincronizar la voz degradada con la de referencia durante ese intervalo.

Se recomienda una duración máxima de silencio al principio y al final de unos 2 s, lo cual puede resultar útil si hay un gran retardo en el sistema.

## 7.9 Filtrado previo

La voz de referencia preparada de conformidad con 7.1 a 7.8 debe filtrarse para tener en cuenta las características de frecuencia en emisión del microteléfono. Debe señalarse que en la Rec. UIT-T P.862 se da por supuesto que la voz de referencia presenta dichas características electroacústicas adecuadamente. Si se parte del supuesto de que la voz de referencia se introduce en las redes como si en realidad se tratara de la salida de un terminal de microteléfono, el UIT-T recomienda que se utilicen las características en emisión del IRS modificadas que se definen en el anexo D/P.830. Este filtrado debería aplicarse después de haber tenido en cuenta todo lo indicado en 7.1 a 7.8.

Al aplicar el filtrado es preciso tener presente la respuesta en frecuencia nominal del sistema sometido a prueba, dado que el filtrado dependerá del lugar en que se introduzca la voz de referencia en el equipo y/o red sometidos a prueba (véase la cláusula 9).

## 7.10 Fondo de ruido

El fondo de ruido presente en la voz de referencia debe ser tan bajo como el que se considera apropiado en las grabaciones realizadas de acuerdo con las Recs. UIT-T P.800 y P.830. Asimismo, es posible añadir un silencio total (por ejemplo, una señal con amplitud digital igual a cero) de modo que las señales de voz de referencia tengan las características adecuadas definidas en 7.1, 7.2, 7.3 y 7.8<sup>7</sup>. Como se dijo en 7.6, esto corresponde al caso en que la voz de referencia es idéntica a la voz fuente.

No obstante, si se prevé la existencia de ruido no deseado en los trayectos de medición descritos en la cláusula 9 o en el fondo de ruido en el propio dispositivo sometido a prueba, debe añadirse deliberadamente un nivel bajo de fondo bajo de ruido blanco, de unos  $-75$  dBov, a la señal de referencia como se indicó anteriormente y almacenarse en formato MIC lineal de 16 bits. El nivel del fondo de ruido debe determinarse entre 0 y 4000 Hz<sup>8</sup>. El ruido con ese nivel no afectará negativamente a los resultados obtenidos de acuerdo con la Rec. UIT-T P.862, pero servirá para suprimir efectivamente la contribución de ese ruido de medición a la nota final [B.3]. Es muy importante añadir este fondo de ruido después de la etapa de filtrado previo descrita en 7.9.

---

<sup>7</sup> El envío de un silencio digital en un circuito de inserción digital, por ejemplo un teléfono RDSI, y luego a un enlace telefónico inalámbrico, pueden producirse efectos secundarios no deseados si la red inalámbrica emplea códecs GSM o 3GPP. Concretamente, al enviar un patrón constante del menor valor positivo de la ley A G.711, +8 lineal (D5 hexadecimal MTC), el códec de voz se reiniciará cada 20 ms, es decir, 50 veces/s. Esto se debe al procedimiento de vuelta a la posición inicial incorporado en el códec a efectos de pruebas del mismo. Aunque el reinicio durante las pausas de voz no es perjudicial para el propio códec, la calidad vocal medida puede diferir de la que se obtendría en la situación normal, en la que no se reinician los códecs. Podría producirse como efecto secundario que el códec no utilizara la transmisión discontinua (DTX) durante las pausas de voz, aunque la red lo hubiese habilitado. Por consiguiente, el efecto de ruido de confort o el posible recorte de voz que pueda introducir el detector de actividad vocal (VAD) no se verificaría en modo alguno. Para resolver este problema concreto, es preciso añadir a las muestras de prueba un fondo de ruido de bajo nivel, unos  $-65$  dBm0 aproximadamente, antes de enviarlas a la red; esto romperá el patrón constante de silencio digital.

<sup>8</sup> Cuando se utiliza una velocidad de muestreo de 16 kHz para la voz de referencia, es preciso poner especial cuidado al determinar el nivel del fondo de ruido.

El nivel vocal activo en el trayecto de inserción de señales del sistema de medición descrito en la cláusula 9 debe calibrarse después de aplicar el filtrado previo<sup>9</sup>.

Debe señalarse que con la inserción propuesta de ruido adicional a la señal de referencia, se obtiene unos resultados más precisos si el ruido no deseado en el trayecto de recepción es un fondo de ruido continuo, pero no se resuelven los problemas que entraña el ruido de confort, el cual se inserta únicamente en las pausas de voz.

### **7.11 Temas de implementación**

Muchas señales incluidas en la prueba de conformidad P.862 no satisfacen los requisitos antes indicados. Para la referida prueba de conformidad esta circunstancia no importa en absoluto ya que el único objetivo es probar la exactitud de la implementación. No obstante, hay que tomar la precaución de que el algoritmo implementado también produce resultados en los casos en que se violan los requisitos definidos en esta Recomendación, ya que de no ser así no puede aplicarse la citada prueba de conformidad.

## **8 Características de la señal degradada que han de evaluarse**

Las señales degradadas son la salida del sistema sometido a prueba que corresponde a la señal de entrada de prueba, incluidos los efectos de la interfaz de medición. En esta cláusula se describen las características de las señales que se almacenan digitalmente como la salida del sistema sometido a prueba para utilizarlas en los cálculos de acuerdo con la Rec. UIT-T P.862. Las características del trayecto de adquisición de señales se analizan en la cláusula 9.

### **8.1 Diferencia en la duración de la voz activa entre la señal de voz de referencia y la de voz degradada**

La duración de la voz activa se define en la Rec. UIT-T P.56.

Para alinear los niveles de las señales de referencia y degradada, en la Rec. UIT-T P.862 se utilizan los niveles RMS de las mismas. Esto significa que el algoritmo puede generar resultados erróneos si falta una parte de señal de voz o si se añade o suprime silencio en la señal degradada.

Cuando se suprima un enunciado de la señal degradada, o se silencie una o varias partes largas de la misma, el nivel de la señal deberá elevarse a un valor superior al real.

Cuando se suprima silencio de la señal degradada, el nivel de la señal deberá reducirse a un valor inferior al real.

Estos problemas afectarán a la cantidad de perturbaciones que presente la señal degradada y, por consiguiente, al resultado de la medición objetiva de la calidad. Si las duraciones de la voz en las señales de referencia y degradada difieren en más de un 25%, el efecto puede ser lo suficientemente grande como para modificar el resultado, especialmente si se han sustituido partes contiguas de larga duración por silencio.

---

<sup>9</sup> Es posible que no pueda introducirse el fondo de ruido en el sistema sometido a prueba si el nivel del fondo de ruido es inferior al mínimo nivel posible del sistema. Por ejemplo, los valores más pequeños para la codificación de ley A son  $\pm 8$  en valores lineales de 16 bit. Por consiguiente, el nivel inferior es de  $-72$  dBov. De lo anterior se desprende que si el nivel de entrada del sistema se calibra a, por ejemplo,  $-30$  dBov, el fondo de ruido de  $-75$  dBov no puede atravesar el sistema y, por tanto, no resuelve el "problema de ruido no deseado". Si se desea verificar niveles de entrada considerablemente superiores que el nivel vocal activo nominal de  $-26$  dBov, y si se aplica un fondo de ruido almacenado de  $-75$  dBov a la muestra de referencia, la aplicación de la Rec. UIT-T P.862 no permite evaluar las posibles degradaciones (por ejemplo, chasquidos, ráfagas de ruido, etc.) que aparezcan durante las pausas de voz, dado que un nivel mayor del fondo de ruido de la muestra degradada puede encubrir las degradaciones de bajo nivel. Ahora bien, este problema se soluciona utilizando un nivel vocal activo nominal de  $-26$  dBov aproximadamente.

## **8.2 Nivel vocal activo**

El nivel vocal activo se define en la Rec. UIT-T P.56.

Aunque este nivel se normaliza al calcular los valores PESQ, se recomienda que el nivel vocal digital almacenado como señales degradados para aplicar el algoritmo PESQ esté alrededor de -30 dBov a fin de evitar los recortes y la distorsión de cuantización. Debe señalarse que no debe utilizarse la Rec. UIT-T P.862 para evaluar los efectos del nivel de recepción/escucha<sup>10</sup>.

## **8.3 Diferencia en la duración del silencio al principio y al final entre la voz de referencia y la degradada**

Para alinear los niveles de las señales de referencia y degradadas, la Rec. UIT-T P.862 utiliza los valores RMS de los niveles de las mismas. Si la señal degradada contiene pausas largas al principio y al final, la alineación de niveles puede no ser la óptima, lo que puede convertirse en un problema si las duraciones de las señales de referencia y degradada difieren en más del 20%<sup>11</sup>.

Por otra parte, en la Rec. UIT-T P.862 no se toman en consideración las distorsiones en la señal degradada que se producen antes del inicio o después del final de la señal de voz activa. Esta señal de voz activa se determina a partir del primer y último puntos en los que la señal de referencia rebasa los 50 dB de SPL aproximadamente.

## **9 Características de los trayectos de inserción y adquisición de señales**

En esta cláusula se describen las características deseadas de los trayectos de inserción y adquisición de señales para realizar mediciones físicas. Los circuitos y el entorno de medición pueden afectar a los resultados de la Rec. UIT-T P.862 si no se toma la precaución de controlar los factores que intervienen. Se ha de tratar de reducir al mínimo el ruido y las interferencias en los trayectos de inserción y adquisición para que éstos no afecten a los resultados.

### **9.1 Influencia de los circuitos de medición y la configuración de pruebas en el trayecto de inserción**

El método preferible es utilizar siempre que sea posible una interfaz definida, como POTS o RDSI, de modo que el equipo de pruebas se calibre para esas interfaces con los niveles nominales de señal recomendados.

