



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

P.800

(08/96)

SERIE P: CALIDAD DE TRANSMISIÓN TELEFÓNICA
Métodos de evaluación objetiva y subjetiva de la calidad

**Métodos de determinación subjetiva de la
calidad de transmisión**

Recomendación UIT-T P.800

(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 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.800

MÉTODOS DE DETERMINACIÓN SUBJETIVA DE LA CALIDAD DE TRANSMISIÓN

Resumen

Esta Recomendación describe los métodos y procedimientos para llevar a cabo evaluaciones subjetivas de la calidad de transmisión. La principal revisión contenida en esta versión de la Recomendación consiste en la adición de un anexo que describe la determinación de índices por categorías de comparación (CCR). Se han efectuado otras modificaciones para ajustar esta Recomendación a la reciente revisión de la Recomendación P.830.

Orígenes

La Recomendación UIT-T P.800, ha sido revisada 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.

Palabras clave

determinación de índices por categorías absolutas, determinación de índices por categorías de degradación, determinación de índices por comparación de categorías, evaluación subjetiva, prueba de conversación, prueba de escucha, prueba subjetiva.

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	3
4	Abreviaturas.....	3
5	Convenios	3
6	Métodos recomendados	3
6.1	Pruebas de opinión sobre la conversación	3
6.2	Pruebas de opinión sobre la escucha.....	4
6.3	Entrevistas y cuestionarios.....	5
6.4	Otras pruebas	5
	Anexo A – Pruebas de opinión sobre la conversación.....	6
A.1	Instalaciones para la prueba.....	6
	A.1.1 Condiciones físicas	6
	A.1.2 Establecimiento de la conexión	11
	A.1.3 Control	11
A.2	Diseño del experimento	11
A.3	Conversación	12
A.4	Procedimiento de la prueba.....	12
	A.4.1 Elección de los participantes	12
	A.4.2 Escalas de opinión	12
	A.4.3 Instrucciones a los participantes	13
	A.4.4 Recopilación de datos.....	14
	A.4.5 Tratamiento de los resultados	14
	Anexo B – Pruebas de escucha – determinación de índices por categorías absolutas (ACR).	15
B.1	Grabaciones de señales en la fuente.....	15
	B.1.1 Entorno de grabación.....	15
	B.1.2 Sistema emisor.....	15
	B.1.3 Sistema de grabación.....	15
	B.1.4 Material de conversación.....	15
	B.1.5 Procedimiento de grabación	16
	B.1.6 Participantes en las pruebas.....	17
	B.1.7 Niveles de la señal vocal.....	17
	B.1.8 Señal de calibración.....	17

	Página
B.2 Selección de las condiciones de circuito.....	17
B.2.1 Niveles de entrada y niveles de escucha.....	17
B.2.2 Participantes en las pruebas.....	18
B.2.3 Condiciones de referencia.....	18
B.2.4 Otras condiciones	18
B.3 Concepción del experimento	18
B.4 Procedimiento de la prueba de escucha	18
B.4.1 Entorno de escucha.....	18
B.4.2 Sistema de escucha	19
B.4.3 Nivel de escucha.....	19
B.4.4 Oyentes	19
B.4.5 Escalas de opinión recomendadas por el UIT-T.....	19
B.4.6 Instrucciones a los participantes	21
B.4.7 Análisis estadístico y presentación de los resultados.....	21
Anexo C – Pruebas de detectabilidad de respuesta cuantificada	22
Anexo D – Método de determinación de índices por categorías de degradación (DCR)	23
D.1 Introducción	23
D.2 Procedimiento de determinación de índices por categorías de degradación (DCR)...	23
D.2.1 Muestras de señales vocales	23
D.2.2 Condiciones de referencia.....	24
D.2.3 Presentación de los estímulos	24
D.2.4 Instrucciones para la prueba	24
D.3 Análisis estadístico	25
Anexo E – Método de determinación de índices por categorías de comparación.....	25
E.1 Introducción	25
E.2 Referencia de calidad.....	26
E.3 Referencias MNRU.....	26
E.4 Presentación a los oyentes.....	26
E.5 Análisis de los datos	26
Anexo F – Método del umbral para comparar sistemas de transmisión con un sistema de referencia.....	27
F.1 Introducción	27
F.2 Procedimiento de prueba.....	27

	Página
F.3 Presentación de las señales	28
F.4 Fuentes vocales	29
F.5 Entorno de escucha	29
F.6 Oyentes	29
F.7 Fiabilidad	29

Introducción

Las redes de telecomunicación modernas proporcionan una extensa gama de servicios vocales que utilizan numerosos sistemas de transmisión. En particular, la rápida expansión de las tecnologías digitales ha hecho crecer la necesidad de evaluar las características de transmisión de los nuevos equipos. En muchas circunstancias es necesario determinar los efectos subjetivos causados por algún nuevo equipo de transmisión o por la modificación de las características de transmisión de una red telefónica. Esta Recomendación describe métodos para obtener evaluaciones subjetivas de los sistemas y componentes de transmisión. La Recomendación G.113 contiene información útil sobre las degradaciones que pueden sobrevenir. La Recomendación P.11 analiza los efectos que pueden tener las degradaciones de la transmisión en los usuarios de las redes y servicios de telecomunicaciones. Los métodos descritos en esta Recomendación pueden utilizarse para estimar los factores de degradación del equipo o las unidades de distorsión de cuantificación que se describen en la Recomendación G.113.

Recomendación P.800¹

MÉTODOS DE DETERMINACIÓN SUBJETIVA DE LA CALIDAD DE TRANSMISIÓN

(Modificada en Helsinki, 1993; revisada en 1996)

1 Alcance

La presente Recomendación constituye una guía para las Administraciones que desean llevar a cabo pruebas subjetivas de la calidad de transmisión en sus propios laboratorios. No se analizan otros tipos de prueba descritos con detalle en diversas Recomendaciones y documentos del UIT-T, a saber:

- a) determinación de equivalentes de referencia y de equivalentes relativos - véase el *Manual sobre Telefonometría*, Ginebra, 1993;
- b) determinación de índices de sonoridad - véase la Recomendación P.78;
- c) determinación de índices de nitidez (valores de las atenuaciones equivalentes para la nitidez - AEN) - véase el *Manual sobre Telefonometría*, Ginebra, 1993.

Tampoco se consideran los diversos tipos de pruebas especializadas que se utilizan en el curso del desarrollo de los elementos del equipo telefónico a fin de determinar el origen de los fallos y defectos, tales como las pruebas de diagnóstico mediante rimas [1] y otras pruebas relacionadas con el estudio de los aspectos específicos de la producción de señales vocales.

La finalidad de esta Recomendación es indicar los métodos considerados adecuados para determinar el grado de satisfacción que cabe prever en el funcionamiento de ciertas conexiones telefónicas.

Los métodos descritos a continuación son de aplicación general, cualquiera que sea la forma que revistan los factores de degradación presentes. Esos factores son, por ejemplo, la pérdida de transmisión (que a menudo depende de la frecuencia); el ruido del circuito; errores de transmisión (errores de bit aleatorios además de la supresión de tramas que se produce en sistemas como los de comunicaciones móviles); el ruido ambiental; el efecto local; el eco para el hablante; diversos tipos de distorsión no lineal, incluida la codificación a baja velocidad binaria, el tiempo de propagación, los efectos perjudiciales de los dispositivos accionados por la voz, las distorsiones de la escala de tiempo debidas a la conmutación de paquetes, y las degradaciones del canal de comunicación que varían en el tiempo, incluidas las de los altavoces. También hay que tener presente las combinaciones de dos o más de estos factores. En las Recomendaciones P.830 (códecs de voz digitales), P.84 (DCME/PCME) y P.85 (dispositivos generadores de voz) se dan más orientaciones para aplicaciones específicas.

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 eue 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.

¹ Antigua Recomendación P.80.

- Publicación 1260 de la CEI: 1995, *Electroacoustics - Octave-band and fractional - Octave-band filters.*
- Publicación 581-5 de la CEI: 1981, *High fidelity audio equipment and systems; Minimum performance requirements - Part 5: Microphones.*
- Publicación 651 de la CEI: 1979, *Sound level meters. (Amendment 1-1993) (Corrigendum March 1994).*
- ISO 266: 1975, *Preferred frequencies for measurements.*
- ISO 1996-1: 1982, *Acoustics - Description and measurement of environmental noise - Part 1: Basic quantities and procedures.*
- ISO 1996-2: 1987, *Acoustics - Description and measurement of environmental noise - Part 2: Acquisition of data pertinent to land use.*
- ISO 1996-3: 1987, *Acoustics - Description and measurement of environmental noise - Part 3: Application to noise limits.*
- Recomendación UIT-T G.113 (1996), *Degradaciones de transmisión.*
- Recomendación G.722 del CCITT (1988), *Codificación de audio de 7 kHz dentro de 64 kbit/s.*
- Recomendación G.726 del CCITT (1990), *Modulación por impulsos codificados diferencial adaptativa (MICDA) a 40, 32, 24 y 16 kbit/s.*
- Recomendación G.728 del CCITT (1992), *Codificación de señales vocales a 16 kbit/s utilizando predicción lineal con excitación por código de bajo retardo.*
- Recomendación UIT-T G.729 (1996), *Codificación de la voz a 8 kbit/s mediante predicción lineal con excitación por código algebraico de estructura conjugada.*
- Recomendación UIT-T P.10 (1993), *Vocabulario de términos sobre calidad de transmisión telefónica y aparatos telefónicos.*
- Recomendación UIT-T P.11 (1993), *Efectos de las degradaciones de la transmisión.*
- Recomendación P.48 del CCITT (1988), *Especificación de un sistema intermedio de referencia.*
- Recomendación UIT-T P.56 (1993), *Medición objetiva del nivel vocal activo.*
- Recomendación UIT-T P.78 (1993), *Método de prueba subjetivo para determinar índices de sonoridad de acuerdo con la Recomendación P.76.*
- Recomendación UIT-T P.810 (1996), *Aparato de referencia para el ruido modulado.*
- Recomendación P.82 del CCITT (1984), *Método para la evaluación del servicio desde el punto de vista de calidad de transmisión de la palabra.*
- Recomendación UIT-T P.830 (1996), *Evaluación de la calidad de funcionamiento subjetiva de los códecs digitales de banda telefónica y de banda ancha.*
- Recomendación UIT-T P.84 (1993), *Método de pruebas de escucha subjetiva para la evaluación de equipos de multiplicación de circuitos digitales y de sistemas de voz por paquetes.*
- Recomendación UIT-T P.85 (1994), *Método para la evaluación subjetiva de la calidad vocal de los dispositivos generadores de voz.*

3 Definiciones

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

3.1 dBov: dB relativos a la sobrecarga de un sistema digital.

3.2 Q: Relación, en dB, de la potencia vocal a la potencia de ruido modulado en el aparato de referencia para ruido modulado, tal como se describe en la Recomendación P.810.

4 Abreviaturas

A los efectos de esta Recomendación, se utilizan las abreviaturas siguientes.

