



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

P.501

(08/96)

SERIE P: CALIDAD DE TRANSMISIÓN TELEFÓNICA

Aparatos para mediciones objetivas

**Señales de prueba para utilización en
telefonometría**

Recomendación UIT-T P.501

(Anteriormente Recomendación del CCITT)

RECOMENDACIONES DE LA SERIE P DEL UIT-T
CALIDAD DE TRANSMISIÓN TELEFÓNICA

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
Aparatos para mediciones objetivas	Serie	P.50 P.500
Medidas electroacústicas objetivas	Serie	P.60
Medidas relativas a la sonoridad vocal	Serie	P.70
Métodos para evaluaciones objetivas y subjetivas 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.501

SEÑALES DE PRUEBA PARA UTILIZACIÓN EN TELEFONOMETRÍA

Resumen

Esta Recomendación describe las señales de prueba utilizables con varias finalidades en telefonometría. Indica una amplia variedad de señales de prueba, desde señales poco complejas hasta señales sumamente complejas que comprenden numerosos parámetros típicos de la palabra. Además de las señales técnicas, como ondas sinusoidales o señales de ruido, se describen otras señales similares a las vocales.

Esta Recomendación expone los principios de construcción de cada tipo de señal de prueba e indica sus propiedades características, como los espectros de densidad de potencia, las funciones de densidad de probabilidad o las respuestas del filtro conformador.

También se describe en general la aplicación típica de las señales de prueba, a modo de orientación con respecto a las reglas generales de aplicación. La descripción detallada de la aplicación figura en las Recomendaciones concretas que describen los procedimientos de medida para cada aplicación.

Para evitar problemas en la construcción de las señales de prueba descritas, esta Recomendación incluye un CD-ROM que contiene todas las señales de prueba.

Orígenes

La Recomendación UIT-T P.501 ha sido preparada por la Comisión de Estudio 12 (1993-1996) del UIT-T y fue aprobada por el procedimiento de la Resolución N.º 1 de la CMNT el 30 de agosto de 1996.

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 N.º 1 de la CMNT (Helsinki, 1 al 12 de marzo de 1993).

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.

© UIT 1996

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.....	1
4 Resumen de las señales de prueba y aplicaciones típicas descritas en esta Recomendación.....	2
5 Tipos de señales de prueba	5
5.1 Señales no similares a la voz (enteramente artificiales)	5
5.1.1 Señales deterministas.....	5
5.1.2 Señales aleatorias.....	7
5.1.3 Señales combinadas aleatorias y deterministas	7
5.2 Señales similares a la voz	8
5.2.1 Señales fuente compuestas (señales compuestas en tiempo).....	8
5.2.2 Ruido modulado similar a la voz.....	21
5.2.3 Señales compuestas en frecuencia (tecnología de tono sonda).....	21
5.2.4 Señales complejas compuestas	22
5.3 Señales de habla.....	28

Recomendación P.501

SEÑALES DE PRUEBA PARA UTILIZACIÓN EN TELEFONOMETRÍA

(Ginebra, 1996)

1 Alcance

La presente Recomendación describe las señales de prueba aplicables a diversas finalidades en telefonometría. La Recomendación ofrece una gran variedad de señales de prueba que va desde las señales de escasa complejidad a las señales de prueba con alto grado de complejidad que incorporan numerosos parámetros típicos de la voz. Además se describen señales técnicas como las ondas sinusoidales o de ruido más parecidas a la voz.

La Recomendación ofrece una visión general sobre la aplicación típica de las señales de prueba descritas. Esta visión general constituye una directriz que da reglas de aplicación generales. La descripción detallada de una aplicación debe no obstante encontrarse en la Recomendación individual que describe los procedimientos de medición para esa aplicación.

Con el fin de evitar problemas en la formación de las señales de prueba descritas, esta Recomendación contiene un CD-ROM en el que se almacenan todas las señales de prueba.

2 Referencias

Las siguientes Recomendaciones 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 P.50 (1993), *Voz artificial*.
- Recomendación UIT-T P.59 (1993), *Habla conversacional artificial*.
- Recomendación UIT-T P.340 (1996), *Características de transmisión de los aparatos telefónicos manos libres*.

3 Definiciones y abreviaturas

A los efectos de la presente Recomendación, se aplican las definiciones siguientes.

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

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

3.3 transformación rápida de Fourier (FFT, *fast Fourier transform*)

3.4 función de transferencia de modulación (MTF, *modulation transfer function*): Señal de modulación, obtenida de la envolvente de una señal de prueba.

3.5 proceso del modelo vocal de Markov (MSMP, *markov speech model process*): Señal que simula la voz utilizando cadenas de Markov adiestrables para la generación de una señal semejante a la voz, teniendo en cuenta el proceso de generación del habla real.

3.6 secuencia de seudorruido, secuencia PN: Ruido pseudoaleatorio con contenido de frecuencias definido, obtenido por la transformación inversa de Fourier de un espectro de frecuencias definido previamente.

3.7 generador de voz simulada (SSG, *simulated speech generator*): Señal que ofrece propiedades semejantes a la voz, construida teniendo en cuenta el proceso de generación del habla real.

3.8 índice de transmisión vocal (STI, *speech transmission index*): Índice que expresa la inteligibilidad del habla, especialmente en condiciones de reverberación, obtenido de la medición de la MTF.

4 Resumen de las señales de prueba y aplicaciones típicas descritas en esta Recomendación

Sistemas lineales, invariantes en tiempo (por ejemplo, aparatos de microteléfono)									
	Onda sinusoidal (Nota 1)	Ruido (Nota 2)	CSS (Nota 3)	Tono sonda	SSG	Rec. P.50	Rec. P.59	MSMP	Habla
Índices de sonoridad	x	x	x	x	x	x	x	x	x
Respuestas de frecuencia	x	x	x	x	x	x	x	x	x
Efecto local del oyente/efecto local del hablante	x	x	x	x	x	x	x	x	x
Distorsión armónica	x		x						
Distorsión		x	x						
Señales fuera de banda	x		x						
Mediciones de nivel	x	x	x	x	x	x	x	x	x
Mediciones de retardo	x	x	x		(x)	(x)		(x)	(x)
Mediciones de eco		x	x	x	x	x	x	x	x
NOTAS x Aplicable (x) Aplicable con ciertas precauciones 1 Incluyendo onda sinusoidal modulada y espectros de Fourier. 2 Incluyendo ruido rosa, blanco y conmutado. 3 Incluyendo diversas combinaciones de señales de sonidos vocalizados y de medición (secuencias de seudorruido, sinusoidales, etc.).									

Sistemas no lineales y/o variantes en tiempo									
	Onda sinusoidal (Nota 1)	Ruido (Nota 2)	CSS (Nota 3)	Tono sonda	SSG	Rec. P.50	Rec. P.59	MSMP	Habla
Índices de sonoridad (valores a largo plazo)		(x)	(x)		(x)	(x)	(x)	(x)	(x)
Índices de sonoridad (valores a corto plazo)		(x)	x						
Respuestas de frecuencia (valores a largo plazo)		(x)	(x)		(x)	(x)	(x)	(x)	(x)
Respuestas de frecuencia (valores a corto plazo)		(x)	x						
Efecto local del oyente/efecto local del hablante (valores a largo plazo)		(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)
Distorsión armónica			(x)						
Distorsión		(x)	(x)		(x)	(x)	(x)	(x)	(x)
Señales fuera de banda			(x)						
Mediciones de nivel		(x)	(x)		(x)	(x)	(x)	(x)	(x)
Mediciones de retardo		(x)	(x)						
Mediciones de eco		(x)	(x)		(x)	(x)	(x)	(x)	(x)
<p>NOTAS</p> <p>x Aplicable</p> <p>(x) Aplicable con ciertas precauciones</p> <p>1 Incluyendo onda sinusoidal modulada y espectros de Fourier.</p> <p>2 Incluyendo ruido rosa, blanco y conmutado, así como la MTF.</p> <p>3 Incluyendo diversas combinaciones de señales de sonidos vocalizados y de medición (secuencias de seudoruido, sinusoidales, etc.).</p> <p>4 Los valores a largo plazo corresponden al comportamiento en estado permanente de los sistemas. Los valores a corto plazo corresponden al comportamiento dinámico de los sistemas.</p>									

Teléfonos manos libres									
	Onda sinusoidal (Nota 1)	Ruido (Nota 2)	CSS (Nota 3)	Tono sonda	SSG	Rec. P.50	Rec. P.59	MSMP	Habla
Índices de sonoridad (valores a largo plazo)		x	x	(x)	x	x	x	x	x
Índices de sonoridad (valores a corto plazo)		(x)	x						
Respuestas de frecuencia (valores a largo plazo)		x	x	(x)	x	x	x	x	x
Respuestas de frecuencia (valores a corto plazo)		(x)	x						
Distorsión armónica			x						
Distorsión		(x)	x						
Señales fuera de banda			x						
Mediciones de nivel		x	x	(x)	x	x	x	x	x
Mediciones de retardo		x	x						
Características de conmutación	(x)	(x)	x	(x)	(x)	(x)	(x)	(x)	(x)
Mediciones de reverberación		x	x						
Características del compensador de eco		(x)	(x)		(x)	(x)		(x)	(x)

NOTAS

x Aplicable

(x) Aplicable con ciertas precauciones

- 1 Incluyendo onda sinusoidal modulada y espectros de Fourier.
- 2 Incluyendo ruido rosa, blanco y conmutado, así como ruido modulado (MTF).
- 3 Incluyendo diversas combinaciones de señales de sonidos vocalizados y de medición (secuencias de seudoruido, sinusoidales, etc.).
- 4 Los valores a largo plazo corresponden al comportamiento en estado permanente de los sistemas.
Los valores a corto plazo corresponden al comportamiento dinámico de los sistemas.

Compensadores de eco									
	Onda sinusoidal (Nota 1)	Ruido (Nota 2)	CSS (Nota 3)	Tono sonda	SSG	Rec. P.50	Rec. P.59	MSMP	Habla
Mediciones de nivel		(x)	x	(x)	x	x	x	x	x
Mediciones de retardo		(x)	x						
Características de conmutación	(x)	(x)	x		(x)	(x)	(x)	(x)	(x)
Comportamiento del ruido de fondo		(x)	x	(x)	x	x	x	x	x
Atenuación del eco, TCL		(x)	(x)	(x)	x	x	x	x	x
Comportamiento en diloquia		(x)	x		(x)	(x)	(x)	(x)	(x)
NOTAS									
x Aplicable									
(x) Aplicable con ciertas precauciones									
1 Incluyendo los espectros de la onda sinusoidal modulada y de Fourier.									
2 Incluyendo el ruido rosa, blanco y conmutado así como el ruido modulado (MTF).									
3 Incluyendo diversas combinaciones de señales vocalizadas y señales de medición (secuencia de seudoruido, sinusoide, etc.).									

5 Tipos de señales de prueba

Se dispone de diferentes señales de prueba con distintos niveles de complejidad, las cuales se han evaluado para diferentes tipos de aplicaciones. Pueden identificarse diferentes grupos de señales según su complejidad, que se reseñan a continuación.

5.1 Señales no similares a la voz (enteramente artificiales)

Las señales no semejantes a la voz son las señales de medición clásicas. Pueden dividirse en señales deterministas y señales aleatorias. Las señales deterministas pueden definirse mediante una fórmula que describe totalmente la señal en el dominio del tiempo o de la frecuencia, mientras que las señales aleatorias pueden ser descritas por la estadística de la señal y su espectro a largo plazo.

5.1.1 Señales deterministas

5.1.1.1 Descripción

- onda sinusoidal;
- onda sinusoidal modulada;
- espectro generado de Fourier.

Las señales deterministas para mediciones en telefonometría se describen típicamente en el dominio de la frecuencia. La descripción general de una señal sinusoidal que podría estar modulada en amplitud o frecuencia es la siguiente:

$$s(t) = [A + \mu_{am} \cdot \cos(2\pi t \cdot f_{am})] \cdot \cos(2\pi t \cdot f_0 + \mu_{fm} \cdot \sin(2\pi t \cdot f_{fm})) \quad [5-1]$$

A Amplitud de la señal

μ_{am} Factor de modulación de la modulación de amplitud

f_{am} Frecuencia de modulación de la modulación de amplitud

- f_0 Frecuencia portadora
- μ_{fm} Factor de modulación de la modulación de frecuencia
- f_{fm} Frecuencia de modulación de la modulación de frecuencia
- t Tiempo

El barrido de frecuencia lineal viene descrito por:

$$s(t) = A(f_0 + st) \cdot e^{j(\pi t(f_0 + st) + \phi_0)} \quad [5-2]$$

- s Velocidad de barrido
- f_0 Frecuencia de arranque

El barrido logarítmico viene descrito por:

$$s(t) = A(f_0 \cdot 10^{t/Td}) \cdot e^{j((2\pi f_0 Td / \ln 10) \cdot 10^{t/Td} + \phi_0)} \quad [5-3]$$

- Td Tiempo invertido en barrer una octava

El espectro sinusoidal múltiple puede describirse mediante su espectro de Fourier y puede expresarse de la manera siguiente:

$$s(t) = \sum_n A_n \cdot \text{sen}(2\pi t \cdot f_n + \phi_n) \quad [5-4]$$

- A_n Amplitud del componente de frecuencia n
- f_n Frecuencia del componente de frecuencia n
- ϕ_n Fase del componente de frecuencia n
- t Tiempo

En 5.2.1.1 se describe una señal de seudoruido, que puede considerarse como una forma especial de señal sinusoidal múltiple adaptada a FFT. Mientras que las señales sinusoidales discretas suelen aplicarse a niveles fijos independientes de la frecuencia, las señales sinusoidales múltiples y de barrido se aplican a menudo con una ponderación de frecuencia que concuerda más estrechamente con el espectro de la voz humana.

5.1.1.2 Aplicación

Siempre pueden utilizarse señales deterministas para determinar las características de transferencia de sistemas lineales invariantes en tiempo, principalmente en el dominio de la frecuencia. Estas señales se utilizan típicamente para determinar la distorsión armónica y la distorsión de intermodulación. La ventaja de estas señales es el fácil tratamiento y determinación de los parámetros de sistema mediante sencillas mediciones de nivel.

La ventaja del barrido logarítmico es que la resolución efectiva en frecuencia es más semejante a la resolución de frecuencias del oído humano.

El barrido lineal permite controlar el resultado de la medición y procesarlo directamente en el dominio de la frecuencia así como en el dominio del tiempo, utilizando técnicas FFT, por ejemplo. Además, el barrido lineal ofrece especialmente oportunidades para suprimir el ruido indeseado y aislar los ecos eléctricos o acústicos.

Deben tratarse con cuidado todas las señales sinusoidales discretas o múltiples que han de utilizarse en dispositivos de medición en los que se emplean técnicas adaptativas. La función de autocorrelación de la señal de medición no debe situarse periódicamente dentro de la ventana de procesamiento del dispositivo sometido a prueba.

5.1.2 Señales aleatorias

5.1.2.1 Descripción

Ruido aleatorio

El ruido aleatorio puede determinarse por sus características estadísticas, el espectro de densidad de potencia a largo plazo, una función de densidad de probabilidades de una y dos dimensiones, o simplemente como una historia en el tiempo en el caso de que se muestree la señal aleatoria. Las señales siguientes se utilizan típicamente en telefonometría:

Ruido blanco

- Características de frecuencia: los límites inferior y superior del espectro de densidad de potencia generada vienen definidos por la aplicación; típicamente se describe el espectro de densidad de potencia a largo plazo.
- Función de densidad de probabilidad: distribución gaussiana típica con un factor de cresta de $11 \text{ dB} \pm 1 \text{ dB}$.

Ruido rosa

- Características de frecuencia: el espectro de densidad de potencia de la señal muestra una disminución de 3 dB/octava, los límites inferior y superior del espectro de densidad de potencia generada vienen definidos por la aplicación, típicamente se describe el espectro de densidad de potencia a largo plazo.
- Función de densidad de probabilidad: distribución gaussiana típica con un factor de cresta de $11 \text{ dB} \pm 1 \text{ dB}$.

Para telefonometría, estas señales suelen modularse por una señal moduladora rectangular (modulación TODO/NADA (ON/OFF)). Parámetros típicos de la modulación son: 250 ms en TODO (ON) y 150 ms en NADA (OFF). Mediante esta modulación se simula de manera muy sencilla la modulación típica de la voz.

5.1.2.2 Aplicación

Se utilizan típicamente señales aleatorias para sistemas lineales invariantes en tiempo con el fin de determinar los niveles en banda ancha o los niveles en fractales de octava. Además pueden utilizarse estas señales para determinar las características de transferencia en el dominio de la frecuencia, tales como la respuesta de frecuencia o los índices de sonoridad.

Se utilizan señales aleatorias para las mediciones siempre que se requieran tiempos de promediación largos (típicamente $> 10 \text{ s}$).

5.1.3 Señales combinadas aleatorias y deterministas

5.1.3.1 Descripción

Mediante la modulación del ruido aleatorio con una señal determinista puede obtenerse la función de transferencia de modulación (MTF) de un sistema. Si, por ejemplo, tomamos una señal de ruido con una intensidad media $\bar{I}(t)$ y la modulamos con una señal sinusoidal de frecuencia f con índice de modulación $m = 1$ obtenemos una señal:

$$I(t) = \bar{I}(t) \cdot (1 + \cos(2\pi ft)) \quad [5-5]$$

5.1.3.2 Aplicación

Si se aplica esta señal a un sistema la función de transferencia de modulación (MTF) [2] puede hallarse midiendo el índice de modulación $m(f)$ de la salida del sistema. Midiendo e interpretando

la MTF de la manera correcta puede predecirse en cierto grado la inteligibilidad del habla en un sistema (véase 5.2.2).

5.2 Señales similares a la voz

Las señales similares a la voz se presentan con diversos grados de complejidad. El grado de complejidad está siempre relacionado con los parámetros típicos del habla simulada por la señal similar a la voz. En general, pueden encontrarse las clases de señales enumeradas a continuación.

Los bloques constitutivos de todas las señales similares a la voz son -además de otros parámetros- sonidos vocalizados y no vocalizados. En general, se requiere que un teléfono o dispositivo procesador de la voz que utilice detectores de palabra se active o se mantenga activado en la presencia de dichos sonidos vocalizados o no vocalizados, en tanto que otras señales puedan impedir o causar una interrupción de la transmisión.

5.2.1 Señales fuente compuestas (señales compuestas en tiempo)

5.2.1.1 Descripción

a) Consideraciones generales

Cuando se forma una señal fuente compuesta, se consideraron especialmente importantes las tres componentes siguientes:

- señal vocalizada para simular las propiedades de la voz;
- señal determinista para la medición de las funciones de transferencia sin errores estadísticos, y con espectros de densidad de potencia constante de la señal de excitación en el dominio de frecuencia que ha de medirse;
- "señal" de pausa que proporciona la modulación de amplitud.

De ahí resultan las características siguientes:

- i) breve periodo de medición;
- ii) posibilidad de inyectar la señal de prueba en el sentido de habla y de escucha al mismo tiempo (funcionamiento dúplex).

La idea básica para la utilización de tal señal es colocar el dispositivo sometido a prueba en un estado bien definido, reproducible, durante el periodo de medición y asegurar que las funciones de transferencia del dispositivo no cambian apreciablemente durante la medición real (régimen casi estacionario). Con mayor precisión, la señal fuente compuesta consta de los componentes siguientes:

1) Señal vocalizada producida por la señal de "voz artificial" conforme a la P.50

La parte de señal vocalizada de la CSS es la señal de acondicionamiento cuyo fin es activar los posibles detectores de palabra en los sistemas controlados por la voz. El motivo de haber elegido una señal vocalizada es que presumiblemente todos los dispositivos diseñados para transmisión del habla responderán rápidamente a un sonido vocalizado. Esta señal tiene que activar el dispositivo sometido a prueba en el sentido de transmisión que ha de medirse. Dado que se conocen exactamente la duración, el principio y el fin de la señal vocalizada, esta señal puede utilizarse para medir el tiempo de conmutación aplicable al sentido de transmisión que se prueba. Por medio de la forma de la señal en el dominio del tiempo, puede determinarse el tiempo de conmutación y el tiempo de retardo del sistema total de acuerdo con la P.340. La duración de la señal asciende a 50 ms.

2) Señal de seudoruido

La señal de medición es la señal de seudoruido (PN) presentada después del sonido de voz artificial vocalizado. Esta señal tiene ciertas características similares al ruido. La magnitud

de su transformada de Fourier es constante con la frecuencia, mientras que la fase es variable. Para las mediciones usualmente sólo tiene interés la magnitud de la función de transferencia, y la fase no es tan importante aunque también puede ser determinada.

La señal se produce de la forma siguiente:

Primeramente se genera un espectro complejo en el dominio de la frecuencia según la ecuación siguiente:

$$H(k) = W(k) \cdot e^{j i_k \cdot \pi} ; k = -M/2, \dots, M/2, \sin 0; i_k \{+1, 0\}, i_k = -i_{-k} \text{ aleatorio} \quad [5-6]$$

El índice M se ajusta al tamaño elegido de la FFT (por ejemplo 2048 puntos). La ecuación muestra que la cantidad de espectro complejo producido es constante para todas las frecuencias si W(k) se elige igual a 1 para todas las frecuencias, mientras que la fase puede ser π o 0 para cada frecuencia, correspondiendo a una secuencia aleatoria. No obstante, para producir una ponderación diferente en el dominio de la frecuencia, W(k) puede ajustarse fácilmente con el fin de generar espectros diferentes durante la secuencia de seudoruido. A continuación, este espectro se transformará en el dominio del tiempo por medio de la transformada inversa de Fourier, obteniendo la señal siguiente:

$$S(n) = \frac{1}{M} \sum_{k=-M/2, k \neq 0}^{M/2} H(k) \cdot e^{j 2\pi \cdot n \cdot k / M} , n = -M/2, \dots, M/2 - 1; \quad [5-7]$$

Se produce así una señal limitada en tiempo (correspondiendo a la longitud de la transformada de Fourier elegida) y que se ajusta correctamente al tamaño de FFT elegido. Si se desea una secuencia temporal más larga puede repetirse cíclicamente la señal. Este método permite obtener secuencias temporales de cualquier longitud.

La duración de esta señal de medición alcanza aproximadamente 200 ms mediante la elección apropiada de M, la velocidad de muestreo y el número de repeticiones.

NOTA 1 - En el caso típico, la longitud de la FFT debe ser corta para los sistemas con parámetros que varíen mucho con el tiempo, tales como las técnicas de compensación, con el fin de obtener una buena y rápida estimación de la función de transferencia variante en el tiempo. Para los sistemas que incorporan técnicas adaptativas, como son los compensadores de eco o compensadores de ruido, puede convenir que el número de M sea mayor (cercano a la duración de señal de 200 ms) con el fin de que la función de autocorrelación de la señal de medición no se sitúe periódicamente dentro de la ventana de procesamiento del dispositivo que se prueba.

NOTA 2 - En lugar de las señales de ruido pseudoaleatorio, pueden ser apropiadas para aplicaciones especiales otras señales como las secuencias de M (secuencias de máxima longitud) u otras secuencias con funciones de autocorrelación perfectas. Para otras aplicaciones como las mediciones de distorsión la señal PN puede sustituirse por señales apropiadas, como las sinusoidales o los ruidos de banda estrecha.

3) Pausa

La pausa tiene dos finalidades. Se necesita una pausa inicial antes de aplicar cualquier señal de medición para colocar en un estado inicial definido los sistemas con funciones de transferencia variantes en tiempo. Para este fin, la pausa debe ser lo más larga posible (>1s). Si, no obstante, el sistema ha de ponerse en un estado constantemente activado (similar a una conversación en curso), las pausas intermedias deben ser más breves (cerca de 100 ms) para proporcionar una modulación de amplitud adecuada a la señal compuesta.

La pausa de la secuencia de CSS debe estar comprendida en el margen de 100 a 150 ms.

NOTA 3 - Puede ampliarse la pausa para mediciones en las que necesite observarse un comportamiento a largo plazo después de la activación. En este caso, la pausa puede prolongarse hasta varios segundos dependiendo de los requisitos de la medición.

Con el fin de conseguir una secuencia a largo plazo exenta de desplazamiento, la secuencia de CSS repetida debe invertirse en amplitud (desfasaje de 180°).

b) *Cálculo y análisis utilizando una señal fuente compuesta*

Cuando la señal fuente compuesta se utiliza para mediciones, la secuencia de sonido vocalizado, señal de seudoruido y pausa puede repetirse cíclicamente. Esto significa que al terminar la pausa la secuencia se inicia otra vez, empezando con un sonido vocalizado. Por este procedimiento pueden producirse secuencias de cualquier longitud.

1) Principio de la calibración acústica y eléctrica, niveles de señal de prueba

Una vez creada una secuencia como la descrita anteriormente, esta señal puede tratarse como una señal de medición normal, por ejemplo, como el ruido rosa conmutado. La calibración del nivel (acústica y eléctrica) se realiza utilizando la secuencia entera que incluye sonidos vocalizados, secuencias de seudoruido y pausas. En principio, puede utilizarse un medidor de valores eficaces normal con una anchura de banda de 20 kHz que funcione con promediación "rápida". El método preferido, no obstante, es el análisis de FFT para los cálculos de nivel. Los parámetros para los cálculos basados en FFT son los siguientes:

- velocidad de muestreo conforme con la elegida para la generación de señal (preferentemente 44,1 kHz o 48 kHz);
- longitud de la FFT conforme con la elegida para la generación de señal;
- ventanas rectangulares;
- ausencia de superposición;
- promediación sobre la **secuencia total (repetida)**, incluyendo los sonidos vocalizados, las secuencias PN y las pausas;
- cálculo del nivel obtenido del espectro de densidad de potencia mediante el cálculo de la FFT (integración de los niveles sobre todas las componentes de frecuencias).

2) Parámetros del análisis

Para la medición de las funciones de transferencia en el sentido de emisión y de recepción en cuanto a índices de sonoridad, etc., se repite también cíclicamente la secuencia de sonido vocalizado, secuencia PN y pausa. El nivel de la secuencia completa se ajusta de manera que el nivel total medido esté de acuerdo con el especificado, como ya se ha descrito.

Todas las mediciones (análisis) se realizan **solamente durante la secuencia PN**. Para el análisis de todos los parámetros de transmisión en el dominio de la frecuencia y de los índices de sonoridad la señal medida y la transformada de Fourier se han de referir siempre a la señal de entrada transformada de Fourier que utiliza los mismos parámetros de análisis. En el caso de las mediciones acústicas, estas señales de entradas se miden en el MRP para obtener las características de emisión. En las mediciones eléctricas la señal medida y la transformada de Fourier se refieren a la señal de entrada inyectada en el trayecto de transmisión digital o en el analógico. Esto puede realizarse a través de una medición en dos canales o aplicando técnicas en las que puede almacenarse la señal de entrada analizada (señal de medición tomada en el MRP o en la interfaz digital). Para el análisis se utilizan los siguientes parámetros:

- Velocidad de muestreo conforme con la elegida para la generación de señal.
- Longitud de FFT de acuerdo con la utilizada para la generación de señal, aplicada a una secuencia PN únicamente.
- Ventana rectangular.

- Superposición permitida entre 0 y 99,9%. Se ha de aplicar la misma superposición a la señal de medición a la entrada del objeto de prueba (MRP o interfaz digital) y a la señal medida a la salida del objeto de prueba en sentido de emisión o de recepción.
- La señal transformada de Fourier medida o en la salida del sentido de emisión o en sentido de recepción se ha de referir a la señal transformada de Fourier en el punto de excitación correspondiente (MRP, entrada digital o analógica).

Otras mediciones pueden necesitar señales sinusoidales o señales de ruido diferentes para la medición de distintos parámetros, como la distorsión. En este caso se sustituye la secuencia PN por la señal correspondiente, por ejemplo, una señal sinusoidal o señal de ruido de banda estrecha. Se calculan como se ha descrito anteriormente los niveles de las señales fuente compuestas completas repetidas cíclicamente, incluyendo el diferente tipo de señal de medición. Las mediciones se realizan utilizando la señal de medición incluida en la señal CS y a través de los cálculos descritos en la norma correspondiente.

5.2.1.2 Realización práctica de una señal fuente compuesta para mediciones de hasta 20 kHz

a) Señal vocalizada para simular las propiedades de la voz

La duración de la señal asciende a 48,62 ms. Durante este periodo cualquier detector de habla deberá haber reconocido la voz y activado el sistema. La señal vocalizada puede describirse como secuencia de palabras de 16 bits con una velocidad de muestreo de 44,1 kHz.

El Cuadro 1 contiene 134 palabras que se han de repetir 16 veces para activar el sistema sometido a prueba durante un tiempo de 48,62 ms.

CUADRO 1/P.501

Muestras (valores ASCII) de la Parte 1 de la señal fuente compuesta (lectura por columnas)

-76	2098	3116	2930	2392	1824	1306	-3462	-7492	-2806	-626
112	2244	3158	2866	2410	1772	1170	-4024	-6414	-2844	-456
298	2360	3180	2808	2430	1742	968	-4590	-5334	-2888	-298
472	2456	3180	2764	2444	1750	702	-5154	-4428	-2898	-130
628	2538	3168	2728	2460	1760	394	-5716	-3772	-2846	
776	2626	3146	2686	2472	1762	76	-6298	-3360	-2698	
916	2730	3132	2632	2452	1736	-244	-6912	-3128	-2460	
1068	2824	3122	2572	2398	1684	-594	-7556	-3002	-2166	
1234	2904	3108	2496	2300	1624	-968	-8194	-2924	-1846	
1398	2964	3096	2432	2178	1572	-1384	-8719	-2870	-1544	
1572	2996	3076	2382	2068	1516	-1846	-8998	-2830	-1274	
1752	3032	3038	2362	1976	1460	-2356	-8898	-2800	-1032	
1932	3072	2992	2368	1892	1390	-2898	-8378	-2792	-818	

b) *Señal de seudorruido*

Los parámetros para la secuencia PN son los siguientes:

Velocidad de muestreo 44,1 kHz, longitud de palabra 16 bits, longitud de la transformada de Fourier 2048 puntos.

$$H(k) = \begin{cases} W(k) \cdot e^{j i_k \pi}; & k = -928, \dots, +928 \sin 0, i_k \{+1, 0\}, \text{aleatorio } i_k = i_{-k} \\ 0 & \text{en otros casos} \end{cases} \quad [5-8]$$

Según la fórmula [5-8] descrita anteriormente la señal de tiempo se calcula por la transformación inversa de Fourier. Esta secuencia se repite 4307 veces para conseguir que la secuencia de medición PN tenga una longitud de 200 ms. El factor de cresta de la secuencia PN es 11 dB ± 1 dB.

Conforme a la resolución de frecuencia de 21,533 Hz (44,1 kHz/2048), existen 928 valores de FFT en la gama de frecuencias de 0 a 20 kHz. Cada valor W(k) es 152680. Está calculado de manera que los niveles dentro de una anchura de banda de 20 kHz sean los mismos para la señal vocalizada sonora y para la secuencia PN.

c) *Pausa*

Se utiliza la pausa como se indica en la descripción general de la señal fuente compuesta. La longitud de la pausa asciende a 101,38 ms con el fin de conseguir una duración de la señal de exactamente 350 ms.

NOTA - Mediante un supermuestreo o submuestreo adecuado pueden conseguirse otras velocidades de muestreo para la secuencia descrita. El filtro de interpolación utilizado para super y submuestreo debe aproximarse al filtro rectangular ideal. La atenuación en la banda de detención debe ser > 60 dB, siendo el rizado en la banda de paso < ±0,2 dB.

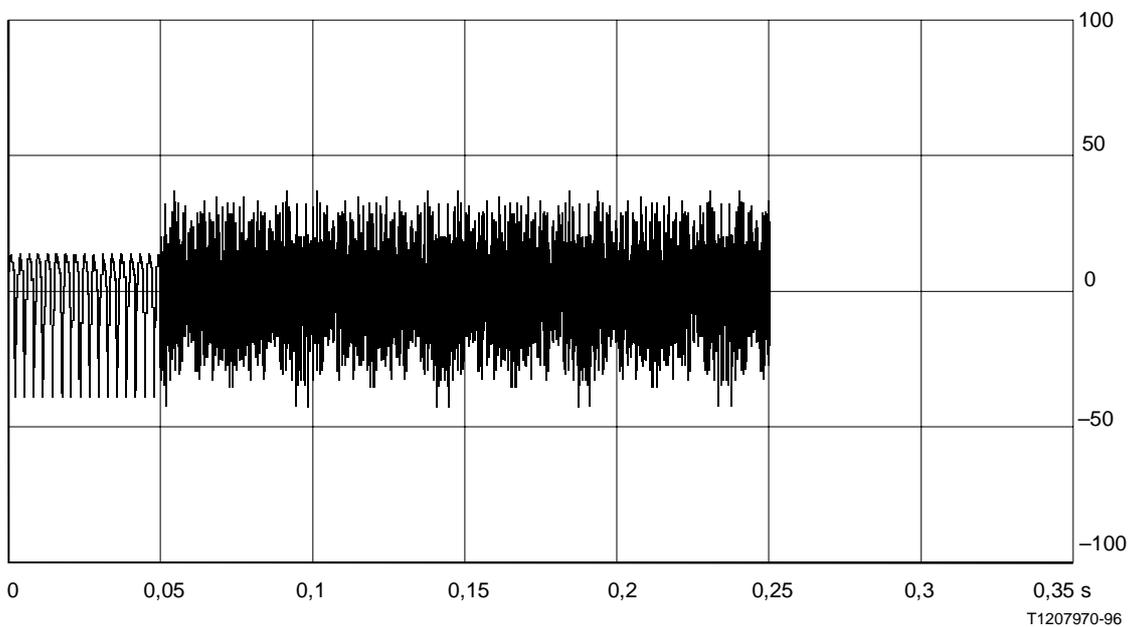


FIGURA 1/P.501

Señal fuente compuesta, señal de tiempo

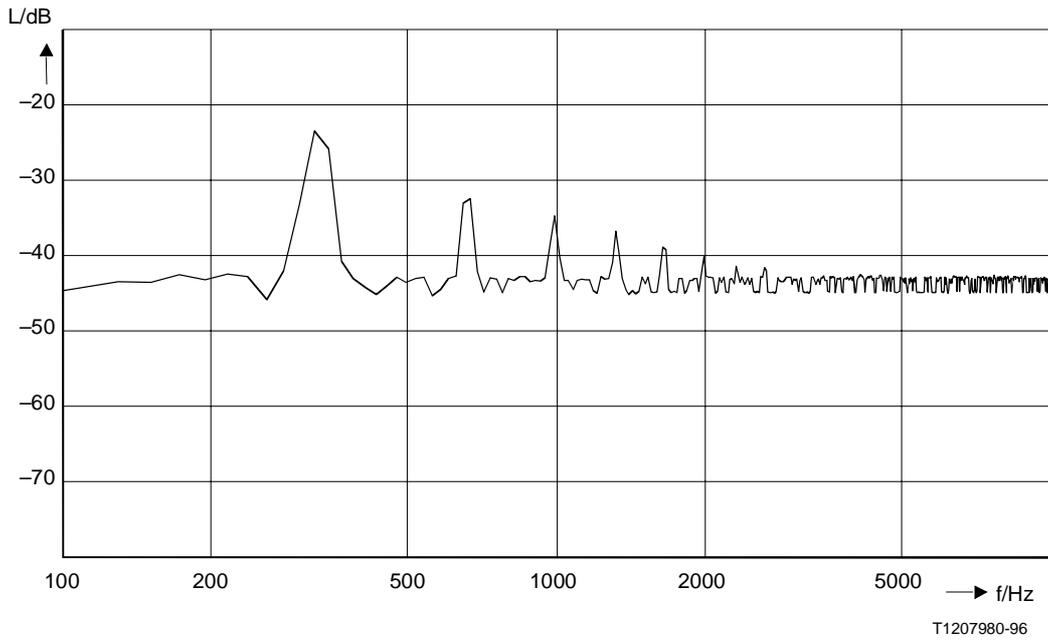


FIGURA 2/P.501

Espectro de densidad de potencia de la señal fuente compuesta (ventana de análisis: Hanning)

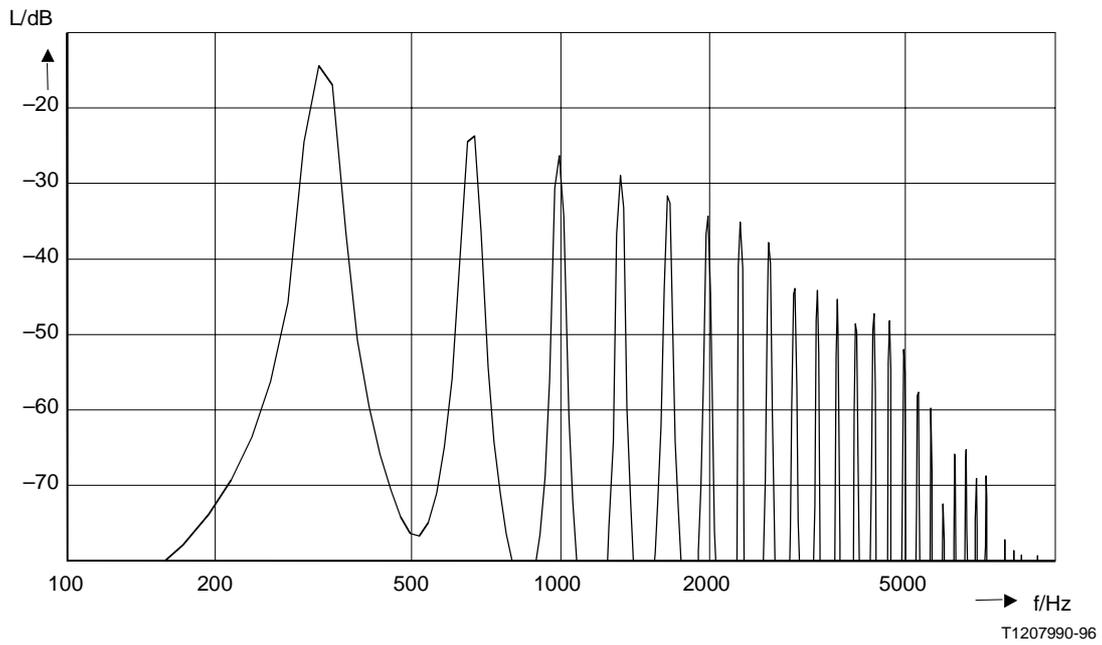


FIGURA 3/P.501

Espectro de densidad de potencia de la señal vocalizada (ventana de análisis: Hanning)

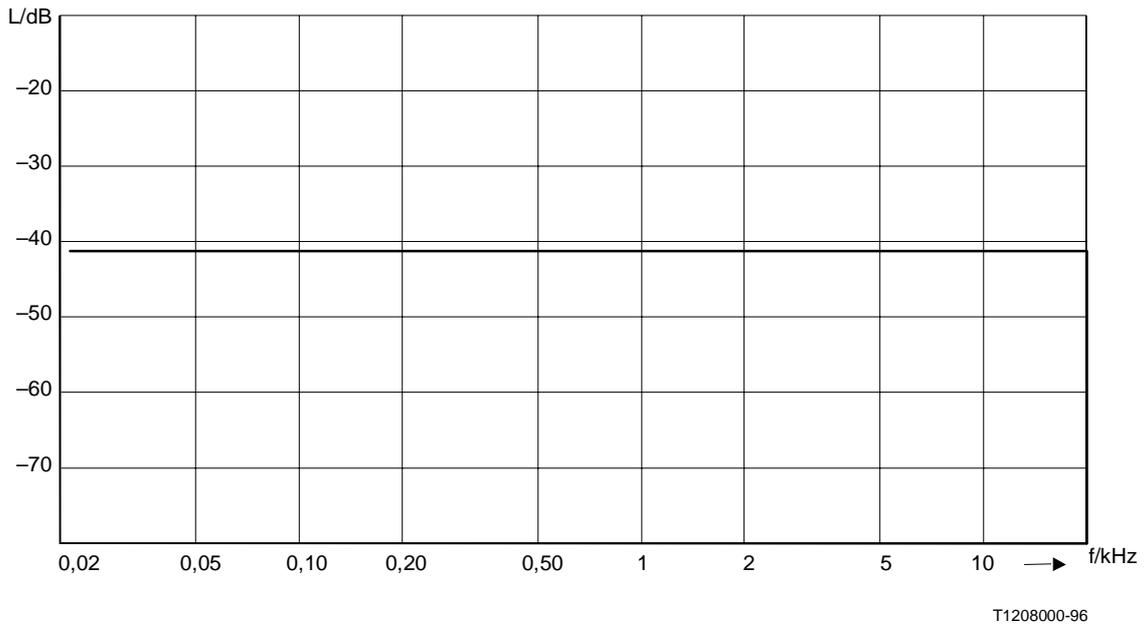


FIGURA 4/P.501

Espectro de densidad de potencia de la secuencia PN (ventana de análisis: rectángulo)

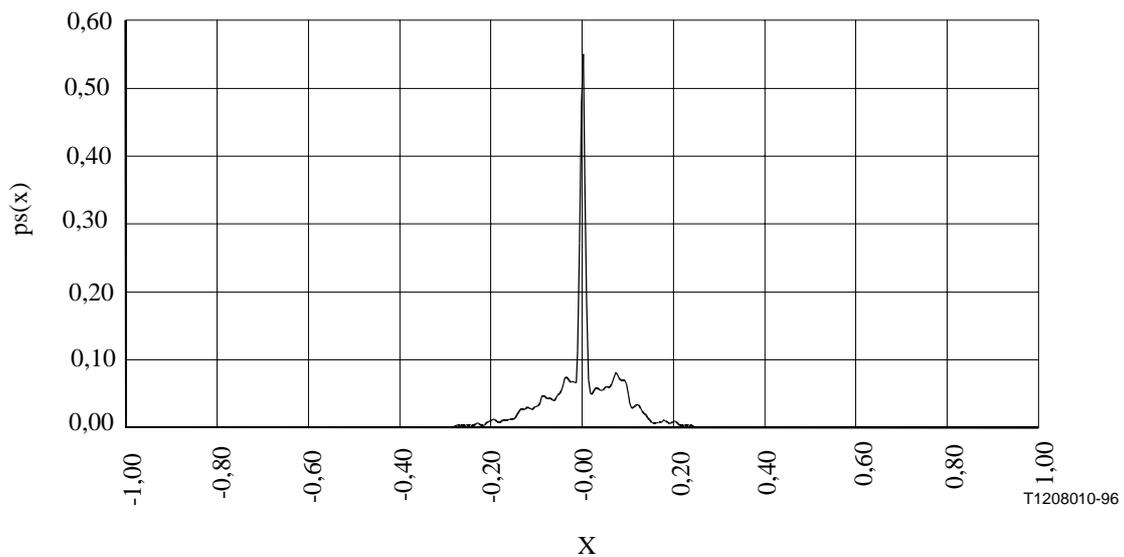


FIGURA 5/P.501

Función de densidad de probabilidad de la señal fuente compuesta según 5.2.1.2

5.2.1.3 Aplicación

La señal descrita anteriormente puede aplicarse a sistemas que se comportan de modo no lineal y variante con el tiempo pero que pueden considerarse en condiciones casi estacionarias durante el breve periodo de la medición. Pueden determinarse parámetros en el dominio de la frecuencia tales como la respuesta de frecuencia, los índices de sonoridad, etc., así como parámetros en el dominio del tiempo tales como los tiempos de conmutación. Si se inserta una señal para la medición de

distorsión en lugar de la secuencia PN (señales sinusoidales o ruido de banda estrecha) pueden determinarse asimismo esos parámetros.

En general, la señal fuente compuesta representa una clase de señales. Si para una aplicación especial se requieren partes más extensas del sonido vocalizado, puede repetirse la secuencia de este sonido hasta que se alcanza la longitud de señal requerida. El mismo procedimiento es aplicable a la secuencia PN y a la pausa. Si se desean aplicaciones especiales de este tipo, deberá describirse el procedimiento de composición de señales en la aplicación correspondiente.

En el caso de sistemas adaptativos que varían sus propiedades de transmisión dependiendo de las características de la señal conviene una señal de baja correlación. Para estos sistemas la longitud de la transformación de Fourier debe ampliarse hasta aproximadamente 200 ms (es decir, FFT de 8192 puntos en vez de la FFT de 2048 puntos). Los parámetros de análisis y generación de la señal tendrán que ajustarse de conformidad con ello: $k = -3715, \dots, 3715$ aleatorios sin 0.

5.2.1.4 Señal fuente compuesta de banda limitada con espectro de densidad de potencia similar a la voz

5.2.1.4.1 Descripción

a) *Señal fuente compuesta para monoloxua*

1) Señal vocalizada de banda limitada

En el Cuadro 1 se indican los valores ASCII para la señal vocalizada descrita en 5.2.1.2, de banda limitada entre 200 Hz y 3,6 kHz. De acuerdo con una velocidad de muestreo de 44,1 kHz los 134 valores ASCII totalizan 3,04 ms. Los valores se leen por columnas (véase el Cuadro 2).

CUADRO 2/P.501

Valores ASCII de la señal vocalizada de banda limitada

-155	948	3224	4000	3129	1440	241	-888	-1853	-6137	-3474
276	1362	3370	4043	3043	1310	190	-957	-2121	-6560	-2508
517	1741	3500	4034	2914	1146	103	-1034	-2414	-6948	-1595
578	2043	3569	3974	2750	965	-9	-1103	-2707	-7301	-802
491	2276	3603	3862	2560	776	-138	-1146	-3017	-7568	
302	2422	3603	3724	2353	603	-267	-1181	-3319	-7732	
86	2500	3595	3577	2155	448	-388	-1190	-3612	-7758	
-103	2552	3586	3439	1991	345	-491	-1198	-3913	-7620	
-207	2595	3595	3336	1853	276	-569	-1215	-4224	-7310	
-198	2655	3638	3267	1750	250	-638	-1259	-4560	-6810	
-60	2758	3724	3224	1672	250	-698	-1327	-4922	-6155	
190	2896	3819	3198	1603	267	-759	-1457	-5301	-5344	
543	3060	3922	3172	1534	267	-813	-1629	-5715	-4439	

Los valores de la señal vocalizada en la gama de frecuencias de 200 Hz a 3,6 kHz se calculan de nuevo de tal manera que el valor medio cuadrático de la señal vocal y el de la secuencia PN sean iguales. La secuencia se repite 16 veces para conseguir una longitud de 48,62 ms.

2) Señal de seudoruido generada utilizando FFT de 2048 puntos

Los parámetros de la secuencia PN son los siguientes:

Velocidad de muestreo 44,1 kHz, longitud de palabra 16 bits, longitud de la transformada de Fourier 2048 puntos.

$$H(k) = \begin{cases} W(k) \cdot e^{j i_k \pi}; & k = -928, \dots, +928 \text{ excepto } 0, i_k \{+1, 0\}, \text{ aleatorio } i_k = -i_{-k} \\ 0 & \text{en otros casos} \end{cases} \quad [5-9]$$

De conformidad con la fórmula [5-7] descrita anteriormente, la señal de tiempo se calcula mediante la transformación inversa de Fourier. Esta secuencia se repite 4307 veces para conseguir una longitud de 200 ms en la secuencia de medición de PN. El factor de cresta de la secuencia PN es 11 dB \pm 1 dB.

Conforme a la resolución de frecuencia de 21,5 Hz (44,1 kHz/2048) existen 928 valores de FFT en la gama de frecuencias de 0 a 20 kHz. Cada uno de los valores W(k) antes del filtrado es de 152680. Se calcula de forma que los niveles dentro de una anchura de banda de 20 kHz sean los mismos para la señal vocalizada y la secuencia PN.

3) Señal de seudoruido generada utilizando FFT de 8192 puntos

De conformidad con la fórmula [5-7] antes descrita, la señal de tiempo se calcula mediante la transformación inversa de Fourier. Esta secuencia se repite 1077 veces para conseguir una longitud de 200 ms en la secuencia de medición de PN. El factor de cresta de la secuencia PN es 11 dB \pm 1 dB.

De acuerdo con la resolución de frecuencias de 5,4 Hz (44,1 kHz/8192) existen 3715 valores de FFT en la gama de frecuencias de 0 a 20 kHz. Cada uno de los valores W(k) antes del filtrado es de 305360. Está calculado de manera que los niveles dentro de una anchura de banda de 20 kHz sean los mismos para la señal vocalizada y la secuencia PN.

Con el fin de conseguir el mismo valor medio cuadrático para la secuencia PN de banda limitada debe aplicarse la función de filtrado expuesta en la Figura 6. El filtro se elige de manera que los niveles de la secuencia PN filtrados y sin filtrar sean iguales.

NOTA 1 – Mediante un supermuestreo o submuestreo apropiado pueden conseguirse otras velocidades de muestreo para la secuencia descrita. El filtro de interpolación utilizado para el super y el submuestreo debe aproximarse a un filtro rectangular ideal. La atenuación en la banda eliminada debe ser > 60 dB, y el factor de ondulación de la banda de paso < \pm 0,2 dB.

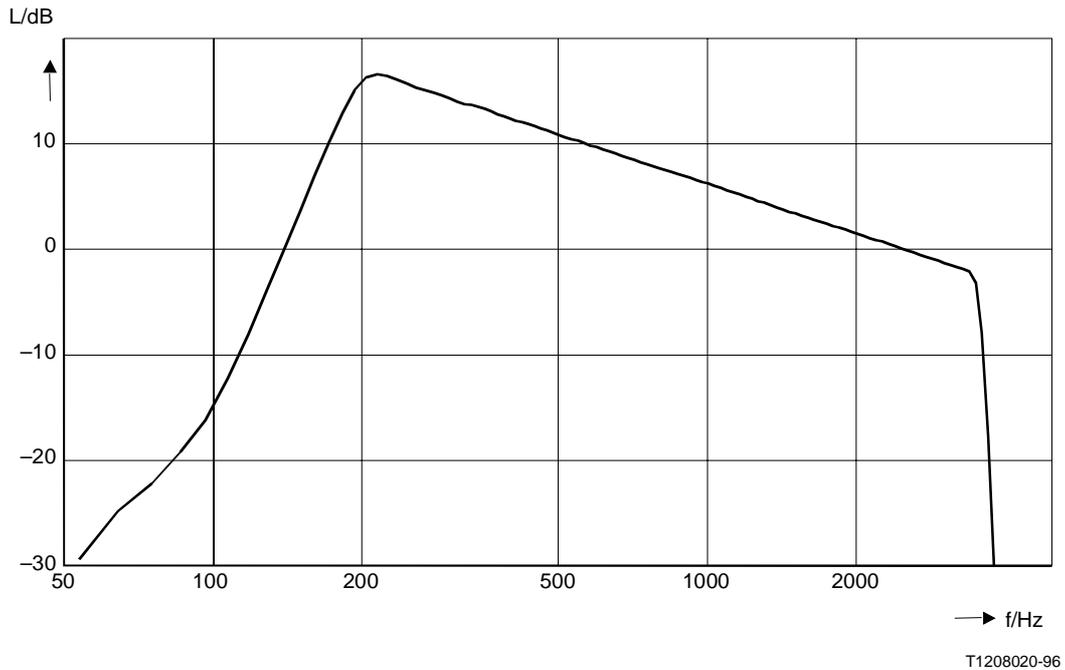


FIGURA 6/P.501

Función de transferencia del filtro para limitar la banda de la secuencia PN

CUADRO 3/P.501

Tabla de las frecuencias críticas del filtro

50 Hz	100 Hz	200 Hz	215 Hz	500 Hz	1 kHz	2,85 kHz	3,6 kHz	3,66 kHz	3,68 kHz
-25,8 dB	-12,8 dB	17,4 dB	17,8 dB	12,2 dB	7,2 dB	0 dB	-2 dB	-20 dB	-30 dB

Para sistemas adaptativos como los compensadores de eco puede ser preferible una secuencia PN más larga con el fin de no tener señales de mediciones correlacionadas dentro de la ventana de adaptación. Para estos sistemas la longitud de la FFT debe extenderse hasta 8192 puntos cuando se utilice la velocidad de muestreo de 44,1 kHz, se ha descrito en el Cuadro 3.

b) *Señal fuente compuesta de banda limitada para diloquia*

Se genera la secuencia de diloquia de la misma manera que la señal de monoloquia. Sin embargo, los tiempos de la señal vocalizada y de la pausa son ligeramente diferentes con miras a conseguir una condición de diloquia típica con dos señales aplicadas simultáneamente, con señal presente solamente en un canal, señales vocalizadas presentes en ambos extremos, así como con señales vocalizadas y no vocalizadas presentes simultáneamente en los diferentes canales. La correlación entre la señal de monoloquia y la señal de diloquia es baja. Ello se consigue eligiendo una señal vocalizada diferente, con una frecuencia de tono diferente y una señal de ruido aleatorio en lugar de la secuencia PN. La duración de la señal vocalizada es de 72,69 ms, la duración de la señal de ruido aleatorio es 200 ms, y la duración de la pausa alcanza los 127,31 ms. Esto da lugar a una longitud total de 400 ms.

1) Señal vocalizada

La señal vocalizada para diloquia se eligió con una frecuencia básica diferente que la señal vocalizada de monoloquia. En el Cuadro 4 figuran los valores para la señal vocalizada en diloquia. El nivel de este sonido vuelve a ser el mismo que en monoloquia. Con una velocidad de muestreo de 44,1 kHz, los 229 valores ASCII representan 5,19 ms. El cuadro se ha de leer por columnas:

CUADRO 4/P.501

Valores ASCII para la señal vocalizada de diloquia de banda limitada

-198	1146	-8292	4827	5853	1422	-1293	-810	-690	-1052	-621
-112	871	-8715	5094	5715	1224	-1302	-793	-724	-1043	-560
-9	560	-9077	5344	5560	1026	-1293	-767	-767	-1043	-509
103	233	-9370	5594	5387	819	-1267	-741	-793	-1052	-457
233	-121	-9542	5827	5215	603	-1250	-698	-819	-1060	-397
388	-491	-9542	6043	5043	388	-1233	-672	-845	-1060	-345
543	-871	-9361	6215	4879	181	-1224	-638	-853	-1060	-276
724	-1250	-8956	6344	4732	9	-1224	-603	-871	-1052	-207
896	-1638	-8327	6413	4586	-181	-1224	-595	-879	-1034	-112
1060	-2043	-7465	6422	4439	-328	-1224	-586	-888	-1017	
1233	-2465	-6396	6379	4276	-448	-1215	-595	-896	-991	
1388	-2896	-5163	6310	4086	-543	-1198	-603	-922	-957	
1517	-3345	-3827	6215	3870	-629	-1172	-621	-948	-931	
1638	-3819	-2448	6120	3629	-707	-1129	-629	-974	-905	
1747	-4310	-1103	6051	3370	-784	-1077	-938	-1009	-888	
1810	-4810	155	6000	3086	-871	-1026	-638	-1026	-862	
1845	-5319	1293	5991	2801	-948	-974	-638	-1052	-845	
1845	-5836	2241	5991	2534	-1026	-922	-638	-1069	-819	
1802	-6353	3034	6000	2267	-1112	-888	-638	-1077	-793	
1707	-6853	3655	6008	2034	-1181	-871	-638	-1069	-767	
1569	-7353	4138	5991	1819	-1241	-845	-647	-1060	-724	
1379	-7836	4517	5939	1612	-1276	-828	-664	-1060	-672	

Con el fin de lograr la longitud requerida de 72,69 ms los valores se han de repetir 14 veces.

2) Ruido aleatorio

Se escoge como ruido aleatorio un ruido gaussiano blanco de banda limitada en 20 kHz. El factor de cresta de las señales es 12 ± 1 dB. El valor medio cuadrático del ruido aleatorio de banda limitada se elige igual al de la señal vocalizada.

Con el fin de limitar la banda del ruido aleatorio entre 200 Hz y 3,6 kHz se utiliza la función de filtro ilustrada en la Figura 6. Esto garantiza el mismo valor medio cuadrático para el ruido aleatorio de banda limitada.

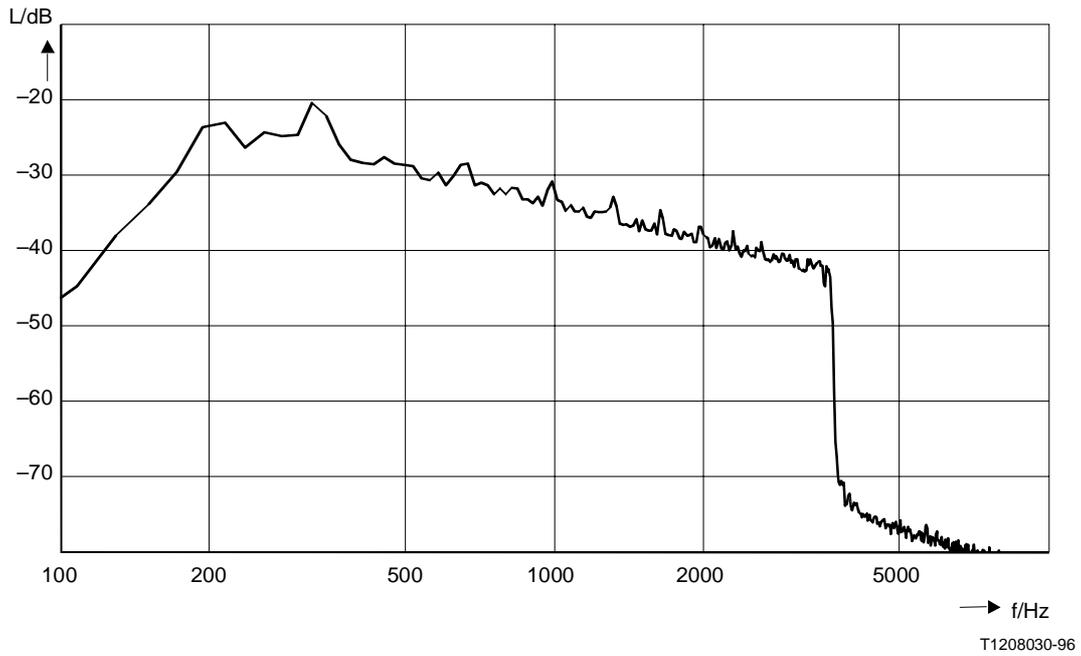


FIGURA 7/P.501

**Espectro de densidad de potencia de la señal fuente compuesta de banda limitada
(señal de monoliquia, ventana de análisis: Hanning)**

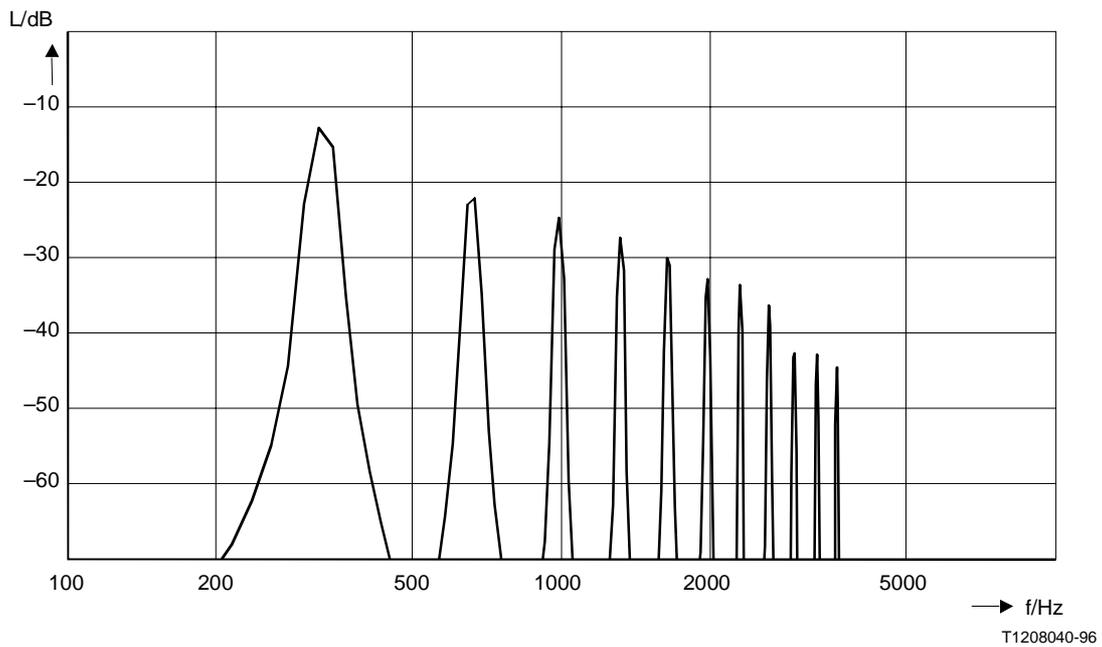


FIGURA 8/P.501

**Espectro de densidad de potencia de la señal vocalizada de banda limitada
(señal de monoliquia, ventana de análisis: Hanning)**

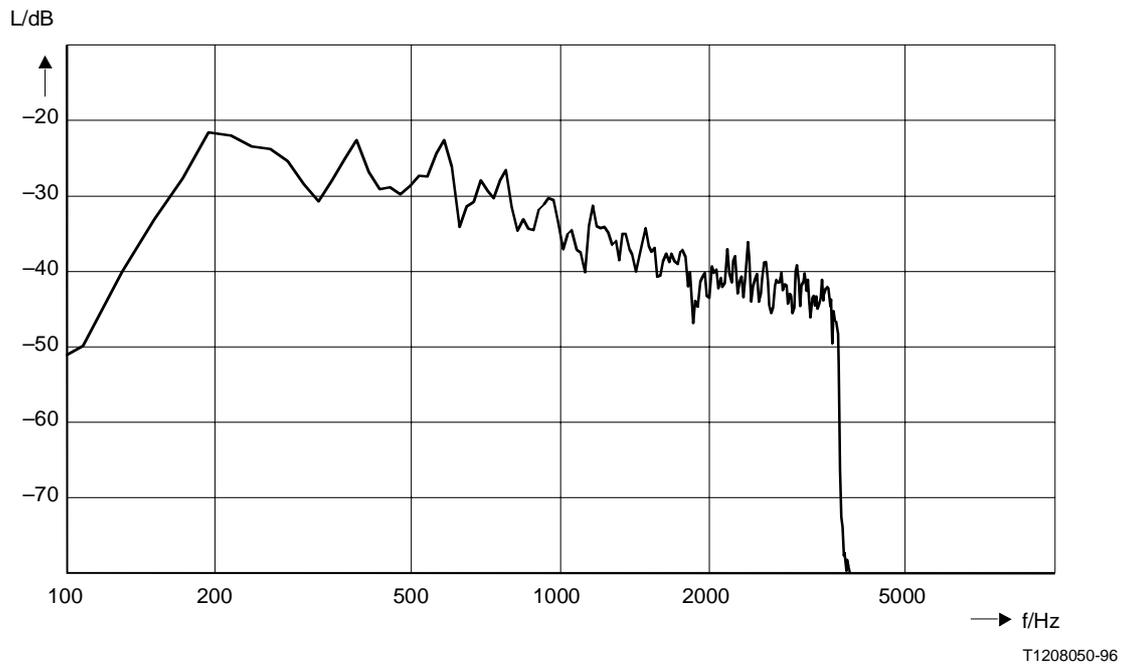


FIGURA 9/P.501

Espectro de densidad de potencia de la señal fuente compuesta de diloquia de banda limitada (ventana de análisis: Hanning)

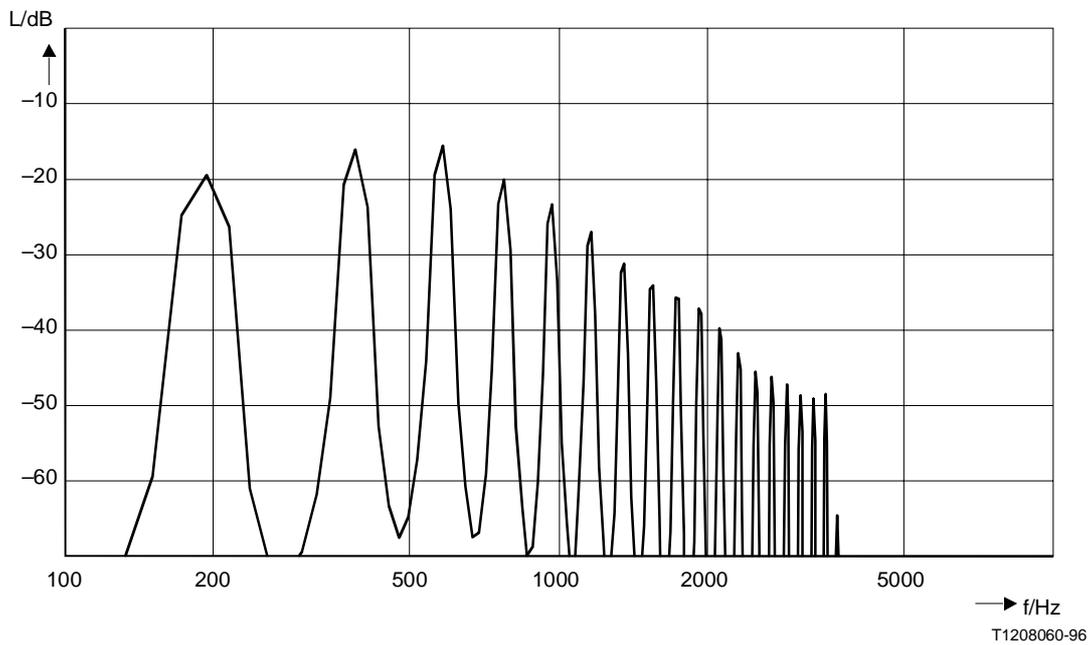


FIGURA 10/P.501

Espectro de densidad de potencia de la señal vocalizada de diloquia de banda limitada (ventana de análisis: Hanning)

NOTA 2 – Mediante un supermuestreo o submuestreo adecuados pueden conseguirse otras velocidades de muestreo para la secuencia descrita. El filtro de interpolación utilizado para el super y submuestreo debe aproximarse a un filtro rectangular ideal. La atenuación en la banda eliminada debe ser > 60 dB, y el factor de ondulación en la banda de paso $< \pm 0,2$ dB.

5.2.1.4.2 Aplicación

La aplicación de las señales fuente compuestas de banda limitada a monoloquia así como a diloquia es válida para todos los casos en que se necesite medir sistemas de banda limitada que funcionen sin linealidad y con variaciones en el tiempo y que necesiten el espectro de densidad de potencia a largo plazo típico de la voz humana. La aplicación típica es la medición de los compensadores de eco de la voz en la red. Para todas las mediciones direccionales se utilizará la señal fuente compuesta de banda limitada en condiciones de monoloquia. En el caso de diloquia se utilizará la señal de diloquia en los dos sentidos de conversación, mientras que la señal de monoloquia se inyecta en el sentido extremo distante.

5.2.2 Ruido modulado similar a la voz

5.2.2.1 Descripción

Como se indicaba en 5.1.3 puede utilizarse la MTF para medir la inteligibilidad del habla de un sistema. Modulando el ruido filtrado en la banda de una octava se puede obtener el MTF en diferentes bandas de octava. Con la ponderación correcta de los índices de modulación en cada banda de octava y frecuencias de modulación diferentes es posible conseguir un índice de transmisión de voz ($0 \leq \text{STI} \leq 1$) que tenga una elevada correlación con la inteligibilidad del habla de un sistema. Puede medirse el STI utilizando una señal compuesta de varias bandas de ruido moduladas simultáneamente. El espectro de densidad de potencia a largo plazo se elige igual al espectro de densidad de potencia del habla. Utilizando la modulación correcta se crea una señal que refleja las características temporales de una conversación en curso. El STI ha demostrado ser un buen instrumento para predecir la inteligibilidad del habla con una extensa gama de distorsiones.

5.2.2.2 Aplicación

Puede utilizarse el STI para medir la inteligibilidad del habla degradada por las distorsiones siguientes:

- ruido;
- filtrado en paso de banda;
- recorte de crestas y más en general una amplia gama de distorsiones no lineales;
- control automático de ganancia;
- reverberación.

5.2.3 Señales compuestas en frecuencia (tecnología de tono sonda)

5.2.3.1 Descripción

Con el fin de determinar las características de transmisión de sistemas telefónicos de variación dinámica puede ser necesario aplicar una señal de acondicionamiento correcta (semejante al habla) simultáneamente a una señal de prueba analítica adecuada. Es por lo tanto esencial que:

- La señal analítica se aplique a un nivel en el cual su influencia sobre el comportamiento dinámico del teléfono que se prueba sea insignificante. Este requisito puede implicar la necesidad de un largo promediado con el fin de obtener la precisión de medición suficiente. Para los sistemas de variación dinámica, ello conduce a una especie de característica de transferencia "media" que no tiene en cuenta los efectos a corto plazo.
- La correlación entre la señal de acondicionamiento y la señal de prueba analítica se mantiene en un mínimo. A menudo, esto se consigue mediante una sencilla separación en el espectro de las dos señales.

En general, no está clara la relación entre la condición real del sistema sometido a prueba y la señal de medición debido a que la señal de medición no está correlacionada con la señal de activación.

5.2.3.2 Aplicación

Se utiliza típicamente este método para determinar las características medias (a largo plazo) de un sistema. Si, por ejemplo, ha de medirse la respuesta de frecuencia media en condiciones de funcionamiento realistas (presencia de reverberación y ruido), puede utilizarse una serie de barridos lineales sinusoidales. A menudo, la señal analítica será un solo tono que también puede utilizarse por ejemplo para la medición de la variación de ganancia temporal causada por la señal de acondicionamiento y medida en la frecuencia del único tono.

5.2.4 Señales complejas compuestas

5.2.4.1 Generador de voz simulado (SSG)

5.2.4.1.1 Descripción

1) *Descripción general*

Para generar una señal que se aproxime a la distribución de amplitud del habla, se modula una señal principal con distribución gaussiana por una señal moduladora especialmente preparada, como se indica en la Figura 11. La señal resultante se conforma para aproximarse al espectro de frecuencias a largo plazo del habla, ilustrado en la Figura 12.

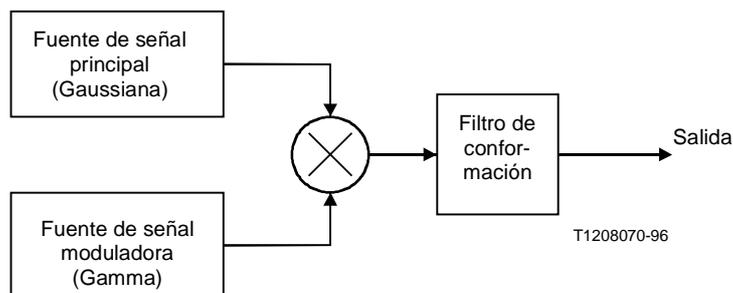


FIGURA 11/P.501

Diagrama de bloques de un generador de voz simulado

2) *Señal principal*

La señal principal consta de ocho segmentos de ruido pseudoaleatorio de 1024 puntos. Cada segmento tiene el mismo espectro de magnitudes pero un espectro de fase diferente, con la fase aleatorizada entre los segmentos y en el interior de los mismos uniformemente de 0 a 360 grados, al objeto de aleatorizar la interacción entre los productos de intermodulación de los componentes espectrales relacionados armónicamente. La duración de cada segmento es 80 ms y están mezclados todos entre sí a través de una ventana en coseno alzado con un segmento de mezclado adicional de 80 ms intercalado. La simultánea desaparición del segmento anterior y aparición del segmento siguiente elimina los fenómenos transitorios que se producirían en los límites entre segmentos. La señal principal completa consta pues de ocho segmentos pseudoaleatorios intercalados con ocho segmentos de mezclado, cada uno de 80 ms de duración y una longitud total de 1,28 s. La deseada conformación de frecuencia para aproximarse a un espectro de habla medio la proporciona un sencillo filtro a la salida.

3) *Señal moduladora*

Las mediciones muestran que una distribución gamma con parámetro $m = 0,545$ ofrece una buena aproximación a la distribución de amplitud instantánea del habla continua. Las características

silábicas pueden representarse por una respuesta de paso bajo que es prácticamente plana hasta aproximadamente 4 Hz (que puede considerarse el punto de -3 dB) seguida por una caída de -6 dB por octava.

La forma de onda final de la señal moduladora se obtuvo empíricamente de la distribución gamma. Variando el periodo de este impulso de manera pseudoaleatoria y ajustando su relación de tiempo de subida y de caída se obtiene una aproximación satisfactoria al espectro de la envolvente de modulación de la voz humana.

4) *Señal combinada*

Con el fin de ampliar el tiempo de repetición de la señal final y extender con mayor uniformidad los máximos de la señal moduladora sobre la secuencia repetida de señal gaussiana, se eligió una relación de 4/255 entre las frecuencias de reloj de muestreo de ambas señales. De este modo la frecuencia de reloj de la señal principal es 12 800 Hz y la frecuencia de reloj de la señal moduladora es aproximadamente de 200,8 Hz. Los tiempos de repetición son: 1,28 s para la señal gaussiana, 10,2 s para la señal moduladora y 326,4 s para la señal final modulada.

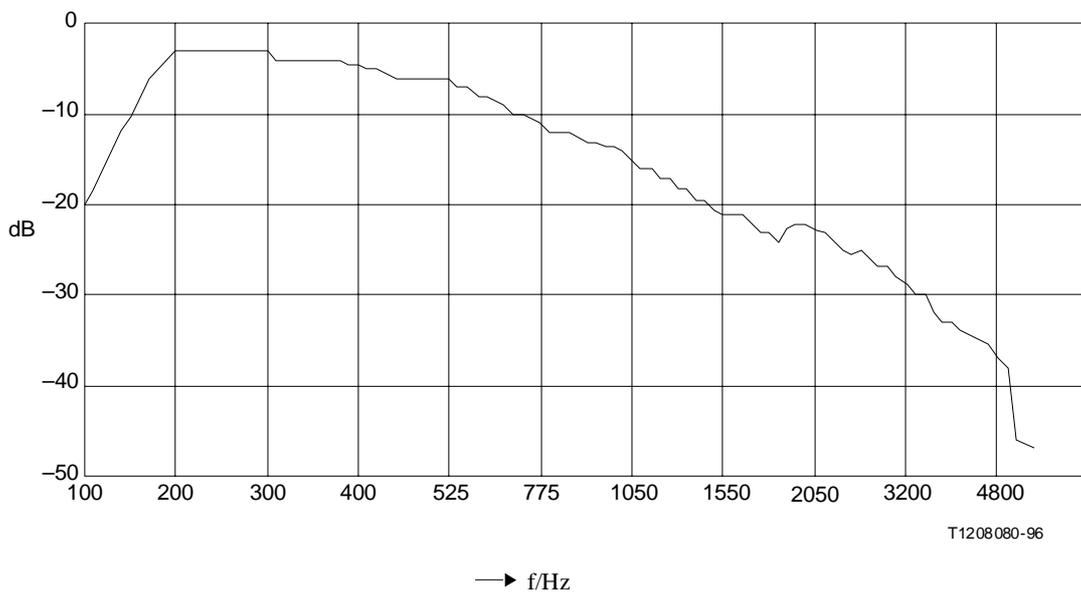


FIGURA 12/P.501

Espectro generado por el simulador de voz

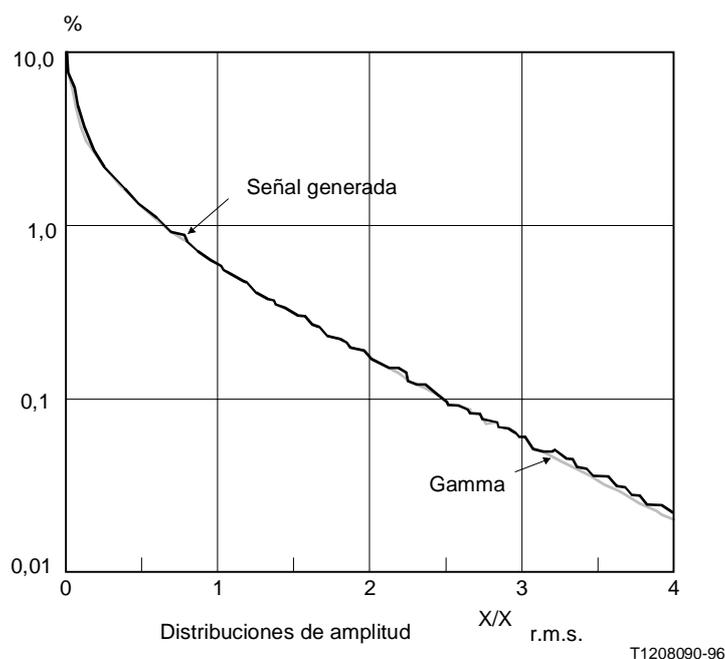


FIGURA 13/P.501

Distribución de amplitudes del simulador de voz

5.2.4.1.2 Aplicación

Las señales del generador de voz simulada (SSG) pueden utilizarse si se requiere una secuencia de habla "típica" para la medición. Análogamente a la señal P.50, la señal SSG representa la voz humana generando parámetros típicos del habla natural mediante un proceso definido. Comparado con el habla real, no se simula ningún lenguaje específico ni voz determinada.

En general se requieren tiempos de promediación largos (>10 s) si se necesita calcular a partir de las mediciones parámetros tales como las respuestas de frecuencia o los índices de sonoridad.

5.2.4.2 Voz artificial UIT-T P.50

En la Recomendación P.50 se describe la señal compuesta compleja similar a la voz más utilizada en telefonometría como voz artificial.

5.2.4.3 Habla conversacional artificial UIT-T P.59

La Recomendación P.59 describe una señal de habla conversacional artificial, compuesta y compleja, que ofrece ráfagas de conversación etc., y además una secuencia de diloquia.

5.2.4.4 Proceso de modelo de habla controlado por cadenas discretas de Markov (MSMP)

5.2.4.4.1 Descripción

1) Consideraciones generales

A continuación se incluye una breve descripción de la versión actual de un proceso de modelo de habla MSMP (proceso de modelo de habla de Markov) propuesto como señal de prueba para las aplicaciones de procesamiento de voz en banda ancha. Esta señal de prueba es una ampliación del modelo conocido para el caso de banda estrecha (MSIRP) [3]. Como se ha mencionado en publicaciones anteriores [3], [4], existían ciertas desviaciones entre el espectro a largo plazo y el espectro de la envolvente de señal del MSIRP y los del habla telefónica natural (300 Hz - 3400 Hz).

Debido a las mejoras del modelo, el espectro de la envolvente de la señal del MSIRP y el MSMP se adaptan más fielmente al del habla en banda estrecha o en banda ancha.

2) *Procedimiento de generación del proceso del modelo de habla de Markov (MSMP)*

El procedimiento de generación del MSMP se ilustra en el diagrama de bloques de la Figura 14. Como muestra este diagrama, el MSMP se constituye como producto de un proceso gaussiano $n(t)$ y un proceso $s(t)$. Este concepto hace posible ajustar por separado la función de densidad de probabilidades de amplitud (PDF) y la autocorrelación del proceso resultante.

Las propiedades que varían en el tiempo del proceso resultante están controladas por cadenas de Markov adiestradas para conseguir estructuras naturales de formantes y de tono. La decisión sobre si una trama de 20 ms de duración es vocalizada (v) o no vocalizada (uv) la realiza la cadena de Markov (mc) que es responsable del valor del tono en esa trama (tono mc). Dicha cadena adiestrada de Markov produce una secuencia natural de 33 valores de tonos diferentes. Uno de estos valores es 0 Hz, lo que indica que la trama actual es no vocalizada. Dependiendo de esta decisión es controlada una cadena de Markov generalizada que produce una secuencia de índices naturales para elegir uno entre 50 filtros formantes para esa trama (formante de mc). Esta cadena de Markov generalizada (formante mc) actúa como la parte oculta de un HMM, que produce una secuencia natural de términos de ganancia y especifica la energía a corto plazo del proceso producido (energía mc).

La rama superior controla la PDF del proceso resultante. Un filtro de paso bajo produce un proceso aleatorio de variación lenta. Este filtro es excitado por un proceso aleatorio ponderado de ruido gaussiano blanco. Los factores de ponderación se obtienen a partir de la energía mc y son constantes durante cada trama. En este lugar debe señalarse una propiedad especial de los procesos resultantes: Las PDF de amplitud de los procesos resultantes son generalmente simétricas. Pero el habla natural tiene una PDF de amplitud asimétrica. Por lo tanto utilizamos dos correspondencias no lineales diferentes para conseguir la PDF deseada del proceso $s(t)$ y conmutar entre las salidas de estas correspondencias no lineales dependiendo del signo del proceso $n(t)$. De este modo se obtiene un proceso aleatorio con la PDF de amplitud deseada. Esta PDF se forma de manera que la multiplicación del proceso $s(t)$ por un proceso gaussiano proporciona la PDF deseada del habla natural.

La rama inferior controla la autocorrelación del proceso $x(t)$. En el caso de una trama no vocalizada, el filtro formante está excitado por un proceso gaussiano no correlacionado con media cero y varianza unidad. En las tramas vocalizadas la excitación podría estar modelada por impulsos separados periódicamente. Debe tenerse en cuenta que la introducción de una frecuencia de tono causa una estructura casi periódica en el proceso resultante. Esto es una contradicción a priori puesto que el proceso resultante tiene que ser un proceso aleatorio. Puede hallarse un compromiso utilizando un filtro de peine con función de transferencia

$$H(z) = \alpha / (1 - a_k z^{-k_0})$$

Obtenemos así un proceso gaussiano que tiene una estructura casi lineal en su espectro como la excitación del filtro formante durante las tramas vocalizadas. Adviértase que el habla natural tiene también una estructura de rayas espectrales aproximada que se corresponde bien con la estructura espectral de la salida del filtro de peine. Las distancias entre estas rayas espectrales representan la frecuencia del tono, que está determinada por k_0 . El valor de k_0 viene modificado por el tono mc, pero es constante durante cada trama vocalizada. La conmutación entre las regiones vocalizadas o no vocalizadas se realiza en dos etapas para conseguir una transición suavizada. De este modo el factor de selectividad a_k del filtro de peine utilizado es igual a 0,6 para la primera y última trama de cada región y 0,95 para todos los demás casos. Finalmente, los filtros formantes están constituidos por un conjunto de 50 filtros en celosía de orden 16. Los coeficientes se obtienen de una lista de códigos, que se optimiza con el algoritmo de Lloyd generalizado [5] para diseño de listas de códigos. Para

suavizar la conmutación entre los filtros formantes, los coeficientes de filtrado reales se actualizan cada 2 ms mediante interpolación lineal entre los dos conjuntos de coeficientes de filtro de las tramas adyacentes.

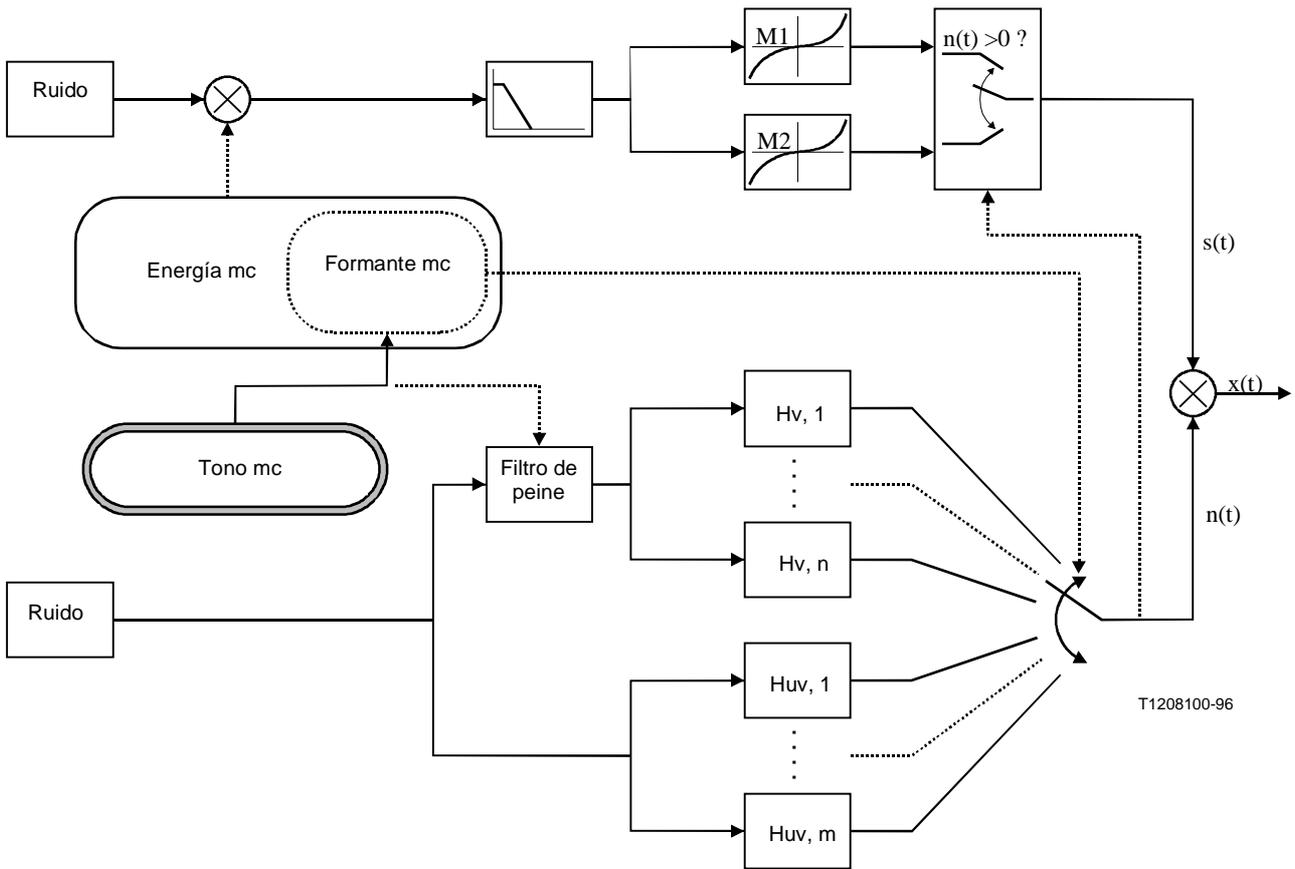


FIGURA 14-1/P.501
Diagrama de bloques del generador MSMP

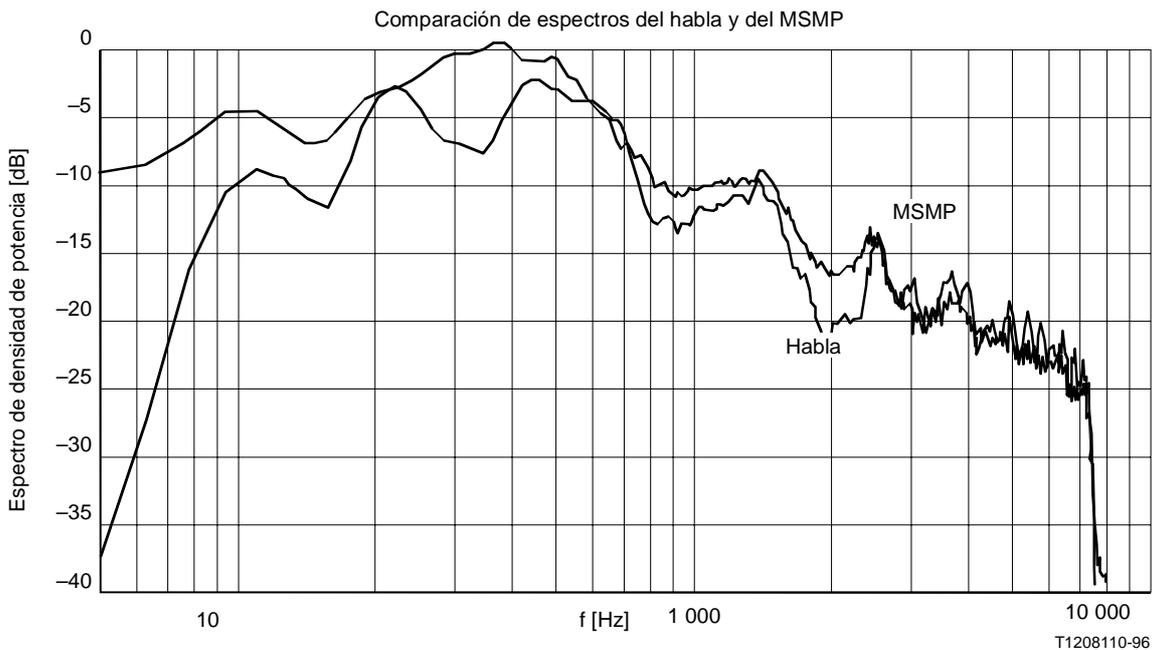


FIGURA 15-2/P.501
Comparación de espectro del habla y del MSMP

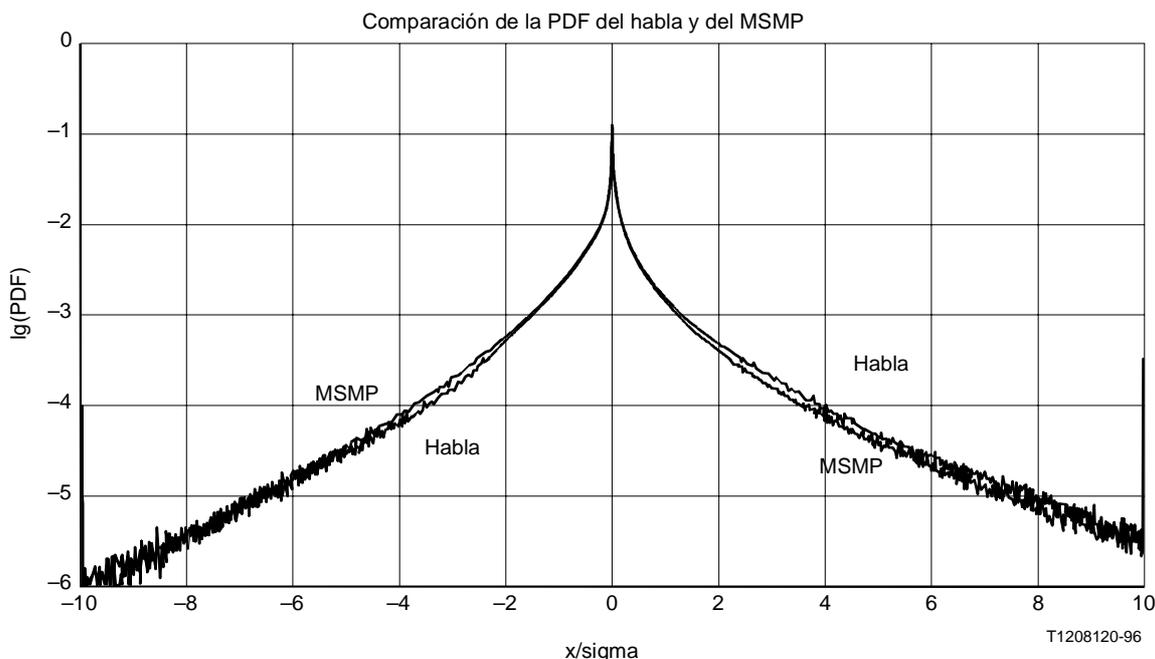


FIGURA 16/P.501

Comparación de la PDF del habla y del MSMP

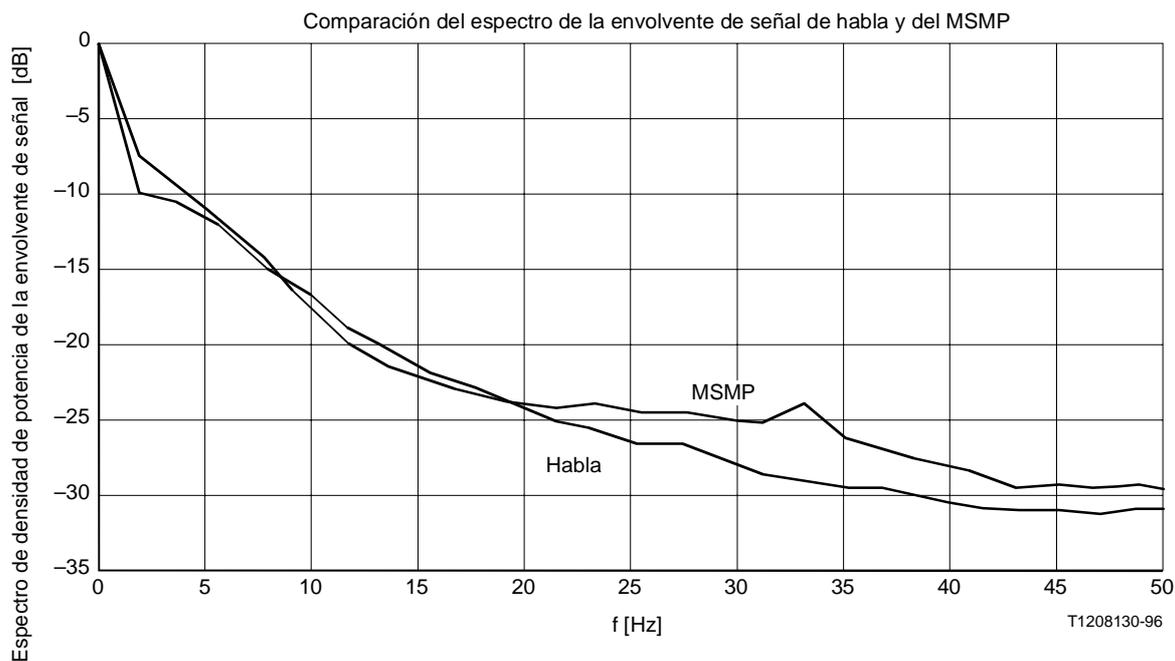


FIGURA 17/P.501

Comparación del espectro de la envolvente de señal del habla y del MSMP

5.2.4.4.2 Aplicación

Pueden utilizarse las señales MSMP si se requiere una secuencia de habla "típica" para la medición. El concepto anterior es aplicable para generar una señal de prueba que se adapta a un idioma especial (inglés o alemán por ejemplo) o a una mezcla de varios idiomas. Para ello solamente tenemos que utilizar otro juego de datos de habla para el adiestramiento de las cadenas de Markov. El MSMP permite fijar no solamente el espectro a largo plazo y la PDF, sino además incluir las características a corto plazo importantes de los cambios "naturales" de formantes y de tono. Este modelo elimina por

una parte las dependencias sobre el material hablado en sus aplicaciones pero por otro lado posibilita la introducción de características especiales requeridas de cierto tipo de habla (variación de características como hombre/mujer, fuerte/débil).

Generalmente se requieren largos tiempos de promediación (>10 s) si se necesita calcular parámetros como las respuestas de frecuencia o índices de sonoridad a partir de las mediciones.

5.3 Señales de habla

En algunas aplicaciones pueden utilizarse señales reales de voz humana (en monólogo o conversación). Si el dispositivo en prueba acusa un marcado comportamiento no lineal y/o variable en el tiempo, los conceptos de respuesta de frecuencia, distorsión y relación señal/ruido ya no son directamente aplicables, o ni siquiera tienen importancia. Puede no obstante investigarse el comportamiento del dispositivo sometido a prueba utilizando fragmentos de habla reales y evaluar la calidad total mediante un modelo de la percepción auditiva humana. Con este modelo puede definirse una "función de respuesta de frecuencia perceptual" o medidas especiales, basadas por ejemplo en parámetros sicoacústicos.

Bibliografía

- [1] GIERLICH (H.W.): A measurement technique to determine the transfer characteristics of hands-free telephones, *Signal Processing*, Vol. 27, No. 3, págs. 281-300, 1992.
- [2] STEENEKEN (H.J.M.), HOUTGAST (T.): A physical method for measuring speech transmission quality, *J. Acoust. Soc. Am.*, Vol. 67, No. 1, págs. 318-326, enero de 1980.
- [3] UIT-T COM 12-14, *Speech and Speech-Model Signals: A Comparison*, UIT-T Study Group 12, Ginebra, mayo de 1993.
- [4] HALKA (U.), HEUTE (U.): Speech-Model Processes Controlled by Discrete Markov-Chains, *Proc. Asilomar Conf. Sig. Syst. Comp.*, págs. 1196-1200, Pacific Grove, EE.UU, 1993.
- [5] LINDE (Y.), BUZO (A.), GREY (R.M.): An Algorithm for Vector Quantizer Design, *IEEE Transactions on Communications*, Vol. COM-28, N° 1, enero de 1980.

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie B	Medios de expresión
Serie C	Estadísticas generales de telecomunicaciones
Serie D	Principios generales de tarificación
Serie E	Red telefónica y RDSI
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión
Serie H	Transmisión de señales no telefónicas
Serie I	Red digital de servicios integrados (RDSI)
Serie J	Transmisiones de señales radiofónicas y de televisión
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	Mantenimiento: 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
Serie P	Calidad de transmisión telefónica
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Equipos terminales y protocolos para los 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 Z	Lenguajes de programación