Si no puede utilizarse una interfaz definida, a menudo se toma como punto de inserción el puerto del microteléfono de un dispositivo terminal, cuyo nivel de entrada requerido no se conoce inicialmente, al tratarse de una interfaz de marca de un determinado fabricante. Aunque existen normas, como la norma norteamericana TIA-810-A, que especifican las características del terminal entre las interfaces acústica y de red, éstas no especifican puntos intermedios tales como la interfaz del microteléfono. La distribución de ganancia y el filtrado dependen del fabricante, e incluso de cada terminal. En algunos casos, estas características pueden configurarse en el dispositivo extremo. Cuando se utiliza el puerto del microteléfono como puerto pruebas de la Rec. UIT-T P.862, el ingeniero de pruebas puede tratar de medir:

- 1) la calidad de funcionamiento del dispositivo terminal y la red conjuntamente;

---

<sup>10</sup> Si se están realizando pruebas sobre una MNRU conforme con la Rec. UIT-T P.810, deben tomarse las debidas precauciones en la ecualización de niveles a fin de preservar el nivel vocal real excluyendo el ruido añadido por la MNRU.

<sup>11</sup> Según las observaciones empíricas, los resultados de aplicar la Rec. UIT-T P.862 a EVRC [B.4] dependen de la alineación concreta de los límites de la trama de codificación con los datos MIC de entrada. El resultado puede variar hasta 0,25 dependiendo de dónde caigan los límites de trama. En el caso de la codificación EVRC, el método para obtener un resultado estable sería medir cada una de las 80 alineaciones posibles y promediar los resultados. Pueden darse situaciones similares con otros procesos DSP.

- 2) la calidad de funcionamiento del propio dispositivo extremo (conectado a una red de referencia); o
- 3) la calidad de funcionamiento de la propia red, con una mínima contribución del dispositivo terminal.

No obstante, en todo caso se desea eliminar las contribuciones de la configuración empleada en la medición.

El ingeniero de pruebas debe asegurarse de que el nivel vocal activo aplicado a la interfaz de marca sea acorde con el nivel de red deseado y la gama dinámica del códec, lo que exige caracterizar adecuadamente la ganancia entre el punto de inserción y el punto de la interfaz entre el terminal y la red en ambos sentidos de transmisión.

Al aplicar las señales de voz a la interfaz de marca, el ingeniero de pruebas debe tener presente el filtrado (por ejemplo, el IRS modificado y la ecualización de frecuencias del transductor) entre las interfaces acústica y de red. Los fabricantes de terminales pueden incorporar cualquier combinación de filtrado acústico, electrónico o digital en ambos lados de la interfaz del microteléfono de marca. Por esa razón, el equipo de pruebas P.862 puede detectar un filtrado total, parcial o ningún filtrado después del punto de inserción. A fin de obtener un resultado preciso, la configuración medida debe incluir el filtrado adecuado para el caso de prueba de que se trate. Análogamente, el filtrado que se aplique a la señal de referencia debe ser idéntico al filtrado aplicado al circuito de pruebas de extremo a extremo.

En este párrafo se explica una técnica ideal que puede emplearse para determinar las características de entrada de la señal de referencia cuando exista la posibilidad de que un filtro de entrada utilizado en el modo de funcionamiento normal del dispositivo terminal de comunicación no se encuentre en el trayecto de medición. Debería utilizarse una boca artificial (por ejemplo la de un simulador de cabeza y torso) para introducir la señal de prueba acústica en el terminal que debería estar conectado a un punto de referencia del extremo distante (por ejemplo, un punto RDSI). El nivel acústico utilizado debería representar la utilización normal del dispositivo terminal, y el nivel de ruido de fondo debería ser inferior a 35 dBA. Esta utilización normal puede reflejar la utilización del micrófono interno o de un equipo manos libres personal, dependiendo de lo que se desee evaluar. La voz artificial debería calibrarse y la posición del terminal debería ser representativa del uso normal. El nivel eléctrico y el contenido en frecuencias deberían adquirirse en un punto de referencia de la conexión de red (por ejemplo el punto extremo de la RDSI). El proceso debería repetirse (con la misma señal de prueba) utilizando la inyección eléctrica en el punto de inyección de pruebas P.862, utilizando el equipo que se utilizará en las pruebas P.862. La señal de entrada debería ajustarse de modo que el nivel eléctrico y el contenido en frecuencias correspondan a los adquiridos durante la inyección acústica. La técnica aquí descrita es el método ideal, y en muchas situaciones puede aplicarse un método aproximado<sup>12</sup>. Si no se utiliza esta técnica, se recomienda que el ingeniero de pruebas consulte las especificaciones del fabricante relativas a las interfaces acústicas y eléctricas con los terminales de comunicación.

## **9.2 Influencia de los circuitos de medición y la configuración de pruebas en el trayecto de adquisición**

Después de atravesar el sistema sometido a prueba, la voz de referencia debe transferirse del punto de adquisición al sistema P.862. Este trayecto de adquisición puede introducir ruido y distorsión, lo que puede afectar a los resultados. El trayecto de adquisición puede presentar problemas tales como bucles de puesta tierra, captación desde conductores de alimentación en c.a., u otras señales en modo común que pudieran estar presente. La captación en la banda puede afectar a los resultados.

---

<sup>12</sup> Esta aproximación es específica del equipo de medición. En el apéndice IV se describe un método posible.



Además, si los niveles de ruido fuera de la banda son lo suficientemente altos puede producirse efecto alias en caso de que el filtrado antialias sea insuficiente.

Para minimizar el ruido causado por los trayectos de inserción y adquisición, se recomienda que la contribución conjunta de ambos sea inferior a  $-70$  dBoV, de modo que se obtenga una SNR de 40 dB y que el resultado de la medición objetiva de la calidad esté determinado exclusivamente por las influencias del sistema sometido a prueba.

Por regla general, unas velocidades de muestreo de variación lenta, una ampliación del tiempo o una compresión del tiempo de la señal transmitida puede dar lugar a unas notas demasiado pesimistas debido a un inadecuado alineamiento en el tiempo.

Si se trata de una transmisión analógica, hay que tener la precaución de que no se produzca una deriva de reloj excesiva entre los convertidores A/D y D/A. Esto podría suceder en los equipos de usuario, especialmente si el hardware no soporta la velocidad de muestreo necesaria y el excitador de la tarjeta de sonido produce una conversión de la velocidad de muestreo del software.

## **10 Análisis de los resultados**

### **10.1 Promedio de los resultados de medición**

Como se subrayó en 7.7, al medir la calidad objetiva deben utilizarse al menos dos voces femeninas y dos masculinas. Antes de promediar o realizar otros cálculos estadísticos, es preciso transformar los resultados de la medición al dominio MOS-LQO (de acuerdo con la Rec. UIT-T P.862.1) y luego promediar para los hablantes y las muestras de voz. Dado que el algoritmo definido en la Recomendación P.862 no es lineal, los resultados de muestras concatenadas no corresponderán con los resultados promediados de esas muestras probadas por separado.

Según se dijo en 6.2, hay dos tipos de aplicación P.862 para las que es necesario utilizar dos métodos de análisis diferentes. En el primer caso, el promedio para los hablantes y muestras de voz debe realizarse antes de continuar con el análisis. Este análisis resulta adecuado para simulaciones de redes controladas con resultados totalmente reproducibles. En el caso de pruebas en una red con tráfico real, es necesario evaluar la calidad para cada muestra, debido a que los canales de transmisión varían de manera no controlada en el tiempo.

### **10.2 Fiabilidad de los resultados de las mediciones de la PESQ**

Se han utilizado numerosas bases de datos para pruebas, validaciones y calibrados P.862 (P.862.1). Según se describe en las Recs. UIT-T P.862 y P.862.1, las bases de datos contenían muestras de voz de diferentes hablantes de ambos sexos, en varios idiomas y que representaban degradaciones de voz generadas por redes simuladas y redes con tráfico real. Por otra parte, las condiciones de red correspondían a aplicaciones fijas, inalámbricas y de VoIP. Para los detalles sobre el contenido de las bases de datos de prueba, véase el apéndice II.

Debe señalarse que los resultados de las mediciones P.862/P.862.1 tienen una fiabilidad del 95% y una precisión controlada y conocida, cuando el algoritmo se utiliza en aplicaciones del mismo tipo que para las que se ha entrenado, probado y validado dicho algoritmo. Dicho de otro modo, los casos de medición han de representar, desde el punto de vista estadístico, el mismo tipo de población de muestras que la de las muestras con las que se ha entrenado, probado, validado y calibrado el algoritmo de P.862/P.862.1, a fin de que los valores de precisión determinados sean válidos. La fiabilidad de los resultados y su precisión se pierde y deja de estar controlada cuando el algoritmo se utiliza para evaluar la calidad de voz en nuevos tipos de tecnologías, con otros tipos de códecs o en nuevas redes reales.

### 10.3 Precisión de los valores de las mediciones PESQ

Para evaluar la calidad de las bases de datos descritas en 10.2 según P.862/P.862.1, se han utilizado tres métricas estadísticas, a saber, el coeficiente de correlación, el error de predicción y la distribución de errores residuales. Según se describió en 10.1, el método de análisis varía en función del tipo de aplicación, de las condiciones de red simuladas controladas y condiciones de pruebas de red con tráfico real o de pruebas de campo.

Al calcular la métrica estadística para todas las condiciones de red simuladas, se han utilizado promedios para cada una de estas condiciones de al menos cuatro veces, dos masculinas y dos femeninas. En el caso de bases de datos de red con tráfico real, la métrica estadística se ha calculado utilizando para cada muestra notas objetivas y subjetivas.

Los resultados de calidad de funcionamiento obtenidos se presentan en los cuadros 1 y 2. Asimismo, se calcularon los límites críticos de confianza del 95% para el coeficiente de correlación y el error de predicción a fin obtener el tope (límite) de correlación inferior del 95% y el tope de error de predicción superior de 95%.

Los resultados se desglosan por tipo de aplicación (por ejemplo, condiciones de red inalámbrica y VoIP simuladas y condiciones de red inalámbrica y VoIP con tráfico real). Por consiguiente, estos valores de precisión expresan la calidad de funcionamiento del algoritmo PESQ cuando se utiliza en alguna de las aplicaciones mencionadas en 10.2.

**Cuadro 1/P.862.3 – Intervalos de confianza para la estimación del coeficiente de correlación y del error de predicción**

Aplicación	N	Métrica	P.862 (PESQ en bruto, obtenida directamente)	P.862.1 (PESQ calibrada)
Datos por simulación (aplicaciones inalámbricas, VoIP y fijas)	1357	R	0,956	0,956
		CI 95%-límite inferior	0,940	0,940
		PE	N.A.	N.A.
		CI 95%- límite superior	N.A.	N.A.
Datos reales (Aplicaciones inalámbricas: GSM EE.UU y UE, CDMA-EE.UU, TDMA-EE.UU, iDEN- EE.UU, AMPS- EE.UU; y aplicación VoIP)	1135	R	0,925	0,926
		CI 95%- límite inferior	0,916	0,917
		PE	0,479	0,462
		CI 95%- límite superior	0,492	0,475

**Cuadro 2/P.862.3 – Distribución del error residual**

Aplicación	MOS bins	<0,25	<0,5	<0,75	<1	<1,25	<1,5	<1,75	<2
Datos reales (Aplicaciones inalámbricas: GSM EE.UU. y UE, CDMA-EE.UU., TDMA-EE.UU., IDEN-EE.UU., AMPS-EE.UU.; y aplicación VoIP)	P.862 CDF (%)	32,51	66,52	90,84	97,97	99,38	99,91	99,91	100
	P.862 prob (%)	32,51	34,09	24,32	7,14	1,41	0,53	0	0,09
	P.862.1 CDF (%)	40,44	70,48	90,33	97,71	99,3	99,7	99,91	100
	P.862.1 Prob (%)	40,44	30,04	19,82	7,4	1,59	0,44	0,18	0,09

#### 10.4 Interpretación de los resultados de precisión

Por definición, el algoritmo de P.862/P.862.1 es un estimador de la opinión subjetiva de la calidad vocal proporcionada con la red sometida a prueba. Debe señalarse que, por esa razón, los resultados de medir la calidad vocal mediante el algoritmo PESQ se ven afectados por los valores de precisión indicados en los cuadros 1 y 2.

Debe señalarse que, como se indicó en 10.2, los valores de precisión son válidos siempre que los casos de medición representen, desde el punto de vista estadístico, la misma población de muestras que la de las muestras presentadas en 10.2.

El límite inferior del 95% del intervalo de confianza del coeficiente de correlación representa las mediciones de acuerdo con P.862/P.862.1 cuya correlación con la opinión subjetiva se prevé que será mayor o igual que el límite inferior del intervalo de confianza del 95% del coeficiente de correlación, con independencia de que las condiciones de la red sean reales (con tráfico real) o simuladas, e independientemente del tipo de red para el que se realizan las pruebas (esto es, inalámbrica, VoIP y fija) (cuadro 1).

La distribución de errores residuales (cuadro 2) representa la función densidad acumulativa (CDF) de los errores absolutos entre las MOS y las notas P.862/P.862.1 y constituye la probabilidad de que el error absoluto sea menor que un valor de error absoluto dado. Por ejemplo, la probabilidad de que el error absoluto sea menor que una MOS de 0,5 es mayor que el 70%, mientras que la probabilidad de que el error sea inferior que una MOS de 0,75 es mayor que el 90%. En el cuadro 2 se facilita también la función densidad de probabilidad (PDF) del error absoluto. Como se prevé, de conformidad con la CDF, la PDF muestra que cuanto menor sea el error absoluto más probabilidad de ocurrencia tiene.

### 11 Informe de los resultados

Según se mencionó en 10.4, dependiendo del tipo de aplicación, es decir, una red simulada o red con tráfico real, las mediciones realizadas con P.862/P.862.1 deben notificarse según la precisión del algoritmo presentada en 10.3 (cuadros 1 y 2).

Se recomienda utilizar el coeficiente de correlación como métrica estadística informativa sobre la calidad de funcionamiento P.862/P.862.1 para una aplicación concreta. Para el informe de los resultados de las mediciones P.862.1 de una aplicación específica se recomienda utilizar el error de predicción junto con la distribución de error residual.

Por regla general, hay que poner en el informe los valores de PESQ promediados, máximos y mínimos, así como el número de mediciones empleadas para calcular el promedio. En el apéndice III se indican algunas recomendaciones precisas para los informes de mediciones PESQ.

Por otra parte, el número de mediciones que alcanzan una determinada nota o gama de notas PESQ puede representarse gráficamente en la forma de una distribución de frecuencias. En los casos en que el sistema sometido a prueba dé resultados relativamente estables de la calidad en escucha, puede emplearse la desviación típica para ayudar a determinar si son necesarias más mediciones para lograr una precisión dada. Esta solución no es válida cuando el sistema sometido a prueba presenta grandes variaciones con el tiempo (por ejemplo, redes VoIP o móviles).

## **12 Orientaciones sobre la utilización de la ampliación a banda ancha (P.862.2) a la P.862**

En principio, la orientación facilitada en este documento es aplicable tanto a la Rec. UIT-T P.862 como a su ampliación a banda ancha, la P.862.2. No obstante, resultan necesarias algunas orientaciones específicas para la ampliación a banda ancha de la Rec. UIT-T P.862.

Las orientaciones que se describen a continuación se refieren principalmente a la característica en emisión del IRS que se aplica a la señal de entrada o a la referencia. En la ampliación a banda ancha, no se recomienda filtrar la señal vocal ni el del ruido ambiente. A esto se hace referencia en 6.2, 7.9 y 7.10.

En cuanto al cálculo del nivel vocal activo de acuerdo con la Rec. UIT-T P.56, se recomienda utilizar la opción de banda ancha de la P.56. A esto se hace referencia en 7.2, 7.4, 8.1 y 8.2.

La inserción de un nivel bajo de fondo de ruido propuesta en 7.10 no se ha evaluado para la ampliación a banda ancha, por lo que no se recomienda.

En las cláusulas 10 y 11 se describe la precisión del método P.862. Las cifras facilitadas sólo son aplicables a la P.862 de banda estrecha.

En 3.6, el nivel 0 dBm0 se describe conforme a la Rec. UIT-T G.711, el cual sólo está disponible para aplicaciones de banda estrecha.

Los dos métodos P.862.1 y P.862.2 se refieren a la escala MOS-LQO definida en la Rec. UIT-T P.800.1. Hay que tener presente también que en el futuro podrá aplicarse un calificador al término MOS-LQO para distinguir los casos de banda ancha y banda estrecha. Los resultados producidos por la Rec. UIT-T P.862.1 se refieren únicamente al contexto de banda estrecha. Los resultados de la Rec. UIT-T P.862.2 se aplican a aplicaciones de banda ancha o una combinación de banda ancha y banda estrecha. Por tanto, no es posible comparar directamente los resultados MOS-LQO de las P.862.1 y P.862.2.

## **Apéndice I**

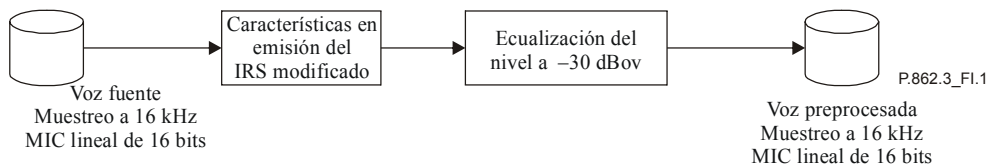
### **Valores de referencia para determinar la calidad objetiva mediante la Rec. UIT-T P.862 para códecs normalizados UIT-T/GSM**

#### **I.1 Los valores de referencia de la Rec. UIT-T P.862.1 se calcularon para las siguientes condiciones códecs/MNRRU a partir de la base de datos de voz del anexo B/P.501:**

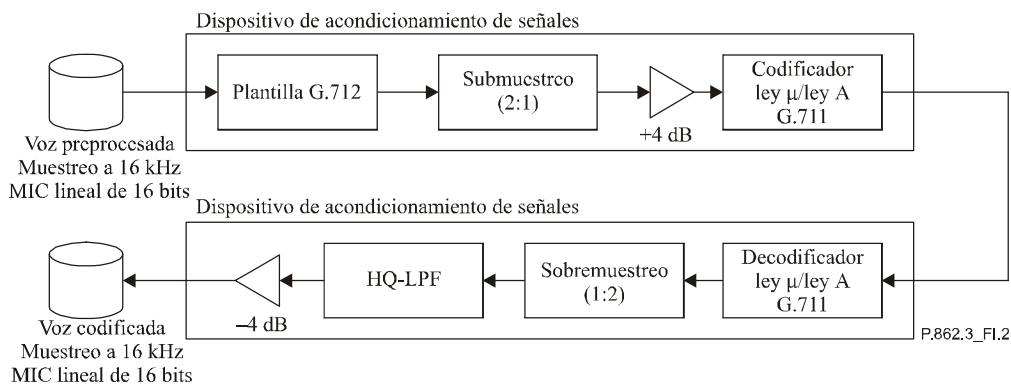
- G.711 ley  $\mu$ , ley A;
- G.726 a 16, 24, 32 y 40 kbit/s;
- G.728;
- G.729;
- G.729 anexo A;
- G.723.1 a 5,3 y 6,3 kbit/s;
- GSM-AMR a 4,75, 5,15, 5,9, 6,7, 7,4, 7,95, 10,2 y 12,2 kbit/s;

- GSM-EFR;
- GSM-FR;
- GSM-HR;
- MNRU (Q = 5, 10, 15, 20, 25, 30, 35, 40 y 45 dB).

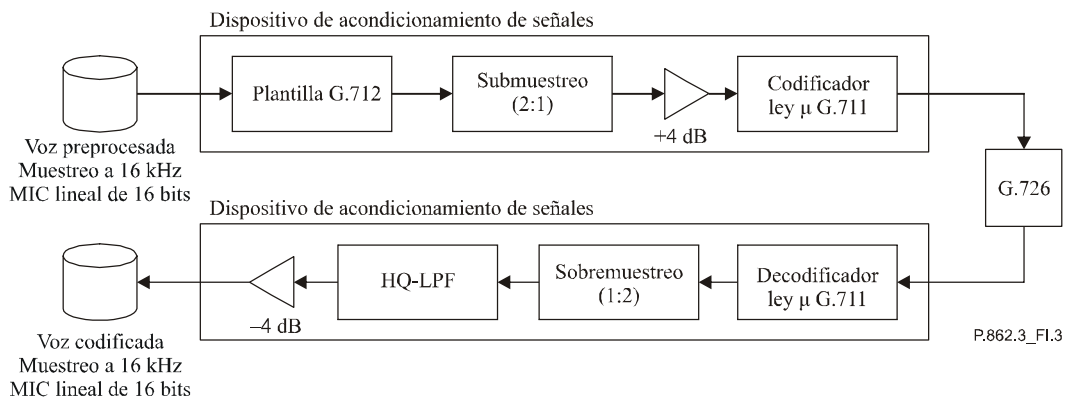
La figura I.1 muestra el preprocesamiento de las señales de voz de referencia. En las figuras I.2, I.3 y I.4 se ilustran los procedimientos de codificación G.711, G.726 y los otros códecs, respectivamente. En la figura I.5 se presenta el procedimiento de procesamiento de MNRU. Los cuadros I.1, I.2 y I.3 contienen los valores de referencia que se obtienen al convertir los valores de PESQ en bruto, obtenidos directamente<sup>13</sup>, a MOS-LQO, utilizando, la Rec. UIT-T P.862.1, para la MNRU y los códecs normalizados del UIT-T. El procesamiento de señales de G.711, G.726 y MNRU se realizó mediante el software procesamiento facilitado en la Rec. UIT-T G.191.



**Figura I.1/P.862.3 – Preprocesamiento de la voz fuente**

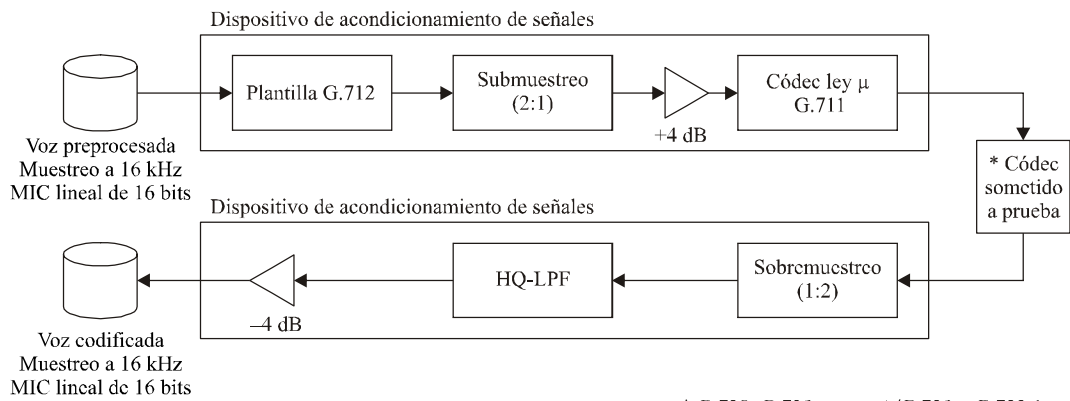


**Figura I.2/P.862.3 – Procesamiento de G.711**

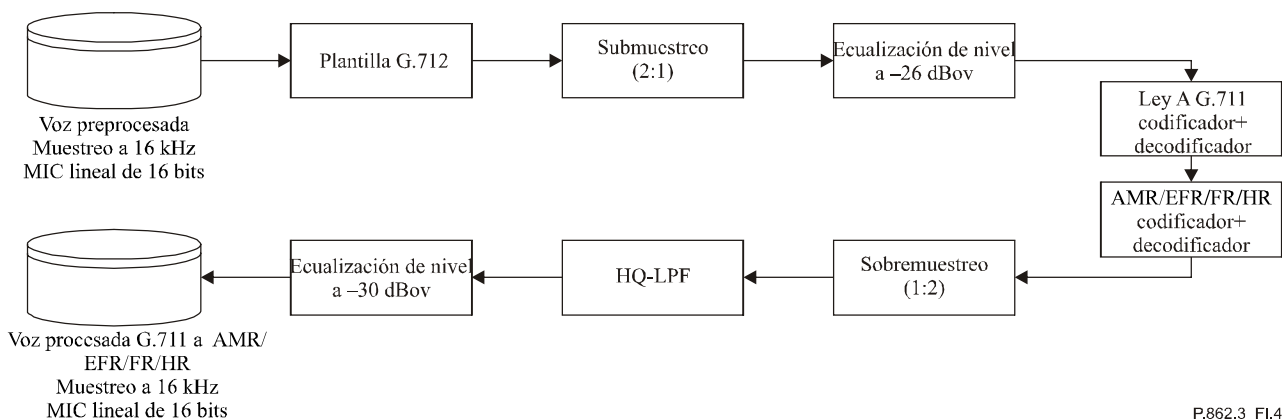


**Figura I.3/P.862.3 – Procesamiento de G.726**

<sup>13</sup> En los códecs GSM-AMR, GSM-EFR, GSM-FR, GSM-HR y ley A G.711, los valores de PESQ en bruto se calcularon a una velocidad de muestreo de 16 kHz.

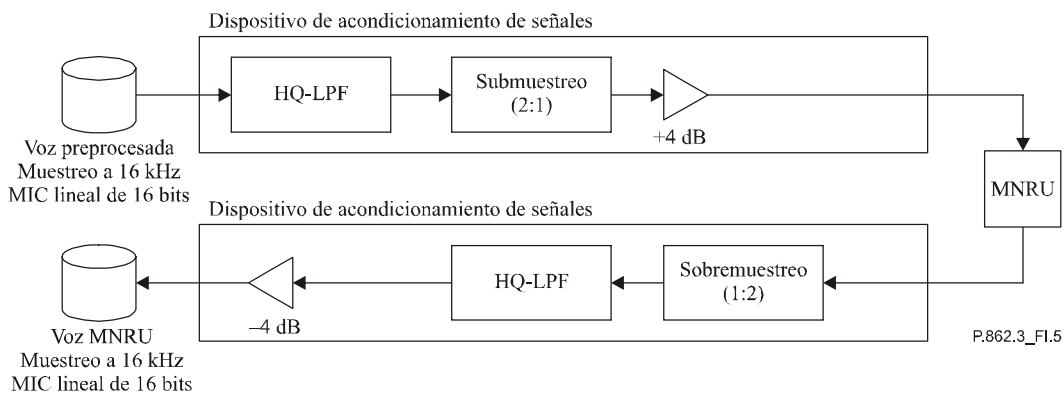


\* G.728, G.729, anexo A/G.729 y G.723.1



P.862.3\_FI.4

**Figura I.4/P.862.3 – Procesamiento de G.728, G.729, anexo A/G.729, G.723.1 y GSM-AMR/EFR/FR/HR**



P.862.3\_FI.5

**Figura I.5/P.862.3 – Procesamiento del MNRU**

A continuación se describe el procedimiento de procesamiento de las figuras I.1 a I.5.

## I.2 Preprocesamiento de la voz fuente

En la parte de preprocesamiento, la señal fuente se hace pasar por un filtro IRS modificado en el lado de emisión y su nivel vocal se ecualiza a -30 dBov como ilustra la figura I.1. Los ficheros de voz preprocesada son señales de entrada a un dispositivo de acondicionamiento de señal. Utilizando el software, facilitado en la Rec. UIT-T G.191, que en adelante denominaremos STL2000, la voz preprocesada puede obtenerse mediante las siguientes instrucciones:

```
$ filter -q mod IRS16 file.inp file.irs
$ sv56demo -q file.irs file.pre 256 1 0 -30
```

### I.3 Procesamiento de la G.711

La voz preprocesada se somete a una plantilla de acuerdo con la Rec. UIT-T G.712 y se submuestra. La voz submuestreada, se ecualiza a  $-26$  dBov y la salida, se convierte en la señal de entrada al codificador G.711. Las instrucciones de procesamiento son las siguientes:

```
$ filter -q -down PCM file.pre tmpfile1
$ sv56demo -q tmpfile1 g711.inp 256 1 0 -26 8000
```

El procesamiento de codificación y decodificación de ley  $\mu$  de la Rec. UIT-T G.711 se realiza mediante la siguiente instrucción STL2000:

```
$ g711demo u lili g711.inp g711.dec
```

La voz decodificada se sobremuestra de 8 a 16 kHz y su nivel se ecualiza a  $-30$  dBov. La voz codificada G.711 puede calcularse mediante las siguientes instrucciones STL2000:

```
$ filter -q -up HQ2 g711.dec tmpfile2
$ scaledemo -q -gain 0.63095 tmpfile2 g711.out
```

### I.4 Procesamiento de la G.726

En el procesamiento G.726, la salida del codificador de ley  $\mu$  G.711 pasa a ser la entrada al codificador G.726 y su salida, es la entrada al decodificador de ley  $\mu$  G.711. La voz codificada G.726 puede obtenerse mediante las instrucciones STL2000 siguientes:

```
$ filter -q -down PCM infile tmpfile1
$ sv56demo -q tmpfile1 tmpfile2 256 1 0 -26
$ g711demo u lilo tmpfile2 tmpfile3
$ g726demo u lolo {40/32/24/16} tmpfile3 tmpfile4
$ g711demo u loli tmpfile4 tmpfile5
$ filter -q -up HQ2 tmpfile5 tmpfile6
$ scaledemo -q -gain 0.63095 tmpfile6 outfile
```

### I.5 Procesamiento de G.728, G.729, anexo A/G.729 y G.723.1

En estos códecs, la señal de entrada del códec sometido a prueba es la salida del códec ley  $\mu$  G.711, es decir, *g711.dec* como se indicó en la figura I.2. Si las salidas del códec sometido a prueba se designan mediante {g728|g729|g729A|G7231}.dec, la voz codificada se obtiene mediante las instrucciones STL2000 siguientes:

```
$ filter -q -up HQ2 {g728|g729|g729A|g7231}.dec tmpfile2
$ scaledemo -q -gain 0.63095 tmpfile2 {g728|g729|g729A|g7231}.out
```

### I.6 Procesamiento de MNRU

La voz preprocesada se submuestra sin modificar la respuesta en frecuencia de la señal de entrada, su nivel se ecualiza a  $-26$  dBov, y su salida, pasa a ser la señal de entrada a la MNRU. La salida de la MNRU se sobremuestra de 8 a 16 kHz y su nivel se ecualiza a  $-30$  dBov. La voz MNRU  $Q$ -dB puede obtenerse mediante las siguientes instrucciones STL2000:

```
$ filter -q -down HQ2 infile tmpfile1
$ sv56demo -q tmpfile1 tmpfile2 256 1 0 -26 8000
$ mnrdemo tmpfile2 tmpfile3 128 1 0 Q
$ filter -q -up HQ2 tmpfile3 tmpfile4
$ scaldemo -q -gain 0.63095 tmpfile3 mnruQ.out
```

**Cuadro I.1/P.862.3 – Valores de referencia P.862.1  
para los códecs normalizados del UIT-T**

Idioma	Nombre del fichero	G.711		G.726				G.728	G.729	G.729A	G.723.1	
		Ley $\mu$	Ley A	16 kbit/s	24 kbit/s	32 kbit/s	40kbit/s				5,3 kbit/s	6,3 kbit/s
Inglés americano	Mujer 1 (0.00-7.97 s).wav	4,46	4,28	2,50	3,34	3,89	4,18	3,95	3,95	3,80	3,65	3,81
	Mujer 2 (0.00-8.06 s).wav	4,45	4,42	3,12	3,76	4,07	4,33	4,27	4,08	3,99	3,67	3,80
	Hombre 1 (0.00-8.44 s).wav	4,49	4,36	2,86	3,82	4,25	4,35	4,19	4,17	4,13	3,90	3,97
	Hombre 2 (0.00-7.96 s).wav	4,47	4,31	2,97	3,80	4,21	4,40	4,22	4,15	4,06	3,95	4,07
	<b>Promedio</b>	<b>4,47</b>	<b>4,34</b>	<b>2,86</b>	<b>3,69</b>	<b>4,11</b>	<b>4,32</b>	<b>4,16</b>	<b>4,09</b>	<b>4,00</b>	<b>3,80</b>	<b>3,92</b>
Chino	Mujer 1 (0.00-10.87 s).wav	4,46	4,48	2,38	3,42	4,21	4,34	3,86	3,72	3,65	3,10	3,33
	Mujer 1b (0.00-13.39 s).wav	4,42	4,43	2,29	3,34	4,07	4,26	3,98	3,80	3,75	3,26	3,49
	Mujer 2 (0.00-13.32 s).wav	4,50	4,50	2,26	3,29	4,02	4,37	4,06	3,88	3,75	3,33	3,53
	Mujer 2b (0.00-13.39 s).wav	4,50	4,50	2,38	3,30	4,03	4,33	4,03	3,87	3,72	3,39	3,59
	Hombre 1 (0.00-12.15 s).wav	4,44	4,48	2,64	3,58	4,23	4,28	4,19	3,83	3,75	3,36	3,51
	Hombre 1a (0.00-12.91 s).wav	4,52	4,51	2,84	3,77	4,21	4,37	4,18	4,06	4,00	3,65	3,89
	Hombre 2 (0.00-12.50 s).wav	4,49	4,48	2,74	3,74	4,30	4,44	4,18	3,99	3,89	3,62	3,78
	Hombre 2b (0.00-12.82 s).wav	4,50	4,40	2,89	3,90	4,29	4,43	4,16	3,95	3,89	3,35	3,55
<b>Promedio</b>	<b>4,48</b>	<b>4,47</b>	<b>2,55</b>	<b>3,55</b>	<b>4,18</b>	<b>4,35</b>	<b>4,08</b>	<b>3,89</b>	<b>3,80</b>	<b>3,38</b>	<b>3,59</b>	
Inglés	Mujer 1 (0.00-8.00 s).wav	4,50	4,49	2,69	3,37	3,88	4,21	3,89	3,72	3,58	3,42	3,59
	Mujer 2 (0.00-8.00 s).wav	4,48	4,46	2,81	3,44	3,92	4,24	4,00	3,91	3,80	3,60	3,67
	Hombre 1 (0.00-8.00 s).wav	4,50	4,45	3,03	3,53	3,96	3,99	4,08	3,88	3,88	3,67	3,81
	Hombre 2 (0.00-8.00 s).wav	4,51	4,48	2,94	3,79	4,24	4,34	4,05	3,73	3,60	3,64	3,83
	<b>Promedio</b>	<b>4,50</b>	<b>4,47</b>	<b>2,87</b>	<b>3,54</b>	<b>4,01</b>	<b>4,20</b>	<b>4,01</b>	<b>3,81</b>	<b>3,72</b>	<b>3,59</b>	<b>3,73</b>
Francés	Mujer 1 (0.00-10.04 s).wav	4,50	4,47	3,06	3,84	4,28	4,42	4,21	3,85	3,77	3,59	3,69
	Mujer 2 (0.00-10.04 s).wav	4,51	4,48	2,76	3,64	4,15	4,41	4,03	3,77	3,64	3,39	3,56
	Hombre 1 (0.00-12.18 s).wav	4,50	4,46	3,09	3,79	4,18	4,32	4,09	3,84	3,82	3,45	3,60
	Hombre 2 (0.00-10.04 s).wav	4,52	4,48	3,33	3,92	4,28	4,40	4,25	4,00	3,91	3,70	3,88
	<b>Promedio</b>	<b>4,51</b>	<b>4,47</b>	<b>3,06</b>	<b>3,80</b>	<b>4,22</b>	<b>4,39</b>	<b>4,15</b>	<b>3,87</b>	<b>3,79</b>	<b>3,54</b>	<b>3,69</b>
Alemán	Mujer 1 (0.00-8.00 s).wav	4,49	4,48	2,68	3,46	4,02	4,27	4,04	3,86	3,69	3,54	3,75
	Mujer 2 (0.00-8.00 s).wav	4,48	4,46	2,84	3,65	4,24	4,40	4,13	4,07	3,89	3,61	3,82
	Hombre 1 (0.00-8.00 s).wav	4,50	4,47	2,99	3,72	4,27	4,41	4,09	3,95	3,87	3,56	3,84
	Hombre 2 (0.00-8.00 s).wav	4,50	4,46	2,86	3,47	4,03	4,35	4,12	4,07	4,01	3,75	3,91
	<b>Promedio</b>	<b>4,50</b>	<b>4,47</b>	<b>2,84</b>	<b>3,58</b>	<b>4,14</b>	<b>4,36</b>	<b>4,09</b>	<b>3,99</b>	<b>3,87</b>	<b>3,62</b>	<b>3,83</b>
Italiano	Mujer 1 (0.00-20.60 s).wav	4,49	4,40	2,38	3,23	3,81	4,25	3,80	3,75	3,62	3,28	3,49
	Mujer 2 (0.00-21.78 s).wav	4,48	4,39	2,72	3,68	4,14	4,34	4,16	3,95	3,87	3,57	3,75
	Hombre 1 (0.00-18.13 s).wav	4,50	4,44	2,61	3,50	4,01	4,33	4,12	3,88	3,81	3,59	3,75
	Hombre 2 (0.00-20.86 s).wav	4,51	4,43	3,05	3,94	4,28	4,41	4,18	4,12	4,05	3,73	3,95
	<b>Promedio</b>	<b>4,49</b>	<b>4,41</b>	<b>2,69</b>	<b>3,60</b>	<b>4,07</b>	<b>4,33</b>	<b>4,07</b>	<b>3,93</b>	<b>3,84</b>	<b>3,55</b>	<b>3,74</b>
Japonés	Mujer 1 (0.00-7.60 s).wav	4,46	4,36	2,22	2,97	3,65	4,11	3,76	3,70	3,61	3,25	3,40
	Mujer 2 (0.00-7.31 s).wav	4,48	4,38	2,45	3,41	4,12	4,40	3,96	3,82	3,73	3,34	3,59
	Hombre 1 (0.00-7.13 s).wav	4,47	4,39	2,36	3,06	3,61	4,20	3,89	3,88	3,74	3,38	3,53
	Hombre 2 (0.00-7.45 s).wav	4,49	4,42	2,95	3,90	4,32	4,45	4,31	4,08	4,00	3,83	3,93
	<b>Promedio</b>	<b>4,47</b>	<b>4,39</b>	<b>2,49</b>	<b>3,35</b>	<b>3,95</b>	<b>4,30</b>	<b>4,00</b>	<b>3,87</b>	<b>3,78</b>	<b>3,46</b>	<b>3,62</b>
Español (latino)	Mujer 1 (0.00-8.00 s).wav	4,47	4,40	2,33	3,05	3,73	4,17	4,02	3,84	3,71	3,42	3,64
	Mujer 2 (0.00-8.00 s).wav	4,40	4,31	2,30	2,92	3,48	4,04	3,77	3,65	3,43	3,07	3,22
	Hombre 1 (0.00-8.00 s).wav	4,46	4,30	2,86	3,64	4,19	4,36	4,05	3,83	3,76	3,69	3,77
	Hombre 2 (0.00-8.00 s).wav	4,49	4,42	2,76	3,72	4,22	4,40	4,09	3,86	3,85	3,60	3,72
	<b>Promedio</b>	<b>4,46</b>	<b>4,36</b>	<b>2,56</b>	<b>3,34</b>	<b>3,93</b>	<b>4,25</b>	<b>3,99</b>	<b>3,80</b>	<b>3,69</b>	<b>3,45</b>	<b>3,60</b>

NOTA – Las casillas sombreadas indican muestras que no cumplen los requisitos de la Rec. UIT-T P.501.



**Cuadro I.2/P.862.3 – Los valores de referencia P.862.1 para el MNRU**

Idioma	Nombre del fichero	MNRU								
		5 dB	10 dB	15 dB	20 dB	25 dB	30 dB	35 dB	40 dB	45dB
Inglés americano	Mujer 1 (0.00-7.97 s).wav	1,80	2,39	3,03	3,68	4,15	4,37	4,47	4,50	4,52
	Mujer 2 (0.00-8.06 s).wav	2,09	2,67	3,24	3,80	4,16	4,35	4,41	4,44	4,44
	Hombre 1 (0.00-8.44 s).wav	1,93	2,58	3,30	4,01	4,32	4,46	4,51	4,52	4,53
	Hombre 2 (0.00-7.96 s).wav	1,99	2,65	3,34	3,90	4,23	4,39	4,45	4,47	4,48
	<b>Promedio</b>	<b>1,95</b>	<b>2,57</b>	<b>3,23</b>	<b>3,85</b>	<b>4,22</b>	<b>4,39</b>	<b>4,46</b>	<b>4,48</b>	<b>4,49</b>
Chino	Mujer 1 (0.00-10.87 s).wav	1,41	1,84	2,45	3,19	3,83	4,20	4,40	4,49	4,52
	Mujer 1b (0.00-13.39 s).wav	1,35	1,74	2,36	3,07	3,66	4,02	4,27	4,36	4,38
	Mujer 2 (0.00-13.32 s).wav	1,46	1,91	2,54	3,33	3,99	4,33	4,46	4,51	4,54
	Mujer 2b (0.00-13.39 s).wav	1,52	2,02	2,72	3,51	4,10	4,35	4,42	4,44	4,45
	Hombre 1 (0.00-12.15 s).wav	1,72	2,29	3,02	3,71	4,16	4,33	4,37	4,38	4,38
	Hombre 1a (0.00-12.91 s).wav	1,78	2,44	3,22	3,91	4,26	4,43	4,51	4,54	4,54
	Hombre 2 (0.00-12.50 s).wav	1,67	2,22	3,02	3,78	4,19	4,36	4,46	4,51	4,52
	Hombre 2b (0.00-12.82 s).wav	1,81	2,46	3,25	3,91	4,29	4,42	4,46	4,47	4,47
	<b>Promedio</b>	<b>1,57</b>	<b>2,10</b>	<b>2,82</b>	<b>3,56</b>	<b>4,08</b>	<b>4,31</b>	<b>4,42</b>	<b>4,47</b>	<b>4,48</b>
Inglés	Mujer 1 (0.00-8.00 s).wav	1,97	2,60	3,27	3,86	4,22	4,41	4,49	4,52	4,53
	Mujer 2 (0.00-8.00 s).wav	1,90	2,51	3,16	3,64	4,06	4,34	4,45	4,50	4,52
	Hombre 1 (0.00-8.00 s).wav	2,41	3,06	3,65	4,11	4,38	4,47	4,51	4,52	4,53
	Hombre 2 (0.00-8.00 s).wav	1,90	2,55	3,32	3,90	4,27	4,45	4,51	4,54	4,54
	<b>Promedio</b>	<b>2,03</b>	<b>2,68</b>	<b>3,35</b>	<b>3,89</b>	<b>4,24</b>	<b>4,42</b>	<b>4,49</b>	<b>4,52</b>	<b>4,53</b>
Francés	Mujer 1 (0.00-10.04 s).wav	1,93	2,55	3,22	3,82	4,23	4,42	4,50	4,53	4,54
	Mujer 2 (0.00-10.04 s).wav	1,74	2,31	3,02	3,70	4,15	4,40	4,50	4,53	4,54
	Hombre 1 (0.00-12.18 s).wav	2,07	2,76	3,47	3,96	4,23	4,41	4,49	4,53	4,54
	Hombre 2 (0.00-10.04 s).wav	2,36	3,14	3,88	4,30	4,46	4,51	4,53	4,54	4,54
	<b>Promedio</b>	<b>2,01</b>	<b>2,68</b>	<b>3,41</b>	<b>3,96</b>	<b>4,27</b>	<b>4,44</b>	<b>4,51</b>	<b>4,53</b>	<b>4,54</b>
Alemán	Mujer 1 (0.00-8.00 s).wav	1,70	2,28	3,00	3,65	4,19	4,41	4,49	4,53	4,54
	Mujer 2 (0.00-8.00 s).wav	1,74	2,26	2,94	3,64	4,13	4,33	4,47	4,51	4,53
	Hombre 1 (0.00-8.00 s).wav	1,88	2,47	3,17	3,75	4,20	4,42	4,51	4,53	4,54
	Hombre 2 (0.00-8.00 s).wav	2,09	2,71	3,40	3,97	4,33	4,46	4,51	4,53	4,54
	<b>Promedio</b>	<b>1,84</b>	<b>2,43</b>	<b>3,13</b>	<b>3,76</b>	<b>4,21</b>	<b>4,41</b>	<b>4,49</b>	<b>4,53</b>	<b>4,54</b>
Italiano	Mujer 1 (0.00-20.60 s).wav	1,52	2,01	2,69	3,44	4,01	4,33	4,46	4,51	4,53
	Mujer 2 (0.00-21.78 s).wav	1,73	2,25	2,92	3,58	4,12	4,38	4,47	4,51	4,53
	Hombre 1 (0.00-18.13 s).wav	1,79	2,43	3,24	3,94	4,33	4,46	4,51	4,53	4,54
	Hombre 2 (0.00-20.86 s).wav	2,09	2,86	3,65	4,16	4,40	4,49	4,52	4,53	4,54
	<b>Promedio</b>	<b>1,76</b>	<b>2,37</b>	<b>3,13</b>	<b>3,80</b>	<b>4,23</b>	<b>4,42</b>	<b>4,49</b>	<b>4,52</b>	<b>4,53</b>
Japonés	Mujer 1 (0.00-7.60 s).wav	1,49	1,94	2,57	3,32	3,92	4,27	4,43	4,50	4,52
	Mujer 2 (0.00-7.31 s).wav	1,48	1,91	2,51	3,27	3,99	4,35	4,48	4,52	4,53
	Hombre 1 (0.00-7.13 s).wav	1,55	2,02	2,65	3,39	4,05	4,36	4,47	4,52	4,53
	Hombre 2 (0.00-7.45 s).wav	1,74	2,27	3,01	3,77	4,24	4,45	4,51	4,53	4,54
	<b>Promedio</b>	<b>1,56</b>	<b>2,03</b>	<b>2,68</b>	<b>3,44</b>	<b>4,05</b>	<b>4,36</b>	<b>4,47</b>	<b>4,52</b>	<b>4,53</b>
Español (latino)	Mujer 1 (0.00-8.00 s).wav	1,49	1,95	2,51	3,16	3,81	4,24	4,43	4,50	4,53
	Mujer 2 (0.00-8.00 s).wav	1,61	2,08	2,62	3,19	3,81	4,19	4,40	4,49	4,52
	Hombre 1 (0.00-8.00 s).wav	2,02	2,70	3,53	4,14	4,39	4,48	4,51	4,52	4,53
	Hombre 2 (0.00-8.00 s).wav	1,65	2,24	3,04	3,73	4,23	4,44	4,51	4,53	4,54
	<b>Promedio</b>	<b>1,67</b>	<b>2,23</b>	<b>2,92</b>	<b>3,59</b>	<b>4,08</b>	<b>4,35</b>	<b>4,46</b>	<b>4,51</b>	<b>4,53</b>

NOTA – Las casillas sombreadas indican muestras que no cumplen los requisitos de la Rec. UIT-T P.501.

**Cuadro I.3/P/862.3 – Valores de referencia P.862 para los códecs normalizados GSM**

Idioma	Nombre del fichero	AMR								EFR	FR	HR
		12,2 kbit/s	10,2 kbit/s	7,95 kbit/s	7,4 kbit/s	6,7 kbit/s	5,9 kbit/s	5,15 kbit/s	4,75 kbit/s			
Inglés americano	Mujer 1 (0.00-7.97 s).wav	3,87	3,75	3,60	3,61	3,52	3,43	3,33	3,18	3,94	3,03	3,20
	Mujer 2 (0.00-8.06 s).wav	4,13	4,07	3,96	3,92	3,79	3,72	3,62	3,49	4,08	3,70	3,54
	Hombre 1 (0.00-8.44 s).wav	4,10	4,03	3,94	3,97	3,88	3,84	3,64	3,50	4,19	3,69	3,46
	Hombre 2 (0.00-7.96 s).wav	4,10	4,06	4,03	4,01	3,91	3,85	3,73	3,70	4,20	3,59	3,72
	<b>Promedio</b>	<b>4,05</b>	<b>3,98</b>	<b>3,88</b>	<b>3,88</b>	<b>3,78</b>	<b>3,71</b>	<b>3,58</b>	<b>3,47</b>	<b>4,10</b>	<b>3,51</b>	<b>3,48</b>
Chino	Mujer 1 (0.00-10.87 s).wav	3,94	3,81	3,46	3,50	3,27	3,15	2,99	3,00	3,98	3,18	2,94
	Mujer 1b (0.00-13.39 s).wav	3,97	3,82	3,52	3,45	3,36	3,22	3,00	2,95	4,04	3,02	2,95
	Mujer 2 (0.00-13.32 s).wav	4,08	3,99	3,71	3,62	3,53	3,38	3,21	3,07	4,12	3,07	3,01
	Mujer 2b (0.00-13.39 s).wav	4,08	4,01	3,74	3,74	3,62	3,40	3,21	3,13	4,10	3,03	3,04
	Hombre 1 (0.00-12.15 s).wav	3,94	3,84	3,66	3,68	3,54	3,45	3,22	3,22	4,06	3,50	3,26
	Hombre 1a (0.00-12.91 s).wav	4,23	4,14	3,95	3,92	3,78	3,65	3,34	3,35	4,20	3,65	3,36
	Hombre 2 (0.00-12.50 s).wav	4,07	3,93	3,78	3,77	3,62	3,46	3,22	3,21	4,15	3,72	3,30
	Hombre 2b (0.00-12.82 s).wav	4,16	4,15	3,94	3,94	3,80	3,64	3,40	3,27	4,23	3,64	3,44
	<b>Promedio</b>	<b>4,06</b>	<b>3,96</b>	<b>3,72</b>	<b>3,70</b>	<b>3,57</b>	<b>3,42</b>	<b>3,20</b>	<b>3,15</b>	<b>4,11</b>	<b>3,35</b>	<b>3,16</b>
Inglés	Mujer 1 (0.00-8.00 s).wav	4,00	3,82	3,63	3,62	3,42	3,31	3,21	3,11	3,99	3,27	3,07
	Mujer 2 (0.00-8.00 s).wav	3,81	3,78	3,65	3,62	3,56	3,49	3,36	3,34	3,78	3,31	3,28
	Hombre 1 (0.00-8.00 s).wav	4,01	3,88	3,75	3,67	3,46	3,57	3,23	3,01	4,03	3,54	3,37
	Hombre 2 (0.00-8.00 s).wav	4,06	3,83	3,73	3,68	3,53	3,48	3,18	2,98	4,10	3,75	3,49
	<b>Promedio</b>	<b>3,97</b>	<b>3,83</b>	<b>3,69</b>	<b>3,65</b>	<b>3,49</b>	<b>3,46</b>	<b>3,25</b>	<b>3,11</b>	<b>3,98</b>	<b>3,47</b>	<b>3,30</b>
Francés	Mujer 1 (0.00-10.04 s).wav	4,11	4,01	3,79	3,83	3,66	3,42	3,33	3,30	4,04	3,49	3,37
	Mujer 2 (0.00-10.04 s).wav	3,91	3,85	3,57	3,54	3,40	3,31	3,22	2,98	3,83	3,19	3,24
	Hombre 1 (0.00-12.18 s).wav	4,00	3,88	3,71	3,72	3,54	3,37	3,23	3,14	4,07	3,49	3,32
	Hombre 2 (0.00-10.04 s).wav	4,11	4,05	3,85	3,91	3,75	3,57	3,37	3,28	4,17	3,84	3,30
	<b>Promedio</b>	<b>4,03</b>	<b>3,95</b>	<b>3,73</b>	<b>3,75</b>	<b>3,59</b>	<b>3,42</b>	<b>3,29</b>	<b>3,18</b>	<b>4,03</b>	<b>3,50</b>	<b>3,31</b>
Alemán	Mujer 1 (0.00-8.00 s).wav	4,08	3,98	3,65	3,60	3,54	3,36	3,15	3,06	4,06	3,40	3,20
	Mujer 2 (0.00-8.00 s).wav	4,21	4,14	3,93	3,88	3,76	3,63	3,53	3,47	4,17	3,54	3,37
	Hombre 1 (0.00-8.00 s).wav	4,12	4,08	3,90	3,88	3,78	3,66	3,55	3,50	4,19	3,82	3,43
	Hombre 2 (0.00-8.00 s).wav	4,17	4,07	3,97	3,92	3,77	3,69	3,65	3,56	4,21	3,70	3,38
	<b>Promedio</b>	<b>4,14</b>	<b>4,07</b>	<b>3,86</b>	<b>3,82</b>	<b>3,72</b>	<b>3,58</b>	<b>3,47</b>	<b>3,40</b>	<b>4,16</b>	<b>3,62</b>	<b>3,35</b>
Italiano	Mujer 1 (0.00-20.60 s).wav	3,80	3,67	3,51	3,41	3,34	3,24	3,12	2,92	3,81	2,83	3,00
	Mujer 2 (0.00-21.78 s).wav	4,09	4,04	3,88	3,86	3,74	3,61	3,40	3,27	4,14	3,29	3,32
	Hombre 1 (0.00-18.13 s).wav	4,03	3,95	3,82	3,78	3,64	3,49	3,32	3,20	4,13	3,37	3,17
	Hombre 2 (0.00-20.86 s).wav	4,23	4,15	4,00	4,07	3,89	3,84	3,60	3,47	4,27	3,58	3,47
	<b>Promedio</b>	<b>4,04</b>	<b>3,95</b>	<b>3,80</b>	<b>3,78</b>	<b>3,65</b>	<b>3,55</b>	<b>3,36</b>	<b>3,22</b>	<b>4,08</b>	<b>3,27</b>	<b>3,24</b>
Japonés	Mujer 1 (0.00-7.60 s).wav	3,89	3,75	3,53	3,41	3,40	3,28	3,08	3,10	3,87	2,92	2,92
	Mujer 2 (0.00-7.31 s).wav	3,92	3,85	3,54	3,62	3,42	3,27	3,20	3,08	3,99	2,82	3,09
	Hombre 1 (0.00-7.13 s).wav	3,87	3,81	3,59	3,50	3,42	3,28	3,19	3,12	3,91	2,89	2,91
	Hombre 2 (0.00-7.45 s).wav	4,18	4,08	3,92	3,94	3,84	3,73	3,59	3,44	4,21	3,63	3,55
	<b>Promedio</b>	<b>3,97</b>	<b>3,87</b>	<b>3,64</b>	<b>3,61</b>	<b>3,52</b>	<b>3,39</b>	<b>3,27</b>	<b>3,19</b>	<b>4,00</b>	<b>3,06</b>	<b>3,12</b>
Español (latino)	Mujer 1 (0.00-8.00 s).wav	4,03	3,96	3,63	3,68	3,50	3,30	3,23	3,16	3,96	3,03	3,20
	Mujer 2 (0.00-8.00 s).wav	3,63	3,48	3,23	3,26	3,05	3,03	2,88	2,80	3,73	2,68	2,72
	Hombre 1 (0.00-8.00 s).wav	3,98	3,61	3,67	3,49	3,51	3,38	3,23	3,15	4,15	3,44	3,19
	Hombre 2 (0.00-8.00 s).wav	4,04	3,87	3,71	3,58	3,52	3,28	3,11	3,14	4,09	3,48	3,36
	<b>Promedio</b>	<b>3,92</b>	<b>3,73</b>	<b>3,56</b>	<b>3,50</b>	<b>3,40</b>	<b>3,25</b>	<b>3,11</b>	<b>3,06</b>	<b>3,99</b>	<b>3,16</b>	<b>3,12</b>

NOTA – Las casillas sombreadas indican muestras que no cumplen los requisitos de la Rec. UIT-T P.501.

## Apéndice II

### Bases de datos de pruebas para P.862/P.862.1

Las bases de datos de pruebas contienen condiciones de red que corresponden a aplicaciones fijas, inalámbricas y VoIP. En las pruebas y análisis P.862 y P.862.1 se ha incluido un amplio conjunto de códecs utilizados en diferentes tecnologías (tales como GSM-FR, GSM-EFR, GSM-AMR, CDMA-EVRC, IS136-ACELP, G.711, G.726, G.728, G.729, JDC-HR).

En el cuadro II.1 se resume el contenido de todas las bases de datos con las que se ha validado la PESQ (P.862 y P.862.1).

**Cuadro II.1/P.862.3 – Resumen del tipo y contenido de las bases de datos**

Descripción	Condiciones
Caracterización a 8 kbit/s interfuncionamiento con normas, Supl. P 23, exp 1	cod./decod. + transc.; códecs: G.711, G.726, G.728, G.729, Is-54, GSM-FR, JDC-HR
Caracterización a 8 kbit/s errores y ruido en el canal, Supl. P 23, exp 3	cod./decod. + trans + Patrones de error + BGN, códecs: G.729
Redes inalámbricas con tráfico real	<ul style="list-style-type: none"> <li>– Redes inalámbricas: IS-136, CDMA, iDEN, AMPS; GSM-EE.UU., GSM-Europa;</li> <li>– Códecs: IS-54, sACELP a 8 kbit/s, QCELP 13 kbit/s, GSM-FR&amp;EFR, CDMA- EVRC</li> </ul>
Códecs, patrones de error, transcodificaciones, ruido	<ul style="list-style-type: none"> <li>– cod./decod. + Patrones de error + BGN; códecs: G.711, G.726, G.728, G.729, GSM-FR</li> <li>– cod./decod. + Patrones de error (niveles C/I)+BGN; códecs: G.711, G.723 + transcodificaciones, ACELP a 8 kbit/s, EVRC, GSM-EFR&amp;FR</li> <li>– cod./decod. + trans + Patrones de error ("bad" &amp; "good"cond) + BGN, ATM/RDSI/POTS; códecs: G.729, G.728, GSM-FR, GSM-HR</li> </ul>
Prueba de ruido de fondo, redes GSM y fijas	<ul style="list-style-type: none"> <li>– cod./decod. + VAD + VQE (NR); códecs: GSM-FR, G.729</li> <li>– cod./decod. + Patrones de error; códecs: GSM-FR, GSM-AMR-HR</li> </ul>
DTX, prueba de VAD y borrado de trama/ráfaga, AMR, redes GSM y fijas	cod./decod. + Patrones de error; códecs: GSM-FR, G.726, G.728, G.729, AMR
AMR + patrones de error en canales GSM esperados y no esperados	cod./decod.+ Patrones de error; códecs: AMR475, AMR590, AMR740, AMR122
AMR + patrones de error con protección de error igual y desigual	cod./decod.+ Patrones de error; códecs: AMR515, AMR740, AMR102, AMR122, AMR475, AMR590, AMR670, AMR795
Patrón de error y pérdida de paquetes en VoIP; ruido	cod./decod. + trans + Patrones de error + Pérdida de paquetes + BGN
Prueba de casos de pérdida de paquetes en VoIP	VoIP: PktInsert, PktDelete, PktMute
Códecs VoIP	VoIP: G.723.1, G.728, G.711, G.729, G.726, PDC-HR
Condiciones VoIP	PLC y gestión de la memoria tampón de fluctuación de fase. (G.711, G.729, G.723 @6.3 &5.3), longitud de paquetes variable
Condiciones VoIP reales	VoIP: PABX Internet, Pasarelas IP

## Apéndice III

### Informe sobre las mediciones P.862/P.862.1

Las mediciones P.862/P.862.1 deben notificarse según la precisión del algoritmo presentada en 10.3 (cuadros 1 y 2).

#### III.1 Informe e interpretación de los resultados PESQ promediados

Como se indicó en 10.1, en el caso de condiciones de prueba de red simuladas y controladas se recomienda realizar promedios para cada condición, con al menos cuatro hablantes.

Se prevé que las notas P.862.1 promediadas tengan un error de predicción máximo del  $\pm 95\%$  del límite superior del intervalo de confianza.

En el caso de pruebas de campo de red, hay situaciones en las que conviene evaluar aproximadamente la calidad vocal de la red dentro de una cierta zona de pruebas, y/o durante una ventana de tiempos. En este caso, las notas P.862.1 se promedian a lo largo de la ruta de pruebas y/o la ventana de tiempos. Se prevé que las notas P.862.1 promediadas tengan un error de predicción máximo del  $\pm 95\%$  del límite superior del intervalo de confianza. Por consiguiente, suponiendo por ejemplo que la red sometida a pruebas sea una red CDMA, se prevé que el error de predicción del promedio para las mediciones P.862.1 promediadas será menor o igual que una MOS de 0,462.

Debe señalarse que para los dos casos indicados anteriormente, a saber, condiciones simuladas y reales, existe un riesgo del 5% de que el error de medición sea superior al 95% del límite superior del error de predicción. Además, según se dijo en 10.2, los casos de medición tienen que ser similares a los indicados en el cuadro 1.

#### III.2 Informe e interpretación de los resultados de las mediciones individuales de la PESQ

Como se indicó en 10.1, en el caso de pruebas de red con tráfico real, es necesario realizar una evaluación de la calidad para cada muestra, debido a que los canales de transmisión varían con el tiempo de forma incontrolada.

Se recomiendan dos procedimientos de elaboración de informes, dependiendo el tipo de mediciones de PESQ, que se utilizan para analizar la calidad vocal de la red.

El primer tipo es la nota P.862.1 promediada, que se describe en III.1 y ofrece una estimación aproximada de la calidad vocal dentro de una cierta zona o durante una ventana de tiempos.

El otro tipo de nota es la medición P.862.1, cuando resulte necesario resolver problemas de la red.

En el caso de mediciones individuales, se recomienda calcular un histograma de las mediciones P.862.1 junto con la escala de notas MOS. Como procedimiento de análisis en este caso, se recomienda tomar un umbral de calidad vocal subjetiva e imponer que un porcentaje de notas mínimo obligatorio supere dicho umbral subjetivo o, lo que es equivalente, una probabilidad mínima obligatoria de que las mediciones P.862.1 en la red sometida a prueba superen el umbral subjetivo. De este modo, la calidad vocal basada en las notas P.862.1 individuales se representa mediante la probabilidad o porcentaje de notas que superan un determinado umbral de calidad vocal subjetiva.

Al calcular esta probabilidad o porcentaje, hay dos tipos de errores que afectan a los resultados. El primero se produce al calcular la probabilidad o el porcentaje. El otro se debe a la distribución de errores residuales P.862/P.862.1. Debido al error de las mediciones P.862.1, no todos los puntos que según las mediciones están por encima (y/o por debajo) de un determinado umbral MOS se encuentran en realidad en ese punto. Según se dijo en 10.4, la distribución de errores residuales (véase el cuadro 2) muestra que cuanto menor sea el error absoluto más probabilidad hay de errores

absolutos mayores. Por consiguiente, las notas P.862.1 que se encuentren más cerca del umbral MOS tienen más probabilidad de caer al otro lado del umbral de lo que en principio se obtiene en las mediciones. Por ejemplo, considérese un umbral de MOS=3 y una red de pruebas CDMA. Las notas P.862.1 entre 3 y 3,1 tienen una probabilidad del 40,44% de estar por debajo del umbral. En cambio, la probabilidad de que las notas P.862.1 entre 3,1 y 3,2 estén por debajo del umbral es de sólo el 30,04%. Análogamente, las notas entre 2,9 y 3 tienen una probabilidad de estar por encima del umbral de MOS = 3 del 40,44%, mientras que la probabilidad de que las notas estén entre 2,8 y 2,9 es de sólo un 30,04%.

Para calcular el error de medición P.862.1 cuando se utilizan notas individuales se recomienda el método que se describe a continuación:

El error de medición P.862.1 debido al error que se produce al evaluar la probabilidad de que las notas P.862.1 rebasen un determinado umbral impuesto de MOS se define como la desviación típica de una distribución binomial, cuya probabilidad de ocurrencia viene dada por:

$$p = \frac{n}{N}$$

siendo n el número de notas que rebasan el umbral y N el número total de mediciones P.862.1. El error de evaluación de la probabilidad, Error1, viene dado por:

$$Error1 = \pm \sqrt{\frac{p \times (1-p)}{N}}$$

y se caracteriza por el 95% del intervalo de confianza:

$$\pm z_{\alpha} \times Error1$$

donde:

$z_{\alpha} = 2$  representa el cuantil gaussiano para una probabilidad de 95% (la distribución normal describe bastante bien el comportamiento de cada una de las notas PESQ).

Se recomienda que el error de medición P.862/P.862.1 debido al error residual del algoritmo PESQ Error2 se calcule como la desviación típica de la distribución binomial que describe el error residual indicado en el cuadro 2.

Cuando se utilicen las notas P.862/P.862.1 individuales, se recomienda calcular el error de medición total de la calidad vocal como la raíz cuadrada de la suma de los cuadrados de los dos errores cuadráticos mencionados antes, esto es Error1 y Error2.

## Apéndice IV

### Método de calibración de las interfaces de marca

NOTA – El método que aquí se describe se utiliza en situaciones en las que no se conocen con exactitud los niveles de señal necesarios y, por lo tanto, no pueden aplicarse los métodos recomendados en la cláusula 9. No debe esperarse obtener la máxima precisión con este método. También es importante que todos los ajustes de nivel se realicen en el dominio analógico o, que se utilice una longitud de palabra suficiente en el dominio digital (de al menos 24 bits).

#### IV.1 Calibración del nivel en emisión (extremo próximo) del equipo de pruebas

Ajústese el nivel de salida del equipo de pruebas a un nivel que esté aproximadamente en la mitad de la gama de funcionamiento de cualquier AGC potencial. Para ello se mide el nivel de señal en el terminal del extremo distante y se varía la atenuación del nivel de señal en el extremo próximo a la mitad de la gama para la cual el nivel en el extremo distante permanece constante. Otra posibilidad es utilizar un teléfono en el extremo distante y ajustar el nivel en el extremo próximo a la mitad de la gama para la que se obtiene un nivel en escucha confortable en el extremo distante. El segundo método también es preferible cuando no se dispone de AGC.

#### IV.2 Calibración del nivel en recepción (extremo distante) del equipo de pruebas

El nivel en el extremo distante debe ajustarse de modo que la atenuación entre el fichero registrado y el fichero de referencia esté cerca de 0 dB.

## BIBLIOGRAFÍA

- [B.1] Objective quality evaluation based on ITU-T Rec. P.862 by using long reference speech (NTT), COM12-D008, enero de 2005.
- [B.2] Objective quality measurement using artificial voice signals (NTT), COM12-D145, septiembre de 2003.
- [B.3] Addition of noise floor to reference speech used in ITU-T Rec. P.862, COM12-D011, enero de 2005.
- [B.4] ANSI/TIA-127-A-2004 (2004), *Enhanced Variable Rate Codec Speech Service Option 3 for Wideband Spread Spectrum Digital Systems*.







## **SERIES DE RECOMENDACIONES DEL UIT-T**

Serie A	Organización del trabajo del UIT-T
Serie D	Principios generales de tarificación
Serie E	Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión, sistemas y redes digitales
Serie H	Sistemas audiovisuales y multimedios
Serie I	Red digital de servicios integrados
Serie J	Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedios
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	Gestión de las telecomunicaciones, incluida la RGT y el mantenimiento de redes
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica, instalaciones telefónicas y redes locales
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Terminales para servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos, comunicaciones de sistemas abiertos y seguridad
Serie Y	Infraestructura mundial de la información, aspectos del protocolo Internet y Redes de la próxima generación
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación