



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

**UIT-T**

SECTOR DE NORMALIZACIÓN=  
DE LAS TELECOMUNICACIONES  
DE LA UIT

**P.502**

(05/2000)

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

Aparatos para mediciones objetivas

---

**Métodos de pruebas objetivas para los sistemas  
de comunicación vocal con señales de prueba  
complejas**

Recomendación UIT-T P.502

(Anteriormente Recomendación del CCITT)

---

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	Serie	P.10
Líneas y aparatos de abonado	Serie	P.30 P.300
Patrones de transmisión	Serie	P.40
<b>Aparatos para mediciones objetivas</b>	<b>Serie</b>	<b>P.50</b>
	<b>Serie</b>	<b>P.500</b>
Medidas electroacústicas objetivas	Serie	P.60
Medidas relativas a la sonoridad vocal	Serie	P.70
Métodos de evaluación objetiva y subjetiva de la calidad	Serie	P.80 P.800
Calidad audiovisual en servicios multimedios	Serie	P.900

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

## **Recomendación UIT-T P.502**

### **Métodos de pruebas objetivas para los sistemas de comunicación vocal con señales de prueba complejas**

#### **Resumen**

Esta Recomendación UIT-T describe métodos y procedimientos de evaluación de terminales complejos, componentes de red y sistemas de transmisión. Los métodos de prueba utilizan principalmente las señales descritas en las Recomendaciones UIT-T P.50, P.59 y P.501. Se describen las posibles repercusiones sobre la percepción subjetiva de la calidad de la voz, para diversas implementaciones técnicas y en distintas situaciones, explicándose los procedimientos de medida pertinentes.

#### **Orígenes**

La Recomendación UIT-T P.502, preparada por la Comisión de Estudio 12 (1997-2000) del UIT-T, fue aprobada por el procedimiento de la Resolución 1 de la CMNT el 18 de mayo de 2000.

#### **Palabras clave**

Calidad vocal, habla simultánea, métodos de análisis, monólogo.

## 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 Conferencia Mundial de Normalización de las Telecomunicaciones (CMNT), 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 CMNT.

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.

## 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 2001

Es propiedad. Ninguna parte de esta publicación puede reproducirse o utilizarse, de ninguna forma o por ningún medio, sea éste electrónico o mecánico, de fotocopia o de microfilm, sin previa autorización escrita por parte de la UIT.

## ÍNDICE

		Página
1	Alcance .....	1
2	Referencias.....	1
3	Definiciones y abreviaturas.....	2
4	Característica de convergencia de los compensadores de eco .....	3
4.1	Percepción subjetiva de la degradación de la calidad vocal .....	3
4.2	Parámetros objetivos relacionados en la modalidad de monólogo .....	4
4.3	Métodos de análisis.....	4
4.3.1	Método de prueba del tiempo de convergencia ( $T_c$ ).....	4
4.3.2	Atenuación de retorno del eco, ponderada en el tiempo ( $ERL_{tst}$ ) – monólogo .....	5
5	Evaluación de la calidad vocal en habla simultánea.....	6
5.1	Percepción subjetiva de la degradación de la calidad vocal .....	6
5.2	Parámetros objetivos relacionados.....	6
5.3	Métodos de análisis.....	7
5.3.1	Método de habla simultánea con señal fuente compuesta (CSS) .....	7
5.3.2	Pruebas en habla simultánea con secuencias combinadas en paralelo .....	10
6	Características de la comprensión-expansión y del AGC .....	20
6.1	Percepción subjetiva de la degradación de la calidad vocal .....	20
6.2	Parámetros objetivos relacionados.....	21
6.3	Métodos de análisis.....	21
7	Calidad de transmisión del ruido de fondo .....	23
7.1	Percepción subjetiva de la degradación de la calidad .....	23
7.2	Parámetros objetivos relacionados.....	23
7.3	Métodos de análisis.....	23
8	Características de conmutación.....	25
8.1	Percepción subjetiva de la degradación de la calidad vocal .....	25
8.2	Parámetros objetivos relacionados.....	25
8.3	Métodos de análisis.....	26
8.3.1	El intervalo de atenuación y las características de conmutación .....	26
8.3.2	El nivel umbral y el tiempo de establecimiento (para el nivel de activación mínimo) .....	27
8.3.3	Tiempo de retención (bloqueo).....	28
8.3.4	Nivel umbral y tiempo de conmutación de RCV a SND (SND a RCV) .....	29

	<b>Página</b>
8.3.5 Características de conmutación en presencia de ruido de fondo .....	30
Anexo A – Explicación detallada de la metodología de prueba de la $ERL_t$ ponderada en el tiempo .....	31
A.1 Algoritmo de la atenuación de retorno del eco .....	31
A.1.1 Atenuación de retorno del eco, ponderada en el tiempo ( $ERL_t$ ).....	31
A.1.2 Audibilidad del eco modelado .....	32
A.1.3 Expresión de los resultados de la $ERL_t$ .....	33
A.1.4 Algoritmo de prueba de la $ERL_t$ .....	34
Anexo B – Filtros de medición en habla simultánea con el método A .....	38
Anexo C – Descripción de la secuencia de acondicionamiento .....	39
C.1 Acondicionamiento del compensador antes del habla simultánea .....	39
C.1.1 Máscaras de actividad para el acondicionamiento en habla simultánea .....	39
C.1.2 Sincronización de las máscaras de actividad para el acondicionamiento en habla simultánea .....	41
C.1.3 Compensación de los filtros de la medición .....	41
Apéndice I – Referencias bibliográficas .....	41
Apéndice II – Ejemplos de evaluaciones .....	42
II.1 Ejemplos de las evaluaciones definidas en la cláusula 5 .....	42
II.1.1 Respuesta en frecuencia durante el habla simultánea .....	42
II.1.2 Variaciones de nivel durante el habla simultánea.....	50
II.1.3 Conmutación durante el habla simultánea.....	52

## **Introducción**

Esta Recomendación UIT-T describe métodos y procedimientos de evaluación de terminales complejos, componentes de red y sistemas de transmisión. Se explican distintos métodos de prueba en función de los diversos parámetros y sistemas a medir. Los métodos de prueba utilizan principalmente las señales descritas en las Recomendaciones UIT-T P.50, P.59 y P.501. Se describen las posibles repercusiones sobre la calidad subjetiva de la voz percibida, para diversas implementaciones técnicas y en distintas situaciones, explicándose los procedimientos de medida pertinentes.

## Recomendación UIT-T P.502

### Métodos de pruebas objetivas para los sistemas de comunicación vocal con señales de prueba complejas

#### 1 Alcance

Esta Recomendación UIT-T pretende definir los métodos de prueba que pueden emplearse para evaluar elementos específicos que influyan en la transmisión de calidad vocal de los terminales o de los sistemas de transmisión vocal. Los métodos descritos en la presente Recomendación UIT-T utilizan las señales de prueba definidas en las Recomendaciones UIT-T P.50, P.59 y P.501.

Esta Recomendación UIT-T presenta una serie de métodos de investigación de ciertos parámetros considerados importantes para la evaluación de los sistemas de comunicación vocal. Los parámetros de calidad de funcionamiento se califican por la percepción subjetiva de la degradación vocal y por los parámetros objetivos relacionados. Se presentan métodos de análisis para cada uno de los parámetros.

#### 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.

- Recomendación UIT-T G.122 (1993), *Influencia de los sistemas nacionales en la estabilidad y el eco para la persona que habla en las conexiones internacionales.*
- Recomendación UIT-T G.168 (2000), *Compensadores de eco de redes digitales.*
- Recomendación UIT-T P.10 (1998), *Vocabulario de términos sobre calidad de transmisión telefónica y aparatos telefónicos.*
- Recomendación UIT-T P.50 (1999), *Voces artificiales.*
- Recomendación UIT-T P.51 (1996), *Boca artificial.*
- Recomendación UIT-T P.56 (1993), *Medición objetiva del nivel vocal activo.*
- Recomendación UIT-T P.57 (1996), *Oídos artificiales.*
- Recomendación UIT-T P.58 (1996), *Simulador de cabeza y torso para telefonometría.*
- Recomendación UIT-T P.59 (1993), *Habla conversacional artificial.*
- Recomendación UIT-T P.340 (2000), *Características de transmisión de los aparatos telefónicos manos libres.*
- Recomendación UIT-T P.501 (2000), *Señales de prueba para utilización en telefonometría.*
- Recomendación UIT-T P.581 (2000), *Uso del simulador de cabeza y torso para pruebas de terminales manos libres.*
- Recomendación UIT-T P.800 (1996), *Métodos de determinación subjetiva de la calidad de transmisión.*

- Recomendación UIT-T P.810 (1996), *Aparato de referencia para ruido modulado*.
- Recomendación UIT-T P.830 (1996), *Evaluación subjetiva de la calidad de funcionamiento de los códecs digitales de banda telefónica y de banda ancha*.
- *Manual sobre telefonometría* de la UIT-T, 2ª edición; Ginebra 1992.
- CEI 60651 (1997), *Sound Level Meters*.

### 3 Definiciones y abreviaturas

En esta Recomendación UIT-T se definen los términos siguientes.

**características del control automático de ganancia:** Características de los sistemas de control automático de ganancia.

**intervalo de atenuaciones ( $a_H$ , *attenuation range*):** Intervalo de las atenuaciones (en dB) introducidas en el sentido de emisión o recepción de un terminal o sistema.

**Atenuación vocal emitida en habla simultánea ( $A_{sdt}$ , *send speech attenuation during double talk*)**

**Atenuación vocal recibida en habla simultánea ( $A_{rdt}$ , *received speech attenuation during double talk*)**

**tiempo de ataque:** Tiempo necesario para la total activación de un trayecto de la transmisión (por parte del compresor-expansor).

**factor de cresta:** Relación del valor de cresta al valor medio cuadrático de una señal.

**compresión-expansión:** Atenuación/amplificación de una señal en función de su nivel.

**señal fuente compuesta (CSS, *composite source signal*):** Señal compuesta en el tiempo por distintos elementos de señal.

**mejora de la atenuación de retorno del eco (ERLE, *echo return loss enhancement*):** Medición para determinar la mejora percibida de la perturbación de las señales de eco.

**atenuación de retorno del eco (ERL, *echo return loss*):** Medición para determinar la perturbación percibida de las señales de eco.

**atenuación de retorno del eco, habla simultánea ( $ERL_{dt}$ , *echo return loss, double talk*):** Medición para determinar la perturbación percibida de las señales de eco en condiciones de habla simultánea.

**atenuación de retorno del eco ponderada en el tiempo – monólogo ( $ERL_{tst}$ , *echo return loss, temporally weighted, single talk*):** Medición para determinar la perturbación percibida de las señales de eco en condiciones de monología, teniendo en cuenta ciertos efectos psicoacústicos.

**atenuación de retorno del eco ponderada en el tiempo – habla simultánea ( $ERL_{tdt}$ , *echo return loss, temporally weighted, double talk*):** Medición para determinar la perturbación percibida de las señales de eco en condiciones de habla simultánea, teniendo en cuenta ciertos efectos psicoacústicos.

**transformada rápida de Fourier (FFT, *fast Fourier transform*).**

**proceso del modelo vocal de Markov (MSMP, *markov speech model process*).**

**procesador no lineal (NLP, *non-linear processor*):** Procesador que se suele utilizar en los compensadores de eco para eliminar el eco residual.

**secuencia de pseudoruido (secuencia PN), (*Pn-pseudo noise sequence*):** Ruido pseudoaleatorio de frecuencias definidas, resultado de la transformada inversa de Fourier de un espectro de frecuencias predefinido.

**RCV:** Sentido de recepción.

**tiempo de liberación:** Tiempo necesario para la total desactivación de un trayecto de la transmisión (por parte del compresor-expansor).

**entrada en recepción ( $R_{in}$ , *receive input*):** Punto de acceso (eléctrico) de recepción del dispositivo objeto de la prueba.

**SND:** Sentido de emisión.

**salida en emisión ( $S_{out}$ , *send output*):** Punto de acceso (eléctrico) de emisión del dispositivo objeto de la prueba.

**atenuación por acoplamiento del terminal (TCL, *terminal coupling loss*):** Atenuación del eco de un terminal medida entre  $R_{in}$  y  $S_{out}$ , que comprende el SLR y el RLR.

**tiempo de convergencia ( $T_c$ , *convergence time*)**

Véase 4.3.1.

**tiempo de retención (bloqueo) ( $T_H$ , *hang-over time*)**

Véase la Recomendación UIT-T P.340.

**tiempo inicial de convergencia ( $T_{ic}$ , *initial convergence time*)**

Véase tiempo de convergencia.

**tiempo de establecimiento ( $T_R$ , *build-up time*)**

Véase la Recomendación UIT-T P.340.

**tiempo de conmutación ( $T_S$ , *switching time*)**

Véase la Recomendación UIT-T P.340.

**nivel umbral ( $V_{TH}$ , *threshold level*)**

Véase la Recomendación UIT-T P.340.

## **4 Característica de convergencia de los compensadores de eco**

Esta cláusula define la característica de convergencia de los compensadores de eco. Se definen los métodos de evaluación de los efectos subjetivos de los distintos parámetros de los compensadores de eco y se plantean métodos objetivos para definir dichos parámetros.

### **4.1 Percepción subjetiva de la degradación de la calidad vocal**

La profundidad de convergencia, o mejora de la atenuación de retorno del eco (ERLE, *echo return loss enhancement*), define la aptitud de un compensador de eco para eliminar las señales devueltas en sentido opuesto al de transmisión en un trayecto de eco. Puede tratarse de eco acústico en el caso de un teléfono manos libres, o eco de híbrido en el caso de una conversión de dos a cuatro hilos. Si la ERLE es pequeña las señales del eco residual resultarán más audibles.

El tiempo de convergencia define la rapidez del compensador de eco para alcanzar un estado estable cuando la señal de retorno del eco residual está suficientemente atenuada sin atenuación de inserción en ninguno de los trayectos de la transmisión vocal. Se trata pues del tiempo necesario para que la atenuación alcance un valor que se diferencie menos de 3 dB de la ERL, *y/o la pérdida de [25] dB*. En los compensadores de eco la convergencia debe ser rápida para evitar que el eco llegue al interlocutor al principio de la conversación.

Pueden generarse ráfagas de eco en condiciones tales que el compensador de eco tenga dificultades para converger sobre un cierto trayecto de eco. La degradación subjetiva es función del nivel y longitud de las ráfagas de eco, de su proximidad, de su frecuencia y del retardo en el trayecto del eco.

## **4.2 Parámetros objetivos relacionados en la modalidad de monólogo**

La calidad de las características de control del eco viene definida por los siguientes parámetros:

- La atenuación del retorno del eco como función del tiempo, definida como atenuación de retorno del eco (ERL).
- La atenuación de retorno del eco ponderada en el tiempo ( $ERL_t$ ).
- El tiempo necesario para la convergencia del AEC, definido como tiempo de convergencia ( $T_c$ ).

En situación de habla simultánea hay que considerar además los siguientes parámetros:

- Calidad de funcionamiento dúplex en función del tiempo.
- Respuesta en funcionamiento dúplex para los anteriores parámetros.
- Respuesta de la atenuación a los problemas ambientales o de la red.

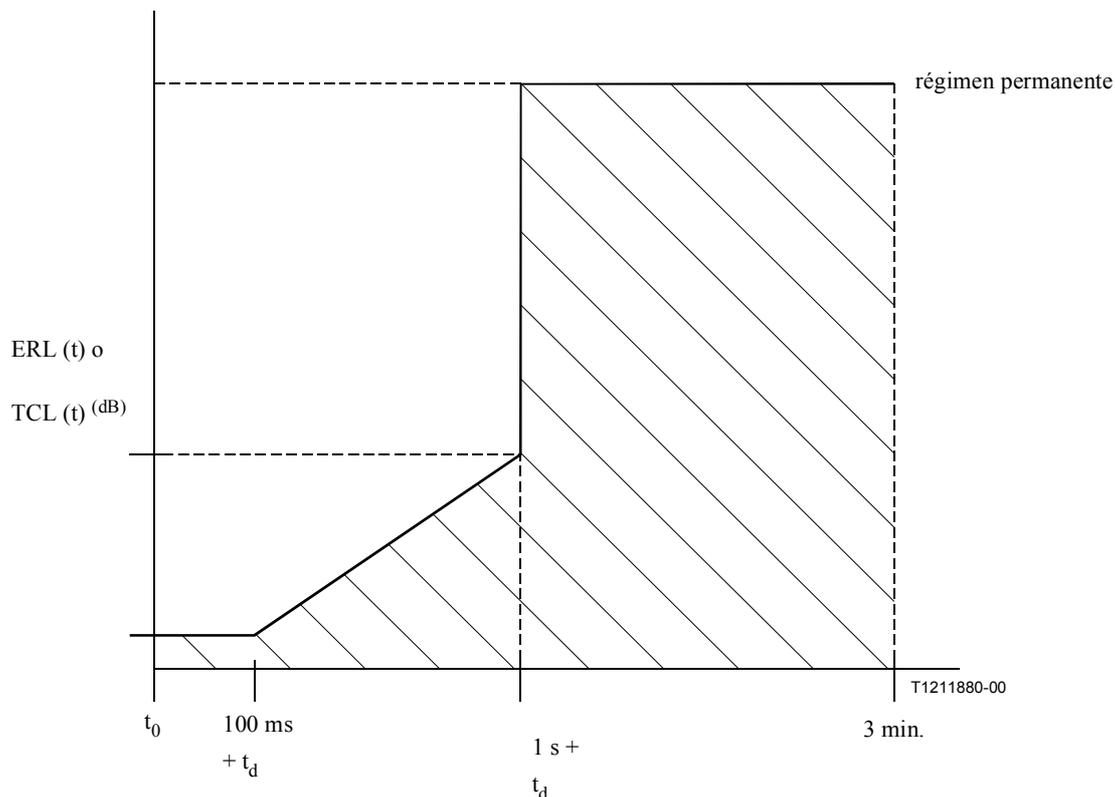
Los parámetros de habla simultánea se tratan en la cláusula 5.

## **4.3 Métodos de análisis**

La subcláusula siguiente plantea ciertas técnicas de medición objetiva que pueden utilizarse en la evaluación de los parámetros definidos anteriormente. La Recomendación UIT-T G.168 describe un montaje adecuado para las pruebas de compensadores de eco vocal de red. Las condiciones de prueba de los teléfonos manos libres se encuentran en la Recomendación UIT-T P.340.

### **4.3.1 Método de prueba del tiempo de convergencia ( $T_c$ )**

La descripción del método de prueba del tiempo de convergencia de los compensadores de eco se encuentra en la Recomendación UIT-T G.168. Para integrar la salida del medidor de nivel se utiliza un filtro de ponderación exponencial con una constante de tiempo de 35 ms ("Impulso", CEI 60651). La señal de salida medida se muestra en forma de gráfica nivel tiempo. Suele definirse un límite en función del tiempo, que no debe sobrepasarse. La figura 1 muestra un ejemplo representativo de dicha gráfica.



**Figura 1/P.502 – Ejemplo de límite característico de la ERL o la TCL en función del tiempo**

La ERL se suele medir en los compensadores de eco de red. La TCL define la atenuación de acoplamiento del terminal, con inclusión de las interfaces acústicas.

NOTA – No hay una definición exacta de  $T_c$ . Una de las definiciones posibles de  $T_c$  sería el tiempo que la ERL tarda en alcanzar un valor de 3 dB superior al del régimen permanente.

#### 4.3.2 Atenuación de retorno del eco, ponderada en el tiempo ( $ERL_{tst}$ ) – monólogo

La medición de la atenuación de retorno del eco, realizada con arreglo a la metodología definida en las Recomendaciones UIT-T G.122 y G.168 constituye el sistema tradicional en el que se basan actualmente los trabajos de cálculo y planificación. La metodología  $ERL_{tst}$ , más parecida a las impresiones subjetivas que tienen en cuenta los efectos temporales del eco, se describe a continuación.

Se mide la atenuación de retorno del eco ponderada en el tiempo desde la interfaz de red. Este método proporciona una medida de las ráfagas de eco y puede emplearse así mismo en vez de otras mediciones de la atenuación del retorno del eco de mayor duración.

El anexo A recoge el pseudocódigo para implementar este método. La señal de prueba a utilizar debe ser lo más parecida a la voz posible. Se podría utilizar otras señales de prueba, pero los resultados podrían ser indebidamente optimistas.

La señal de prueba se aplica en  $R_{in}$  durante 30 segundos (véase por ejemplo la figura 10), para que las distintas unidades funcionales (y especialmente el compensador de eco acústico) alcancen el régimen permanente. En las mediciones de controladores de eco acústicos, la única señal aplicada al micrófono, o micrófonos, es la de retorno acústico procedente del altavoz o altavoces.

Se graban las señales eléctricas en  $R_{in}$  y  $S_{out}$  durante el minuto siguiente. Se alinean en el tiempo las grabaciones de  $R_{in}$  y  $S_{out}$  añadiendo a la señal  $R_{in}$  el retardo del sistema entre  $R_{in}$  y  $S_{out}$ . El valor,

función del tiempo, de  $ERL_{tst}$ , es la diferencia (en dB) entre el nivel de la señal en  $R_{in}$  y en  $S_{out}$  calculada mediante el algoritmo del anexo A.

NOTA – Los trayectos del eco pueden variar durante la medición, ya que dependen del entorno y de cómo se utilicen los equipos.

## **5 Evaluación de la calidad vocal en habla simultánea**

La situación más crítica en una conversación es la de habla simultánea. Los equipos que efectúen algún tipo de tratamiento no lineal o variable en el tiempo pueden degradar considerablemente la calidad vocal y en particular los parámetros análogos a la "capacidad de habla simultánea" (percibida subjetivamente).

### **5.1 Percepción subjetiva de la degradación de la calidad vocal**

Los efectos más molestos en habla simultánea son los siguientes:

- las frases, palabras y sílabas interrumpidas o transmitidas parcialmente durante el habla simultánea, poco antes o poco después de ésta;
- la transmisión de voz o de ruido de fondo cuyo nivel varíe en el tiempo, provocando molestas "variaciones de nivel durante el habla simultánea";
- el eco durante el habla simultánea.

Las situaciones más críticas se presentan en los intervalos de tiempo inmediatamente anteriores o posteriores al habla simultánea. En estos periodos el autoenmascaramiento de la propia voz ya no resulta eficaz (véase Zwicker [5]).

Cuando se utilice la compensación de eco, es preciso que la convergencia sea rápida para facilitar con prontitud el habla simultánea al principio de la llamada. Esto significa que se puede suprimir antes el procesador no lineal, permitiendo el habla simultánea plena. La profundidad de convergencia determina la audibilidad del eco residual en habla simultánea. Si se utiliza una cierta atenuación conmutada para aumentar la atenuación de retorno del eco, puede producirse una atenuación de la voz audible en habla simultánea. Pueden aparecer ráfagas de eco en habla simultánea si los errores de detección provocan la divergencia del compensador de eco. Esto produciría ráfagas de eco bastante sonoras.