ACR	Determinación de índices por categorías absolutas (<i>absolute category rating</i>)
BER	Proporción de bits erróneos (<i>bit error rate</i>)
CCR	Determinación de índices por categorías de comparación (<i>comparison category rating</i>)
CMOS	Nota media de opinión sobre las comparaciones (<i>comparison mean opinion score</i>)
DCR	Determinación de índices por categorías de degradación (<i>degradation category rating</i>)
DMOS	Nota media de opinión sobre las degradaciones (<i>degradation mean opinion score</i>)
FER	Velocidad de borrado de trama (<i>frame erasure rate</i>)
IRS	Sistema intermedio de referencia (Recomendación P.48) (<i>intermediate reference system</i>)
MIC	Modulación por impulsos codificados
MICDA	Modulación por impulsos codificados diferencial adaptativa
MNRU	Aparato de referencia para ruido modulado (Recomendación P.810) (<i>modulated noise reference unit</i>)
MOS	Nota media de opinión (<i>mean opinion score</i>)
SNR	Relación señal-ruido (<i>signal-to-noise ratio</i>)

5 Convenios

La evaluación subjetiva de los sistemas y equipos de telecomunicación puede en principio llevarse a cabo utilizando métodos de pruebas subjetivas de escucha solamente o de conversación. En el terreno práctico, las pruebas de sólo escucha tal vez sean el único método factible de pruebas subjetivas durante el desarrollo de nuevos equipos de transmisión o servicios de telecomunicación. Esta Recomendación describe los procedimientos recomendados para los métodos de evaluación subjetiva de conversación y de sólo escucha.

6 Métodos recomendados

6.1 Pruebas de opinión sobre la conversación

Las pruebas de conversación en laboratorio pretenden, en la medida de lo posible, reproducir en el laboratorio las condiciones reales de servicio de los usuarios telefónicos. Para ello es preciso seleccionar adecuadamente las condiciones del circuito y los participantes, y llevar a cabo las pruebas de manera apropiada.

Es importante especificar y establecer correctamente las condiciones simuladas en la prueba, realizar mediciones precisas antes y después de cada experimento, prever los dispositivos auxiliares necesarios, como los de marcación y tono de llamada, y llevar registros fieles de los resultados de cada prueba. En el Anexo A figura una descripción detallada del método, así como las consideraciones y precauciones necesarias.

6.2 Pruebas de opinión sobre la escucha

No está previsto que las pruebas de opinión sobre la escucha alcancen el mismo grado de realismo que las pruebas de conversación y por consiguiente las restricciones son menos rigurosas en algunos aspectos; no obstante, la artificialidad que hay que aceptar exige un control estricto de muchos parámetros que en las pruebas de conversación se permite que alcancen su propio equilibrio.

El método de prueba recomendado para pruebas de escucha solamente es el de determinación de «índices de categorías absolutas» (ACR, *absolute category rating*) que se describe en el Anexo B; este método se adapta al de juicios de categoría recomendado para las pruebas de conversación (véase el Anexo A), que se ha adoptado en parte por las mismas razones. Los índices de categoría se aplican a grupos breves de frases sin relación, cada uno de los cuales pasa a través de un cierto número de procesos normalizados, así como a través de los procesos bajo prueba. Este método es bien conocido y se ha aplicado a conexiones telefónicas analógicas y digitales y a dispositivos de telecomunicación como los códecs digitales. Por ejemplo, en los trabajos realizados en las Recomendaciones G.726 (MICDA a 32 kbit/s), G.728, G.729 y G.722, los laboratorios de varios países llevaron a cabo pruebas subjetivas utilizando el mismo método en condiciones físicas semejantes y con sistemas de transmisión idénticos, y los resultados mostraron un alto grado de coherencia.

Otros métodos utilizados normalmente son el método de detectabilidad con respuesta cuantificada, el de determinación de índices por categoría de degradación (DCR, *degradation category rating*), el de determinación de índices por categorías de comparación (CCR, *comparison category rating*) y el método de umbral.

El Anexo C describe las pruebas de detectabilidad con respuesta cuantificada que son adecuadas para evaluar los valores de umbral de ciertas cantidades y sus probabilidades asociadas. Por ejemplo, este es el mejor método para determinar el nivel por encima del cual la interferencia ocasionada por una sola frecuencia tiene una determinada probabilidad de que sea objetable o detectable, o la probabilidad de que la diafonía en una gama dada de niveles sea inteligible.

Una alternativa al método de índices por categorías absolutas es el denominado método de determinación de índices por categorías de degradación (DCR), que se describe con detalle en el Anexo D. El método DCR compara el sistema bajo prueba con una referencia fija de alta calidad y evalúa la degradación mediante una escala de cinco notas (desde «inaudible» hasta «muy molesta»). Este método es adecuado cuando la degradación es pequeña (especialmente las degradaciones digitales). Por consiguiente, puede ser especialmente útil para evaluar, por ejemplo, algoritmos destinados a procesamiento digital de la palabra. El método DCR puede así servir para optimizar el sistema, una vez demostrado, por los métodos descritos en los Anexos A y B, que la conexión más desfavorable que contiene la degradación en cuestión se encuentra dentro de límites aceptables.

El Anexo E describe una variante del procedimiento DCR denominada determinación de índices por categorías de comparación (CCR). Como en el DCR, el método CCR compara el sistema sometido a prueba con una referencia fija de alta calidad (en el caso CCR en una escala que va desde "mucho mejor" a "mucho peor"). Este procedimiento puede ser especialmente adecuado para sistemas que mejoran la calidad de la señal vocal recibida (por ejemplo, sistemas de compensación de ruido).

El método de umbral, que también resulta adecuado para la optimización del sistema, se describe en el Anexo F. Comparando directamente el sistema sometido a prueba con un sistema de referencia,

como el aparato de referencia para ruido modulado (MNRU, *modulated noise reference unit*) descrito en la Recomendación P.810, es posible determinar el valor de la condición de referencia (Q en los procesos digitales), igualándolo a la calidad de funcionamiento del sistema sometido a prueba.

En 2.6 del *Manual sobre Telefonometría* figura información sobre otros tipos de métodos de pruebas subjetivos que incluyen métodos de escala.

Las pruebas de escucha tienen aplicación directa en la evaluación de los sistemas de transmisión física que son esencialmente unidireccionales, entre los que se encuentran los circuitos de radiodifusión, los sistemas de avisos públicos y los de anuncios grabados, donde puede haber degradaciones de la escucha tales como atenuaciones, ruido y distorsión.

Con ciertas reservas pueden aplicarse los resultados de las pruebas de sólo escucha a la evaluación de las conversaciones efectuadas a través de un sistema bidireccional, por ejemplo, la conexión a una red telefónica pública con conmutación. Se deben tener debidamente en cuenta los efectos de los siguientes factores adicionales:

- degradaciones de la palabra (por ejemplo, efecto local y eco);
- degradaciones de la conversación (por ejemplo, tiempo de propagación y mutilación de la señal vocal a causa de los dispositivos activados por la voz).

Los anexos a la presente Recomendación proporcionan información sobre la preparación del material de conversación, el procesamiento de dicho material, los principios del experimento (incluida la elección de las condiciones del circuito), el procedimiento de la prueba de escucha y el tratamiento de los resultados.

6.3 Entrevistas y cuestionarios

Si se dispone del personal numeroso que ello exige y la importancia del estudio lo justifica, la calidad de transmisión puede determinarse mediante «observaciones en condiciones de servicio». En la Recomendación P.82 se indica la forma de llevar a cabo estas observaciones, incluidas las preguntas que deben hacerse a los abonados entrevistados. Para mantener un grado elevado de precisión, es preciso realizar al menos 100 entrevistas para cada condición.

Un inconveniente que presenta el método de observación en servicio, para muchos efectos, es el hecho de que apenas se pueden controlar las características detalladas de las conexiones telefónicas objeto de prueba. Sin embargo, este método permite hacer una estimación global del comportamiento del equipo en un entorno real.

En 2.5.8.3 del *Manual sobre Telefonometría* figura más información al respecto.

6.4 Otras pruebas

En la referencia [2] aparece información sobre un método que supera los inconvenientes de la técnica de entrevistas indicada en 6.3 y mantiene la mayoría de sus ventajas. Este método, denominado SIBYL, permite cursar una pequeña parte de las comunicaciones habituales de un usuario a través de dispositivos especiales que modifican la calidad normal de la transmisión de acuerdo con un programa de prueba. Cuando se trata de esa forma a una comunicación, se solicita al participante que manifieste su opinión marcando una cifra de una serie de ellas. De esta manera todos los resultados se almacenan en el computador de control y se mantiene el secreto de las comunicaciones.

Anexo A

Pruebas de opinión sobre la conversación

A.1 Instalaciones para la prueba

A.1.1 Condiciones físicas

A.1.1.1 Cabinas de prueba

Los dos participantes se sientan en cabinas separadas insonorizadas, situadas cerca del punto desde el cual se controla el experimento. El volumen de la sala no debe ser inferior a 20 m³, con un tiempo de reverberación menor de 500 ms (normalmente entre 200 y 300 ms), para los sistemas convencionales del tipo microteléfono, o para sistemas de casco telefónico, y no debe ser inferior a 30 m³ para sistemas de manos libres (debe prestarse más atención si el tiempo de reverberación es una variable del experimento).

Las dimensiones interiores de la cabina deben ser tales que se reduzca al mínimo el efecto de la onda estacionaria. Una relación típica de dichas dimensiones es 5:4:3.

Los materiales de construcción de las salas deben garantizar una atenuación suficiente del ruido externo, a fin de satisfacer los requisitos de A.1.1.2.1.

Las cabinas deben decorarse convenientemente para lograr un ambiente natural.

A.1.1.2 Ruido

A.1.1.2.1 Umbral de ruido

El nivel de ruido ambiente (cuando no se introduce deliberadamente ruido ambiental) debe mantenerse lo más bajo posible. Por razones prácticas, tales como renovación regular del aire en la cabina, el objetivo es un límite superior NC25 [3] o NR25 (véase la Recomendación ISO 1996). Estos valores se aproximan a los niveles de ruido en los hogares (dormitorios), hospitales y bibliotecas.

A.1.1.2.2 Ruido ambiental

El ruido ambiental se introduce con el espectro necesario (por ejemplo, el espectro Hoth para representar el caso típico de ruido en una habitación – véase A.1.1.2.2.1) al nivel necesario (por ejemplo, 50 dBA) medido con un sonómetro de precisión conforme a la Publicación 651 de la CEI, empleando la «ponderación A» y la característica «rápida» del medidor. Si en el mismo experimento distintas conversaciones exigen diversos niveles de ruido de sala, deben tomarse precauciones para evitar que las transiciones sean demasiado evidentes a los participantes. Idealmente, el ruido de sala debe cambiarse sólo cuando los participantes se encuentran fuera de las salas insonorizadas. Si no es posible, los cambios de nivel se realizan de forma gradual (a un ritmo no superior a 4 dB por segundo) cuando no está en curso ninguna conversación experimental y la atención de los participantes está ocupada, por ejemplo, para comunicarse con el operador.

En A.1.1.2.2.1 y A.1.1.2.2.2 figuran los espectros con características a largo plazo apropiadas.

En algunas circunstancias es preciso utilizar un ruido cuyo nivel o espectro fluctúe, como las grabaciones magnetofónicas de ruido de oficina o de tráfico reales. En tales casos debe asegurarse que las características estadísticas son estables cuando se promedian a lo largo de un periodo de tiempo razonablemente corto, por ejemplo un minuto.

Se recomienda medir al menos dos veces el espectro y el nivel de ruido, al principio y al final del experimento. Toda variación significativa en las dos mediciones, al compararlas, debe ser evaluada por el experimentador, pues puede que arroje dudas sobre la validez del experimento.

Es fundamental que los altavoces y amplificadores sean capaces de reproducir fielmente el ruido necesario.

A.1.1.2.2.1 Ruido de sala

El ruido de sala tendrá un espectro de densidad de potencia correspondiente a la publicada por Hoth [4]. En el Cuadro A.1 se indica la densidad espectral con un nivel ajustado para producir una lectura de 50 dBA en un sonómetro conforme con la Publicación 651 de la CEI. Esto se ilustra en la Figura A.1. Este espectro es independiente del nivel, o sea que para 40 dBA, el nivel en cada banda será inferior en 10 dB al indicado en el Cuadro A.1. En este cuadro consta también información sobre la potencia en cada banda de un tercio de octava.

CUADRO A.1/P.800

Espectro de frecuencias del ruido de sala

Frecuencia (Hz)	Densidad espectral (dB SPL/Hz)	Anchura de banda $10 \log_{10} \Delta f$ (dB)	Potencia total en cada banda de 1/3 de octava (dB SPL)	Tolerancia (dB)
100	32,4	13,5	45,9	±3
125	30,9	14,7	45,4	
160	29,1	15,7	44,9	
200	27,6	16,5	44,1	
250	26,0	17,6	43,6	
315	24,4	18,7	43,1	
400	22,7	19,7	42,3	
500	21,1	20,6	41,7	
630	19,5	21,7	41,2	
800	17,8	22,7	40,4	
1000	16,2	23,5	39,7	
1250	14,6	24,7	39,3	
1600	12,9	25,7	38,7	
2000	11,3	26,5	37,8	
2500	9,6	27,6	37,2	
3150	7,8	28,7	36,5	
4000	5,4	29,7	34,8	
5000	2,6	30,6	33,2	
6300	-1,3	31,7	30,4	
8000	-6,6	32,7	26,0	

NOTAS

- 1 La señal eléctrica de entrada, por ejemplo de ruido blanco, tendrá una banda limitada a las bandas de un tercio de octava centradas en las frecuencias preferidas según la ISO (ISO 266) entre 100 Hz y 8000 Hz, y los bordes de la banda se conformarán a los filtros descritos en la Publicación 1260 de la CEI.
- 2 El ruido de sala acústico es difícil de controlar en baja frecuencia, sobre todo en la región no especificada por debajo de 100 Hz, debido a las dimensiones de las cabinas de prueba típicas, a la mediocre atenuación de estas cabinas y a la influencia de ruidos externos (por ejemplo, aparatos de climatización). Por ello es conveniente elegir una cabina de prueba que reduzca al mínimo estos niveles de presión sonora en baja frecuencia.

A.1.1.2.2 Ruido en el interior de vehículos

Se ha recomendado utilizar dos espectros para el ruido en el interior de vehículos [5] y [6]. Quedan adecuadamente representados por curvas simplificadas [7]: un espectro para los vehículos en movimiento y el otro para los vehículos estacionados. En el Cuadro A.2 se indican las densidades espectrales junto con información sobre la potencia en cada banda de un tercio de octava. La densidad espectral correspondiente a los vehículos en movimiento está representada en la Figura A.2 a) y la correspondiente a los vehículos estacionados en la Figura A.2 b). Estos espectros son independientes del nivel.

NOTA - Los espectros de ruido que figuran en el Cuadro A.2 deben considerarse provisionales. Están en estudio especificaciones más detalladas.

El Cuadro A.3 recoge los valores calculados del nivel de presión sonora no ponderada a diversas velocidades, en las bandas de frecuencias de un tercio de octava recomendadas por la ISO, centradas en 63 Hz a 8000 Hz.

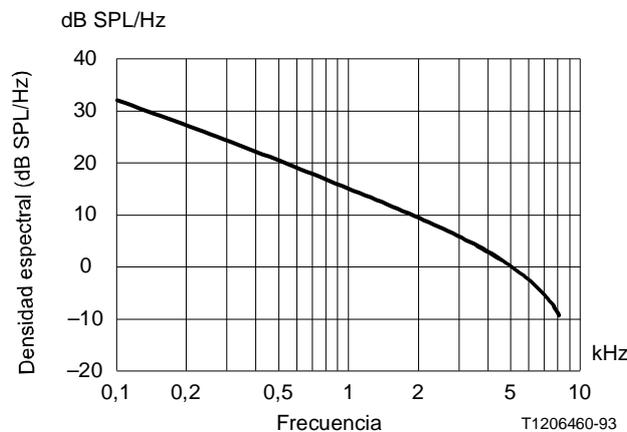


FIGURA A.1/P.800

Densidad espectral del ruido de sala

CUADRO A.2/P.800

Espectro del ruido en el interior de vehículos

Frecuencia (Hz)	Densidad espectral (dB SPL/Hz)		Anchura de banda $10 \log_{10} \Delta f$ (dB)	Potencia total en cada banda de un tercio de octava (dB SPL)		Tolerancia (dB)
	En movimiento	Estacionado		En movimiento	Estacionado	
63	72,3	58,3	11,7	84,0	70,0	
80	69,3	55,0	12,7	82,0	66,7	
100	66,5	49,8	13,5	80,0	63,3	
125	63,3	45,1	14,7	78,0	60,0	
160	60,3	42,0	15,7	76,0	56,7	
200	57,5	36,8	16,5	74,0	53,3	
250	54,4	34,7	17,6	72,0	52,3	

CUADRO A.2/P.800

Espectro del ruido en el interior de vehículos

Frecuencia (Hz)	Densidad espectral (dB SPL/Hz)		Anchura de banda $10 \log_{10} \Delta f$ (dB)	Potencia total en cada banda de un tercio de octava (dB SPL)		Tolerancia (dB)
	En movimiento	Estacionado		En movimiento	Estacionado	
315	51,3	32,6	18,7	70,0	51,3	±3
400	48,3	30,6	19,7	68,0	50,3	
500	45,4	28,7	20,6	66,0	49,3	
630	42,3	26,6	21,7	64,0	48,3	
800	39,3	24,6	22,7	62,0	47,3	
1000	36,5	22,8	23,5	60,0	46,3	
1250	33,3	20,6	24,7	58,0	45,3	
1600	30,3	18,6	25,7	56,0	44,3	
2000	27,5	16,8	26,5	54,0	43,3	
2500	24,4	14,7	27,6	52,0	42,3	
3150	21,3	12,6	28,7	50,0	41,3	
4000	18,3	10,6	29,7	48,0	40,3	
5000	15,4	8,7	30,6	46,0	39,3	
6300	12,3	6,6	31,7	44,0	38,3	
8000	9,3	4,6	32,7	42,0	37,3	

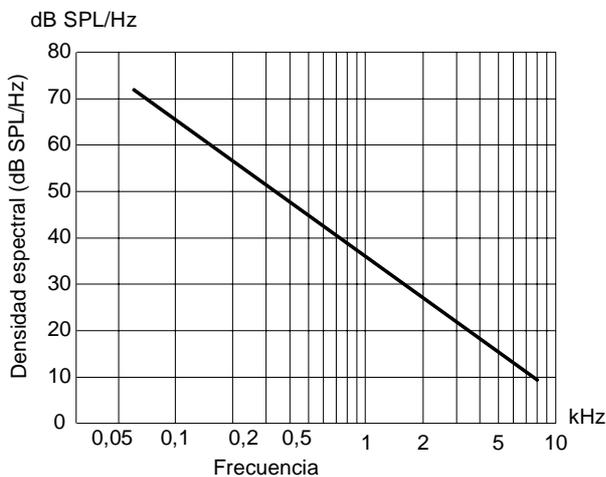
CUADRO A.3/P.800

Niveles de presión sonora calculados de los espectros

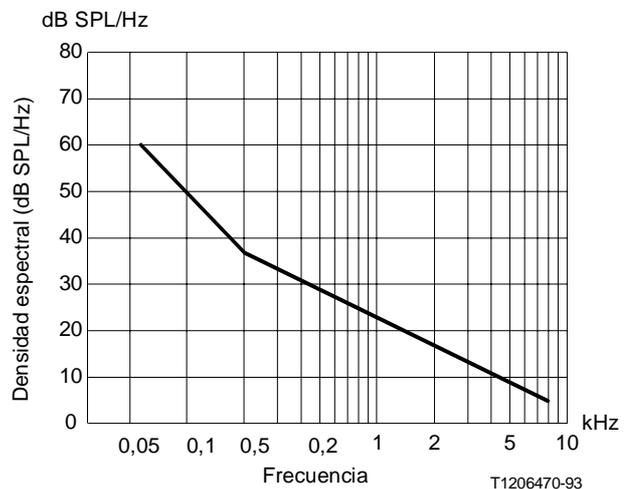
Espectro	Nivel de presión sonora no ponderado (dB SPL)
En movimiento 30 km/h	80
80 km/h	85
110 km/h	90
Estacionado	75

NOTAS a los Cuadros A.2 y A.3:

- Estos valores son válidos para los vehículos típicos. Estos niveles pueden tener que disminuirse en el caso de los vehículos de lujo y aumentarse en el caso de los vehículos más ruidosos.
- Debido a las dificultades prácticas que plantea la generación de niveles tan altos de presión sonora a baja frecuencia y teniendo en cuenta que la palabra normal no contiene ninguna energía aparente por debajo de unos 63 Hz, gama de frecuencias en la cual el oído es también relativamente insensible, conviene probablemente limitar el espectro de ruido recomendado a las frecuencias superiores a 63 Hz. No obstante, debe tenerse presente que las vibraciones en frecuencias bajas y medias tienen importantes efectos fisiológicos y psicológicos, los cuales deben estudiarse en su propio plano.
- La señal eléctrica de entrada, por ejemplo de ruido blanco, tendrá una banda limitada a las bandas de un tercio de octava centradas en las frecuencias preferidas según la ISO (ISO 266) entre 63 y 8000 Hz, y los bordes de la banda se conformarán a los filtros descritos en la Publicación 1260 de la CEI.
- El ruido de sala acústico es difícil de controlar en baja frecuencia, sobre todo en la región no especificada por debajo de 63 Hz, debido a las dimensiones de las cabinas de pruebas típicas, a la mediocre atenuación de estas cabinas y a la influencia de ruidos externos (por ejemplo, aparatos de climatización). Por ello es conveniente elegir una cabina de prueba que reduzca al mínimo estos niveles de presión sonora indeseados en baja frecuencia.



a) Densidad espectral para vehículos en movimiento



b) Densidad espectral para vehículos estacionados

FIGURA A.2/P.800

Densidad espectral del ruido en vehículos

A.1.1.3 Posición de la medición del ruido

Se recomienda efectuar las mediciones del nivel de presión sonora (SPL, *sound pressure level*) en las cabinas de prueba (véase A.1.1.1) en las siguientes condiciones:

- los muebles deben estar en sus correspondientes posiciones;
- no debe haber ningún sujeto o personal de prueba presente;
- el SPL se medirá a una distancia vertical de 740 mm por encima del centro del asiento de la silla donde se encuentra el participante con un medidor conforme a la Recomendación P.54 que utilice ponderación «A»;
- el espectro del ruido ambiental se medirá en bandas de un tercio de octava centradas en las frecuencias preferidas definidas en ISO 266 y debe permanecer dentro de las tolerancias especificadas; por ejemplo ± 3 dB para el ruido de Hoth (véase A.1.1.2.2.1);
- en las salas donde va a realizarse más de un experimento, la diferencia en dBA para todas las posiciones de los sujetos de prueba no debe variar más de ± 2 dB.

NOTA – Se sugiere una distancia mínima entre cada altavoz y la posición de medición de 1,5 m.

A.1.2 Establecimiento de la conexión

Al establecer la conexión en el laboratorio deben tenerse en cuenta los siguientes elementos:

- los aparatos telefónicos;
- el instante de establecimiento de la llamada inicial;
- la representación en el laboratorio de las conexiones telefónicas.

Se recomienda medir al menos dos veces la característica de sensibilidad/frecuencia de la conexión; al principio y al final del experimento. Toda variación significativa en las dos mediciones, al compararlas, debe ser evaluada por el experimentador, pues puede que arroje dudas sobre la validez del experimento.

En 2.5.8.2 del *Manual sobre Telefonometría* figura información detallada al respecto.

A.1.3 Control

El control se puede realizar de diversas formas, pero en general se emplean tres técnicas:

- *Sistema de intercomunicación* - Es esencial para que el participante y el experimentador puedan comunicarse entre sí.
- *Control visual* - Tiene dos objetivos; el primero es la seguridad y el segundo la observación de las peculiaridades del participante; por ejemplo, la forma en que sostiene el microteléfono.
- *Cintas magnetofónicas y sistema de grabación* - Facilita la obtención de diversas informaciones útiles como la duración de la llamada, la tensión de la señal vocal, y la actividad vocal.

En 2.5.8.2 del *Manual sobre Telefonometría* figura más información al respecto.

A.2 Diseño del experimento

Existen una serie de métodos adecuados para diseñar los experimentos; entre los que cabe citar los cuadrados latinos, los cuadrados de Youden, los bloques incompletos equilibrados y la aleatorización con réplica, entre otros. El experimentador debe decidir el método a utilizar teniendo en cuenta el número de las condiciones de prueba, la precisión de los resultados y la capacidad para evaluar correctamente las conclusiones.

Entre los diseños más adecuados figuran los de tipo cuadrado grecolatino $n \times n$, una detallada descripción de estos diseños figura en 2.5.8.2 del *Manual sobre Telefonometría*.

A.3 Conversación

Debe hacerse todo lo posible para que las conversaciones tengan sentido y utilidad y para que los participantes puedan sacar el máximo provecho de las posibilidades de transmisión del circuito de prueba.

La regla general es que cada conversación tenga un comienzo y un final naturales. A menos que sea absolutamente necesario, la conversación no debe finalizar nunca en medio de la prueba (con excepción de las pruebas de conversación simplificada, descritas en 2.5.8.2 del *Manual sobre Telefonometría*).

En 2.5.8.2 d) de dicho *Manual* figura un ejemplo de conversación.

A.4 Procedimiento de la prueba

A.4.1 Elección de los participantes

Los sujetos que toman parte en las pruebas de conversación se eligen al azar entre los usuarios telefónicos, estableciéndose dos condiciones:

- a) no deben haber participado directamente en trabajos relativos a la evaluación de la calidad de transmisión de los circuitos telefónicos, o tareas similares como la codificación de señales vocales; y
- b) no deben haber participado en ninguna prueba subjetiva en los últimos seis meses y en ninguna prueba de conversación en los últimos doce meses.

Si la población disponible es demasiado restringida, debe tenerse en cuenta esta circunstancia al extraer conclusiones de los resultados.

No se toman medidas para equilibrar el número de participantes masculinos y femeninos, a menos que así lo requiera la forma del experimento. Los participantes se agrupan de forma arbitraria por parejas en el programa establecido antes de la prueba y permanecen emparejados mientras dure la misma.

A.4.2 Escalas de opinión

El UIT-T recomienda las siguientes escalas de opinión.

A.4.2.1 Escala de opinión sobre la conversación

Pueden utilizarse varias escalas de cinco notas para juzgar las categorías con diversos fines. Desde el punto de vista de los participantes, es muy importante la presentación y los textos de las escalas de opinión y deben seguir las normas elaboradas tras muchos años de experiencia. La siguiente escala de opinión es la utilizada más frecuentemente en aplicaciones del UIT-T y los textos deben adecuarse al idioma utilizado, lo que quizá origine pequeños cambios con respecto al texto original en inglés.

Al final de cada conversación se obtiene de cada participante esta escala de categorías.

Opinión sobre la conexión que acaba de utilizar

Excelente

Buena

Regular

Mediocre

Mala

El experimentador atribuye los siguientes valores a las opiniones:

Excelente = 5 Buena = 4 Regular = 3 Mediocre = 2 Mala = 1

y todo el procesamiento estadístico posterior se lleva a cabo en función de estas cifras. La media aritmética de cualquier conjunto de estas notas se denomina nota media de opinión sobre la conversación y se representa por el símbolo MOS_C (o cuando no se dispone de sufijos, por el símbolo MOS_c).

NOTA – Anteriormente se utilizaban las siguientes equivalencias: Excelente = 4, Buena = 3, Regular = 2, Mediocre = 1, Mala = 0. Cuando se utilicen los resultados de anteriores experimentos, debe tenerse en cuenta que hay que incrementar las notas medias en una unidad para su comparación adecuada con las notas obtenidas actualmente. Esta es la única diferencia en el procesamiento numérico que puede aplicarse.

A.4.2.2 Escala de dificultad

Se trata de una respuesta binaria obtenida de cada participante al final de cada conversación.

¿Ha experimentado usted o su interlocutor alguna dificultad para hablar o escuchar a través de la conexión?

Sí

No

El experimentador atribuye los siguientes valores a las respuestas:

Sí = 1 No = 0

La cantidad evaluada (porcentaje de respuestas «Sí») se denomina de dificultad porcentual o porcentaje «Dificultad» y se expresa mediante el símbolo %D. El tanto por uno correspondiente se expresa por el símbolo d; es decir, $\%D = 100d$.

NOTA – A menudo es preciso conocer la naturaleza de la dificultad, para lo cual el experimentador solicita al participante que describa con sus propias palabras la percepción de la dificultad.

Desde el punto de vista de los participantes en los experimentos, son muy importantes la disposición y denominación de la escala de opinión y deben seguir las normas elaboradas tras muchos años de experiencia (véase A.4.3).

A.4.2.3 Otras escalas de opinión

Pueden ser adecuadas otras escalas de opinión que son variantes de los métodos denominados «estimación de la magnitud» y «adaptación mutua de modalidad» [8]. Las respuestas en estas escalas pueden ser una de las siguientes:

- a) una de una serie numérica de categorías denominadas 1, 2, 3, 4 ó 5 (señalada de esa forma al participante), pero cuyas descripciones se relacionan sólo con la primera y la última, para identificar la dimensión subjetiva;
- b) una señal numérica en una escala que va desde uno a una cifra mucho mayor que cinco, por ejemplo 10 ó 100; o
- c) una longitud proporcional a alguna propiedad (por ejemplo, la calidad), señalada de forma manual en una línea recta determinada.

En 2.6.2 del *Manual sobre Telefonometría* figura una descripción de los métodos experimentales.

A.4.3 Instrucciones a los participantes

Los participantes reciben instrucciones a su llegada en la primera visita. Es normal que el participante haya recibido una carta antes de esa visita en la que figura información no técnica sobre el experimento y se indica lo que se espera de él. En el Cuadro 3/2.5 de 2.5.8.2 del *Manual sobre Telefonometría* aparece un ejemplo de carta.

A su llegada se les pregunta si han leído y comprendido la carta. Se aclara cualquier duda y se les da la oportunidad de hacer preguntas. Se realiza una demostración del funcionamiento de las cabinas insonorizadas y de sus instalaciones y se informa a los participantes del número de llamadas que efectuarán en esta visita. En las siguientes visitas se indica a los participantes únicamente que el procedimiento será el mismo utilizado la vez anterior, variando posiblemente el número de llamadas. En el Cuadro 4/2.5, 2.5.8.2, del *Manual sobre Telefonometría* aparece un ejemplo de algunos detalles de operación necesarios en un experimento.

A.4.4 Recopilación de datos

Los niveles de la señal vocal y los datos correspondientes, como las duraciones y los factores de actividad pueden obtenerse a partir de las grabaciones pero normalmente se miden en línea, mediante medidores controlados por computador, y se almacenan directamente en los ficheros de éste para su posterior análisis.

El experimentador recibe dos respuestas subjetivas de cada participante y por cada conversación. Los datos esenciales son las notas de opinión sobre la conversación y la decisión sobre la dificultad. Estas respuestas pueden obtenerse utilizando cualquier medio adecuado, incluidos lápiz y papel, pulsadores electrónicos, teclados o terminales de pantalla táctil. Un ejemplo de formato de respuesta figura en el Cuadro 5/2.5, 2.5.8.2 del *Manual sobre Telefonometría*.

A.4.5 Tratamiento de los resultados

Se trata de un tema muy amplio del que sólo se ofrece un breve resumen.

Cada conversación da lugar a dos opiniones sobre ella en la escala: Excelente - Buena - Regular - Mediocre - Mala (que corresponden respectivamente a las notas 5, 4, 3, 2, 1), dos votos en la escala de dificultad (con las notas 1 = Sí; 0 = No), dos niveles de señal vocal activa medidos y un valor de duración. En algunos casos particulares, también puede obtenerse información relativa a otras variables; por ejemplo, pueden efectuarse grabaciones de vídeo para observar la forma en que los participantes sostienen los microteléfonos o pueden obtenerse otros datos a partir de los formularios de opinión o de las grabaciones de audio.

Para cada condición de prueba debe calcularse el promedio de las notas de opinión. Deben evaluarse los límites de confianza y llevarse a cabo pruebas de significado mediante técnicas convencionales de análisis de la varianza.

Las hipótesis normales relativas al análisis de la varianza se cumplen casi por completo en el caso de escala de opinión, en el del nivel de la señal vocal activa y en la mayoría del resto de variables de interés; pero no se satisfacen, especialmente la hipótesis de que la varianza residual es constante, en el caso de una variable binaria como la de la escala de dificultad. A pesar de ello, la experiencia confirma la observación efectuada en otros campos [9], en el sentido de que la técnica del análisis de la varianza es lo suficientemente consistente como para ofrecer resultados razonables, aun cuando se produzcan desviaciones muy importantes con respecto a las condiciones ideales desde el punto de vista estadístico. Los resultados de la primera etapa del análisis de la varianza de las escalas de dificultad se deben considerar con cierta reserva; pero una vez confirmado que no hay anomalías sin explicación en los resultados ni problemas inexplicables con el análisis correspondiente de los resultados MOS_C , puede emprenderse abiertamente la segunda etapa (análisis detallado de los promedios de las combinaciones fin-condición) con ayuda de una transformación matemática.

En 2.5.9 del *Manual sobre Telefonometría* figura una descripción detallada del análisis.

Si es preciso, puede completarse el examen de los datos mediante gráficos que ilustren la nota media de opinión en función del parámetro sometido a prueba; por ejemplo, MOS_C en función de la atenuación del circuito. El eje de ordenadas del gráfico debe ser siempre MOS_C .

Anexo B

Pruebas de escucha – determinación de índices por categorías absolutas (ACR)

B.1 Grabaciones de señales en la fuente

Para eliminar variaciones no deseadas en la fuente de las señales vocales, en primer lugar se deben grabar o almacenar muestras de señales vocales con las propiedades normalizadas deseadas, de la forma siguiente.

B.1.1 Entorno de grabación

El hablante debe sentarse en una sala silenciosa, de un volumen entre 30 y 120 m³ y con un tiempo de reverberación inferior a 500 ms (preferentemente, entre 200 y 300 ms). El nivel de ruido de sala debe ser inferior a 30 dBA, sin crestas dominantes en el espectro.

Las características del ruido de sala deben señalarse de la forma más completa posible; por ejemplo, en dBA, en forma de espectro a largo plazo y como distribución amplitud-tiempo. Conviene grabar una muestra de 30 segundos del ruido de sala para realizar investigaciones más detalladas, si es preciso.

B.1.2 Sistema emisor

Cualquiera que sea el sistema emisor elegido; por ejemplo, sistema de telefonía local o un sistema intermedio de referencia (IRS, *intermediate reference system*) especificado en la Recomendación P.48, el sistema debe calibrarse de acuerdo con la Recomendación pertinente (por ejemplo, la Recomendación P.64) y la característica sensibilidad-frecuencia en emisión debe indicarse en su totalidad. El Anexo D/P.830 describe el "IRS modificado" que se ha considerado apropiado para la evaluación de las conexiones enteramente digitales que utilizan códecs de voz.

Se recomienda medir al menos dos veces la característica de sensibilidad en emisión de la conexión; al principio y al final del experimento. Toda variación significativa de las dos mediciones, al compararlas entre sí, debe ser evaluada por el experimentador puesto que podría arrojar dudas sobre la validez del experimento.

B.1.3 Sistema de grabación

El sistema de grabación debe ser de alta calidad (de estudio) y puede adoptar una de las siguientes formas:

- a) Magnetófono convencional de dos pistas. Debe indicarse el tipo de ecualización, pero se recomienda utilizar el de la CEI. En todo momento deben utilizarse cintas de alta calidad (de bajo nivel de ruido y baja transferencia magnética).
- b) Procesador de audio digital de dos canales equipado con un magnetoscopio de alta calidad o un equipo de cinta magnética de audio digital (DAT, *digital audio tape*).
- c) Sistema de almacenamiento digital controlado por computador.

El tercer sistema es el más adecuado y resulta muy flexible, pero por razones de orden práctico a menudo se elige uno de los otros dos. En estos sistemas una de las dos pistas debe emplearse para grabar la conversación y la otra para insertar señales de control con nivel y frecuencia determinados, a fin de evitar problemas de diafonía.

B.1.4 Material de conversación

El material de conversación debe consistir en frases sencillas, breves y con significado, elegidas al azar y fáciles de entender (por ejemplo extraídas de publicaciones no técnicas o de periódicos). Con estas frases deben confeccionarse listas de forma aleatoria, de tal manera que no haya ninguna

conexión evidente entre el significado de una frase y la siguiente. Deben evitarse las frases muy cortas y muy largas pues el objetivo es que cada frase pueda incluirse en un intervalo de tiempo de 2 a 3 segundos. En el Cuadro B.1 figuran ejemplos de frases.

El experimentador debe decidir el número de frases que es necesario en cada grupo para constituir una muestra de señales vocales. Se recomienda un mínimo de dos y un máximo de cinco. También tiene su importancia el intervalo de tiempo entre las frases, durante el cual puede escucharse el ruido del circuito y pueden ajustarse los procesos de adaptación a nuevos estados. Conviene grabar los grupos más largos que puedan necesitarse, puesto que siempre es posible obtener grupos más breves copiando o volviendo a reproducir parte de las frases más largas.

Los grupos se combinan en listas de cinco o diez grupos cada una, de modo que cada lista completa puede utilizarse como una serie de muestras sometidas al mismo tratamiento, pero variando el nivel de escucha o algún otro parámetro cuando se la reproduzca.

CUADRO B.1/P.800

Ejemplos de material vocal (en inglés)

You will have to be very quiet.
There was nothing to be seen.
They worshipped wooden idols.
I want a minute with the inspector.
Did he need any money?

B.1.5 Procedimiento de grabación

Se recomienda utilizar el siguiente plan de grabación, ampliamente extendido.

Las señales vocales se graban mediante un micrófono lineal y un amplificador de bajo nivel de ruido con respuesta en frecuencia uniforme, como indica la Publicación 581-5 de la CEI. El micrófono se coloca a una distancia entre 140 mm y 200 mm de los labios del participante. En ciertas aplicaciones puede ser necesario utilizar un filtro cuando se detectan sonidos producidos por la respiración del hablante.

Las mismas señales vocales pueden ser grabadas simultáneamente de la salida de emisión de un sistema intermedio de referencia (IRS, véase la Recomendación P.48), sosteniendo el microteléfono de forma normal. Si la investigación en curso así lo requiere, puede utilizarse otro aparato telefónico en lugar del IRS.

Se emplean de forma simultánea dos sistemas de grabación distintos: uno para grabar las señales vocales de banda ancha en un canal y otro para grabar las señales vocales telefónicas en el canal correspondiente. El segundo canal de cada sistema de grabación se utiliza para grabar las señales de control, como se indica en 2.5 del *Manual sobre Telefonometría*.

Este sistema de grabación doble asegura que la misma señal vocal se graba de dos maneras (señal vocal telefónica y señal vocal de banda ancha). Normalmente sólo se necesita una de ambas en los experimentos, pero en ocasiones es preciso utilizar las dos y, en cualquier caso, es ventajoso poder realizar mediciones comparativas de ambas versiones.

Durante la grabación debe observarse el nivel vocal activo, definido en la Recomendación P.56. Durante el proceso de grabación se mantiene el nivel vocal activo de ambos sistemas de grabación entre 20 y 30 dB por debajo del punto de sobrecarga del sistema de grabación para cada frase medida por separado. Cualquier grupo de frases en las que no se cumpla esta condición debe volverse a grabar.

Se recomienda que la relación entre el nivel vocal activo y el nivel de ruido ponderado sofométricamente (para la definición véase 8.2.3/P.830), SNR(p), en el medio de grabación tenga un valor > 40 dB, tomándose como objetivo el valor de 50 dB.

Todas las muestras de señales vocales utilizadas en un experimento pueden ser distintas; esta condición es fundamental para las pruebas de esfuerzo en la escucha y deseable para otros tipos de prueba.

B.1.6 Participantes en las pruebas

Deben participar cuantas personas se hayan previsto en la planificación del experimento (véase B.3).

Los hablantes deben pronunciar las frases de manera fluida pero sin énfasis y no deben presentar deficiencias de dicción como tartamudeo; deben adoptar un volumen de voz que les resulte confortable y que puedan mantener de forma constante.

B.1.7 Niveles de la señal vocal

Una vez completadas, las grabaciones se reproducen y se mide el nivel vocal activo de cada frase con un medidor conforme a la Recomendación P.56. A continuación, se vuelven a grabar las listas (anuncios, frases y tonos de control) en un segundo sistema introduciendo los ajustes de ganancia necesarios para que cada grupo de frases tenga el nivel vocal activo normalizado indicado más adelante, manteniendo las relaciones de tiempo adecuadas entre las frases y las señales de tono en el otro canal.

Para las señales vocales en banda estrecha, el nivel normalizado se obtiene midiendo y ajustando directamente la señal de banda estrecha grabada; el valor objetivo recomendado es -26 dB ($+ 0,5$ dB) con respecto al nivel de saturación de cresta del sistema de grabación. El tono de calibración tiene un nivel cuyo valor eficaz es igual al nivel activo medio de la señal vocal grabada por segunda vez.

Para la señal vocal en banda ancha, debe tenerse en cuenta la utilización que va a hacerse de las grabaciones. A veces conviene adoptar los mismos niveles que para la señal vocal telefónica, pero si la grabación está destinada a la reproducción por un altavoz o por una boca artificial, cada uno de los niveles vocales previstos debe ser tal que se mantenga la igualdad a la salida de la cadena de reproducción electroacústica.

B.1.8 Señal de calibración

Al principio de cada grabación, se inserta en la etapa de grabación posterior un tono de 20 segundos o más (a efectos de calibración) con un nivel cuya relación con el nivel vocal activo medio se conoce (lo más conveniente es que ambos niveles sean iguales). Este tono de calibración es normalmente de 1000 Hz pero puede ser de otra frecuencia si las grabaciones van a reproducirse a través de sistemas (tales como ciertos tipos de codificadores de subbanda) que responden a 1000 Hz de una forma especial.

Este tono puede utilizarse posteriormente para ajustar los niveles vocales de entrada medios (véase B.4.3).

B.2 Selección de las condiciones de circuito

B.2.1 Niveles de entrada y niveles de escucha

Al seleccionar las condiciones del circuito debe prestarse una atención especial a:

- la gama de niveles de entrada;
- la gama de niveles de escucha:
 - no hay un nivel de escucha óptimo universal;

- en la práctica aparecerán varios niveles de escucha;
- deben hacerse consideraciones sobre la comparación;
- pueden producirse interacciones.

En 2.5.8.1 del *Manual sobre Telefonometría* figura una explicación detallada de estos aspectos.

B.2.2 Participantes en las pruebas

Como los procesos complejos a menudo afectan de forma distinta a las voces femenina y masculina, el plan del experimento debe contemplar dos tipos de voces como factor de equilibrio; las notas para las señales vocales femenina y masculina deben evaluarse por separado y sólo se promediarán si producen interacciones y efectos importantes que no sean estadísticamente distintos.

Además, para disminuir el riesgo de que los resultados dependan en gran medida de las peculiaridades de las voces elegidas, es fundamental utilizar más de una voz masculina y más de una voz femenina, de acuerdo con un plan equilibrado.

B.2.3 Condiciones de referencia

Cada experimento debe incluir condiciones de referencia para poder comparar adecuadamente los experimentos efectuados en diferentes laboratorios o en el mismo laboratorio pero en momentos distintos. Tales condiciones de referencia dependerán del parámetro que va a evaluarse. Para un sistema digital las condiciones de referencia pueden incluir el aparato de referencia para ruido modulado (MNRU), conforme a la Recomendación P.810; en otros casos, son apropiadas otras degradaciones controladas (por ejemplo, relación señal/ruido, véase 8.2.3/P.830).

B.2.4 Otras condiciones

Además de los requisitos indicados en B.2.1 a B.2.3 inclusive deberán cumplirse otras condiciones, según el objetivo de la prueba. Por ejemplo, el ruido de sala puede ser una variable, así como la tasa de errores en los bits en un sistema digital o el desvanecimiento de Rayleigh en un sistema de radiocomunicaciones.

B.3 Concepción del experimento

Para el diseño del experimento se aplican los mismos principios indicados en A.2.

Además, debe tenerse en cuenta:

- a) los requisitos indicados en B.2;
- b) el efecto del orden de presentación.

Para un grupo determinado de participantes, la prueba viene limitada en tiempo por la máxima duración posible de la sesión sin que aparezca el cansancio. Si el experimento es demasiado amplio para realizarse en una sesión, conviene subdividirlo en dos o más sesiones. Idealmente ninguna sesión debe durar más de 20 minutos y en ningún caso debe rebasar los 45 minutos.

B.4 Procedimiento de la prueba de escucha

B.4.1 Entorno de escucha

La sala de escucha debe satisfacer las mismas condiciones que la sala de grabación (véase B.1.1) con la excepción de que el ruido ambiental (véase A.1.1.2.2) debe ajustarse al nivel apropiado. Véanse en A.1.1.2.2.1 y A.1.1.2.2.2 ejemplos del espectro de ruido.

Se recomienda medir al menos dos veces el espectro y el nivel de ruido; al principio y al final del experimento. Toda variación significativa en ambas mediciones, al compararlas entre sí, debe ser evaluada por el experimentador pues puede arrojar dudas sobre la validez del experimento.

B.4.2 Sistema de escucha

Sea cual fuere el sistema de escucha elegido (por ejemplo, sistema telefónico local, sistema intermedio de referencia como especifica la Recomendación P.48, sistema de altavoces), debe calibrarse de acuerdo con la Recomendación pertinente (por ejemplo, la Recomendación P.64) y debe indicarse detalladamente la característica de sensibilidad de recepción en función de la frecuencia. El Anexo D/P.830 describe el "IRS modificado" que se ha considerado apropiado para la evaluación de las conexiones enteramente digitales utilizando códecs vocales.

Se recomienda medir al menos dos veces esta característica de la conexión; al principio y al final del experimento. Toda variación significativa de las dos mediciones, al compararlas entre sí, debe ser evaluada por el experimentador pues puede arrojar dudas sobre la validez del experimento.

B.4.3 Nivel de escucha

La ganancia del sistema debe ajustarse de tal forma que el tono de calibración (véase B.1.8) obtenido de las cintas procesadas produzca el nivel de escucha requerido.

Las variaciones en el nivel de escucha requeridas por el diseño del experimento pueden modificarse

- a) utilizando atenuadores/amplificadores en el sistema de escucha; o
- b) se las puede incluir en la etapa de procesamiento o grabación.

No se recomienda el segundo método porque es difícil mantener una relación señal/ruido lo bastante elevada a niveles bajos y porque la flexibilidad y la variedad de la aleatorización se reducen notablemente.

El nivel de escucha debe registrarse siempre; en 2.5 del *Manual sobre Telefonometría* figura información al respecto.

B.4.4 Oyentes

Los participantes en las pruebas de escucha se escogen al azar entre la población que utiliza normalmente el servicio telefónico, fijando las siguientes condiciones:

- a) que no hayan participado directamente en trabajos relacionados con la evaluación de la calidad de transmisión de los circuitos telefónicos o tareas afines, como codificación de la palabra;
- b) que no hayan participado en pruebas subjetivas, de cualquier naturaleza, al menos durante los seis meses precedentes, ni en pruebas de opinión sobre la escucha al menos durante un año; y
- c) que no hayan oído antes la misma lista de frases.

Si la población disponible es excesivamente restringida, debe tenerse en cuenta este hecho al sacar conclusiones de los resultados.

En algunos casos puede ser necesario seleccionar a los participantes en las pruebas; en el Anexo B/P.78 aparece un método al respecto.

B.4.5 Escalas de opinión recomendadas por el UIT-T

Pueden usarse diversas escalas de evaluación de 5 notas para distintos propósitos. La disposición y el enunciado de las escalas de opinión consideradas por los participantes en los experimentos son muy importantes y deben seguir las normas a las que se ha llegado tras muchos años de experiencia. Las

siguientes escalas de opinión son las utilizadas con más frecuencia por el UIT-T, y deben utilizarse las denominaciones equivalentes en cada idioma, de forma que las variaciones con respecto al texto original inglés sean mínimas:

a) **Escala de calidad de escucha**

<i>Calidad de la señal vocal</i>	<i>Nota</i>
Excelente	5
Buena	4
Regular	3
Mediocre	2
Mala	1

La magnitud evaluada a partir de las notas (nota media de opinión sobre la calidad de escucha) se representa por el símbolo MOS.

b) **Escala de esfuerzo de escucha**

La denominación de las escalas de opinión de esfuerzo en la escucha es muy importante. Sin ellas el resto de descripciones pueden provocar graves equívocos.

<i>Esfuerzo necesario para comprender el significado de las frases</i>	<i>Nota</i>
Audición perfecta; ningún esfuerzo	5
Cierta atención es necesaria; ningún esfuerzo apreciable	4
Esfuerzo moderado	3
Esfuerzo considerable	2
Significado incomprensible, aun con el mayor esfuerzo	1

La magnitud evaluada a partir de las notas (nota media de opinión de esfuerzo en la escucha) se representa por el símbolo MOS_{LE}, pero cuando no se dispone de notación con sufijos, se utiliza el símbolo MOS_{le}.

c) **Escala de sonoridad preferida**

<i>Sonoridad preferida</i>	<i>Nota</i>
Mucho mayor que la preferida	5
Mayor que la preferida	4
Preferida	3
Menor que la preferida	2
Mucho menor que la preferida	1

La magnitud evaluada a partir de las notas (nota media de opinión de sonoridad preferida) se representa por el símbolo MOS_{LP}, pero cuando no se dispone de notación con sufijos se utiliza el símbolo MOS_{lp}.

NOTA – En 2.6 del *Manual sobre Telefonometría* y en el Informe 751 del CCIR, Volumen VIII.3, 1986, aparecen ejemplos de escalas subjetivas alternativas que pueden utilizarse únicamente si las tres escalas de opinión anteriores no satisfacen las necesidades del experimentador.

B.4.6 Instrucciones a los participantes

En el Cuadro B.2 aparece un ejemplo de instrucciones típicas a los participantes. Estas instrucciones deben darse (también de forma oral, si es preciso) antes de iniciar el experimento. Cuando el participante ha comprendido las instrucciones, debe escuchar la lista preliminar y dar su opinión. No debe indicarse a los participantes si las muestras preliminares incluyen los mejores o peores ejemplos de la gama, ni agotar la gama de condiciones que esperan escuchar. Tras la lista preliminar, debe darse el tiempo suficiente para responder a las posibles preguntas de los participantes. Deben contestarse las preguntas relativas al procedimiento o al significado de las instrucciones, pero cualquier pregunta de tipo técnico se responderá de la forma siguiente: «No podemos decirle nada al respecto hasta que haya finalizado el experimento».

CUADRO B.2/P.800

Ejemplo de instrucciones a los participantes

EXPERIMENTO DE ESCUCHA N° ...

En este experimento usted escuchará grupos breves de frases a través del microteléfono y debe dar su opinión sobre la calidad de la señal vocal que escuche.

Sobre la mesa, frente a usted, hay una caja con cinco pulsadores iluminados. Al encenderse todas las lámparas, oírás ... frases. Escúchelas y cuando se apaguen las lámparas pulse el botón correspondiente a la siguiente escala para indicar su opinión.

ESFUERZO NECESARIO PARA COMPRENDER EL SIGNIFICADO DE LAS FRASES

- 5 Audición perfecta, ningún esfuerzo.
- 4 Cierta atención es necesaria; ningún esfuerzo apreciable.
- 3 Esfuerzo moderado.
- 2 Esfuerzo considerable.
- 1 Significado incomprensible, aun con el mayor esfuerzo.

El botón pulsado permanecerá iluminado durante un breve tiempo, posteriormente se apagará y habrá una breve pausa antes de que todas las lámparas vuelvan a encenderse para evaluar el siguiente grupo de ... frases.

Después de ... grupos (para los que se solicita una opinión) habrá una pausa más larga. En esta prueba habrá un total de ... grupos, y un número similar en la prueba o pruebas siguientes.

Le agradecemos su colaboración en este experimento.

B.4.7 Análisis estadístico y presentación de los resultados

En cada nivel de escucha debe calcularse, para cada condición la media numérica (de todos los participantes) y estos valores medios deben someterse a una inspección inicial (para determinar los efectos derivados de las voces masculina y femenina).

No se recomienda calcular para cada condición las desviaciones típicas. Deben evaluarse los límites de confianza y realizarse pruebas de significado mediante técnicas de análisis de varianza convencionales.

NOTA – Anteriormente se utilizaban con frecuencia equivalencias, por ejemplo: Excelente = 4, Buena = 3, Regular = 2, Mediocre = 1, Mala = 0. Siempre que se utilicen los resultados de experimentos anteriores debe tenerse en cuenta que las notas medias deben incrementarse en una unidad para que sean comparables con los resultados de los últimos experimentos; se trata de la única diferencia en el proceso numérico que puede aplicarse.

El método de análisis de las escalas de opinión indicado en B.4.5 se ajusta a los principios señalados en A.4.5.

Como ayuda adicional para examinar los resultados, deben elaborarse gráficos que muestren la nota media de opinión en función del parámetro sometido a prueba; por ejemplo, MOS en función de la atenuación de circuito. En el gráfico el eje de ordenadas debe ser siempre la nota MOS.

Debe calcularse con cuidado el promedio de las notas de los participantes masculinos y femeninos, pero ello no implica la validez de este paso para el estudio detallado y la interpretación de los resultados, a menos que las pruebas de significado lo justifiquen.

Anexo C

Pruebas de detectabilidad de respuesta cuantificada

El método más adecuado para obtener información sobre la detectabilidad o alguna propiedad análoga de un sonido (como el eco) en función de cierta magnitud objetiva (como el nivel de escucha), es un método de respuesta cuantificada similar en principio al descrito en 2.2 del *Manual sobre Telefonometría*. La principal diferencia es que la respuesta del participante no es una decisión en forma «Referencia» o «Prueba» (designación del circuito con volumen sonoro más elevado entre dos circuitos), sino una opinión en una escala como la siguiente [10]:

Escala de opiniones sobre la detectabilidad

- A Objetable
- B Detectable
- C No detectable

entiéndase que B significa «detectable pero no objetable».

Se pueden utilizar escalas de este tipo, generalmente con tres notas, en diversas pruebas de respuesta cuantificada; por ejemplo, la escala indicada anteriormente puede emplearse cuando el estímulo es un eco, reverberación, efecto local, mutilación de las señales vocales por la conmutación o tonos interferentes, en tanto que la diafonía y posiblemente el eco en ciertas circunstancias pueden juzgarse según la escala inteligible-detectable-no detectable.

A veces es admisible considerar estas opiniones como notas con los valores 2, 1, 0 respectivamente, y tratarlas del mismo modo que se tratarían las notas de opinión sobre la escucha o la conversación. Sin embargo, esto es a menudo insatisfactorio porque las decisiones sobre una escala como la relativa a detectabilidad (véase lo indicado anteriormente) no son realmente equivalentes de respuestas basadas en una escala continua, como pueden considerarse legítimamente que son las opiniones basadas en escalas como la de «sonoridad preferida» (véase B.4.5), sino que comprenden efectivamente dos dicotomías distintas (por ejemplo, detectable/no detectable y objetable/no objetable), las cuales aunque no son dependientes entre sí, pueden desencadenar procesos psicológicos distintos. En otros términos, la objetabilidad o la inteligibilidad difieren en especie, y no sólo en grado, de la detectabilidad. Por esta razón, es más ventajoso como método de análisis expresar la probabilidad de respuesta de acuerdo con cada dicotomía por separado, en función de alguna variable objetiva, mediante ecuaciones «probit» o «logit», y utilizando a continuación los cuantiles u otros parámetros como base de comparación entre condiciones de circuito, de forma análoga a la utilizada para aplicar índices de nitidez.

La realización de experimentos de este tipo se asemeja a la de las pruebas de esfuerzo en la escucha (véase el Anexo B), pero hay algunas diferencias. En particular, conviene que la primera presentación de la señal en cada serie de frases se haga con un nivel de escucha elevado, de forma que el oyente no tenga ninguna duda en cuanto al tipo de señal que debe evaluar. Cuando intervengan factores como el efecto local o el eco, se pedirá al participante que además de escuchar, hable.

Como se describe en la Recomendación P.78, con los participantes en estos experimentos se suelen realizar mediciones audiométricas sencillas, de manera que los resultados puedan expresarse con relación a sus umbrales de audición.

En [11], figuran ejemplos de aplicación de estas técnicas.

A veces se investiga el ruido, el desvanecimiento y otras perturbaciones mediante respuestas basadas en una escala que contiene muchas más notas; por ejemplo [12]:

- A *Inaudible* – Ruido completamente indetectable.
- B *Apenas audible* – El ruido puede apenas detectarse escuchando con mucha atención.
- C *Ligero* – Ruido detectable, pero no molesto.
- D *Moderado* – Ruido ligeramente molesto.
- E *Bastante elevado* – El ruido causa una molestia apreciable.
- F *Elevado* – El ruido es muy molesto, pero la llamada podría proseguir.
- G *Intolerable* – El ruido es tan elevado que habría que abandonar la llamada o solicitar al operador que cambie la línea.

Estas escalas se aproximan más a las del tipo de cuantificación continua, como la escala de sonoridad preferida, y pueden tratarse de forma similar.

Anexo D

Método de determinación de índices por categorías de degradación (DCR)

D.1 Introducción

El método de determinación de índices por categorías absolutas (ACR, *absolute category rating*) descrito en el Anexo B suele tener poca sensibilidad para diferenciar los circuitos de buena calidad. Con una versión modificada del procedimiento ACR denominada procedimiento de determinación de índices por categorías de degradación (DCR, *degradation category rating*) [13] se logra una sensibilidad más elevada. Este procedimiento se ha adaptado de las Recomendaciones del CCIR [14] para efectuar evaluaciones de circuitos de buena calidad. El procedimiento DCR, que utiliza en particular una escala de molestias y una referencia de calidad elevada antes de evaluar cada configuración, parece adecuado para evaluar señales vocales de buena calidad.

D.2 Procedimiento de determinación de índices por categorías de degradación (DCR)

D.2.1 Muestras de señales vocales

Cada configuración se evalúa mediante las notas de opinión sobre muestras vocales emitidas por al menos cuatro participantes. Cada muestra debe estar formada por dos frases separadas aproximadamente por 0,5 s de silencio. Esas dos muestras (S1, S2), que comprenden cuatro frases distintas, deben elegirse entre un grupo más amplio compuesto por frases fonéticamente equilibradas, de forma que la nota media obtenida al evaluar los circuitos de referencia (por ejemplo, MNRU para procesos digitales) por medio de estas frases sea aproximadamente igual a la obtenida por medio del conjunto más amplio. Por consiguiente, el conjunto está formado por ocho muestras definidas de la forma siguiente:

- el participante T1 lee las muestras S1, S2;
- el participante T2 lee las muestras S1, S2;
- el participante T3 lee las muestras S1, S2;

- el participante T4 lee las muestras S1, S2;
- etc.

Esto produce una repetición de las dos muestras durante la prueba. Se considera que esto no es un factor decisivo para el procedimiento en el que se evalúa una degradación con respecto a una referencia, en particular en los casos de buena calidad telefónica, donde la inteligibilidad de las señales vocales es casi perfecta. La utilización de muestras distintas para cada configuración, como a menudo sucede en los experimentos ACR, (en los que se confunden los efectos del orador y de la oración) puede ser uno de los motivos para explicar la falta de sensibilidad del método ACR.

Se permiten algunas variaciones en este esquema básico: aumentar el número de participantes, o combinar los efectos de la frase y del participante. Sin embargo, es importante que cada configuración se evalúe en el mismo grupo.

D.2.2 Condiciones de referencia

Deberán incluirse las condiciones de referencia; por ejemplo, para procesos digitales es conveniente incorporar un ruido multiplicativo con valores de Q comprendidos entre 10 y 30 dB con un mínimo de cuatro valores.

Debe elegirse una referencia de calidad para insertarla antes de cada calificación. Habitualmente se emplean las condiciones originales, es decir, muestras sin más degradación que la introducida por los sistemas de emisión y por las limitaciones de la anchura de banda de frecuencias. Por consiguiente, la elección de la referencia de calidad depende de la aplicación; es decir, para la telefonía convencional, la señal de fuente está limitada a una anchura de banda de 3,4 kHz, para la telefonía de banda ancha la limitación de la anchura de banda es de 7 kHz y para señales sonoras de alta calidad la señal tiene una limitación de banda de 15 ó 20 kHz.

D.2.3 Presentación de los estímulos

Los estímulos se presentan a los oyentes por pares (A-B) o por pares repetidos (A-B-A-B), siendo A la muestra de referencia de alta calidad y B la misma muestra procesada por el sistema sometido a evaluación. La finalidad de la muestra de referencia es fijar cada calificación de los oyentes. Se incluyen algunos «pares nulos» (A-A), al menos uno por participante, para verificar el grado de fijación. El empleo de una referencia y de juicios subjetivos con respecto a esa referencia es un procedimiento muy común en sicoacústica. Tiende a mejorar la sensibilidad de la evaluación global de los oyentes. Las muestras A y B deben estar separadas por un intervalo de 0,5-1 s. En un procedimiento de pares repetidos (A-B-A-B), la separación entre los dos pares debe ser de 1-1,5 s.

El efecto del orden habitualmente observado en las pruebas de escucha de una muestra (prueba ACR, por ejemplo) no se observa en el procedimiento DCR. De esa forma puede emplearse un solo orden aleatorio de presentación. Por consiguiente, la prueba básica y las condiciones de referencia serán ocho veces el número de condiciones nominales (cuatro participantes \times dos muestras).

D.2.4 Instrucciones para la prueba

Deben darse instrucciones a los participantes para que clasifiquen las condiciones según una escala de categorías de degradaciones de cinco notas, a saber:

- 5 Degradación inaudible
- 4 Degradación audible, pero no molesta
- 3 Degradación ligeramente molesta
- 2 Degradación molesta
- 1 Degradación muy molesta

La magnitud obtenida de las notas (nota media de opinión sobre las degradaciones) se representa por el símbolo DMOS.

D.3 Análisis estadístico

Las sensibilidades pueden cuantificarse mediante pruebas estadísticas de comparación múltiple. Si es preciso efectuar una comparación *a posteriori* de los circuitos, puede aplicarse eficazmente la prueba de Tukey [15] de la diferencia honradamente significativa (HSD, *honestly significant difference*). La prueba HSD está diseñada para efectuar todas las comparaciones por pares entre las medias y determinar el grado de significación de las diferencias en los valores medios.

Anexo E

Método de determinación de índices por categorías de comparación

E.1 Introducción

El método de determinación de índices por categorías de comparación (CCR) es semejante al método de determinación de índices por categorías de degradación (DCR) descrito en el Anexo D. En cada prueba se presentan a los oyentes un par de muestras vocales. En el procedimiento DCR, se presenta primero una muestra de referencia (no procesada), seguida por la misma muestra vocal procesada por alguna técnica. En el método DCR los oyentes siempre evalúan el grado en que se ha *degradado* la muestra procesada (la segunda) con relación a la muestra no procesada (la primera). En el procedimiento CCR, se elige al azar en cada prueba el orden de las muestras procesada y no procesada. En la mitad de las pruebas, la muestra no procesada va seguida por la muestra procesada. En las pruebas restantes se invierte el orden. Los oyentes utilizan la escala siguiente para calificar la calidad de la segunda muestra con relación a la de la primera:

La calidad de la segunda comparada con la calidad de la primera es:

- 3 Mucho mejor
- 2 Mejor
- 1 Ligeramente mejor
- 0 Aproximadamente igual
- 1 Ligeramente peor
- 2 Peor
- 3 Mucho peor

En realidad los oyentes proporcionan dos juicios con una sola respuesta: "¿qué muestra tiene mejor calidad?" y "¿en qué medida?". Los métodos DCR y CCR son particularmente útiles para evaluar el comportamiento de los sistemas de telecomunicación cuando su entrada está deteriorada por el ruido de fondo. No obstante, una ventaja del método CCR sobre el DCR es la posibilidad de evaluar el procesamiento de la señal vocal que o bien degrada o mejora la calidad de la voz.

La cantidad evaluada a partir de las puntuaciones (nota media de opinión sobre las comparaciones) viene representada por el símbolo CMOS.

NOTA - Debe tenerse cuidado cuando se utiliza el método CCR. Ciertos laboratorios han encontrado que el método es útil para evaluar los sistemas de reducción de ruido. Sin embargo, cuando se utilizó este método en las recientes evaluaciones subjetivas del códec G.729 (8 kbit/s), se encontró que el método era demasiado sensible al evaluar el comportamiento del códec para señales vocales mezcladas con ruido de fondo.

E.2 Referencia de calidad

La muestra de referencia (no procesada) (referencia de calidad o conexión directa) se presenta antes o después de la señal procesada o degradada. La muestra de referencia se genera utilizando el mismo orador y material vocal que se utilizó para la muestra procesada. Esta muestra de referencia será degradada por el mismo ruido (si existe) y procesada a través de los mismos procesos preliminares, tales como característica del transmisor, compansión logarítmica, etc. De este modo habrá una referencia de calidad distinta para cada una de las condiciones de prueba.

E.3 Referencias MNRU

Deben incluirse las condiciones de referencia MNRU para calibrar la escala de calificación. Estas referencias de ruido multiplicativo se utilizan sin mezcla adicional con los ruidos ambientales.

E.4 Presentación a los oyentes

Cada una de las muestras de voz se presenta al oyente a través de la condición de referencia de calidad y a través de un códec de prueba o condición de referencia (por ejemplo, Recomendación G.726, MNRU). Además, debe incluirse un "par nulo" para cada una de las referencias de calidad. En estas pruebas la referencia de calidad se presenta dos veces.

Los oyentes deberán evaluar la calidad de la segunda muestra con relación a la calidad de la primera muestra. Esta calificación se verifica sobre la escala de 7 notas indicada en E.1. En el Cuadro E.1 se dan unos ejemplos de instrucciones a los oyentes.

E.5 Análisis de los datos

Se ha de poner cuidado en el análisis de los datos procedentes de un experimento CCR. Dado que la mitad de las pruebas para cualquier condición de pruebas se presentan en el orden (no procesado, procesado) y la otra mitad se presentan en el orden opuesto, un sencillo promedio de las notas numéricas debe proporcionar un CMOS de aproximadamente 0 para todas las condiciones. Es necesario recodificar los datos primarios. Si el orden de presentación es (procesado, no procesado) entonces el signo de la nota numérica debe invertirse ($-1 \rightarrow 1$, $-2 \rightarrow 2$, ... , $2 \rightarrow -2$, $1 \rightarrow -1$). Las notas recodificadas pueden utilizarse para calcular la CMOS, las desviaciones típicas, etc. Los resultados se presentan, pues, en el orden de (no procesado, procesado). Puede también realizarse un análisis de varianza apropiado u otras pruebas estadísticas sobre las notas recodificadas. No obstante, no puede presumirse que las notas de opinión sobre las comparaciones representen una escala de intervalos lineal. Por consiguiente, tal vez necesiten aplicarse en su lugar estadísticas de escalas ordinales.

Ejemplo de instrucciones a los participantes**INSTRUCCIONES A LOS OYENTES****Prueba de determinación de índices por categorías de comparación****"Evaluación de la influencia de los diversos ruidos ambientales sobre la calidad de los diferentes sistemas telefónicos"**

En este experimento usted oírá parejas de muestras de voz que han sido registradas mediante diversos equipos telefónicos experimentales. Usted escuchará estas muestras a través del microteléfono que tiene enfrente.

Lo que usted va a oír es un par de frases, un corto periodo de silencio, y otro par de frases.

Usted evaluará la calidad del segundo par de frases en comparación con la calidad del primer par de frases.

Usted escuchará cuidadosamente cada par de muestras. Después, cuando se encienda la luz verde, por favor registre su opinión sobre la calidad de la segunda muestra comparada con la calidad de la primera muestra, utilizando la escala siguiente:

La calidad de la segunda comparada con la calidad de la primera es:

- 3 Mucho mejor
- 2 Mejor
- 1 Ligeramente mejor
- 0 Aproximadamente igual
- 1 Ligeramente peor
- 2 Peor
- 3 Mucho peor

Usted dispondrá de 5 segundos para registrar su respuesta pulsando el botón correspondiente a la opinión elegida. Habrá una breve pausa antes de la presentación del siguiente par de frases.

Empezaremos con una breve sesión práctica para familiarizarle con el procedimiento de pruebas. Las pruebas reales tendrán lugar durante sesiones de 10 a 15 minutos.

Anexo F**Método del umbral para comparar sistemas de transmisión con un sistema de referencia****F.1 Introducción**

Por comparación directa de un sistema de transmisión con uno de referencia es posible evaluar el comportamiento del sistema sometido a prueba en términos de característica de degradación del sistema de referencia que puede variarse y ajustarse a valores definidos. Un ejemplo de dicha característica es la relación señal/ruido (para la definición véase 8.2.3/P.830), SNR(p). El método descrito a continuación conduce a un umbral de igualdad definido como el nivel de preferencia del 50% entre el MNRU y el sistema digital.

F.2 Procedimiento de prueba

Se utiliza un procedimiento de prueba de escucha solamente. Se presenta a los oyentes un par de señales compuesto por una señal de referencia y una señal de prueba, y se les pide que indiquen cuál de las dos señales del par consideran que tiene más calidad (índice de preferencia). La equivalencia subjetiva se define como el valor de referencia correspondiente al punto de intersección de la curva

de regresión de las notas de preferencia al nivel de preferencia del 50%. En la Figura F.1 se presenta un ejemplo de la relación señal/ruido equivalente obtenida con notas de preferencia hipotéticas.

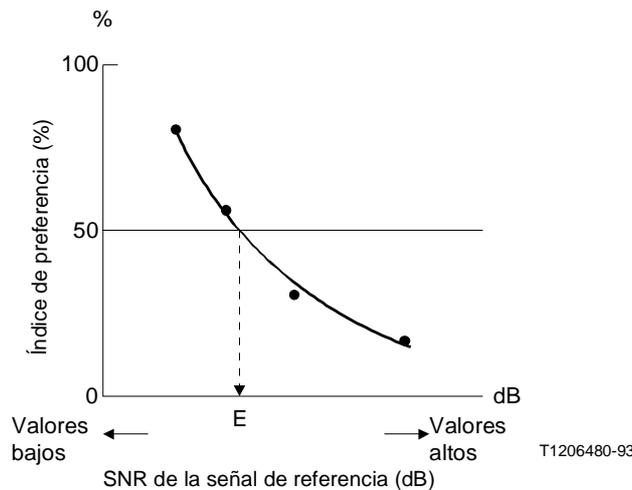


FIGURA F.1/P.800

Ejemplo de un umbral de equivalencia E con notas de preferencia hipotéticas

F.3 Presentación de las señales

La señal de referencia A y la señal de prueba B se disponen en igual número de pares A-B y pares B-A, y se presentan en orden aleatorio. Se introducen en el trayecto de referencia varios niveles de degradación espaciados, por ejemplo, a intervalos de 2 dB, de manera que la gama de notas de preferencia se extienda del 20% al 80%, con la preferencia del 50% situada en el centro de la gama de degradaciones. En la Figura F.2 se muestra un diagrama de temporización de la presentación.



T1206490-93

FIGURA F.2/P.800

Cronograma de la presentación

Se pide al participante que emita un juicio y responda «A es mejor» o «B es mejor» (elección forzada). Se prohíben las respuestas «A es igual a B», o «no hay diferencia». La duración de la presentación debe limitarse a unos seis minutos para no cansar a los oyentes. Pueden presentarse más muestras de escucha después de un periodo de descanso adecuado. Se recomiendan al menos dos, o preferentemente cuatro a cinco repeticiones (de presentaciones idénticas).

NOTA – Si el sistema de referencia está disponible en soporte físico y puede modificarse fácilmente la característica de degradación entre presentaciones, puede emplearse un procedimiento simplificado. En este

caso, el equilibrado para obtener una calidad percibida igual lo efectúa el participante. El ajuste se hace durante la pausa entre los pares. La señal de referencia se presenta siempre en primer lugar. La presentación continúa hasta que el participante indica que se ha alcanzado el umbral de igualdad.

F.4 Fuentes vocales

Es necesario emplear frases cortas pronunciadas al menos por dos hombres y dos mujeres, y preferentemente cuatro o seis de cada; se requieren frases diferentes para cada hablante. La duración debe ser de 2,5 a 5 segundos para señales vocales e inferior a 10 a 15 segundos para señales musicales. Deben evitarse chasquidos al comienzo y al final de las muestras. Debe utilizarse un micrófono lineal de anchura de banda suficiente para registrar las señales de la fuente en una cámara anecoica que tenga un ruido ambiente inferior a 20 dBA y un tiempo de reverberación inferior a 0,3 segundos, de 125 a 8000 Hz. Si se utiliza un equipo digital de grabación, el nivel de ruido de cuantificación debe ser inferior al nivel de ruido en MIC lineal de 14 bits.

F.5 Entorno de escucha

En la prueba de escucha debe utilizarse un sistema de alta fidelidad para la reproducción de sonido. Cuando la escucha se efectúa con altavoces, el equipo de reproducción debe ser de calidad de estudio, y la sala de escucha ajustarse al Informe 797 del CCIR o a la Publicación 268-13 de la CEI. Si se utilizan auriculares, es preferible la escucha dióptica (biaural). La anchura de banda debe ser por lo menos igual a la del sistema digital sometido a prueba.

F.6 Oyentes

Aunque es preferible seleccionar los oyentes con arreglo a la descripción del método ACR (véase el Anexo B), no es una condición estricta en la prueba de comparación por parejas. Si el objeto de la prueba de escucha es obtener las opiniones de oyentes no especializados, se necesitan participantes no especializados. Sin embargo, si no fuera éste el objeto de la prueba, puede recurrirse a oyentes especializados y ampliarse la prueba de escucha aumentando el número de repeticiones para cada oyente. El número mínimo de oyentes es de seis, y preferentemente debe ser de 12 ó más. Varios participantes pueden escuchar simultáneamente, pero deben asegurar que sus respuestas se obtengan independientemente.

F.7 Fiabilidad

Dado que las variaciones en la nota de preferencia en las pruebas subjetivas se supone que se ajustan a una distribución en t, la anchura de variación de las notas r que produce un 95% de fiabilidad para la nota u ($0 \leq u \leq 1$) a lo largo de las (n) pruebas (número de repeticiones para cada par de presentación multiplicado por el número de participantes y el número de señales de la fuente) viene dada por la ecuación (E-1).

$$r = \pm t(n - 1, 0,05) \cdot \sqrt{u(1 - u) / (n - 1)} \quad (\text{E-1})$$

NOTA – Cabe esperar que el método del umbral ofrezca resultados estables y fiables, incluso en sistemas de alta calidad con poca degradación.

La degradación puede introducirse en el sistema de referencia, por ejemplo, añadiendo ruido blanco. Para los sistemas digitales se recomienda el ruido multiplicativo definido en la Recomendación P.810 (MNRU). Para codificadores de voz digitales de banda ancha se recomienda la utilización de un MNRU de banda ancha, como el descrito en la Recomendación P.810. En algunas circunstancias puede ser más adecuado utilizar ruido conformado en vez de ruido blanco.

Bibliografía

- [1] VOIERS (W.D.): Evaluating processed speech using the Diagnostic Rhyme Test, *Speech Technology*, Volumen 1, N.º 4, páginas 30 a 39, enero-febrero de 1983.
- [2] Suplemento N.º 5 de la Recomendación P.74 del CCITT *Método de prueba subjetiva SIBYL, Libro Rojo*, Volumen V.
- [3] BERANEK (L.L.): Noise and Vibration Control, *McGraw-Hill*, páginas 564 a 566, 1971.
- [4] HOTH (D.F.): Room noise spectra at subscribers' telephone locations, *J.A.S.A.*, Volumen 12, páginas 99-504, abril de 1941.
- [5] Cuestión 24/XII del CCITT, Contribución COM XII-Nº 120, *Ruido en el interior de vehículos automóviles ligeros*, periodo de estudios 1981-1984.
- [6] Cuestión 24/XII del CCITT, Contribución COM XII-Nº 134, *Internal vehicle noise spectra*, periodo de estudios 1981-1984.
- [7] Contribución COM XII-Nº 208 del CCITT, *Comparison of the results of vehicle noise submitted by France and BT*, Periodo de Estudios 1981-1984.
- [8] STEVENS (S.S.): Psychophysics: Introduction to its perceptual, neural, and social prospects, *John Wiley and Sons*, 1975.
- [9] CLARINGBOLD (P.J.): The within-animal bioassay with quantal responses, *Journal of the Royal Statistical Society, Series B*, Volumen 18, Nº 1, páginas 133 a 137, 1956.
- [10] RICHARDS (D.L.): Telecommunication by speech, subcláusula 3.5.3 y subcláusula 4.5.1, *Butterworths*, Londres, 1973.
- [11] Ibid, subcláusulas 3.5.3 y 4.5.1.
- [12] Ibid, subcláusulas 4.2.1.6.
- [13] COMBESCURE (P.) y otros: Quality evaluation of speech coded at 32 kbit/s by means of degradation category ratings, *Proc. ICASSP 82 (International Conference on Acoustics, Speech and Signal Processing)*, Volumen 2, París, mayo de 1982.
- [14] Documento 11/17 del CCIR, *Evaluación subjetiva de la calidad de las imágenes de televisión (UER)*, periodo de estudios 1978-1982.
- [15] TUKEY (J.W.): The problem of multiple comparisons, *Ditton*, Princetown University, Ed 1953.
- [16] GABRIELSSON (A): Statistical treatment of data from listening tests on sound-reproducing systems, Report TA Nº 92, *KTH Karolinska Institutet*, Department of Technical Audiology, S-10044 Estocolmo, Suecia, noviembre de 1979.
- [17] Publicación 268-13 de la CEI, Anexo 3, subcláusula 3.3 (versión resumida de [16]).

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