### **5.2 Parámetros objetivos relacionados**

Los parámetros objetivos relacionados son los siguientes:

- los tiempos de establecimiento (en habla simultánea);
- los tiempos de retención/bloqueo [tiempos de desconexión] (en habla simultánea);
- los tiempos de conmutación (en habla simultánea);
- el intervalo de atenuación (en habla simultánea);
- la distribución de la atenuación (en habla simultánea);
- las respuestas en frecuencia;
- los índices de sonoridad;
- la variación de nivel en habla simultánea (características de compresión-expansión).

Deben considerarse además los siguientes parámetros, asociados principalmente a las implementaciones del compensador de eco:

- El tiempo de convergencia durante el habla simultánea.
- La atenuación de retorno del eco (modalidad habla simultánea): calculada a partir del nivel de eco residual durante el habla simultánea.

- La atenuación de la voz emitida (véase así mismo el intervalo de atenuación) durante el habla simultánea: calculada a partir de la atenuación vocal debida a la introducción de la atenuación conmutada.
- La atenuación de retorno del eco, ponderada en el tiempo, en condiciones de habla simultánea: calculada a partir de del nivel ponderado de las ráfagas de eco.

NOTA – En la realización de mediciones objetivas ha de observarse en todo caso si las respuestas en frecuencia, los índices de sonoridad y las variaciones del nivel o niveles se miden poco antes del periodo de habla simultánea, tras ésta o durante la misma.

### **5.3 Métodos de análisis**

Hay varios métodos de evaluación de la calidad de funcionamiento de un sistema en habla simultánea. La explicación siguiente da una idea de las distintas tecnologías existentes. El método descrito en 5.3.1 es de tipo general y no supone ninguna implementación técnica específica del dispositivo objeto de la prueba. Los métodos de 5.3.2 a 5.3.4 suponen una implementación con compensador de eco.

#### **5.3.1 Método de habla simultánea con señal fuente compuesta (CSS)**

##### **5.3.1.1 Construcción de la señal**

Se puede llevar a la práctica un método de medida en habla simultánea, utilizando la señal de prueba de la figura 2. Dicha señal de prueba consiste en una serie de señales fuente compuesta sin correlacionar (Recomendación UIT-T P.501) que se inyectan simultáneamente en el sentido de la emisión y en el de la recepción. La secuencia de prueba se construye de tal manera que empiece con un nivel alto en el sentido de la emisión y se introduzca un nivel bajo en el sentido de la recepción. El nivel de cada secuencia de origen compuesto se reduce en 0,5 dB en el sentido de la emisión y se incrementa en 0,5 dB en el sentido de la recepción. La diferencia total de nivel entre un "paquete" de origen compuesto máximo y uno mínimo en cada sentido es de 20 dB (30 dB en las aplicaciones de red). Para los terminales manos libres pueden seleccionarse los siguientes intervalos de nivel:

Sentido de la recepción: de –38 dBm a –18 dBm

Sentido de la emisión: de –4,7 dBPa –24,7 dBPa

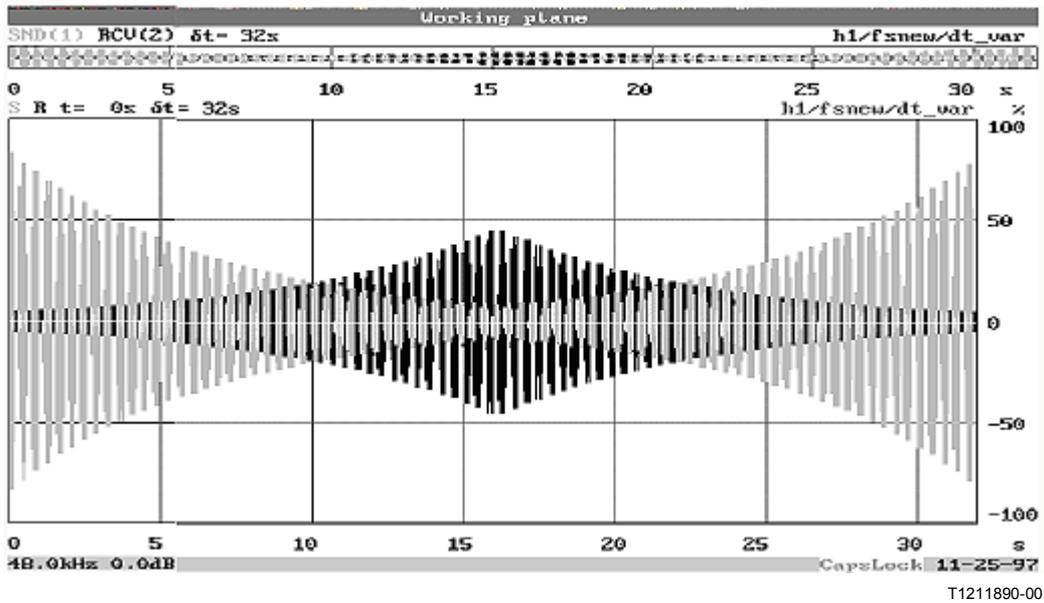
Para las mediciones en la red durante el habla simultánea se pueden escoger, por ejemplo, las siguientes variaciones de nivel:

Sentido de la recepción: de –40 dBm a –10 dBm

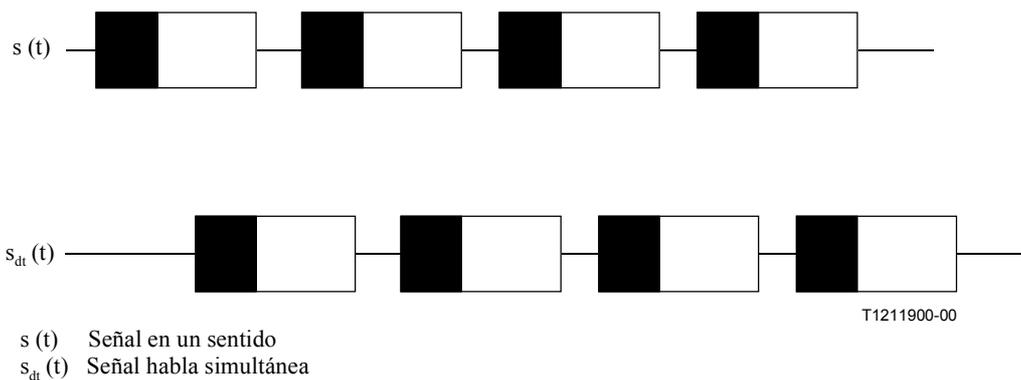
Sentido de la emisión: de –10 dBm a –40 dBm

Es evidente que se permiten distintas variaciones de nivel. Todos los intervalos de nivel dependen del intervalo dinámico a evaluar.

La secuencia típica tiene una estructura simétrica, o sea, tras un mínimo en el sentido de la emisión el nivel vuelve a aumentar mientras que en el sentido de la recepción el nivel de la señal disminuye de nuevo. La estructura simétrica de la señal permite así mismo evaluar el comportamiento simétrico del dispositivo objeto de la prueba.



**Figura 2/P.502 – Sinopsis de la señal de prueba en habla simultánea**



**Figura 3/P.502 – Esquema de la secuencia completa de medición con detalle del solapamiento de la señal en los sentidos de emisión y recepción, disposición fundamental**

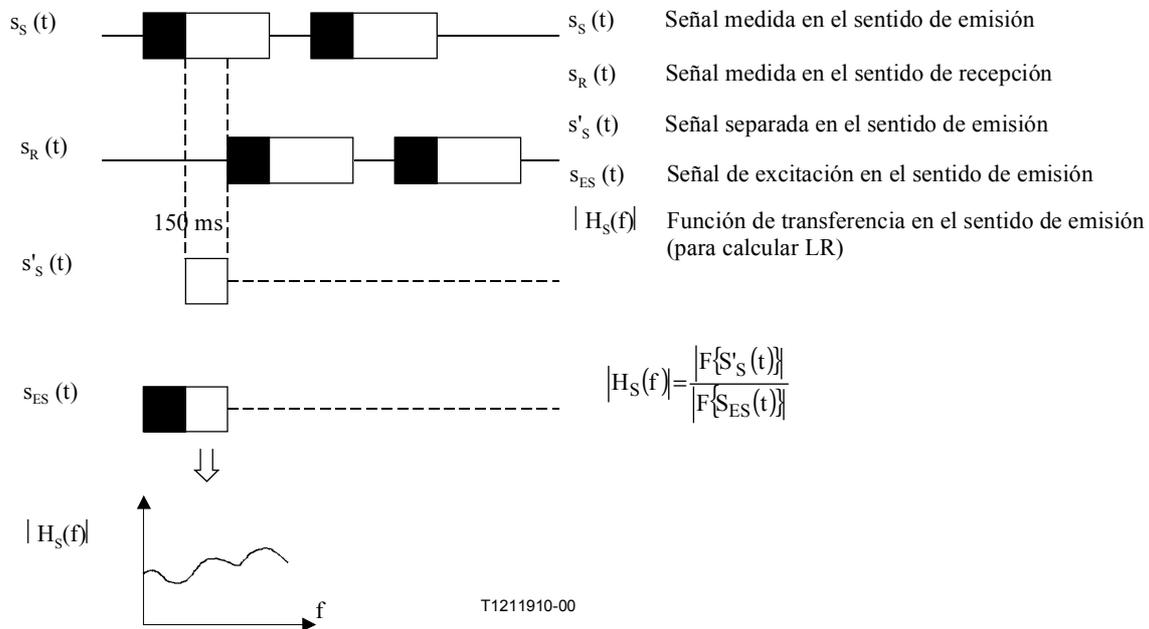
La figura 3 muestra más detalladamente la estructura de la señal. Se puede observar que el solapamiento de las secuencias es sólo parcial. El sonido vocalizado (en negro) se solapa siempre con el final de la secuencia de ruido pseudoaleatorio (en blanco) del canal opuesto. La secuencia se construye de modo tal que durante las pausas en el sentido de recepción puede medirse el de emisión y durante las pausas en el sentido de emisión puede evaluarse el de recepción. Esto puede ser útil por ejemplo cuando se trate de dispositivos analógicos en los que no sea posible el desacoplo suficiente entre la emisión y la recepción por la limitación de las capacidades.

Igualmente se puede construir una secuencia que comience en el nivel de excitación alto en el sentido de recepción y en el nivel de excitación bajo en el sentido de emisión, cuando se trate de evaluar puntos de comienzo de niveles distintos.

En general hay que observar que la estructura de la señal mostrada en las figuras 2 y 3 es un ejemplo de la relación en el tiempo entre los sentidos de emisión y recepción. Es posible utilizar, desde luego, otros intervalos de tiempo [por ejemplo, pausas más largas, secuencias de pseudoruido (pn) de mayor duración y distintos tipos de señales CS] en función de los requisitos de la medición a realizar.

### 5.3.1.2 Procedimientos de evaluación

Esta secuencia ofrece muchas posibilidades en las evaluaciones de habla simultánea. Antes de la evaluación real de un dispositivo hay que compensar el retardo entre la señal de excitación y la señal a medir. En la segunda fase, las señales a medir en cada sentido se separan y se refieren a la señal de excitación. La figura 4 muestra el fundamento de este método cuando se evalúan parámetros en el dominio de las frecuencias (por ejemplo, con la transformada de Fourier).



**Figura 4/P.502 – Fundamento de la separación de la señal y determinación de las características de transferencia**

NOTA – Hay que procurar que el análisis proporcione una estimación válida de las respuestas en frecuencia, los índices de sonoridad, etc. Los algoritmos de codificación implicados pueden conducir a estimaciones incorrectas de las funciones de transferencia, entre otras, y pueden exigir señales de medida o procedimientos de análisis específicos.

Como la señal de medida es un ruido pseudoaleatorio de secuencia CSS, se pueden calcular los siguientes parámetros a partir de 150 ms de la misma:

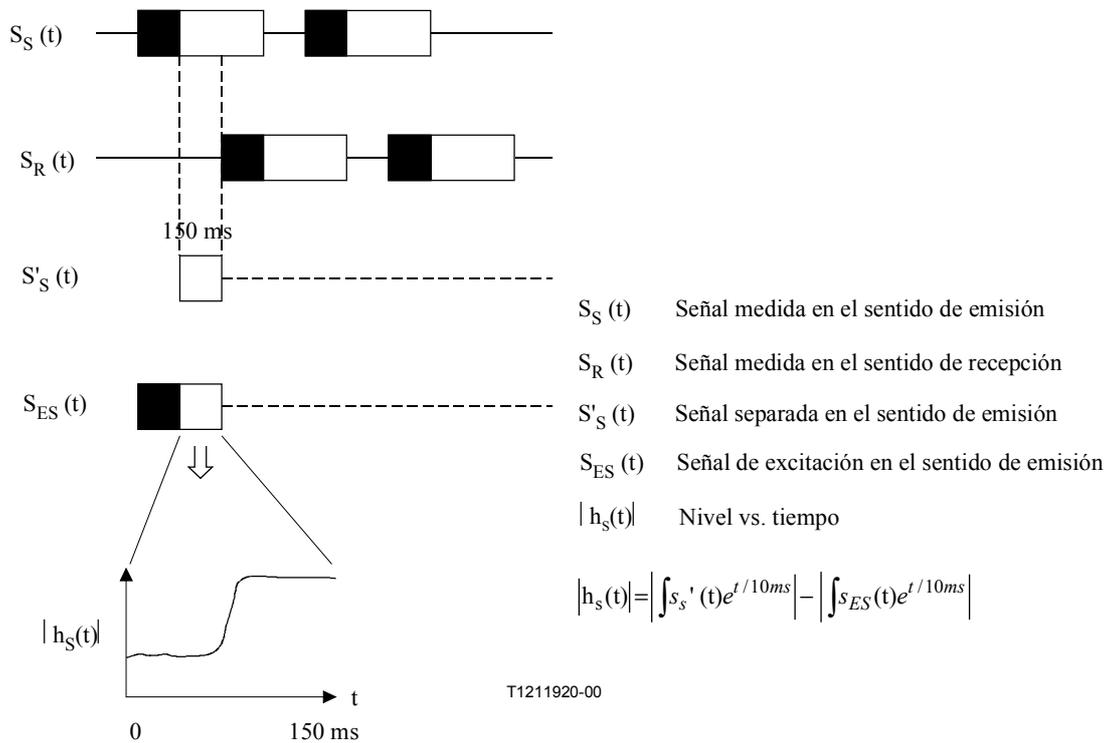
- las respuestas en frecuencia,
- los índices de sonoridad,
- la atenuación a corto plazo (cuando se trate de dispositivos de conmutación de nivel),
- la atenuación a largo plazo (cuando la medición se efectúe en distintos instantes de la secuencia).

Los tiempos de conmutación pueden evaluarse directamente en el dominio del tiempo debido a la exactitud de la definición de las características de la señal:

- el tiempo de establecimiento (debido al solapamiento de las señales, sólo si  $TR > 50$  ms);
- los tiempos de conmutación.

Los tiempos de conmutación se determinan a partir de la evolución del nivel en el tiempo con una constante de tiempo convenientemente pequeña, que suele ser como mínimo 10 veces inferior al tiempo de conmutación del sistema objeto de la prueba. La observación de la señal de salida durante los periodos en que sólo hay una señal presente, permite evaluar con gran precisión los tiempos de conmutación y las variaciones de nivel. La figura 5 presenta un ejemplo genérico de este

procedimiento. Aunque en este ejemplo la señal de salida se refiere a la de entrada, no siempre es necesaria esta referenciación ya que el nivel esperado durante los periodos de evaluación se calcula a partir de del nivel de la componente pseudoaleatoria de la CSS.



NOTA – La constante de tiempo de 10 ms es sólo un ejemplo, pudiéndose utilizar distintas constantes de tiempo en función de las exigencias del análisis.

**Figura 5/P.502 – Fundamento de la separación de la señal y determinación de las constantes de tiempo, este ejemplo muestra los tiempos de conmutación en habla simultánea**

Si fuese necesario evaluar el nivel en función de la frecuencia, se podría solapar la transformada de Fourier de la señal de salida medida o la de Wavelet. Este análisis permitiría evaluar los tiempos de conmutación o la variación del nivel en función de la frecuencia. Hay que procurar que la definición de las ventanas y la longitud de la ventana de análisis sean adecuadas. La longitud de la ventana de análisis debe ser inferior al intervalo de tiempo disponible (por la estructura de la señal y el solapamiento).

### 5.3.2 Pruebas en habla simultánea con secuencias combinadas en paralelo

Las pruebas en habla simultánea suelen imponer restricciones contrapuestas sobre el tipo de señal a utilizar. Frente a las señales combinadas en secuencia descritas en 5.3.1, las secuencias combinadas en paralelo permiten la evaluación durante el habla simultánea real. Debe adoptarse el principio de carácter general de que dichas señales sean ortogonales o se puedan separar de la señal a transmitir inicialmente por medio de filtros. La metodología de la prueba se explica a continuación.

En las pruebas de atenuación de retorno del eco, la señal habla simultánea inyectada en el  $R_{in}$  (o simulador de boca) del compensador debe ser tan parecida a la señal de acondicionamiento como sea posible. Es característico de los compensadores que congelen la adaptación durante el habla simultánea. Por ejemplo, si la señal habla simultánea en  $R_{in}$  fuese diferente de la señal de

acondicionamiento, lo normal sería que el eco residual fuese elevado de manera poco realista. Esta restricción indica que la señal habla simultánea en  $R_{in}$  debe ser la misma que la señal de acondicionamiento en  $R_{in}$  en la medición de la atenuación de retorno del eco.

Por desgracia, la utilización exclusiva de ficheros de sonido en habla simultánea no resulta aceptable. Las partes correlacionadas entre ambos "interlocutores" invalidarían algunos resultados de la prueba: partes del habla de un interlocutor pueden parecer el eco del otro cuando las partes estén correlacionadas. Otro problema es que el inicio del habla simultánea debe detectarse con la máxima precisión en las pruebas de atenuación y de mutilación. Esto sería muy difícil de definir en pruebas repetitivas con distintos ficheros de sonido, pero resulta muy fácil con tonos.

### 5.3.2.1 Método A de construcción de la señal

Para obviar estas dificultades, se utilizan ambos tipos de señal; la voz como señal de acondicionamiento y los tonos para definir con precisión el comienzo del habla simultánea. El modo de empleo dependerá de las pruebas concretas. Las señales de voz utilizadas durante el acondicionamiento se mantienen durante el habla simultánea, si es necesario. El tono sinusoidal se mezcla con el habla o se inyecta aisladamente para proporcionar una referencia fácilmente medible en las pruebas de atenuación o un comienzo del habla simultánea fácilmente definible en las pruebas de mutilación. El empleo de filtros de ranura o de paso banda en  $S_{out}$  (o salida de recepción) a la frecuencia del tono, permite supervisar sólo el tono o sólo la voz.

Cuando el tono está mezclado con la voz, la potencia del tono debe ser representativa de la potencia media vocal en dicha frecuencia a largo plazo, para evitar la repercusión sobre el compensador de las desviaciones brutas de la energía espectral respecto a la señal de acondicionamiento. La Recomendación UIT-T P.50 indica una relación espectral media (se utilizan valores de tercios de octava). Se recomienda utilizar los tonos del cuadro 1. Su potencia se define como el número de dB por debajo de la energía media de la voz activa del fichero de voz, cuando la medición se realiza con arreglo a la Recomendación UIT-T P.56.

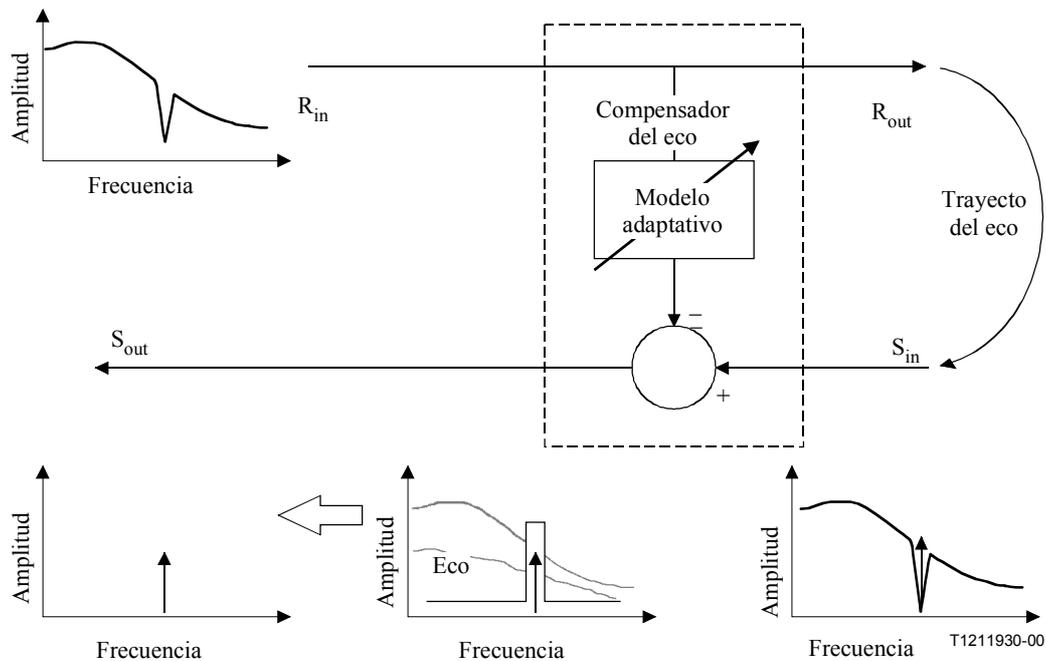
**Cuadro 1/P.502**

Frecuencia del tono	Nivel relativo del tono (dB) por debajo del nivel nominal de la voz
500	9
1 000	14
1 750	18
2 500	22

### 5.3.2.2 Método A de prueba de la atenuación en habla simultánea

#### 5.3.2.2.1 Atenuación vocal en emisión durante el habla simultánea ( $A_{sdt}$ )

El ejemplo de la figura 6 determina la atenuación del habla simultánea en el sentido de emisión. Este concepto es fácilmente aplicable al sentido de recepción invirtiendo las señales y supervisando la salida de recepción.



**Figura 6/P.502 – Fundamento del método de prueba A de la atenuación en habla simultánea**

Esta metodología se explica a continuación.

Cuando haya compensadores de eco, el objeto probado se desactiva (si es posible) y se acondiciona como explica el anexo C. El "interlocutor activo antes del habla simultánea" es el simulador de boca cuando se trate de teléfonos manos libres o  $S_{in}$ . El "interlocutor que inicia el habla simultánea" es  $R_{in}$ .

Obsérvese que las señales en  $R_{in}$  y  $S_{in}$  del simulador de boca aparecen filtradas en ranura a la frecuencia del tono. Este filtro de ranura no está presente durante todo el periodo de acondicionamiento, sino justo antes del habla simultánea y durante el resto de la medición. La idea es mezclar un tono en el simulador de boca (todavía en monoloxua) justo antes del habla simultánea, supervisar el nivel de su valor eficaz, hacer que  $R_{in}$  entre en habla simultánea y continuar supervisando el nivel del tono.

La atenuación en habla simultánea es la diferencia entre el nivel de tono antes del habla simultánea y durante ésta. El tono se discrimina aplicando a  $S_{out}$  un filtro paso banda con la frecuencia del tono. Prosiguiendo con las mediciones durante el habla simultánea, se pueden determinar las características de conmutación y entre ellas el ritmo de inserción y la profundidad. El ritmo de supresión de la atenuación puede determinarse asimismo volviendo a reducir la máscara de actividad del "interlocutor que inicia el habla simultánea" una vez estabilizada la profundidad de la atenuación.

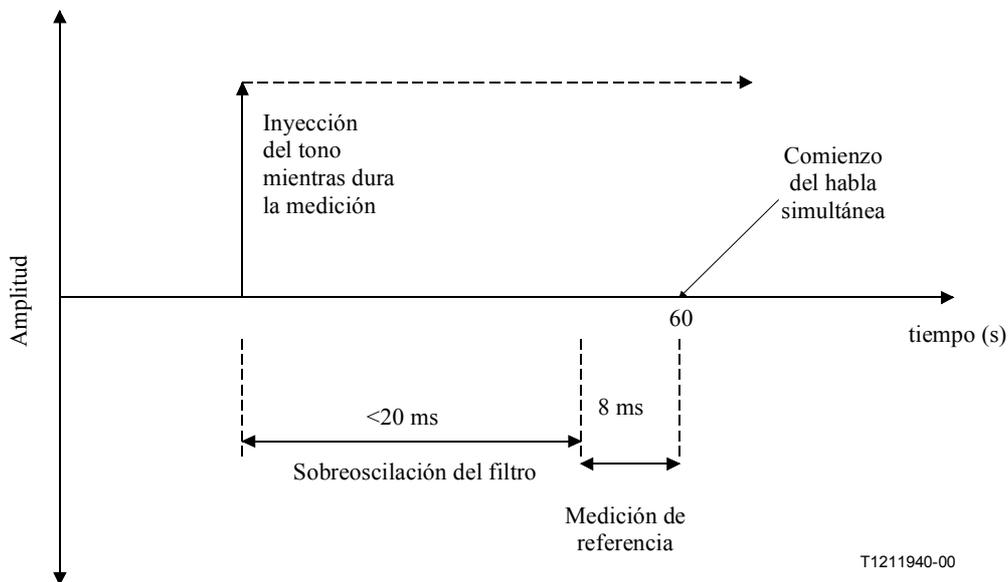
Se describen a continuación las características del filtro de ranura. Hay que aplicar este filtro en  $R_{in}$  para que el eco a la frecuencia del tono no influya en la medición de éste. El filtro de ranura debe presentar la atenuación suficiente para lograr que la voz a la frecuencia del tono quede suficientemente amortiguada y no repercuta en el nivel del tono en el simulador de boca. La anchura de banda del filtro de ranura debe ser lo suficientemente pequeña para reducir al mínimo la repercusión sobre las frecuencias vecinas de modo que las señales no sean demasiado distintas de las de acondicionamiento. El anexo B describe ejemplos de tipos de filtro.

El filtro paso banda tiene restricciones parecidas. Al igual que el filtro de ranura, debe proporcionar atenuación suficiente fuera de banda para que el nivel de tono no quede afectado por la voz ni el eco. Debe tener así mismo una respuesta al impulso suficientemente corta para reducir al mínimo la repercusión en el dominio del tiempo, ya que el ritmo de atenuación es también objeto de la medición.

La respuesta del filtro de ranura al impulso no influye en la medición ya que el nivel del tono mezclado en el simulador de boca es suficientemente grande para enmascarar cualquier sobreoscilación del filtro.

El valor eficaz del tono debe medirse con una ventana móvil de 8 ms para suavización. La ventana se desplaza en saltos de 4 ms para lograr solapamientos de 4 ms entre puntos adyacentes con objeto de suavizar los resultados.

La temporización de la medición debe ser precisa teniendo en cuenta la posible sobreoscilación del filtro paso banda. El tono debe inyectarse a los 60 segundos, menos el tiempo de sobreoscilación del filtro (que se supone  $<20$  ms), menos 8 ms, una vez terminada la sobreoscilación del filtro. El retardo de activación en el sentido de medida provoca mediciones parciales durante el final de la sobreoscilación del filtro paso banda. Mientras el retardo de activación sea pequeño (provisionalmente  $<5$ ms), la sobreoscilación será pequeña no debiendo repercutir en la medición.



**Figura 7/P.502**

Cuando en el AEC se utilicen técnicas de subbandas, o la técnica utilizada sea desconocida, es aconsejable repetir la prueba para cada una de las frecuencias mostradas. En muchos casos, es posible que los resultados de las pruebas de atenuación a una frecuencia cualquiera no sean indicativos de la calidad subjetiva. Como la voz presenta una densidad espectral de energía predominantemente en el intervalo de frecuencias bajas, cabe esperar que la atenuación conmutada sea más perceptible en dicho intervalo. Si la profundidad de la atenuación es independiente de la frecuencia, resulta aconsejable utilizar frecuencias de prueba mayores, ya que los filtros necesarios tendrán menos repercusión en los niveles vocales globales.

El método presentado permite medir la atenuación en función del tiempo tras introducir el habla simultánea a una frecuencia específica. El resultado de este método puede depender significativamente de la naturaleza exacta de la señal vocal utilizada, especialmente al comenzar el habla simultánea. Puede existir asimismo una dependencia de la frecuencia del tono de medición, que es una senoide introducida en el habla real para crear habla simultánea.

A continuación se facilita una descripción detallada de los pasos a seguir.

El compensador se acondiciona como se explica en el anexo C, con la máscara de actividad "interlocutor activo justo antes del inicio del habla simultánea", aplicada a  $S_{in}$  (o en el punto de referencia de boca o MRP cuando se miden terminales). La máscara "Interlocutor iniciando el habla simultánea" se aplica a  $R_{in}$ . Se ejecuta la prueba descrita anteriormente.

Aplicando la ventana móvil de promediación de 8 ms sobre la sinusoide medida en  $S_{out}$ , la función del tiempo  $A_{SDT}$  es la diferencia (en dB) entre el primer promedio de 8 ms antes del habla simultánea y cada promedio de 8 ms tras ésta.

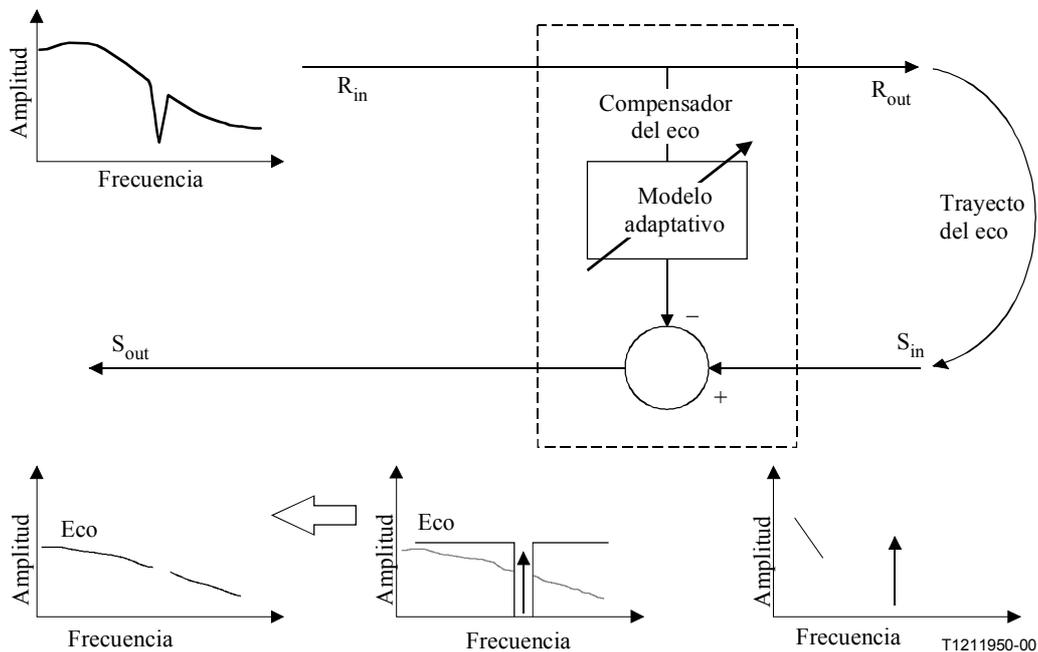
### 5.3.2.2 Atenuación vocal recibida en habla simultánea ( $A_{rdt}$ )

NOTA – En el caso de medición de teléfonos manos libres, la señal de salida recibida debe medirse estando el micrófono de medición tan próximo como sea posible al altavoz para facilitar la discriminación en el dominio acústico.

Se ejecutará la prueba como se explicó antes para la  $A_{sdt}$ , sustituyendo recepción por emisión y viceversa. Por consiguiente, se intercambiarán las señales de recepción y envío y los resultados se supervisan a la salida del receptor.

### 5.3.2.3 Método A de prueba de la atenuación de retorno del eco durante habla simultánea

El ejemplo mostrado en la figura 8 muestra cómo se determina la atenuación de retorno del eco mirando hacia el terminal desde la red. Este concepto es fácilmente aplicable a la atenuación del trayecto del eco del interlocutor invirtiendo las señales y supervisando la salida de recepción.



**Figura 8/P.502 – Fundamento del método A de prueba de la atenuación de retorno del eco**

Se desactiva el aparato y se acondiciona de acuerdo con lo descrito en el anexo C. El "interlocutor activo antes del habla simultánea" es  $R_{in}$ . El "interlocutor que inicia el habla simultánea" es el simulador de boca en el puerto  $S_{in}$ .

El filtro de ranura se aplica a  $R_{in}$  como se explica en 5.3.2.2. Se introduce el habla simultánea aplicando un tono (solamente) a la boca del simulador. Dicho tono se aplica durante 20 s. La

atenuación de retorno del eco se calcula filtrando en ranura el tono en primer lugar y midiendo a continuación el eco residual mediante el método de cálculo de la ERL o de la  $ERL_t$  descrito en el anexo A. La utilización de un tono garantiza la determinación del momento exacto de introducción del habla simultánea.

Una vez terminada el habla simultánea, puede prolongarse la medición de la atenuación de retorno del eco durante 10 segundos para obtener la recuperación tras el habla simultánea. Una vez transcurrido dicho tiempo, debe guardarse un segundo de silencio. De este modo, se puede medir el ruido en el trayecto del eco. Puede suponerse que el ruido y el eco no están correlacionados, y que el ruido es estacionario, el ruido medido en el último segundo puede restarse del eco más ruido durante el habla simultánea para obtener el eco durante el habla simultánea.

La temporización de la medición debe ser precisa ya que se conoce el retardo en el trayecto del eco. Este retardo alinea adecuadamente la fuente y el eco.

El procedimiento detallado de prueba es el siguiente:

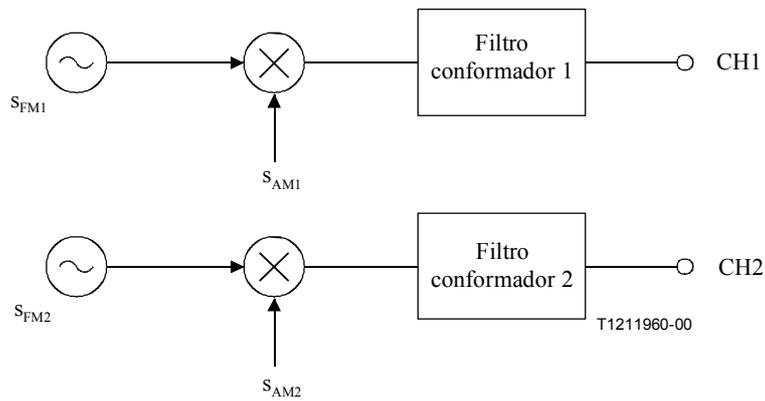
#### **Atenuación de retorno del eco – habla simultánea ( $ERL_{dt}$ ), ponderada en el tiempo – habla simultánea ( $ERL_{tdt}$ )**

Cuando existan compensadores de eco, todas las unidades funcionales de compensación de eco se desactivarán previamente, activándose a continuación. Para la prueba de caja negra, el sistema se conecta en ausencia de estímulos acústicos en las interfaces. Se acondiciona el compensador y se aplica a  $R_{in}$  la máscara "interlocutor activo inmediatamente antes del inicio del habla simultánea". En  $S_{in}$  (o sea, en el simulador de boca) se aplica la máscara "interlocutor que inicia el habla simultánea". Las pruebas se realizan con arreglo al anexo A.

Se registran las señales eléctricas en  $R_{in}$  y  $S_{out}$  durante los 20 segundos de aplicación del tono. Se alinean en el tiempo los registros de  $R_{in}$  y  $S_{out}$ , añadiendo el retardo del sistema entre  $R_{in}$  y  $S_{out}$ . Se calcula la  $ERL_{dt}$  mediante el procedimiento tradicional de cálculo de la atenuación del eco con arreglo a la Recomendación UIT-T G.122 o bien la  $ERL_{tdt}$ . La función del tiempo  $ERL_{tdt}$  es la diferencia (en dB) entre el nivel de la señal en  $R_{in}$  y la señal en  $S_{out}$  calculada con arreglo al anexo A, durante los primeros 20 segundos del habla simultánea.

#### **5.3.2.4 Método B de construcción de la señal**

El método B utiliza secuencias ortogonales generadas por un conjunto de sinusoides moduladas de modo parecido a la voz, conformadas espectralmente. El fundamento general de construcción se encuentra en la Recomendación UIT-T P.501 y se muestra a continuación en la figura 9.



$$s_{FM1,2}(t) = \sum A_{FM1,2} * \cos(2\pi t n * F_{01,2}); \quad n = 1, 2, \dots$$

$$s_{AM1,2}(t) = \sum A_{AM1,2} * \cos(2\pi t F_{AM1,2});$$

**Figura 9/P.502 – Generación de la señal de prueba de dos canales para evaluaciones en habla simultánea con señales de MA y MF**

El cuadro 2 describe los ajustes tipo.

**Cuadro 2/P.502**

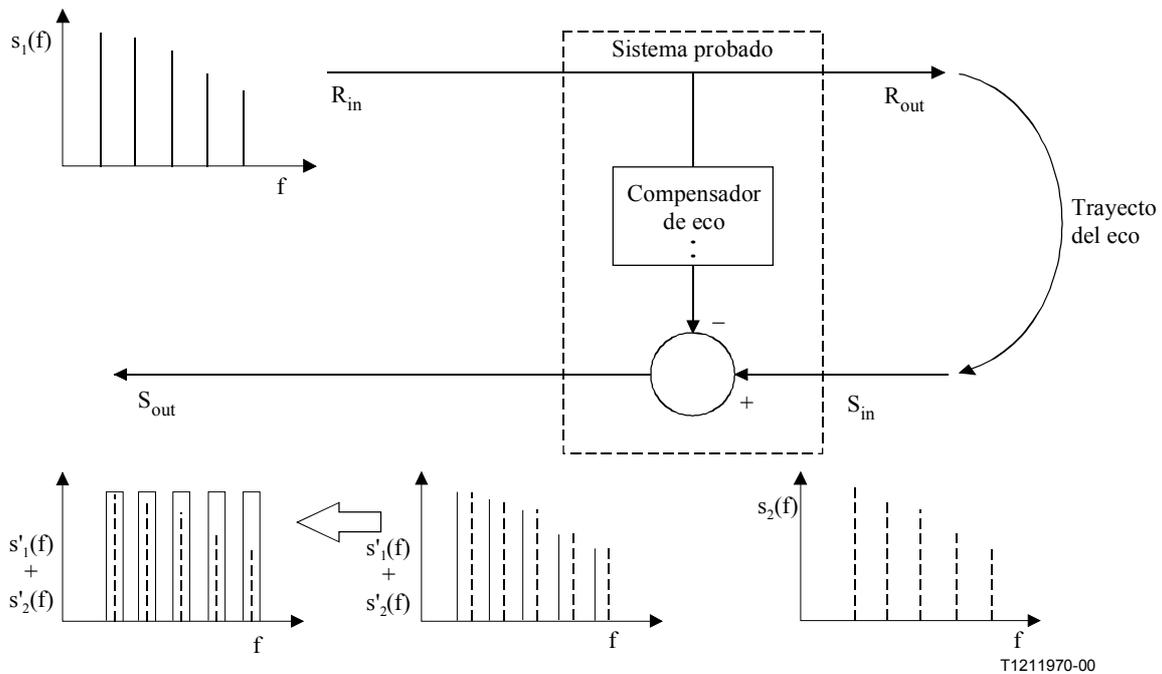
	$f_{AM}$	$f_{FM}$	$F_0$	<b>Filtro conformador</b>
Canal 1 (CH 1)	$f_{AM1} = 3 \text{ Hz}$	$f_{FM1} = 5 \text{ Hz}$	$F_{01} = 270 \text{ Hz}$	LP, 5 dB/oct.
Canal 2 (CH 2)	$f_{AM2} = 3 \text{ Hz}$	$f_{FM2} = 5 \text{ Hz}$	$F_{02} = 290 \text{ Hz}$	LP, 5 dB/oct.

Para mayor información véase la Recomendación UIT-T P.501.

### 5.3.2.5 Método B de prueba de la atenuación en habla simultánea

#### 5.3.2.5.1 Atenuación vocal de emisión durante el habla simultánea ( $A_{sdt}$ )

El ejemplo de la figura 10 determina la atenuación de habla simultánea en el sentido de emisión. Al igual que ocurría con el método A, éste concepto es fácilmente aplicable al sentido de recepción invirtiendo las señales y supervisando la salida de recepción.



**Figura 10/P.502 – Fundamento del método B de prueba de la atenuación en habla simultánea**

La metodología se explica a continuación.

Cuando haya compensadores de eco, el objeto sometido a prueba se desactiva (si es posible), y se acondiciona con arreglo al anexo A. El "interlocutor activo antes del habla simultánea" es el simulador de boca cuando se trata de teléfonos manos libres o la  $S_{in}$ . El "interlocutor que inicia el habla simultánea" es  $R_{in}$ .

La señal de habla simultánea se estudia con arreglo al siguiente análisis.

Para separar la señal eco del habla simultánea se necesita bien sea un filtrado específico bien sea un tratamiento posterior del análisis de la FFT, ya que los espectros de la señal y del habla simultánea son análogos al espectro de un filtro de peine al que se hubiese aplicado una modulación específica. Conviene tener en cuenta la frecuencia central  $f_{mid}$  de cualquier componente de frecuencia, la correspondiente frecuencia de modulación  $f_{mod}$  y las formas de filtro de la función ventana de la transformada de Fourier. De utilizar la solución del filtro, se deberá calcular el ancho de banda de cada filtro de modo que:

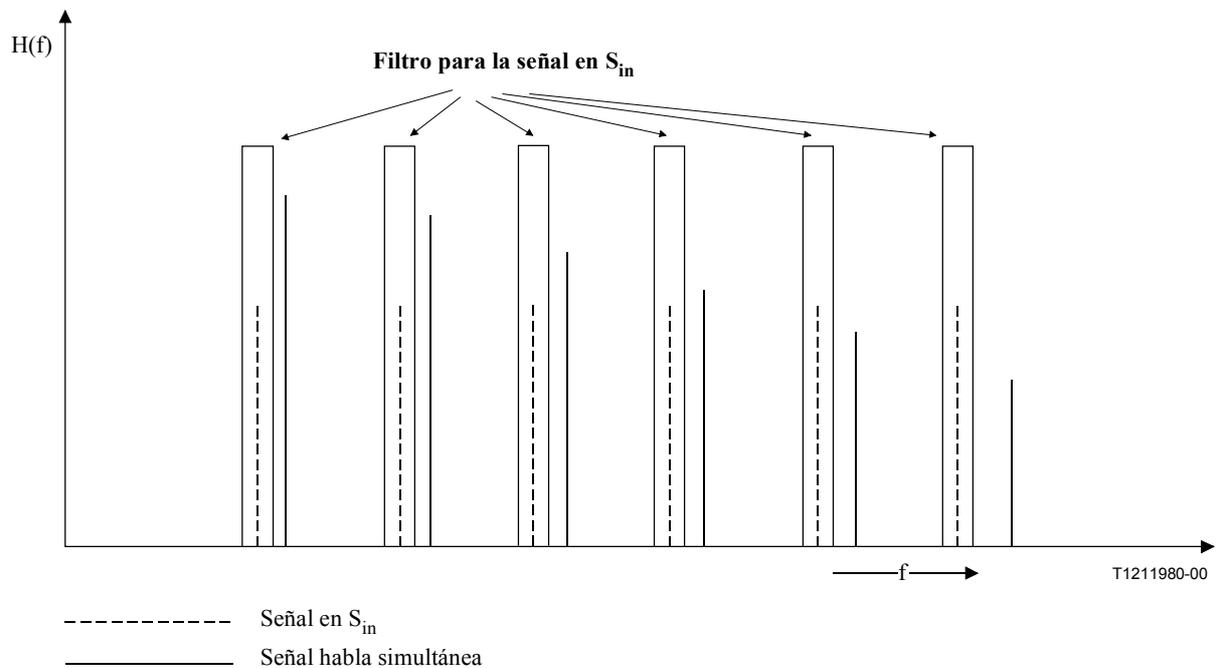
$$f_u = f_{mid} - f_{mod} (fm)$$

$$f_{c0} = f_{mid} + f_{mod} (fm)$$

La atenuación de la banda suprimida debe ser como mínimo 10 dB mayor que el nivel mínimo a medir dentro de la banda de paso. Lo mismo ocurre con el análisis de la transformada de Fourier de la señal de eco medida. En este caso hay que tener en cuenta el efecto de "arrastré" de la función ventana. Con objeto de que la separación entre la señal eco y la señal habla simultánea sea suficiente en el dominio de bajas frecuencias, debe adoptarse una longitud mínima de 8 k para la FFT (frecuencia de muestreo de 44,1 ó 48 kHz) equivalente a una ventana de 170 ms de tiempo, aproximadamente.

La Recomendación UIT-T P.501 contiene una serie de frecuencias de excitación tipo para la señal de monoloquia y la de habla simultánea.

El fundamento del análisis se muestra en la figura 11.



**Figura 11/P.502 – Esquema de la separación de las componentes de la señal en  $S_{in}$**

La atenuación en habla simultánea es la diferencia de nivel entre el tono antes del habla simultánea y durante ésta. La secuencia de sonidos emitidos se discrimina aplicando el filtro o el procedimiento de la FFT descrito con anterioridad. Prolongando la medición durante el habla simultánea, se pueden determinar las características de conmutación, entre ellas el ritmo de inserción y la profundidad, tanto en el tiempo como en frecuencia según el método de análisis adoptado. También puede determinarse el ritmo de supresión de la atenuación reduciendo de nuevo la máscara de actividad para el "interlocutor que inicia el habla simultánea", una vez estabilizada la profundidad de la atenuación.

Como se conoce el retardo del trayecto del eco, la temporización de la medición debe ser precisa. Este retardo alinea adecuadamente la fuente y el eco.

#### **5.3.2.5.2 Atenuación vocal recibida durante el habla simultánea ( $A_{rdt}$ )**

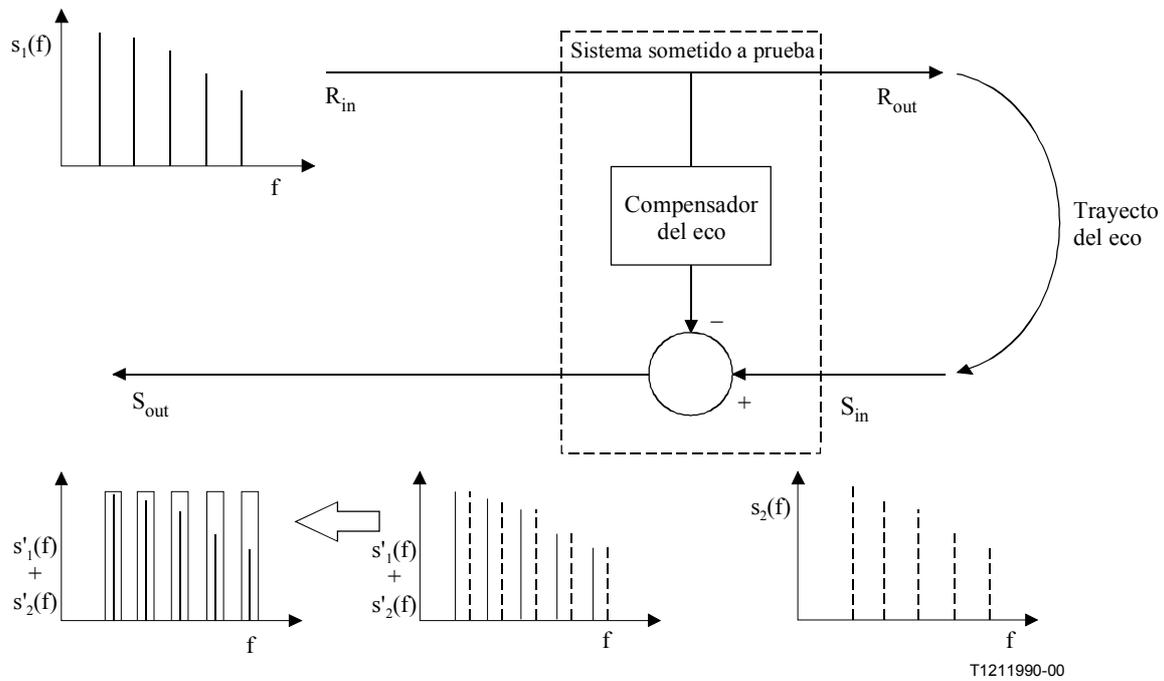
NOTA – Cuando se midan teléfonos manos libres, la señal de salida recibida debe medirse estando el micrófono de medida tan próximo al altavoz como sea posible para facilitar la discriminación en el dominio acústico.

Se realizará la prueba análogamente a la descrita para la  $A_{sdt}$ , sustituyendo recepción por emisión y viceversa. Por consiguiente, se intercambiarán las señales de recepción y emisión y se supervisarán los resultados a la salida del receptor.

La explicación detallada de los distintos pasos (1 a 4) para la evaluación de la "atenuación vocal emitida durante el habla simultánea ( $A_{sdt}$ )" y la "atenuación vocal recibida durante el habla simultánea ( $A_{rdt}$ )" se encuentra en 5.3.2.2.

#### **5.3.2.6 Método B de prueba de la atenuación de retorno del eco durante el habla simultánea**

El ejemplo mostrado en la figura 12 describe cómo se determina la atenuación de retorno del eco mirando hacia el terminal desde la red. Este concepto se aplica sin dificultad a la atenuación del trayecto del eco del interlocutor invirtiendo las señales y supervisando la salida de recepción.

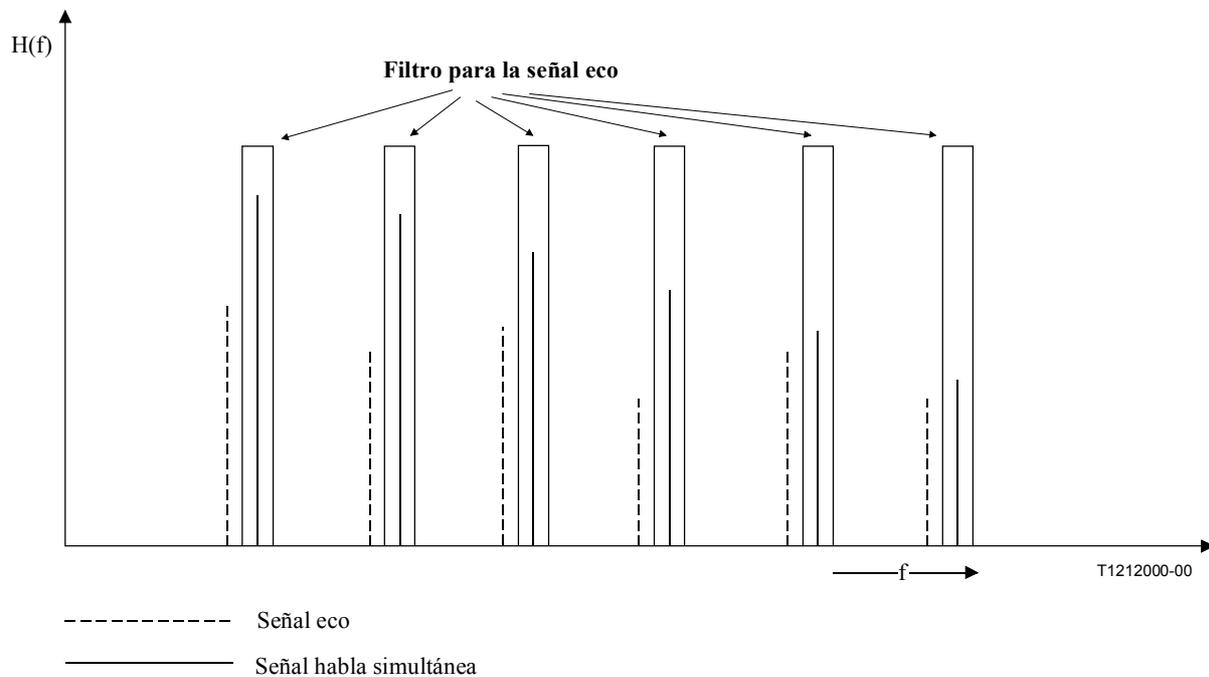


**Figura 12/P.502 – Fundamento del método B de prueba de la atenuación de retorno del eco**

Esta metodología se explica a continuación.

Si hay compensadores de eco, se desactiva el aparato y se acondiciona de acuerdo con el anexo C. El "interlocutor activo antes del habla simultánea" es  $R_{in}$ . El "interlocutor que inicia el habla simultánea" es el simulador de boca o el puerto  $S_{in}$ .

La señal habla simultánea se analiza durante el periodo de habla simultánea según el análisis descrito en la subcláusula anterior. En vez de analizar la señal de habla simultánea se analiza la del eco, aplicando el juego de filtros adecuado. Véase la figura 13.



**Figura 13/P.502 – Esquema de la separación de las componentes de eco de la señal habla simultánea**

Una vez terminada el habla simultánea, puede prolongarse la medición de la atenuación de retorno del eco durante 10 segundos para obtener la recuperación tras el habla simultánea. Una vez transcurrido dicho tiempo, debe guardarse un segundo de silencio. De este modo, se puede medir el ruido en el trayecto del eco. Puede suponerse que el ruido y el eco no están correlacionados, y que el ruido es estacionario, el ruido medido en el último segundo puede restarse del eco más ruido durante el habla simultánea para obtener el eco durante el habla simultánea.

Como se conoce el retardo en el trayecto del eco, la temporización de la medición debe ser precisa. Este retardo alinea adecuadamente la fuente y el eco.

La explicación detallada de los distintos pasos (1 a 4) para la evaluación de la "atenuación vocal emitida durante el habla simultánea ( $A_{sdt}$ )" y la "atenuación vocal recibida durante el habla simultánea ( $A_{rdt}$ )" se encuentra en 5.3.2.3.

## 6 Características de la compresión-expansión y del AGC

La compresión-expansión y el AGC sirven para evitar la sobrecarga de los sistemas, para compensar las variaciones de los niveles vocales y para ecualizar los niveles vocales de la red. En cualquier caso el objetivo es la mejora de la "calidad" vocal o la de las señales de entrada en los dispositivos que manejan señales vocales, tales como compensadores de eco, detectores de voz u otros.

### 6.1 Percepción subjetiva de la degradación de la calidad vocal

Los dispositivos de compresión-expansión, los de AGC y en general cualquier dispositivo que introduzca una amplificación variable en el tiempo de la señal vocal, provocan fluctuaciones del nivel vocal que pueden degradar la calidad vocal. En general, la influencia percibida subjetivamente depende del intervalo de atenuación, de las constantes de tiempo y de las características de control del dispositivo en cuestión.

## 6.2 Parámetros objetivos relacionados

Los parámetros objetivos relacionados son los siguientes:

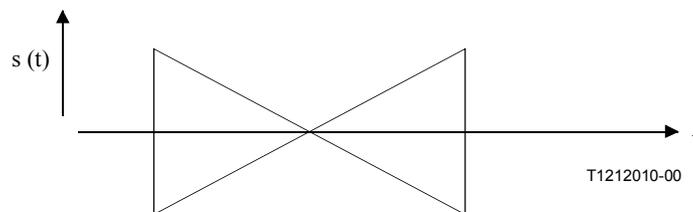
- el intervalo de control del compresor-expansor;
- el intervalo de atenuación (intervalo de ajustes de nivel);
- las constantes de tiempo.

## 6.3 Métodos de análisis

En general, deben utilizarse distintos tipos de señal de prueba para determinar la calidad de funcionamiento de los dispositivos de compresión-expansión y del AGC. Además de las señales artificiales de prueba, deben utilizarse señales vocales o semejantes a éstas, tal como la voz artificial (Recomendación UIT-T P.50) o la señal MSMP descrita en la Recomendación UIT-T P.501.

Las figuras 14 y 15 muestran señales de prueba generadas por una repetición periódica de señales vocales. Estas señales pueden utilizarse para medir los ajustes de nivel en los sistemas que se comportan de modo similar ante la repetición periódica del sonido vocalizado y ante la voz real. Además se puede utilizar la voz artificial.

La figura 14 representa una señal de entrada con un nivel constantemente creciente (decreciente). En la figura 15 sin embargo, los niveles de la señal se ajustan en ciertos periodos.



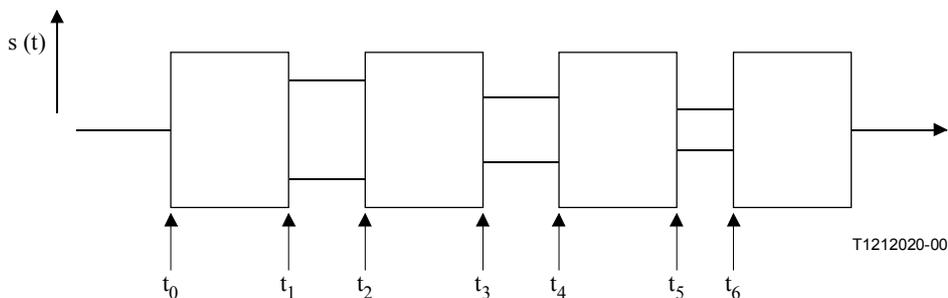
**Figura 14/P.502 – Estructura de la señal de prueba para determinar los ajustes de nivel (nivel de entrada en cambio constante)**

Los parámetros propuestos para la señal de la figura 14 se indican en el cuadro 3.

**Cuadro 3/P.502**

Sentido	Generación de la señal	Nivel máximo	Nivel mínimo	Variación del nivel
<b>SND</b> (emisión)	Sonido vocalizado, repetido periódicamente	-16 dBm (-3,0 dBPa para terminales)	Por debajo del umbral mínimo de ruido	Lineal
<b>RCV</b> (recepción)	Sonido vocalizado, repetido periódicamente	-16 dBm	Por debajo del umbral mínimo de ruido	Lineal

La duración total de la señal puede establecerse en 10 s.



**Figura 15/P.502 – Estructura de la señal de prueba para determinar los ajustes de nivel**

Los parámetros propuestos para la señal de la figura 15 se consignan en el cuadro 4.

**Cuadro 4/P.502**

Sentido	Generación de la señal	Nivel de la señal durante ( $t_1 - t_0$ ) ( $t_3 - t_2$ ) ( $t_5 - t_4$ ) ⋮	Nivel de la señal durante ( $t_2 - t_1$ )	Nivel de la señal durante ( $t_4 - t_3$ )	Nivel de la señal durante ( $t_6 - t_5$ )
<b>SND</b> (emisión)	Sonido vocalizado repetido periódicamente	-16 dBm (-3,0 dBPa para terminales)	-21 dBm (-8,0 dBPa para terminales)	-26 dBm (-13,0 dBPa para terminales)	-31 dBm (-18,0 dBPa para terminales)
<b>RCV</b> (recepción)	Sonido vocalizado repetido periódicamente	-16 dBm	-21 dBm	-26 dBm	-31 dBm

La duración de la señal en los periodos sencillos puede hacerse igual a 2,5 s.

La señal de la figura 14 sirve para determinar el:

– **Intervalo de ajustes de nivel en función del nivel de la señal de entrada**

Se aplica una señal como la descrita (figura 14). El análisis de la señal de salida se realiza sobre la evolución de ésta en el tiempo, refiriendo la señal de salida medida a la señal de excitación alineada en el tiempo. Las constantes de tiempo deben encontrarse en el intervalo 10-125 ms. Hay que tratar de evitar los resultados de las mediciones erróneas producidos por el carácter no vocal de la señal de prueba.

El resultado teórico del análisis es una curva plana en función del tiempo. Si se utiliza el AGC o la compresión-expansión la salida no debe desviarse más de  $\pm 3$  dB de la salida medida media. Las constantes de tiempo del AGC, cuando se utilice éste, deben ser más bien lentas ( $>100$  ms). Si se detecta la compresión-expansión, los tiempos de ataque deben ser bastante pequeños (10-50 ms) aunque los tiempos de liberación deben ser suficientemente grandes. Si se detecta un valor de compresión-expansión o del AGC superior a  $\pm 3$  dB, resulta imprescindible la evaluación subjetiva.

La señal de la figura 15 sirve especialmente para determinar la:

– **Duración de los ajustes de nivel**

En este análisis se debe utilizar la señal de la figura 15. El análisis se realiza igual que el descrito anteriormente. Esto exige que el análisis se haga sobre la variación del nivel en el

tiempo, refiriendo la señal medida a la de excitación alineada en el tiempo y mostrando el resultado en forma de gráfica en función del tiempo. Para obtener una buena resolución, la constante de tiempo para este análisis debe encontrarse en el intervalo 5-10 ms.

Teóricamente no deben observarse diferencias de nivel. Si se observa AGC o compresión-expansión, deben estar en un intervalo inferior a  $\pm 3$  dB. Por consiguiente la duración medida debe ser la descrita anteriormente.

## **7 Calidad de transmisión del ruido de fondo**

Al juzgar la calidad de transmisión del ruido de fondo, el oyente lo percibe como una señal. En general, cuando se transmite la voz, efectos como éste pueden influir en la percepción subjetiva de la calidad. Este parámetro cobra cada vez mayor importancia debido a que los modernos sistemas de telecomunicación se utilizan cada vez más en ambientes ruidosos.

### **7.1 Percepción subjetiva de la degradación de la calidad**

La influencia más característica se observa en el sentido de emisión para el oyente en el extremo remoto cuando se transmite ruido de fondo. En general, influyen en la calidad percibida:

- las fluctuaciones de nivel del ruido de fondo,
- las interrupciones de la transmisión de ruido,
- las modulaciones indeseadas producidas por el tratamiento de la señal.

La influencia puede ser distinta:

- en la modalidad de reposo,
- con voz en el extremo remoto,
- con voz en el extremo próximo,

cada una de estas situaciones debe estudiarse por separado.

### **7.2 Parámetros objetivos relacionados**

En las siguientes descripciones de análisis, el ruido de fondo se toma como señal de prueba. Los efectos percibidos subjetivamente pueden describirse mediante los siguientes parámetros:

- el intervalo de atenuación,
- la atenuación en el sentido SND,
- las características de conmutación,
- el nivel de activación mínimo en el sentido SND,
- la respuesta en frecuencia,
- la sensibilidad de la detección del ruido de fondo (nivel de activación, nivel absoluto, fluctuaciones de nivel).

Además, influyen en la calidad:

- el diseño del NLP o limitadores centrales junto con compensadores de eco,
- el diseño de los sistemas de reducción de ruido.

### **7.3 Métodos de análisis**

En general, la simulación de ruido de fondo puede hacerse con una señal continua (con espectro conformado) u otra señal más compleja que represente las condiciones reales (por ejemplo, el murmullo de voces de una oficina). En estos casos el ruido de fondo debe caracterizarse por su densidad espectral de potencia a largo plazo y el nivel medio aplicado durante la medición.

En las siguientes pruebas se mide la señal de ruido de fondo y no se considera por tanto componente perturbadora, realizándose los análisis sobre ella. La calidad de transmisión del ruido de fondo (desde el extremo próximo en el sentido SND) puede evaluarse en la modalidad de reposo, con habla en el extremo remoto y con habla próxima.

En todos estos casos son importantes los siguientes parámetros:

- la sensibilidad de la detección del ruido de fondo en cuanto al nivel de activación,
- el nivel absoluto de la señal de ruido transmitida,
- las fluctuaciones de nivel de la señal de ruido transmitida.

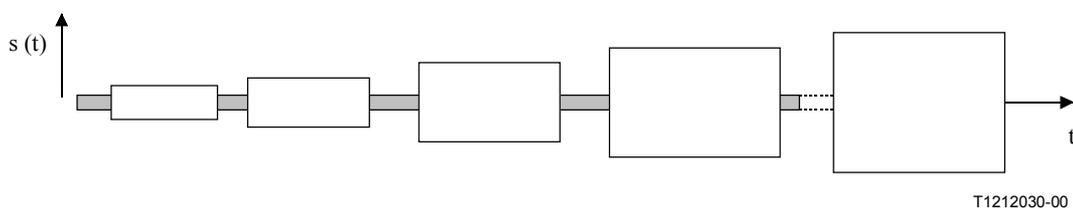
Como las evaluaciones auditivas de los requisitos de las propiedades de transmisión de ruido de fondo continúan realizándose, no existe por el momento una descripción detallada completa de los métodos de análisis.

En la **modalidad de reposo**, la transmisión del ruido de fondo puede medirse, por ejemplo, aplicando un campo de ruido difuso con las características  $H_{oth}$  o utilizando una señal de ruido de fondo adecuada, tal como el murmullo de voces, el ruido de coches u otros que representen situaciones características. El nivel de la señal se aplica, por ejemplo, durante un periodo de 20 segundos, comenzando con un nivel de excitación de 50 dB<sub>SPL</sub> (A) o el nivel correspondiente medido en la interfaz eléctrica. El nivel puede entonces incrementarse, por ejemplo, en 3 dB volviendo a aplicarse a durante 20 segundos. Esta secuencia de medida se construye con bloques de 20 segundos de ruido de  $H_{oth}$  incrementando cada bloque en 3 dB. Al aplicar cada porción de 20 segundos no debe detectarse ninguna variación audible del ruido de fondo. Los valores exactos de las constantes de tiempo de audibilidad no están aún disponibles. Cuando se midan fluctuaciones superiores a  $\pm 3$  dB con respecto al régimen permanente, hay que realizar una evaluación subjetiva.

En general la señal de ruido de fondo debe ser audible en cada momento.

El ruido de fondo transmitido debe ser lo más pequeño posible. Sin embargo es preciso evitar los efectos indeseados de los algoritmos de reducción de ruido. Quedan en estudio simulaciones de ruido de fondo más realistas.

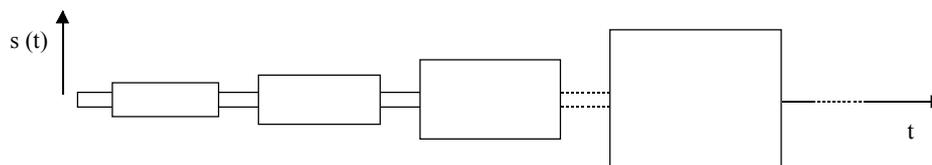
La siguiente estructura de señal puede utilizarse para evaluar la calidad de transmisión del ruido de fondo en el sentido de emisión, coincidiendo con el **habla en el extremo remoto**. La figura 16 muestra una señal de ruido continuo aplicada al extremo próximo (sentido de emisión, de color gris) y la simulación de la voz en el extremo remoto en el sentido de recepción (color blanco, pueden emplearse ráfagas de CSS). La medición se lleva a cabo en el sentido de emisión. En la figura 16 el nivel de las ráfagas CSS es variable y la simulación del ruido de fondo se aplica con un nivel constante.



NOTA – La línea de puntos indica la repetición o alargamiento de la señal de prueba para conseguir la longitud adecuada para la medición.

**Figura 16/P.502 – Ejemplo de estructura de señal de prueba para evaluar la calidad de la transmisión de ruido de fondo en el sentido de emisión (con simulación de habla en el extremo remoto)**

Puede utilizarse una estructura de señal parecida para determinar la calidad de la transmisión del ruido de fondo coincidente con habla en el extremo próximo (véase la figura 17). En este caso, el ruido de fondo y la simulación de la señal vocal (puede utilizarse también una CSS) se aplican y se miden en el mismo sentido (contrariamente a lo que muestra la figura 16), es decir en sentido de emisión.



T1212040-00

NOTA – La línea de puntos indica la repetición o alargamiento de la señal de prueba para conseguir la longitud adecuada para la medición.

**Figura 17/P.502 – Ejemplo de estructura de señal de prueba para evaluar la calidad de la transmisión de ruido de fondo en sentido de emisión (con simulación de voz en el extremo próximo)**

La señal de ruido de fondo y las ráfagas de la CSS se dibujan con el mismo color blanco para indicar que ambas componentes se inyectan en el mismo sentido.

## 8 Características de conmutación

La conmutación puede repercutir en la calidad de transmisión vocal en diversas situaciones y condiciones: en monoloxua y en habla simultánea estando activos el interlocutor del extremo próximo y el del extremo remoto. En cualquier caso podrían truncarse las sílabas y hasta perderse o interrumpirse palabras y frases enteras.

### 8.1 Percepción subjetiva de la degradación de la calidad vocal

La percepción subjetiva de la degradación de la calidad vocal puede describirse como

- silencios;
- sílabas ausentes;
- palabras o frases incompletas.

Los individuos suelen interpretar como silencios las perturbaciones introducidas. Suele producirse interacción entre el eco y la conmutación.

### 8.2 Parámetros objetivos relacionados

Los parámetros objetivos relacionados son harto conocidos. La Recomendación UIT-T P.340 describe parámetros básicos tales como el tiempo de establecimiento, el de retención (bloqueo), el de conmutación, etc. Los parámetros objetivos más importantes son:

- el intervalo de atenuación;
- los tiempos de conmutación ( $T_S$ ), de bloqueo (retención) ( $T_H$ ) y de establecimiento ( $T_R$ );
- la atenuación en sentido SND/RCV en habla simultánea;
- el nivel de activación mínimo para conmutar del sentido de emisión al de recepción y viceversa, y

– la atenuación del eco.

Los parámetros fundamentales de la conmutación vocal son: El nivel umbral ( $V_{TH}$ ), el tiempo de establecimiento ( $T_R$ ), el tiempo de retención (bloqueo) ( $T_H$ ), el tiempo de conmutación ( $T_S$ ), y el intervalo de atenuación ( $a_H$ ). Si se escogen adecuadamente los parámetros de conmutación, se puede reducir al mínimo la degradación de la calidad vocal provocada por la conmutación vocal. Si dichos valores, en especial los tiempos de conmutación, no fuesen los adecuados podrían presentarse efectos de mutilación grave y pérdida de las consonantes iniciales y finales del habla.

Hay que escoger los niveles umbrales de modo que la conmutación no quede interrumpida por fuentes de ruido aleatorio (ambiental) en ningún extremo de la conexión. Además, los efectos del ruido ambiental, del local o de la red, sobre el umbral no deben empeorar la calidad de funcionamiento. Los niveles de ruido ambiente pueden utilizarse para mejorar la calidad umbral de funcionamiento, ya que los interlocutores tienden a hablar más alto en los ambientes ruidosos que en los silenciosos.

El tiempo de establecimiento debe ser suficientemente pequeño para que no se pierdan las componentes vocales en régimen transitorio, aunque no tanto como para que la supresión de la atenuación de inserción resulte ruidosa.

El tiempo de retención (bloqueo) debe ser suficientemente grande para cubrir las pausas vocales, evitando así la conmutación intermitente no deseada antes de que el interlocutor inicial haya acabado, aunque también suficientemente corta para permitir la entrada razonable del segundo interlocutor.

El tiempo de conmutación de un estado activo al siguiente debe equilibrarse para optimizar la simulación del dúplex completo. El tiempo de conmutación es función así mismo del tiempo de establecimiento y del tiempo de retención (bloqueo).

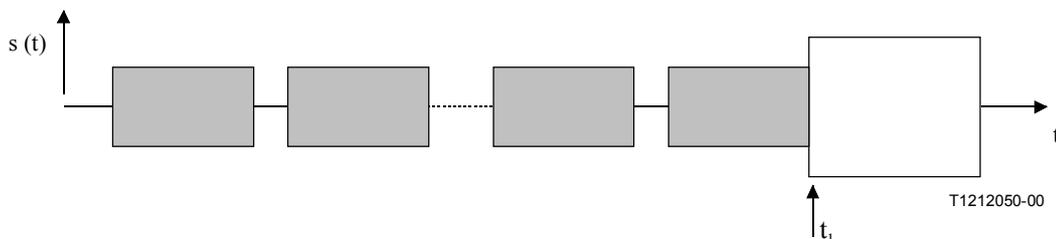
El intervalo de atenuación se obtiene de la diferencia entre el nivel máximo a plena activación y el mínimo alcanzado justo tras la inversión de la transmisión.

### 8.3 Métodos de análisis

Todos los niveles consignados en los cuadros 5 a 7 se refieren al MRP cuando se trata de niveles acústicos o al punto eléctrico de referencia en los demás casos. Cuando se trata de acceso eléctrico y no acústico, los niveles se encierran entre corchetes.

#### 8.3.1 El intervalo de atenuación y las características de conmutación

Uno de los parámetros más importantes, especialmente cuando se trata de montajes con dispositivos de conmutación de nivel, es el **intervalo de atenuación**. Este parámetro puede calcularse con la estructura de la señal de prueba de la figura 18.



NOTA – La línea de puntos indica la repetición o alargamiento de la señal de prueba para conseguir la longitud adecuada para la medición.

**Figura 18/P.502 – Estructura de la señal de prueba para la medición del intervalo de atenuación**

Para activar un trayecto de transmisión se utiliza una secuencia de ráfagas de CSS repetida periódicamente (en gris), simulando el habla. Al final de una de las ráfagas de CSS, señalada por  $t_1$  en el eje de tiempos, la señal de medida se aplica al trayecto opuesto (en blanco). Esta última consiste en la repetición periódica de un sonido vocalizado.

Los valores tipo se muestran en el cuadro 5.

**Cuadro 5/P.502**

	Señal de medida	Nivel de la señal de medida	Señal de activación (en sentido contrario)	Nivel de la señal de activación (en sentido contrario)
Conmutación <b>RCV → SND</b>	Sonido vocalizado en sentido SND, repet. periódica	-3 dBPa -16,7 dBm	CSS en RCV	-18,3 dBm (incl. pausas)
Conmutación <b>SND → RCV</b>	Sonido vocalizado en sentido RCV, repet. periódica	-16,7 dBm	CSS en SND	-4,7 dBPa (-18,3 dBm) (incl. pausas)

Pueden medirse los siguientes parámetros:

- **Intervalo de atenuación**

El intervalo de atenuación se mide activando el sentido contrario antes de medir el intervalo de atenuación en el sentido de la prueba. El intervalo de atenuación se define como la diferencia entre los niveles mínimo y máximo de la señal de prueba transmitida, referida a la señal de excitación. La medición se realiza sencillamente evaluando la variación del nivel en el tiempo. Las constantes de tiempo para esta medición suelen estar en el entorno de 5 ms.

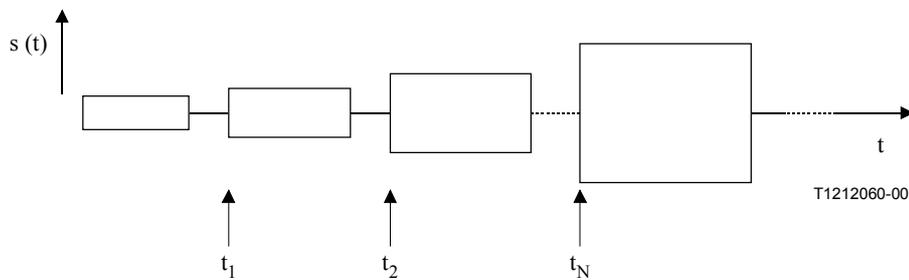
Los límites del intervalo de atenuación de los distintos tipos de terminales manos libres figuran en la Recomendación UIT-T P.340. Estos límites podrían aplicarse a otros dispositivos con conmutación.

- **Características de conmutación (de las señales semejantes al habla), por ejemplo, tiempos de conmutación**

Las definiciones y los límites genéricos de las características de conmutación se encuentran en la Recomendación UIT-T P.340. Las mediciones se realizan básicamente del modo explicado anteriormente, lo que significa que primero se activa el sentido opuesto y después el sentido que se prueba (véase la señal de prueba de la figura 18). El nivel de la señal de prueba transmitida se mide en función del tiempo. Las constantes de tiempo a aplicar en la medición se encuentran así mismo en el entorno de 1 ms.

### 8.3.2 El nivel umbral y el tiempo de establecimiento (para el nivel de activación mínimo)

La estructura de señal de la figura 19 representa elementos de señal de niveles crecientes. El **nivel de activación mínimo** para conmutar en el sentido RCV o SND, desde la modalidad de reposo, puede determinarse mediante dichas secuencias. Para esta señal resultan adecuados periodos de CSS (como simulación de habla) con niveles crecientes.



NOTA – La línea de puntos indica la repetición o alargamiento de la señal de prueba para conseguir la longitud adecuada para la medición.

**Figura 19/P.502 – Estructura de la señal de prueba para determinar el nivel de activación mínimo**

Los valores tipo se han elegido como se muestra en el cuadro 6.

**Cuadro 6/P.502**

	<b>Duración activa/ duración pausa</b>	<b>Nivel del primer periodo</b>	<b>Diferencia de nivel entre dos periodos</b>
CSS para conmutar en <b>sentido SND</b>	248,62 ms/ 451,38 ms	-23 dBPa (nota) (-36,7 dBm)	1 dB
CSS para conmutar en <b>sentido RCV</b>	248,62 ms/ 451,38 ms	-36,6 dBm (nota)	1 dB
NOTA – Estos niveles deben ser suficientemente pequeños para conseguir un amplio intervalo de medida.			

Si las señales transmitidas se miden con referencia a la señal original, es posible determinar el nivel de activación mínimo, pudiéndose analizar la activación al principio de cada ráfaga de señal ( $t_1, t_2, \dots, t_N$ ).

Los parámetros que pueden determinarse mediante esta señal son los siguientes:

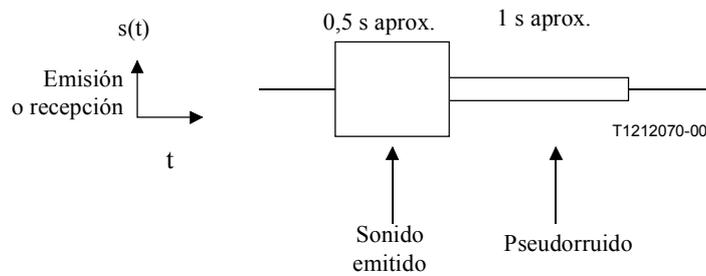
- **Nivel umbral (para señales semejantes al habla)**  
La secuencia de medición se muestra en la figura 19. Para determinar el nivel de activación mínimo sólo hace falta un sencillo análisis de la variación del nivel en el tiempo. La constante de tiempo de la medición se toma entre 1 y 5 ms, midiéndose la variación de nivel de la señal en el tiempo. Se muestra la señal de excitación. El nivel de excitación mínimo para que se active el dispositivo objeto de la prueba puede calcularse a partir de la diferencia de nivel de las partes activas de la señal de origen compuesto. Como se conoce el nivel de excitación, puede calcularse el nivel umbral mínimo.
- **Tiempos de establecimiento (para señales semejantes al habla, dependientes del nivel)**  
El análisis es básicamente análogo al descrito en 8.3.1, con excepción de la constante de tiempo que aquí es 1 ms. Por consiguiente, el tiempo de conmutación se determina calculando la curva de variación del nivel en el tiempo.

### 8.3.3 Tiempo de retención (bloqueo)

La transición de la activación al reposo puede representarse inyectando una señal de activación (por ejemplo, sonido vocalizado de CSS) en un cierto sentido, seguida de otra de nivel inferior en el mismo sentido, que no activa el teléfono manos libres (señal de ruido) (véase la figura 20). La

segunda parte de la señal así medida indica la atenuación, pudiendo calcularse el tiempo de retención (tiempo de desconexión) a partir de ésta.

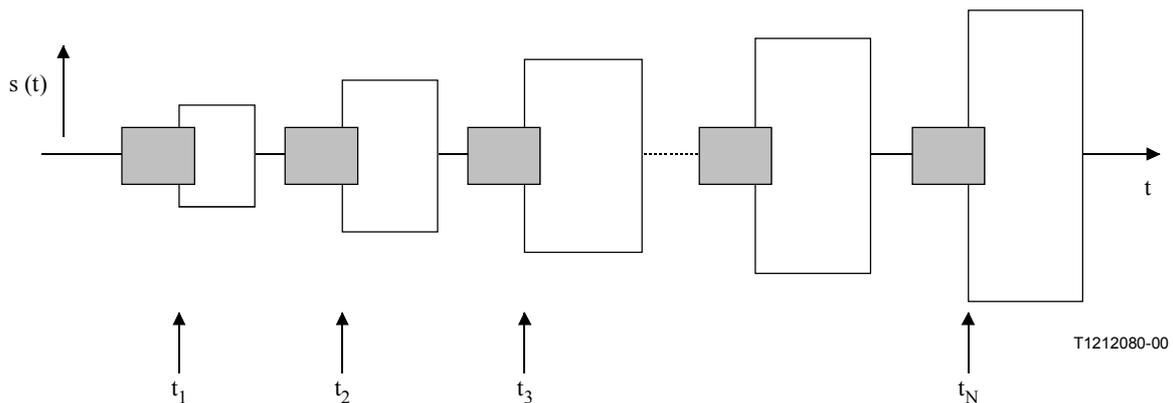
La duración del sonido emitido es 0,5 s con objeto de lograr la estabilidad final del sistema. Esto corresponde a los niveles normales. Si se necesitase realizar la evaluación en función del nivel, se podrían seleccionar los niveles definidos en 8.3.2. La segunda parte de la señal (la señal de ruido) tiene una duración de 1 s. El nivel debe tomarse suficientemente pequeño para no activar el equipo. Los niveles propuestos para aplicar son:  $-34,7$  dBPa para los terminales en emisión,  $-50$  dBm para el acceso eléctrico en emisión y  $-50$  dBm para la recepción.



**Figura 20/P.502 – Medición de la respuesta a la desconexión**

### 8.3.4 Nivel umbral y tiempo de conmutación de RCV a SND (SND a RCV)

Cuando haya que medir los **niveles umbrales para conmutar** de RCV a SND (y viceversa, o sea: del sentido SND a RCV), pueden utilizarse las señales de prueba mencionadas con modificaciones de menor importancia. Como puede observarse en la figura 21, se necesita una señal adicional en el sentido de transmisión opuesto (en gris). El nivel de la señal que se mide (en blanco) se incrementa también periódicamente. Cuando haya que determinar las características de conmutación aplicando señales semejantes a la voz, resulta conveniente utilizar periodos de CSS para las señales de la figura 21. Estas señales tampoco deben estar correlacionadas.



NOTA – La línea de puntos indica la repetición o alargamiento de la señal de prueba para conseguir la longitud adecuada para la medición.

**Figura 21/P.502 – Estructura de la señal de prueba para determinar el nivel de activación mínimo para efectuar la conmutación**

Los valores adecuados se indican en el cuadro 7.

**Cuadro 7/P.502**

	<b>Duración activa/ duración en pausa</b>	<b>Nivel del primer periodo</b>	<b>Diferencia de nivel entre dos periodos</b>	<b>Nivel (parte activa) en el sentido de transmisión opuesto</b>
CSS para conmutar al <b>sentido SND</b>	248,62 ms/ 451,38 ms	-13 dBPa (-26,7 dBm)	1 dB	-16,7 dBm (RCV)
CSS para conmutar al <b>sentido RCV</b>	248,62 ms/ 451,38 ms	-26,7 dBm	1 dB	-3 dBPa (SND) (-16,7 dBm)

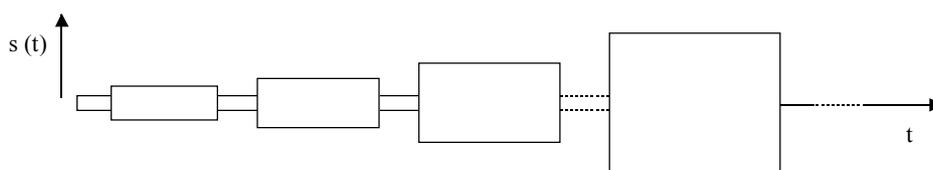
Una vez más es posible analizar la activación al principio de las ráfagas de señal ( $t_1, t_2, \dots, t_N$ ).

Por otra parte, pueden realizarse las mismas pruebas con una simulación de ruido de fondo aplicada al trayecto de transmisión opuesto. Los parámetros susceptibles de evaluación son los siguientes:

- **El nivel de activación mínimo para provocar la conmutación (con señales semejantes a la voz)**  
El nivel de activación mínimo para provocar la conmutación se determina del mismo modo que el nivel de activación mínimo, con la única diferencia de que la evaluación se realiza durante las pausas de la secuencia de habla simultánea. (Véanse los tiempos  $t_1, t_2, \dots$  en la figura 19).
- **Los tiempos de conmutación (cambio)**  
El análisis se realiza de manera análoga a la descrita para los tiempos de conmutación necesarios para la activación mínima (8.3.2).

### 8.3.5 Características de conmutación en presencia de ruido de fondo

La estructura de la señal de la figura 22 puede utilizarse para determinar las **características de conmutación en presencia de ruido de fondo**. En este caso se aplican simultáneamente al mismo canal una señal semejante a la voz (CSS) y una simulación de ruido de fondo. Para la CSS pueden utilizarse los parámetros de los cuadros anteriores.



T1212090-00

NOTA – La línea de puntos indica la repetición o alargamiento de la señal de prueba para conseguir la longitud adecuada para la medición.

**Figura 22/P.502 – Estructura de la señal de prueba para determinar el nivel de activación mínimo en presencia de ruido de fondo**

Los parámetros a determinar son los siguientes:

- **El nivel de activación mínimo en presencia de ruido de fondo**  
Para efectuar esta medición se aplica un ruido de fondo (además de la señal de excitación, véase la figura 22). Se utiliza ruido de  $H_{0th}$  o ruido de fondo característico (a ser posible sin

grandes fluctuaciones de nivel) cuyo nivel convenga a la utilización normal del dispositivo (por ejemplo, la del teléfono manos libres sometido a prueba). Para los teléfonos de oficina el nivel típico está comprendido en la gama de  $-54$  a  $-44$  dBPa(A). Para otras aplicaciones resultarían convenientes otros tipos de ruido de fondo con niveles distintos, por ejemplo: ruido de automóvil.

El nivel de activación mínimo se determina de manera similar a la descrita para el nivel de activación mínimo sin ruido de fondo, escogiéndose las secuencias de la señales de medida de manera similar.

- **El tiempo de establecimiento en presencia de ruido de fondo**

Una vez más se utiliza el mismo procedimiento que el descrito anteriormente para evaluar los tiempos de conmutación para el nivel umbral mínimo (8.3.2). La única diferencia es la presencia de ruido de fondo aplicado de la manera antes descrita.

## ANEXO A

### Explicación detallada de la metodología de prueba de la $ERL_t$ ponderada en el tiempo

#### A.1 Algoritmo de la atenuación de retorno del eco

A continuación se presenta el método de medición de la atenuación de retorno del eco ponderada en el tiempo,  $ERL_t$ . Este método exige la grabación de las señales fuente y eco mientras dure la medición, realizándose el tratamiento de la misma posteriormente. Es posible utilizar técnicas de medición en tiempo real, aunque no estén descritas en la presente Recomendación UIT-T.

No conviene bloquear el compensador en las pruebas de la ERL. Ciertas pruebas realizadas con señales no estacionarias han puesto de manifiesto que los tiempos de convergencia y la consiguiente ERL convergida, una vez "desbloqueado" el compensador, dependían del instante de bloqueo de aquel.

##### A.1.1 Atenuación de retorno del eco, ponderada en el tiempo ( $ERL_t$ )

La ERL ponderada en el tiempo, o  $ERL_t$  tiene por objeto:

- Proporcionar una medida, la  $ERL_t$ , ponderada psicoacústicamente, de la atenuación de retorno del eco dependiente del tiempo, con la presencia de crestas.
- Ofrecer una estimación del número de ráfagas de eco potencialmente rechazables, y de la atenuación de retorno del eco, ponderada psicoacústicamente, durante las ráfagas.

En primer lugar se filtra la señal de eco para asemejarla a la selectividad en frecuencia del oído humano a un nivel de sonoridad de 30 fones, como se explica en A.1.2. La potencia del eco queda así ponderada asemejándose a la respuesta del oído humano.

A continuación puede aplicarse la reducción al ruido y pueden sincronizarse el fichero del eco y el del estímulo. La reducción del ruido permite realizar la medición de éste y su substracción de la señal de eco más ruido, para obtener una mejor estimación del eco por separado. Esta medición debería realizarse durante dos segundos, como mínimo, una vez terminada toda la actividad del estímulo. El eco y la fuente se convierten en tramas de 4 ms promediadas en potencia proporcionando la resolución idónea y la inmunidad a los errores de sincronización.

Si el estímulo está inactivo, el algoritmo se limita a saltar dicha trama, pasando a la siguiente trama de eco y estímulo. Si el estímulo es declarado activo, la trama del eco se compara con un umbral para determinar si se trata de un "evento" de eco. Estos eventos se ponderan acto seguido utilizando un modelo psicoacústico.

Puede calcularse la  $ERL_t$  con un umbral de  $-65$  dB (5 dB por encima del umbral mínimo de ruido de la ley A o la ley  $\mu$ ). Análogamente, en el caso de la ley A, el umbral debe encontrarse 5 dB por encima del umbral mínimo de ruido. El algoritmo de la prueba real en pseudocódigo se explica en detalle en A.1.4.

### A.1.2 Audibilidad del eco modelado

El algoritmo para modelar la audibilidad del eco considera tres aspectos fundamentales del comportamiento auditivo humano:

- 1) La selectividad en frecuencia del oído humano a niveles de sonoridad de 30 fones (respuesta "Fletcher-Munson" equivalente a 30 dB a 1 kHz) [1].

Se ha escogido el valor de 30 fones porque representa el nivel de eco de los terminales de microteléfono que se aproximan a las especificaciones de la atenuación de acoplamiento (calculadas mediante el análisis de planificación de la atenuación) sin alcanzarlas. El cambio de 20 a 50 fones proporciona básicamente la misma ponderación dentro de la banda de telefonía. Se utiliza un filtro de ponderación A.

Obsérvese que el empleo de esta característica de ponderación tan exacta supone la audición del tipo auricular/microteléfono, o la respuesta "presión media audible" (MAP, *mean audible pressure*). La escucha en campo libre, como ocurre en el caso de los terminales manos libres, exigiría la ponderación "campo audible medio" (MAF, *mean audible field*) de Robinson y Dadson, aunque hay poca diferencia. Se utilizará la ponderación MAP para reflejar del mejor modo posible la forma habitual de empleo del microteléfono.

La atenuación media del filtro con ruido blanco es 1,3 dB medida con la  $ERL_s$  o la  $ERL_t$ . Si las señales no son estacionarias la atenuación será función del tiempo.

- 2) La tendencia del oído a combinar la sonoridad de las señales secuenciales aunque sean discretas en el tiempo ("combinación temporal") suele ocurrir cuando ambas señales están separadas por un periodo de silencio inferior a 20 ms [2], [3], [4]. Cuando dos ráfagas de eco están separadas por un periodo de inactividad inferior a 20 ms, se consideran un sólo evento de eco de mayor duración, a efectos de la sonoridad. Esto rige hasta que la separación entre eventos es de 20 ms, caso en el que se vuelve a declarar otro evento de eco. Esto se puede asimilar a un bloqueo de 20 ms para el evento de eco actual. Durante este periodo de bloqueo, las potencias del estímulo y del eco no se consideran parte del evento. La figura A.1 proporciona un ejemplo de combinación temporal.

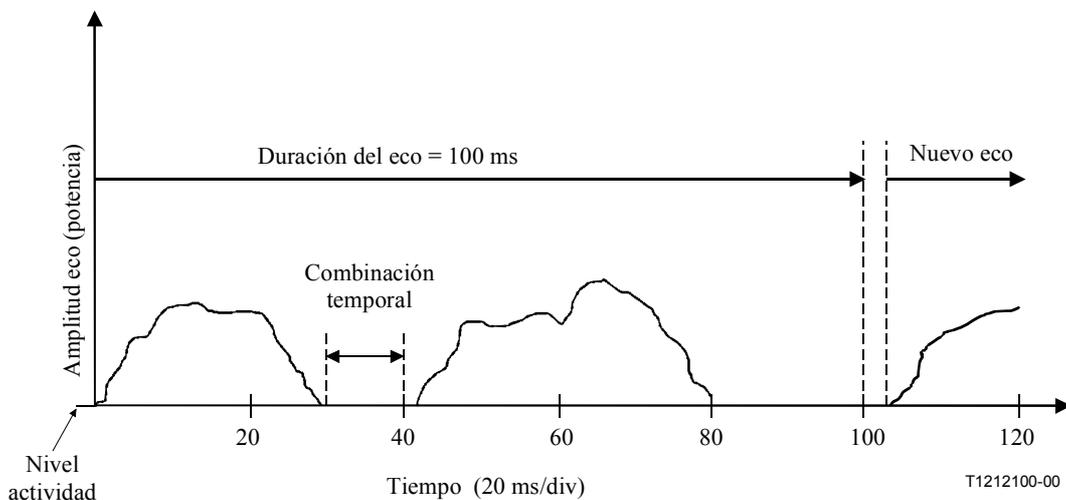


Figura A.1/P.502

- 3) La duración total del evento de eco tras una combinación temporal se mide de acuerdo con el comportamiento de integración temporal natural del oído humano. La duración total comprende todas las separaciones entre eventos que son capturados por la combinación temporal, pero no el bloqueo final de 20 ms.

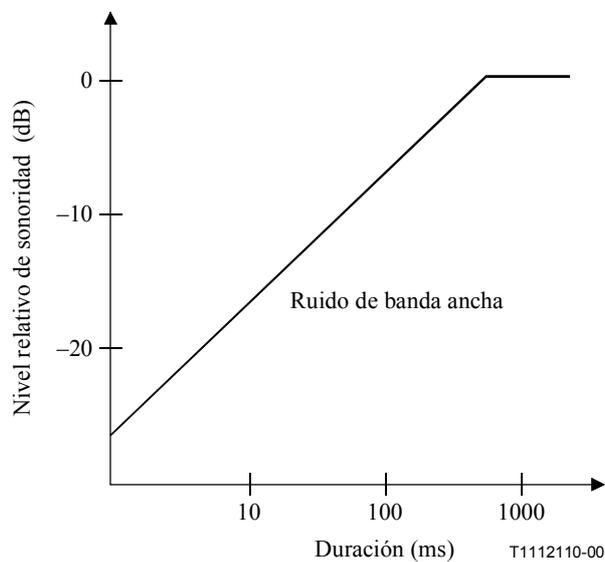
Si la duración es inferior a 750 ms, se reduce el nivel del evento para tener en cuenta el comportamiento de integración temporal del oído humano. Los estudios realizados sobre la audición con ruido han permitido definir la relación entre ambos mediante la siguiente ecuación:

$$\text{Ponderación de integración en el tiempo} = -23 + 8 \log(t) \text{ en dB}$$

siendo  $t$  = duración total del evento de eco (ms),  $t < 750$  ms.

Obsérvese que la relación es ligeramente diferente cuando se trata de tonos, sin embargo se ha considerado que el ruido es mucho más parecido a la auténtica esencia del eco que una senoide.

La figura A.2 muestra una representación gráfica de la ponderación en el tiempo.



**Figura A.2/P.502**

Si la duración es mayor que 750 ms, el nivel del evento total no se pondera. Obsérvese que los resultados de la prueba demuestran que las ráfagas de eco de menos de 750 ms suelen presentarse con cierta frecuencia en los compensadores de eco.

### **A.1.3 Expresión de los resultados de la $ERL_t$**

Los métodos tradicionales de cálculo de la ERL deducen su valor de la potencia del eco referida a la potencia de la fuente mientras dura la medición. Según este método, la potencia final ponderada del eco durante cada evento se refiere a la potencia de la señal fuente durante dicho evento, obteniendo la " $ERL_t$  activa",  $AERL_t$ , de cada evento. El eco se refiere a la señal fuente sólo durante el evento, porque así es como nuestro oído compararía el eco.

La atenuación ponderada de retorno del eco activo promediada a largo plazo se obtiene sumando la potencia de todo el eco ponderado durante los eventos activos y comparándola con la potencia de la fuente observada sólo durante los eventos. El resultado es la " $ERL_t$  activa a largo plazo".

Para comparación con los métodos tradicionales de ERL, se suma la potencia de todo el eco ponderado durante los eventos, luego se refiere a la potencia total de la fuente medida para toda duración de la medición. El resultado es la "ERL<sub>t</sub> a largo plazo".

Obsérvese que la terminología empleada en los resultados de ERL<sub>t</sub> coincide con la Recomendación UIT-T P.56.

Otras estadísticas recopiladas proporcionan la AERL<sub>t</sub> máxima y mínima, la desviación típica ("sigma") de la AERL<sub>t</sub>, la media de la AERL<sub>t</sub> y el número total de eventos de eco (los eventos combinados debidos al efecto "Haas" se consideran un solo evento). También se proporcionan: el número de eventos de eco por minuto, el tanto por ciento de habla libre de eventos de eco, el número de eventos de < 750 ms, la longitud media del evento y la duración de la inactividad de la fuente.

#### A.1.4 Algoritmo de prueba de la ERL<sub>t</sub>

El método ERL<sub>t</sub> para evaluar la atenuación de retorno de eco de un terminal con modelado psicoacústico y para predecir la presencia de ecos potencialmente inaceptables, tiene en cuenta tres aspectos fundamentales de la audición humana:

- la selectividad en frecuencia del oído humano (respuesta "Fletcher Munson");
- la suma en el tiempo del nivel de los eventos separados menos de 20 ms (efecto "Haas");
- la integración en el tiempo de los estímulos de menos de 750 ms.

Los detalles de implementación del algoritmo se presentan a continuación.

Se utiliza una señal fuente como la de las cláusulas 4 y 5. Se recomienda utilizar señales de estímulo semejantes al habla, porque los resultados que se obtienen son los más representativos de las condiciones reales. La salida del sistema consiste siempre en ecos o ruidos que atraviesan el sistema sin ser compensados.

Hay que grabar en formato digital tanto el estímulo como el eco. Hay que obtener las entradas del usuario en función del tipo de aparato (analógico o digital) y las pruebas de EPD<sub>n</sub>, del habla simultánea y del monólogo. Deben utilizarse parámetros de calibración para ajustar la escala de las tramas del eco y el estímulo a valores absolutos, y el procesamiento híbrido debe haber compensado el eco híbrido en los aparatos analógicos de dos hilos.

El fichero de estímulo y el del sonido se procesarán como valores de potencia promediados durante tramas de 4 ms. Las sucesivas tramas del fichero de estímulo se denominarán  $x_i$ , y las tramas del eco se representarán por  $y_i$ , siendo  $i = 1, 2, 3...$  el índice de la trama. Las tramas intermedias correspondientes a un "evento de eco" se representarán por  $x_k$ , e  $y_k$ , siendo  $k = 1, 2, 3...$  el índice del evento de eco, que se restaura al terminar un evento y comenzar otro.

Las estadísticas elaboradas durante la medición de la ERL<sub>t</sub> proporcionan la ERL<sub>t</sub> activa a largo plazo (ALERL<sub>t</sub>, *active long term ERL<sub>t</sub>*), la ERL<sub>t</sub> a largo plazo (LERL<sub>t</sub>, *long term ERL<sub>t</sub>*), la ERL<sub>t</sub> activa máxima y mínima (MINERL, MAXERL), su desviación típica y su media, y el número total de eventos de eco (considerándose los eventos combinados debidos al efecto "Haas" como un solo evento) (NEVENTS). También proporcionan: el número de eventos por minuto (NEVMIN), el porcentaje de habla libre de eventos de eco (PER), el número de eventos < 750 ms (N750), la longitud media del evento (AVGEVENT), y la duración de la inactividad del estímulo (DUR). La terminología empleada en los resultados de ERL<sub>t</sub> coincide con la de la Recomendación UIT-T P.56. La duración de la inactividad del estímulo no se incluye en los resultados función del tiempo.

## El algoritmo ERL<sub>t</sub>

- *Paso 1* (Optativo pero recomendable)  
Calcular la correlación ente el fichero del estímulo y el del eco para ajustar con precisión EPD<sub>n</sub>. Adoptar el criterio de que la cresta de correlación presente se produce en EPD<sub>n</sub> salvo que exista otra cresta de correlación posterior cuya magnitud sea por lo menos 10 dB superior. Esta directriz es aproximada y se inspira en los estudios subjetivos sobre la detección del retardo con varios impulsos.
- *Paso 2*  
Alinear el fichero del eco y el del estímulo suprimiendo el retardo igual a EPD<sub>n</sub> del fichero del eco.
- *Paso 3*  
Las muestras individuales de eco se filtran para ajustar la presión media audible al contorno de sonoridad de 30 fones. Este ajuste puede hacerse con precisión ( $\pm 1$  dB entre 200 Hz y 2500 Hz) utilizando un filtro paso alto de primer orden con el punto de  $-3$  dB en 800 Hz.
- *Paso 4*  
Si puede suponerse que el ruido en el trayecto del eco es estacionario y no está correlacionado con el eco, el ruido se mide durante 2 segundos una vez terminada la actividad de la fuente y del eco. A continuación el eco se resta del eco más ruido para obtener una estimación más exacta del eco por separado.
- *Paso 5*  
Las muestras se convierten en valores absolutos utilizando los datos de calibración. Las muestras de estímulo se combinan en tramas promediadas en potencia de 4 ms representadas por  $x_i$ . Las muestras de eco, ponderadas y con el ruido filtrado se combinan para formar tramas promediadas en potencia de 4 ms representadas por  $y_i$ .
- *Paso 6 – Empezar a calcular la atenuación de retorno del eco*  
Inicializar las variables:  
 $i = 0$  (contador de tramas);  
 $j = 0$  (contador de tramas mientras la señal está inactiva);  
 $n_{k=0}$  (número de tramas en el actual evento de eco);  
NSAMPS = 0 (número acumulado de tramas de la totalidad de eventos);  
HAAS = 0 (contador hasta 20 ms);  
 $e_{i=0}$  (suma actual de la potencia total del eco de todos los eventos una vez ponderados, correspondiente al contador de tramas  $i$ );  
 $p_{i=0}$  (suma actual de la potencia total del estímulo durante la medición, correspondiente al contador de tramas  $i$ );  
 $e_{k=0}$  (suma actual de la potencia del eco durante el evento de eco específico tras la ponderación, correspondiente al contador de tramas  $k$ );  
 $s_{k=0}$  (suma actual de la potencia del estímulo durante el evento de eco específico tras la ponderación, correspondiente al contador de tramas  $k$ );  
WEIGHT = 0 (ponderación en el tiempo del evento más reciente);  
LEVENT = 0 (nivel de la atenuación de retorno del eco del evento más reciente, tras la ponderación);  
NEVENT = 0 (número total de los eventos de eco);  
N750 = 0 (número total de los eventos de eco  $< 750$  ms);

MINERL = 75 (nivel mínimo de la atenuación de retorno del eco de todos los eventos);  
 MAXERL = 0 (nivel máximo de la atenuación de retorno del eco de todos los eventos);  
 EVENT[NEVENT] = 0 (inicializar a cero la serie de todos los niveles de atenuación de evento (en dB) que se utiliza para calcular la desviación típica);  
 TEMPSK = 0 (suma actual de la potencia del estímulo durante todos los eventos);  
 SUM = 0 (se utiliza para calcular la desviación típica);  
 SQ = 0 (se utiliza para calcular la desviación típica).

• *Paso 7*

Incrementar el contador de tramas y leer la potencia media de eco durante 4 ms  $y_i$ , y la potencia media de estímulo durante 4 ms,  $x_i$ ; si no hay más entradas válidas y ambos ficheros de la medición están completos, ir al paso 8.

1  $i = i + 1$  (salvo la última  $i$ , después ir al paso 8).

Sumar las potencias de estímulo:

$$p_i = p_i + x_i$$

¿Es el estímulo bastante intenso para que sea válido el cálculo de la atenuación del eco? Si no es así, desecha la trama actual y pasa a la siguiente.

4 Si  $x_i < (\text{nivel valor eficaz del estímulo a largo plazo} - 25 \text{ dB})$

$$j = j + 1$$

$$i = i + 1$$

Ir a 4.

De lo contrario:

Comparar el eco con el umbral:

Si  $y_i < -65 \text{ dB}$  {5 dB por encima del umbral mínimo de ruido de la ley A o la ley  $\mu$ }

Incrementar el contador de eventos de trama:

$$k = k + 1$$

Incrementar la longitud del evento de trama incluyendo las pausas  $< 20 \text{ ms}$ :

$$n_k = n_k + 1 + \text{HAAS}$$

Poner a cero el contador Haas:

$$\text{HAAS} = 0$$

Acumular la potencia de eco del evento:

$$e_k = e_k + y_i$$

Acumular la potencia del estímulo durante el evento:

$$s_k = s_k + x_i$$

Ir a 1.

De lo contrario:

¿No ha habido eventos en los últimos 20 ms?

Si  $k = 0$

$$\text{HAAS} = 0$$

Ir a 1.

De lo contrario:

Ha habido eventos en los últimos 20 ms:

$$HAAS = HAAS + 1$$

¿Han transcurrido 20 ms sin evento tras un evento reciente?

$$\text{Si } HAAS * 4 < 20$$

Ir a 1.

De lo contrario:

Ha terminado un evento, añadir uno al contador de eventos:

$$NEVENT = NEVENT + 1$$

Incrementar el contador de duración total de eventos sumando la duración total en tramas del evento más reciente:

$$NSAMPS = NSAMPS + n_k$$

¿La duración del evento más reciente fue  $< 750$  ms?

$$\text{Si } n_k * 4 < 750$$

Calcular la ponderación de integración temporal para el evento de eco más reciente:

$$WEIGHT = 8 * \log_{10}(n_k * 4) - 23$$

Incrementar el contador del número de eventos ponderados en el tiempo:

$$N750 = N750 + 1$$

De lo contrario:

Calcular la atenuación de retorno del eco ponderada del evento más reciente en dB:

$$LEVENT = 10 * \log_{10}(s_k / e_k) - WEIGHT$$

Guardar las atenuaciones de eco mínima y máxima en dB:

$$\text{SI } LEVENT < MINERL; MINERL = LEVENT$$

$$\text{SI } LEVENT > MAXERL; MAXERL = LEVENT$$

Guardar la atenuación de retorno del eco del evento más reciente en dB para calcular más adelante la desviación típica:

$$EVENT(NEVENT) = LEVENT$$

Hacer lineal la atenuación de retorno del eco del evento más reciente; recalcular la potencia del eco lineal ponderado:

$$e_k = s_k / (10^{(LEVENT/10)})$$

Acumular toda la potencia de los eventos de eco para calcular  $ALERL_t$  y  $LERL_t$ :

$$e_i = e_i + e_k$$

Acumular toda la potencia de los estímulos para calcular  $ALERL_t$ :

$$TEMPSK = TEMPSK + s_k$$

Poner a cero las variables de los eventos de eco:

$$k = 0$$

$$n_k = 0$$

$$WEIGHT = 0$$

$$HAAS = 0$$

$$e_k = 0$$

$$s_k = 0$$

Ir a 1.

- *Paso 8*

Calcular la ERL<sub>t</sub> activa a largo plazo (ALERL<sub>t</sub>), la ERL<sub>t</sub> a largo plazo (LERL<sub>t</sub>), el número de eventos de eco por minuto (NEVMIN), el porcentaje de habla libre de eventos de eco (PER), la longitud media del evento (AVGEVENT) y la duración de la inactividad del habla (DUR).

NOTA – Comprobar el valor nulo de e<sub>i</sub> antes del cálculo; si e<sub>i</sub> = 0, hacer ALERL<sub>t</sub> y LERL<sub>t</sub> iguales a 100 dB.

$$\text{ALERL}_t = 10 * \log_{10}(\text{TEMPSK}/e_i)$$

$$\text{LERL}_t = 10 * \log_{10}(p_i/e_i)$$

$$\text{NEVMIN} = 60 * \text{NEVENT} / ((i-j) * 0,004) \quad \{\text{número de eventos por minuto}\}$$

$$\text{PER} = 100 * ((i-j) - \text{NSAMPS}) / (i-j) \quad \{\text{porcentaje de habla libre de eco}\}$$

$$\text{AVGEVENT} = \text{NSAMPS} * 4 / \text{NEVENT} \quad \{\text{longitud media del evento en milisegundos}\}$$

$$\text{DUR} = j ** 0,004$$

Calcular la desviación típica analizando la serie EVENT que contiene la atenuación de retorno del eco de cada evento; cada evento, independientemente de su duración, se somete a la misma ponderación para el cálculo de la desviación típica; la conclusión es que la transición entre eventos discretos es menos aceptable que su duración.

Bucle j de 1 a NEVENT:

$$\text{SUM} = \text{SUM} + \text{EVENT}(j)$$

$$\text{SQ} = \text{SQ} + \text{EVENT}(j) ** 2$$

FINBUCLE

$$\text{SIGMA} = \text{SQRT}(\text{SQ} / \text{NEVENT} - [\text{SUM} / \text{NEVENT}] ** 2)$$

Calcular la media de los eventos:

$$\text{MEAN} = \text{SUM} / \text{NEVENT}$$

- *Paso 9*

Mostrar los resultados de las estadísticas:

Imprimir ALERL<sub>t</sub>, LERL<sub>t</sub>, MINERL, MAXERL, NEVENT, NEVMIN, PER, N750, AVGEVENT, DUR, SIGMA, MEAN

## ANEXO B

### Filtros de medición en habla simultánea con el método A

Los ensayos en habla simultánea exigen la utilización de filtros de ranura y de paso banda a diversas frecuencias. A continuación se presenta un cuadro de valores aconsejables.

Se definen los siguientes términos

- fpl: frecuencia inferior a la que la banda de paso o banda suprimida está en –3 dB;
- fpu: frecuencia superior a la que la banda de paso o banda suprimida está en –3 dB;
- fsl: frecuencia inferior a la que la banda de paso o banda suprimida está en –aten dB;
- fsu: frecuencia superior a la que la banda de paso o banda suprimida está en –aten dB;
- aten: atenuación total nominal del filtro;
- aten: atenuación total real del filtro;

- rizo: rizo del filtro en dB ( $\pm$ );
- ganancia: ganancia del filtro paso banda (lineal) en la banda de paso;
- orden: orden del filtro paso banda en derivaciones (frecuencia de muestreo 8 kHz ). El tiempo de sobreoscilación del filtro paso banda es orden veces 125  $\mu$ s. En el filtro de banda suprimida el orden se refiere al de la bicuadrática (elíptica).

Tipo de filtro	FIR paso banda 500 Hz	FIR paso banda 1 kHz	FIR paso banda 1,75 kHz	FIR paso banda 2,75 kHz	IIR banda suprimida 500 Hz	IIR banda suprimida 1 kHz	IIR banda suprimida 1,75 kHz	IIR banda suprimida 2,5 kHz
fpl	495	990	1 733	2 475	400	800	1 450	2 100
fpu	505	1 010	1 767	2 525	610	1 250	2 100	2 950
fsl	435	900	1 611	2 302	435	900	1 610	2 300
fsu	570	1 100	1 900	2 715	570	1 100	1 900	2 715
aten	30	30	30	30	30	30	30	30
aten (real)	31	29,5	34	34	30	40		
rizo	1	1	3	1	1,5	1,5	1,5	1,5
ganancia	0,92	0,9	0,78	0,99				
orden	160	100	80	60	6	6	6	6

Conviene tener en cuenta que el tiempo de sobreoscilación del filtro paso banda influirá en la medición. Las mediciones deben empezar cuando haya terminado la sobreoscilación provocada por la aplicación inicial. Esto es necesario para poder realizar mediciones de referencia puras en las pruebas de atenuación y en las de mutilación. El mayor tiempo de sobreoscilación es de 20 ms para el filtro paso banda de 500 Hz. Como la ventana de promediación para la medición de las pruebas de atenuación es de 8 ms, hay que insertar el filtro paso banda  $20 + 8 = 28$  ms antes del inicio del habla simultánea, o bien  $60 - 0$  (véase la figura 7).

## ANEXO C

### Descripción de la secuencia de acondicionamiento

#### C.1 Acondicionamiento del compensador antes del habla simultánea

La información básica acerca de la temporización de una conversación se encuentra en la Recomendación UIT-T P.59: Secuencias habladas, pausas, habla simultánea y silencio recíproco. Las máscaras de eco citadas a continuación figuran en dicha Recomendación.

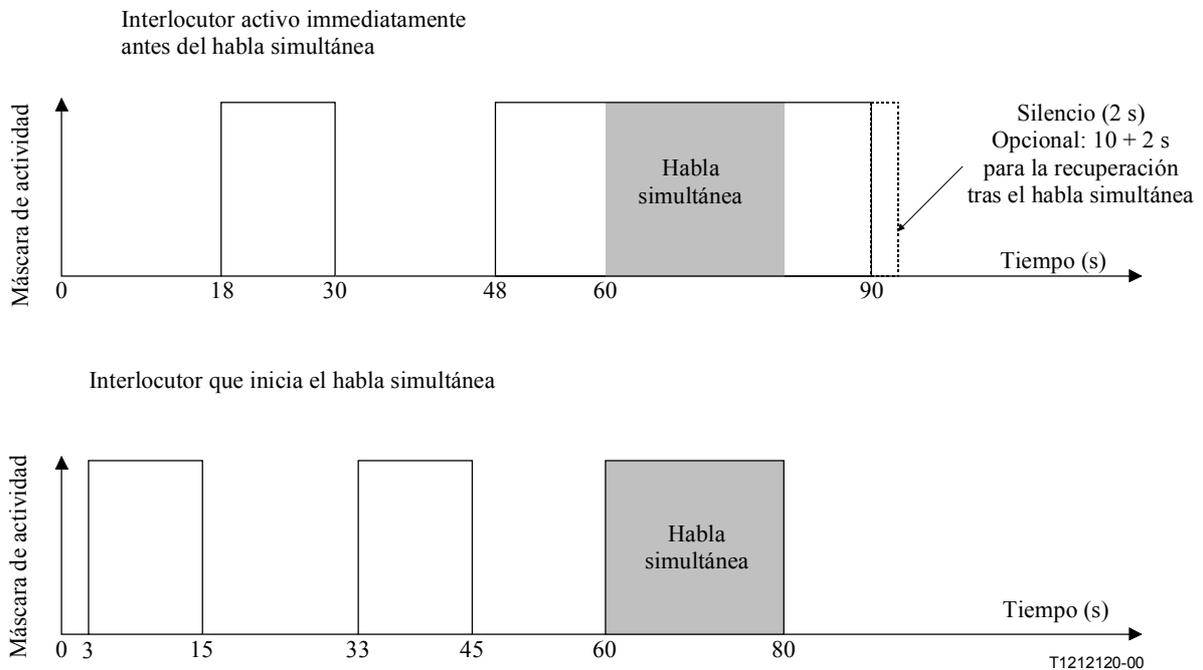
##### C.1.1 Máscaras de actividad para el acondicionamiento en habla simultánea

A continuación se definen las máscaras de amplitud exacta junto con las características de amplitud de la señal en habla simultánea. Cada tipo de prueba en habla simultánea tiene unos requisitos específicos de duración y de amplitud de la señal. Para la atenuación de retorno del eco, la duración debe ser suficientemente larga para captar cualquier divergencia, pero no tanto como para resultar una carga para los recursos de memoria del sistema de la prueba o como para que el tiempo de computación resulte inaceptable. Los ensayos han demostrado que 20 segundos es un valor aceptable para la duración de la prueba de la atenuación del retorno del eco. Una vez terminada el habla simultánea (y habiendo quedado inactivo el interlocutor que inició el habla simultánea), puede

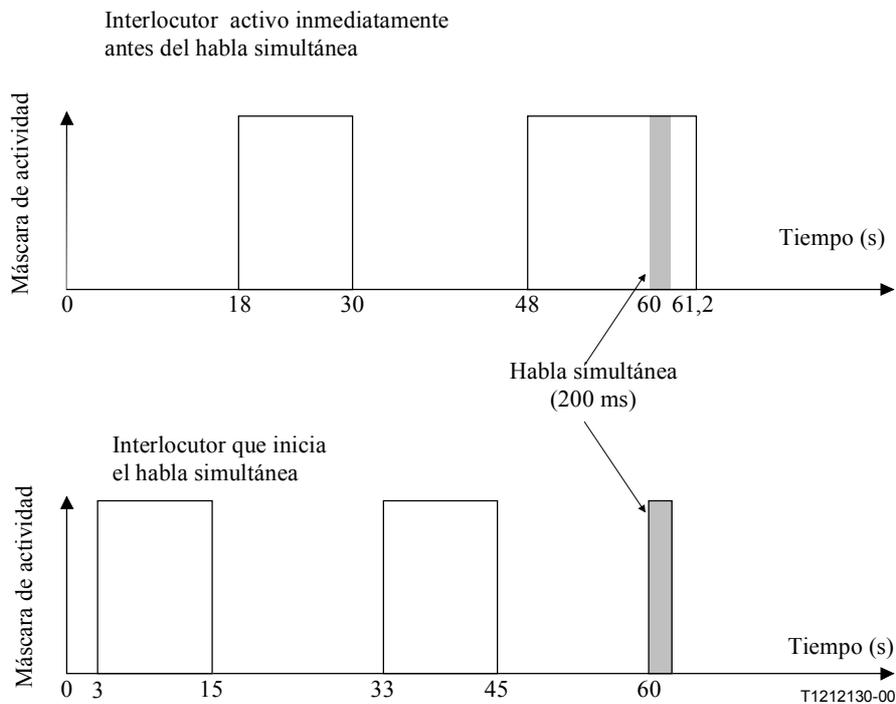
proseguir la medición de la atenuación de retorno del eco durante 10 segundos (permaneciendo activo el interlocutor que lo estaba inmediatamente antes del habla simultánea) para medir la recuperación tras el habla simultánea. Una vez transcurrido dicho tiempo hay que observar dos segundos de silencio. De este modo, se puede medir el ruido en el trayecto del eco. Si se supone que el ruido y el eco no están correlacionados y que el ruido es estacionario, el ruido medido en dichos dos segundos puede restarse del eco más ruido durante el habla simultánea para obtener una medida más exacta del eco en habla simultánea.

La duración del habla simultánea durante las pruebas de atenuación y mutilación de la misma puede ser mucho más pequeña. Como todas las constantes de tiempo objeto de estudio deben ser inferiores a 200 ms, la duración del habla simultánea se establece en 200 ms. El análisis prosigue (permaneciendo activo el interlocutor que lo estaba inmediatamente antes del habla simultánea) durante un segundo una vez finalizada el habla simultánea para que las pruebas de atenuación puedan medir cualquier supresión de la atenuación al volver a introducir el monólogo. En las pruebas de atenuación y mutilación, no es necesario realizar estimaciones y correcciones del ruido.

Las figuras C.1 y C.2 muestran las máscaras recomendables.



**Figura C.1/P.502 – Máscara de actividad para la prueba de la atenuación de retorno del eco**



**Figura C.2/P.502 – Máscara de actividad de las pruebas de atenuación y mutilación**

### C.1.2 Sincronización de las máscaras de actividad para el acondicionamiento en habla simultánea

Hay que sincronizar la temporización de las máscaras para evitar el habla simultánea prematura. Esto supone tener en cuenta el retardo de 1,5 ms en el trayecto aéreo entre la boca artificial y el HFT. La señal del simulador de boca debe empezar 1,5 ms antes que la señal  $R_{in}$ , lo que es posible atrasando 1,5 ms el fichero de estímulo en recepción. El comienzo del habla simultánea se define como el momento en que se observa actividad de emisión válida en la posición del micrófono y actividad de recepción válida en  $R_{in}$ .

En las gráficas de las máscaras de actividad, el punto de comienzo  $t = 0$  se refiere al principio del fichero aplicado a  $R_{in}$ . El punto de comienzo de la señal del simulador de boca puede considerarse  $t = -1,5$  ms, aunque el terminal las interpretará como sincronizadas. En la coordenada de 60 s, se introduce el habla simultánea y comienza la prueba correspondiente.

### C.1.3 Compensación de los filtros de la medición

Los métodos de prueba en habla simultánea exigen la inserción de filtros en el trayecto de audio. Dichos filtros influirán en la resolución en el dominio del tiempo y en el momento exacto de comienzo de la prueba del habla simultánea. Empleando filtros cuyos tiempos de sobreoscilación sean conocidos, se puede poner la medición en estado de espera mientras desaparece la sobreoscilación del filtro.

## APÉNDICE I

### Referencias bibliográficas

- [1] HEARING, GULICK, GESCHIEDER, FRISNA: *Oxford University Press*, 1989.
- [2] DAVIS (D.), DAVIS (C.): The LEDE Concept, *JAES*, 1985.
- [3] OLIVE (S.): The Detection of Reflections, *JAES*, 1987.

- [4] OLIVE (S.): Modification of Timbre by Resonance, *JAES*, 1988.
- [5] ZWICKER (E.), FASTL (H.): Psychoacoustics, *Springer Verlag*, 1990.
- [6] Enhancements of hands-free telecommunications, *Esprit Consortium, Annals of telecommunications*, 49 Nos. 7-8, 1994.
- [7] Methodology of Evaluation and Standards, Deliverable 1.2, *Freetel*, Julio de 1993.
- [8] GIERLICH (H.W.): The auditory perceived quality of hands-free telephones: auditory judgements, instrumental measurements and their relationship. *Speech Communication 20*, pp. 241-254, Octubre de 1996.
- [9] Subjective valuation procedures for hands-free telephones – Double talk performance. *ITU-T Contribución COM 12-5*, Ginebra, abril de 1997.
- [10] Subjective evaluation of hands-free telephones using conversational tests, specific double talk tests and listening only tests. *ITU-T Contribución COM 12-6*, Ginebra, abril de 1997.
- [11] Double talk measurements for hands-free telephones: Measurement proposals and measurement results. *ITU-T Contribución COM 12-32*, Ginebra, febrero de 1998.

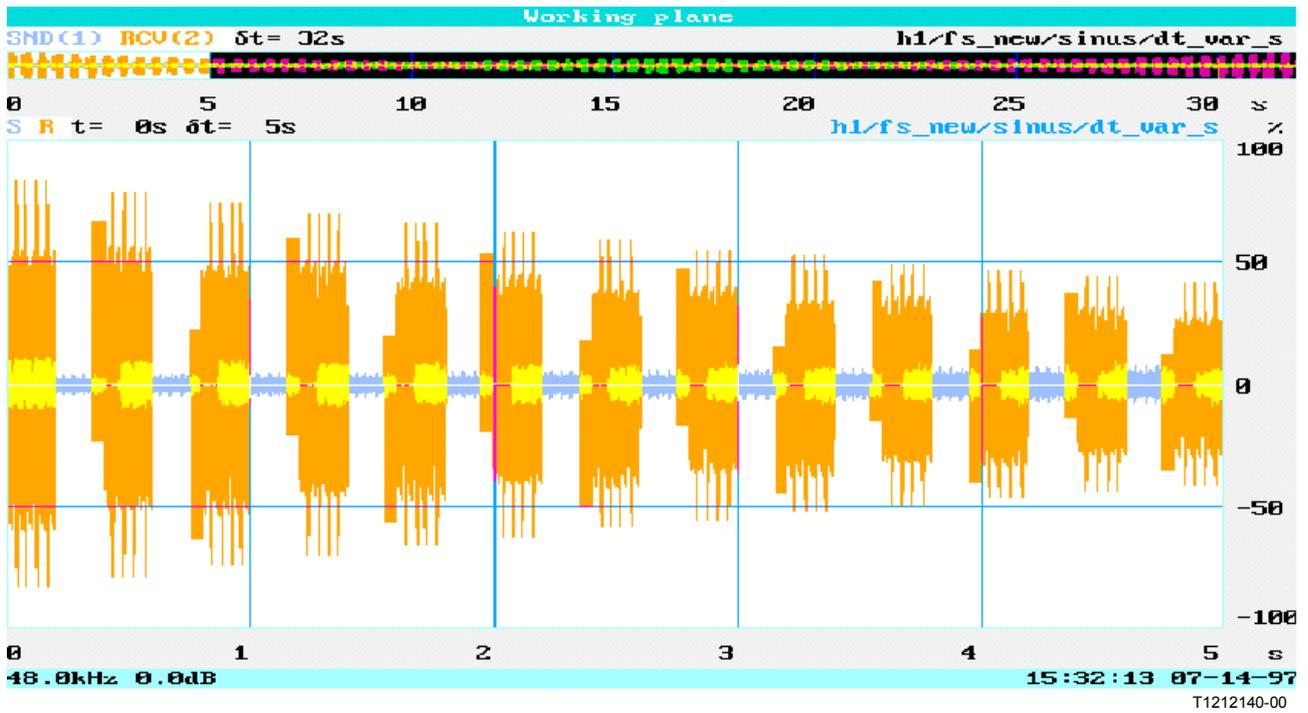
## APÉNDICE II

### Ejemplos de evaluaciones

#### **II.1 Ejemplos de las evaluaciones definidas en la cláusula 5**

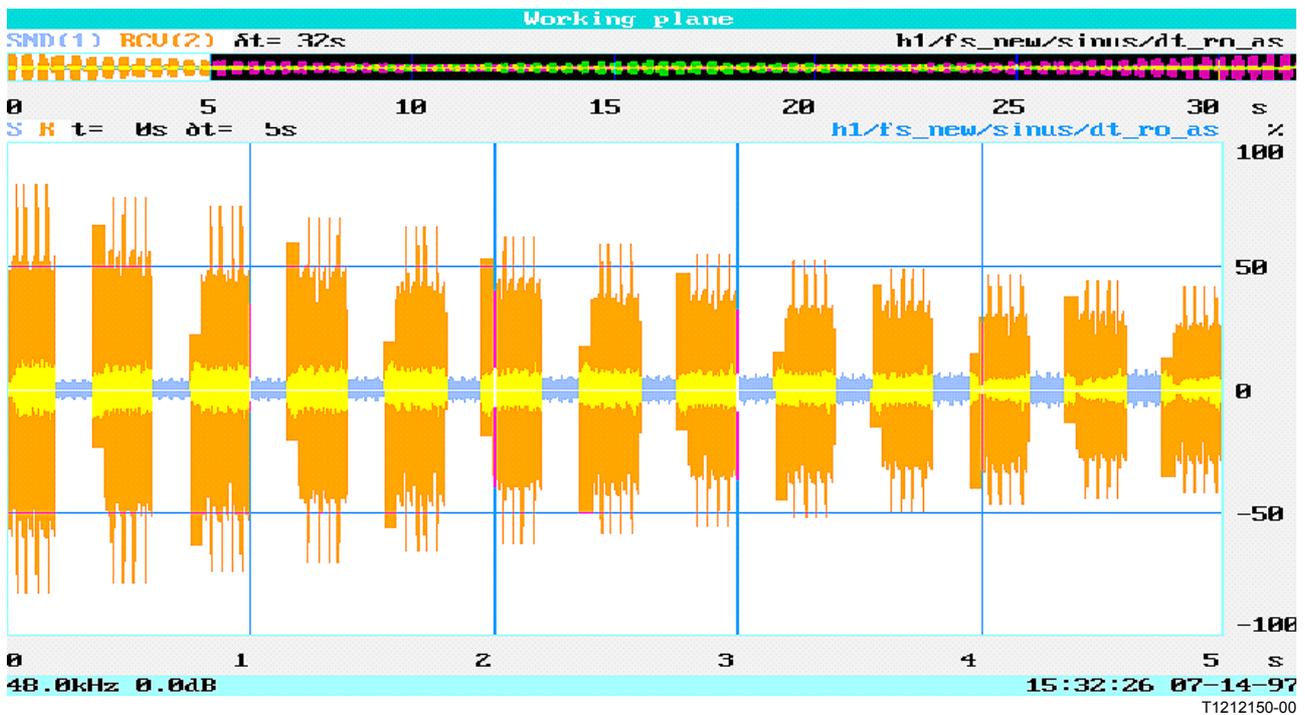
##### **II.1.1 Respuesta en frecuencia durante el habla simultánea**

En la figura II.1 se muestran algunos ejemplos de aplicación del procedimiento, inspirados en la evaluación de los teléfonos manos libres.



La señal de excitación en el sentido de emisión se muestra en tono oscuro.  
 La señal medida en el sentido de emisión se muestra en tono claro.  
 Las pausas contienen señal de habla simultánea transferida del altavoz del HTF a su micrófono.

**Figura II.1/P.502 – Sentido de emisión en un HFT de conmutación de nivel**

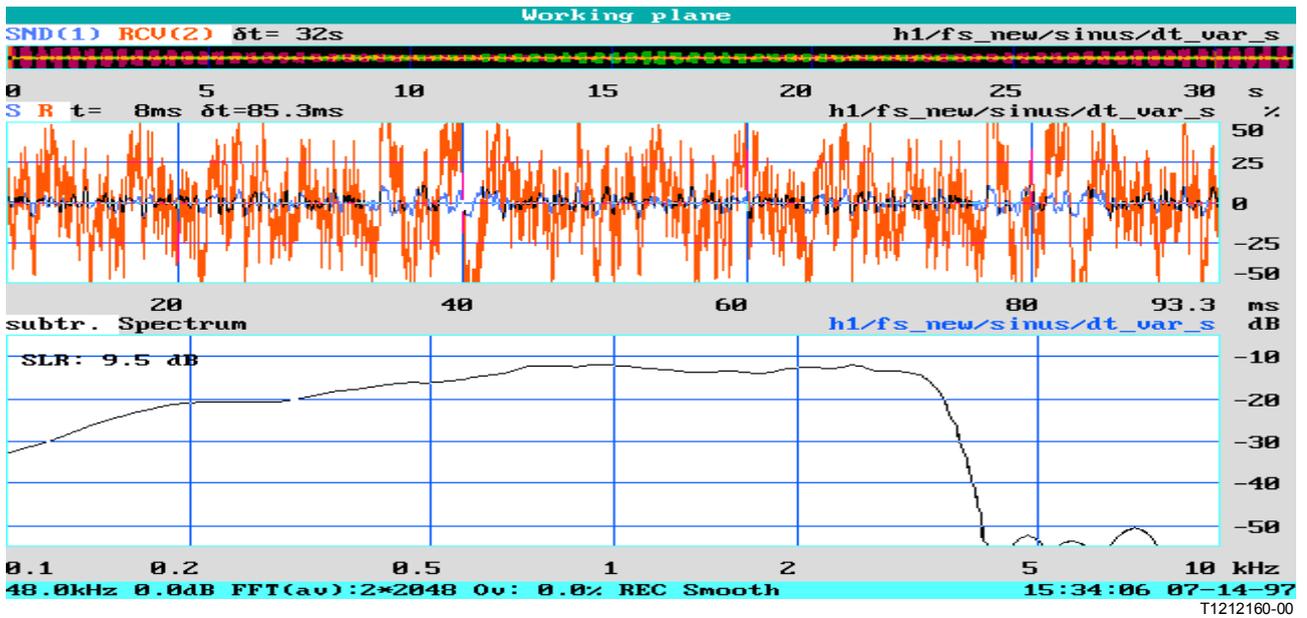


La señal de excitación en el sentido de emisión se muestra en tono oscuro.  
 La señal medida en el sentido de emisión se muestra en tono claro.  
 Las pausas contienen señal de habla simultánea transferida del altavoz del HFT a su micrófono.

**Figura II.2/P.502 – Sentido de emisión en un HFT con compensador de eco y conmutación de nivel**

Las figuras II.1 y II.2 muestran el resultado obtenido con este tipo de señales de prueba para evaluar el sentido de emisión de dos teléfonos individuales manos libres. Para la medición propiamente dicha sólo son de importancia los periodos en los que está presente exclusivamente la señal en el sentido de emisión. Esto sucede de acuerdo con la figura 3 y un periodo de 150 ms que empieza 50 ms después de la señal de activación del sonido emitido. Durante dicho intervalo sólo está presente la señal de emisión. Las figuras II.1 y II.2 muestran los resultados de la medición sólo como secuencia en el tiempo durante los primeros 5 s.

Las figuras II.3, II.4 y II.5 muestran ahora el resultado de la respuesta en frecuencia y de las mediciones del SAR del teléfono manos libres de conmutación de nivel en el sentido de emisión.



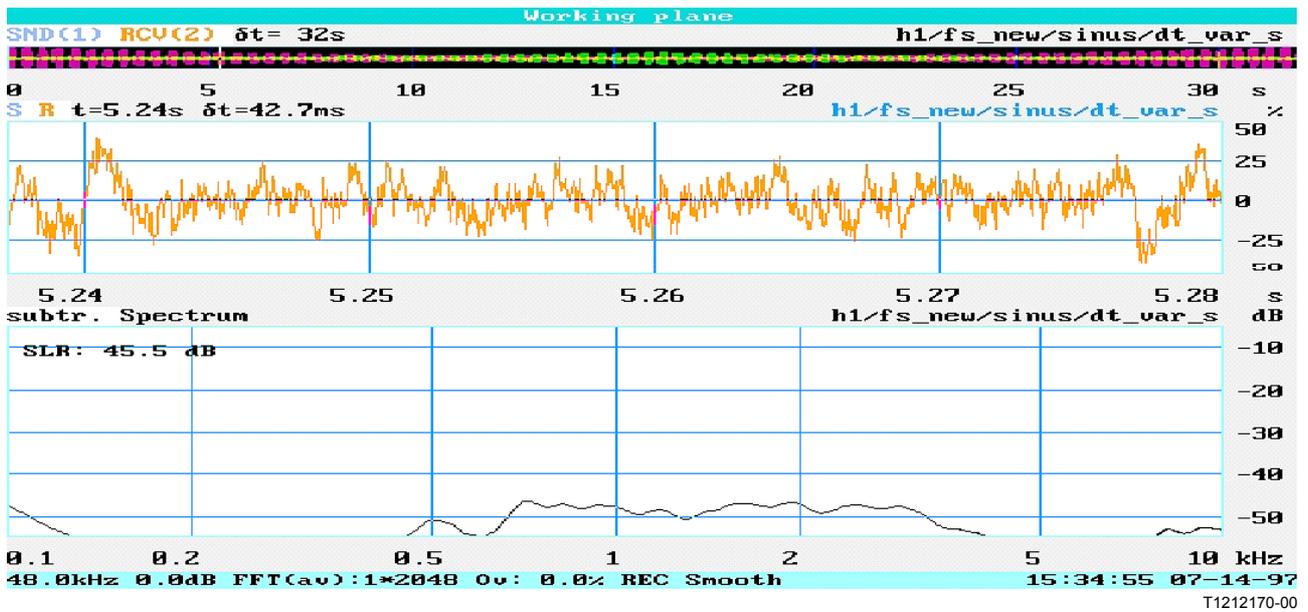
SLR = 9,5 dB

Medición inicial con un nivel de señal de  $-4,7$  dBPa.

La parte superior representa la señal histórica en el tiempo.

La inferior la respuesta en frecuencia y el índice de sonoridad.

**Figura II.3/P.502 – HFT con conmutación de nivel: Características de transferencia e índices de sonoridad**



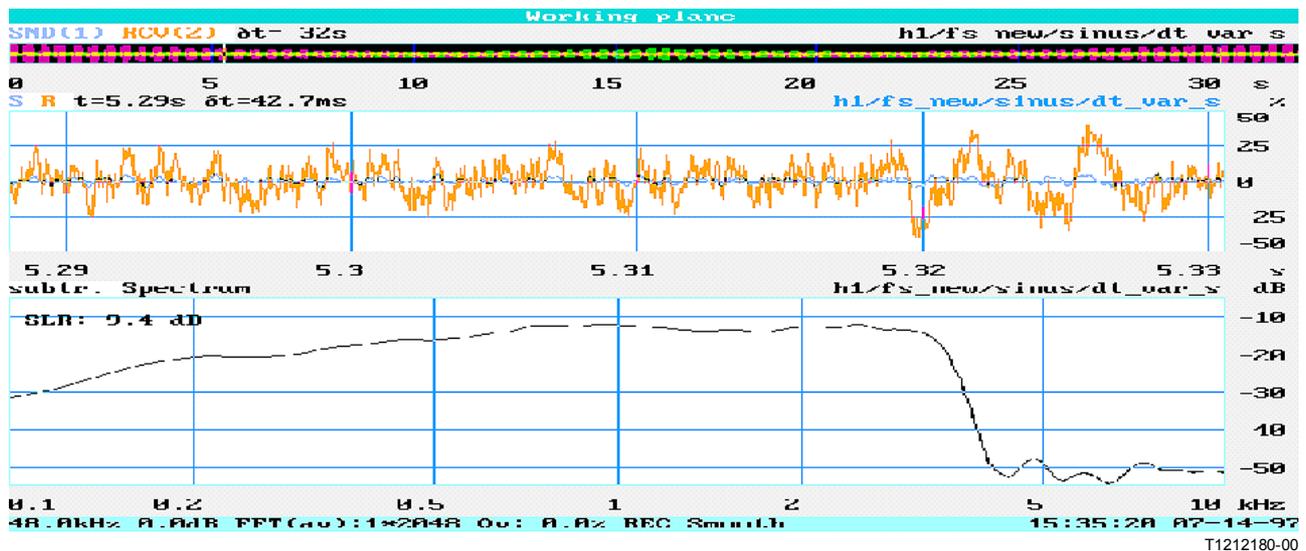
SLR = 45,5 dB

Medición tras ~5 s, siendo el nivel de excitación de la señal de -9,5 dBPa, medición realizada antes de la activación del teléfono.

Parte superior: señal histórica en el tiempo.

Parte inferior: respuesta en frecuencia e índice de sonoridad.

**Figura II.4/P.502 – HFT de conmutación de nivel: Características de transferencia e índices de sonoridad**



SLR = 9,4 dB

Medición tras ~5 s, siendo el nivel de excitación de la señal de -9,5 dBPa, medición realizada antes de la activación del teléfono.

Parte superior: señal histórica en el tiempo.

Parte inferior: respuesta en frecuencia e índice de sonoridad.

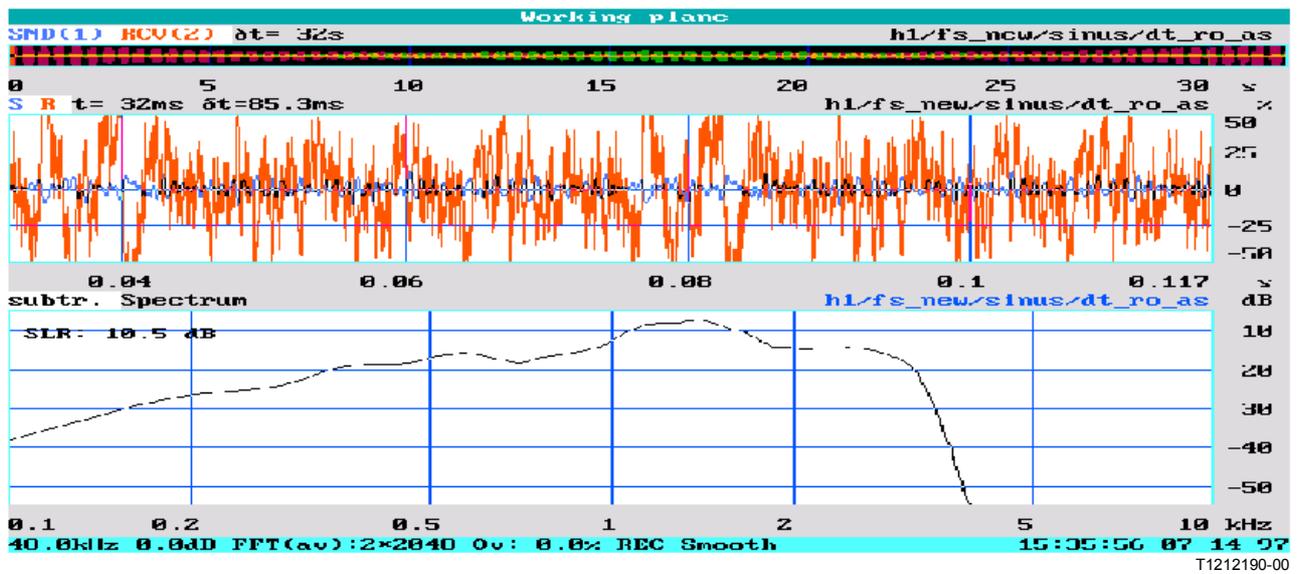
### Figura II.5/P.502 – HFT de conmutación de nivel: Características de transferencia e índices de sonoridad

Las figuras II.3 a II.5 muestran cómo reacciona el dispositivo de conmutación de nivel ante este tipo de señal en el sentido de emisión. El análisis pone de manifiesto que el índice de sonoridad de emisión al comienzo de la secuencia, con el sentido de la emisión totalmente activado, es 9,5 dB. Durante los periodos en que la activación del sistema no es completa, el índice de sonoridad en emisión es 45,5 dB (véase asimismo la respuesta en frecuencia medida en dicho instante). La comparación de las respuestas en frecuencia permite afirmar que la atenuación del teléfono es de unos 30 dB. No se observan características dependientes de la frecuencia.

Se puede también medir el tiempo de activación de este dispositivo en función del nivel de emisión y recepción. Por ejemplo, tras ~5 s, es de unos 90 ms. El nivel de emisión en este instante es -9,5 dBPa y el de recepción -31,5 dBm.

Es posible calcular análogamente las características de transferencia, índices de sonoridad y tiempos de conmutación para otras combinaciones de niveles.

Las figuras II.6 a II.8 muestran el mismo tipo de evaluación, aunque esta vez en un dispositivo con compensador de eco y conmutación de nivel. Como este teléfono no conmuta durante las secuencias de habla simultánea, se han evaluado secuencias al comienzo, tras 5 s y tras 16 s.



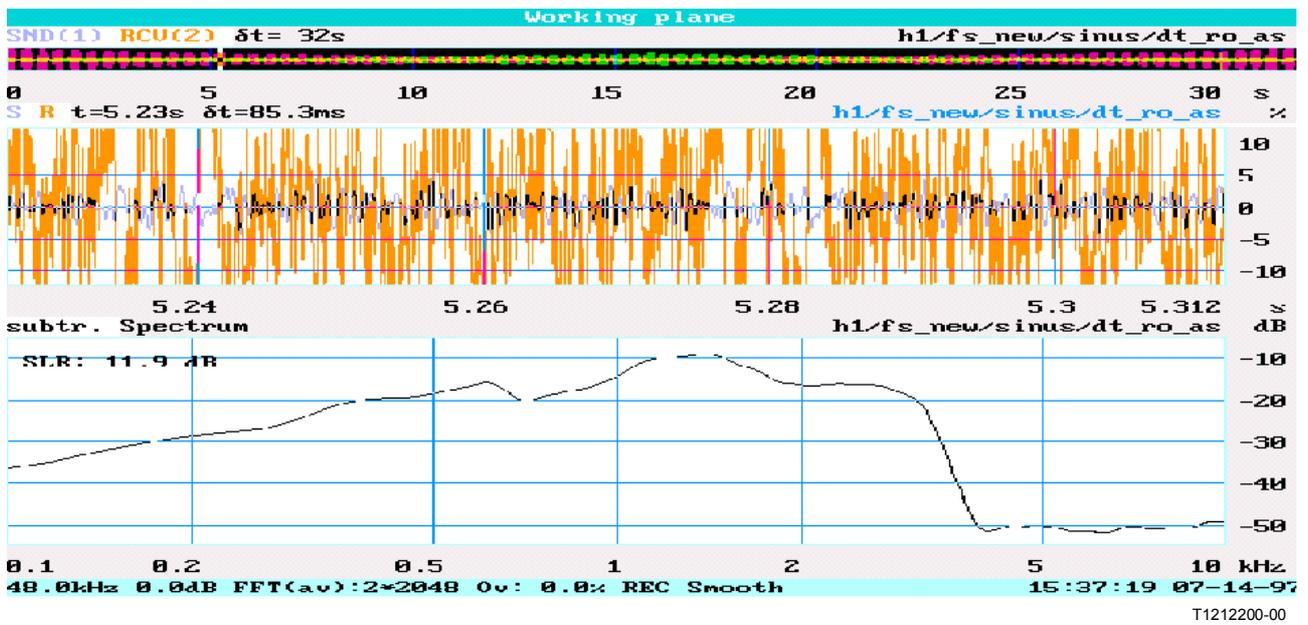
SLR = 10,5 dB

Medición inicial con un nivel de señal de  $-4,7$  dBPa.

La parte superior representa la señal histórica en el tiempo.

La parte inferior la respuesta en frecuencia y el índice de sonoridad.

**Figura II.6/P.502 – HFT con conmutación de nivel y compensación de eco: Características de transferencia e índices de sonoridad**



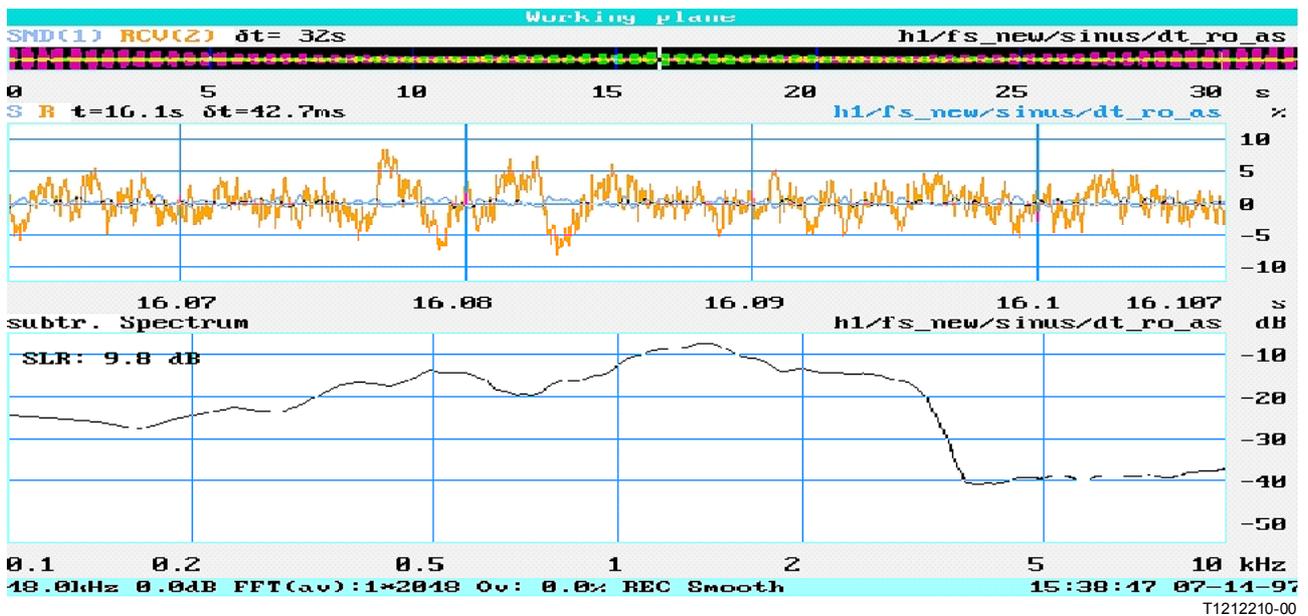
SLR = 11,9 dB

Medición tras ~5 s, siendo el nivel de excitación de la señal de  $-9,5\text{ dBPa}$ , medición realizada antes de la activación del teléfono.

Parte superior: señal histórica en el tiempo.

Parte inferior: respuesta en frecuencia e índice de sonoridad.

**Figura II.7/P.502 – HFT con conmutación de nivel y compensación de eco: Características de transferencia e índices de sonoridad**



SLR = 9,8 dB

Medición tras ~5 s, siendo el nivel de excitación de la señal de -24,7 dBPa, medición realizada tras la activación del teléfono.

Parte superior: señal histórica en el tiempo.

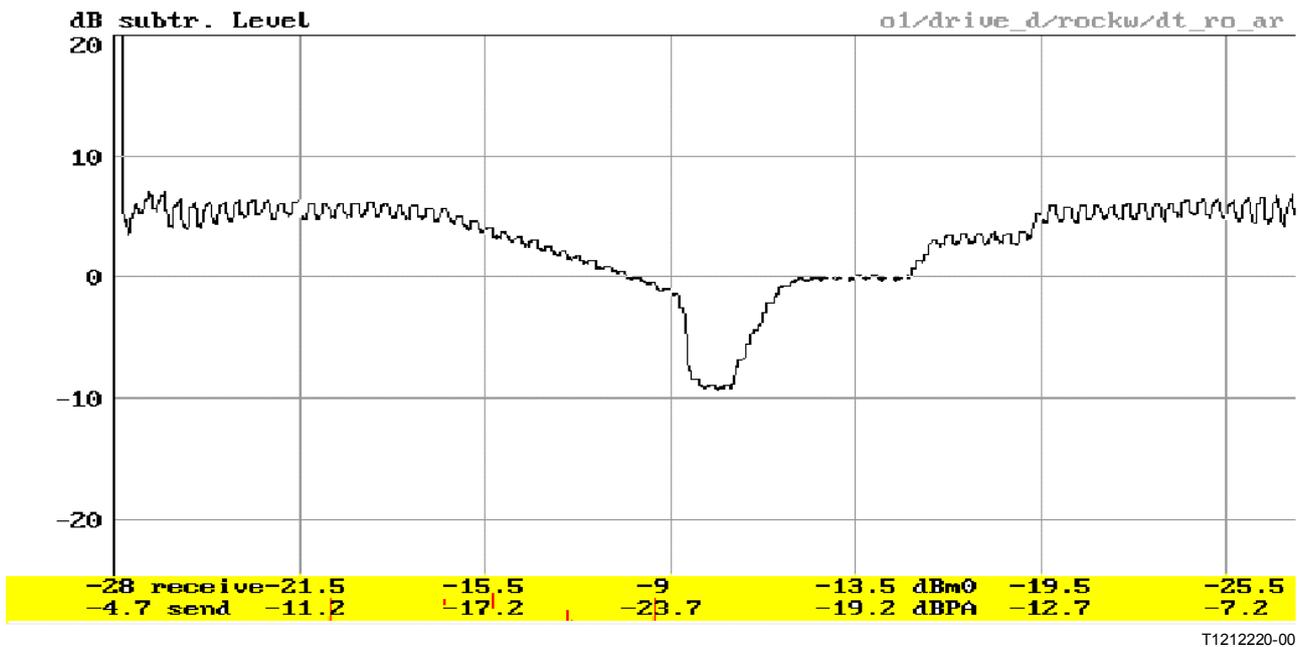
Parte inferior: respuesta en frecuencia e índice de sonoridad.

**Figura II.8/P.502 – HFT con compensación de eco y conmutación de nivel: Características de transferencia e índices de sonoridad**

La evaluación de este aparato pone de manifiesto la estabilidad de las características de transferencia y de los índices de sonoridad, en cualquier situación, en todo el intervalo de niveles comprendido entre -4,7 dBPa y -24,7 dBPa en presencia de la señal habla simultánea, no observándose desconexión. Sin embargo los índices de sonoridad y las respuestas en frecuencia permiten observar claramente ciertos efectos de la compresión expansión. El índice de sonoridad no es siempre estable. Hay una variación del índice de sonoridad (función de la señal) de unos 2 dB.

### II.1.2 Variaciones de nivel durante el habla simultánea

La figura II.9 permite observar el comportamiento general del HFT en condiciones de habla simultánea. Puede verse cómo la amplificación del trayecto de recepción depende del nivel de señal en recepción y en emisión. La figura 2 muestra la señal de prueba utilizada en esta evaluación.

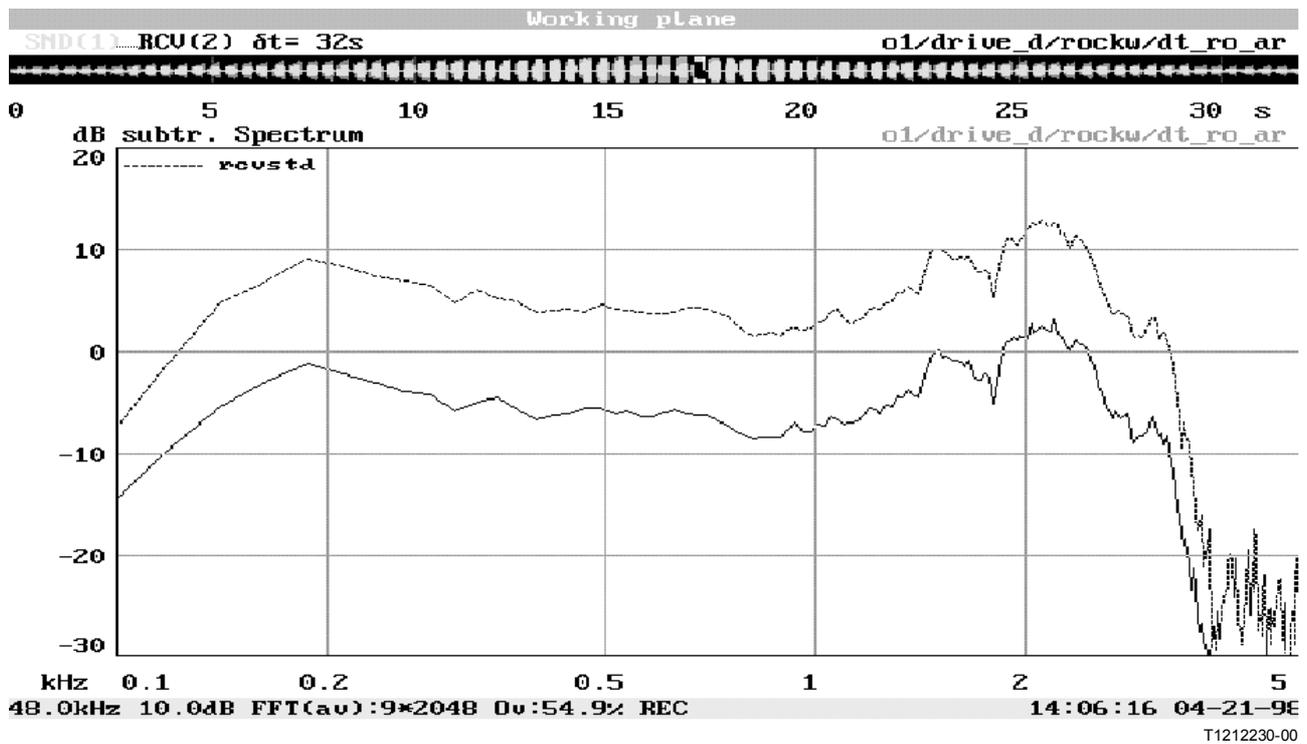


Variación de nivel:  
 En recepción: -28 dBm0 ... -8 dBm0 ... -28 dBm0  
 En emisión: -4,7 dBPa ... -24,7 dBPa ... -4,7 dBPa

**Figura II.9/P.502 – Variación de la amplificación en el sentido de recepción durante el habla simultánea**

Es evidente que existe un AGC asimétrico que tendrá un efecto manifiesto en la fluctuación de la sonoridad percibida subjetivamente. Al crecer los niveles de recepción se observa una suave reducción de la amplificación, seguida de una repentina atenuación de 10 dB aproximadamente. Al volver a reducir el nivel aparecen 3 etapas de atenuación: 10 dB, 12,5 dB y 15 dB, de las que la etapa de 10 dB en particular tendrá una repercusión importante en la calidad percibida. La variación de nivel es bastante elevada (15 dB) y comporta un cambio brusco de amplificación.

La figura II.10 muestra una evaluación más detallada de lo que ocurre con este tipo de señal de prueba. La figura II.10 muestra las respuestas en frecuencia e índices de sonoridad (a niveles de la señal de recepción de -28 dBm y -8 dBm0) que se miden en el sentido de recepción durante el habla simultánea utilizando la pausa de la señal de habla simultánea. En este caso, la respuesta en frecuencia y el índice de sonoridad se midieron efectuando el análisis con 88 ms de señal. Este ejemplo pone de manifiesto la variación del nivel en función de la frecuencia del HFT objeto de la prueba.



RLR = 6,3 dB con:  
 señal de excitación de alto nivel (-4,7 dBPa) en la emisión y de bajo nivel (-28 dBm0) en recepción.

RLR = 19,5 dB con:  
 señal de excitación de alto nivel (-24,7 dBPa) en la emisión y de bajo nivel (-8 dBm0) en recepción.

**Figura II.10/P.502 – Respuestas en frecuencia e índices de sonoridad en el sentido de emisión durante el habla simultánea**

### II.1.3 Conmutación durante el habla simultánea

La figura II.11 muestra un ejemplo para el sentido de emisión. En el sentido de emisión se produce la conmutación de nivel durante el habla simultánea a niveles de emisión inferiores a -10,7 dBPa. Para determinar la importancia subjetiva de esta conmutación, hay que determinar asimismo las constantes de tiempo, la ganancia de conmutación y la variación de dichos parámetros en función del nivel. En la figura II.11 pueden observarse el tiempo de conmutación y su ganancia, y el aumento de la amplificación en 10 dB, 35 ms después del habla simultánea. El tiempo de conmutación es de unos 10 ms.

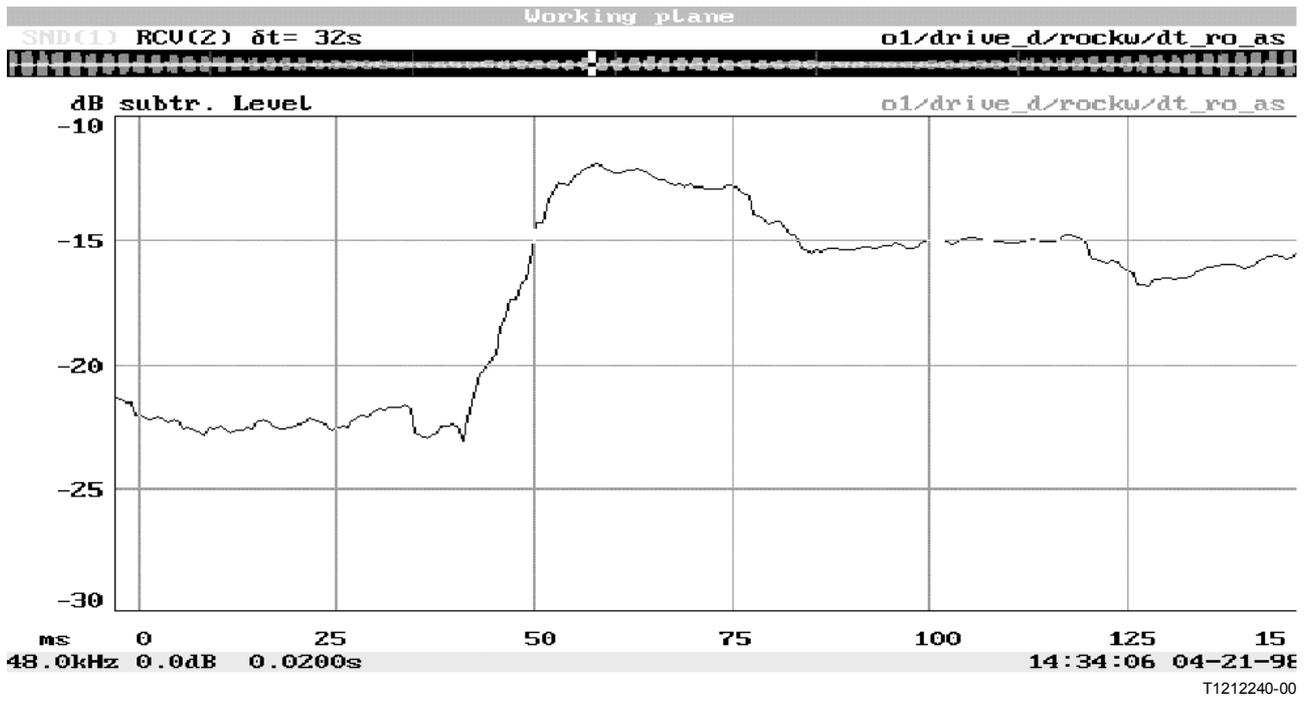


Figura II.11/P.502 – Tiempo de conmutación y variación de nivel durante el habla simultánea

## SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie B	Medios de expresión: definiciones, símbolos, clasificación
Serie C	Estadísticas generales de telecomunicaciones
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 multimedia
Serie I	Red digital de servicios integrados
Serie J	Transmisiones de señales radiofónicas, de televisión y de otras señales multimedia
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	RGT y mantenimiento de redes: sistemas de transmisión, circuitos telefónicos, telegrafía, facsímil y circuitos arrendados internacionales
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
<b>Serie P</b>	<b>Calidad de transmisión telefónica, instalaciones telefónicas y redes locales</b>
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 y comunicación entre sistemas abiertos
Serie Y	Infraestructura mundial de la información y aspectos del protocolo Internet
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación