

## RECOMENDACIÓN UIT-R BS.1116-1\*

**MÉTODOS PARA LA EVALUACIÓN SUBJETIVA DE PEQUEÑAS DEGRADACIONES EN LOS SISTEMAS DE AUDIO INCLUYENDO LOS SISTEMAS DE SONIDO MULTICANAL**

(Cuestión UIT-R 85/10)

(1994-1997)

La Asamblea de Radiocomunicaciones de la UIT,

*considerando*

- a) que las Recomendaciones UIT-R BT.500, UIT-R BS.562, UIT-R BT.710 y UIT-R BT.811 establecen algunos métodos para evaluar la calidad subjetiva de los sistemas de audio y de vídeo;
- b) que las pruebas de escucha subjetivas permiten evaluar el grado de molestia que causa al oyente cualquier degradación de la señal deseada durante su transmisión entre la fuente de origen y el oyente;
- c) que los métodos objetivos clásicos pueden no ser adecuados para evaluar los esquemas de codificación de audio avanzados y que se están desarrollando métodos de evaluación objetiva mediante percepción para probar la calidad del sonido de los sistemas de sonido;
- d) que es importante utilizar métodos normalizados para el intercambio, la compatibilidad y la evaluación correcta de los datos de prueba;
- e) que la introducción de los nuevos sistemas avanzados de audiodigital que explotan las propiedades psicoacústicas, especialmente con pequeñas degradaciones, exige la aplicación de nuevos métodos de evaluación subjetiva;
- f) que la aparición de nuevos sistemas de sonido estereofónico multicanal, con o sin acompañamiento de imagen, requiere el empleo de nuevos métodos de evaluación subjetiva, incluyendo las condiciones experimentales,

*recomienda*

- 1 que para la evaluación subjetiva de las pequeñas degradaciones en los sistemas de audio, incluidos los sistemas de sonido multicanal (con o sin imagen), se utilicen los procedimientos de prueba, evaluación y presentación de resultados que figuran en el Anexo 1.

## ANEXO 1

**1 Consideraciones generales****1.1 Alcance**

La idea es servirse de esta Recomendación para evaluar sistemas que introducen degradaciones tan pequeñas que no son detectables si no se controlan rigurosamente las condiciones experimentales y se efectúa el correspondiente análisis estadístico. La aplicación de esta Recomendación a los sistemas que introducen degradaciones relativamente importantes y fácilmente detectables, conduciría a un derroche de tiempo y de esfuerzos y podría producir, por otra parte, resultados menos fiables que una prueba más simple. Esta Recomendación constituye la referencia básica para las demás Recomendaciones, que pueden contener condiciones especiales adicionales con respecto a las de este Anexo o requisitos menos estrictos.

---

\* Esta Recomendación debe señalarse a la atención del Grupo ad hoc de la Organización Internacional de Unificación de Normas/Grupo de Expertos sobre imágenes en movimiento (ISO/MPEG) – Audio.

## 1.2 Contenido

Este Anexo consta de 11 puntos en los que se detallan los diversos aspectos de las pruebas:

- 1 Consideraciones generales
- 2 Diseño del experimento
- 3 Selección de los oyentes
- 4 Método de prueba
- 5 Atributos
- 6 Material de programa
- 7 Dispositivos de reproducción
- 8 Condiciones de escucha
- 9 Análisis estadístico
- 10 Presentación de los resultados de los análisis estadísticos
- 11 Contenido de los informes de prueba.

También se incluyen varios Apéndices que contienen pautas para la selección de oyentes experimentados, así como un ejemplo de las instrucciones que se dan a los participantes en las pruebas.

El Apéndice 4 es un Glosario de palabras comunes que se utilizan con un cierto significado técnico.

## 2 Diseño del experimento

Para reunir información adecuada en un dominio de interés científico se utilizan muchos tipos de estrategias de investigación. En la evaluación subjetiva de las pequeñas degradaciones de los sistemas de audio, se utilizarán los métodos experimentales más formales. Los experimentos subjetivos se caracterizan, en primer lugar, por la manipulación y el control real de las condiciones del experimento y, en segundo lugar, por los datos cuantitativos procedentes de observadores humanos.

Es necesario llevar a cabo una planificación y diseño detallados del experimento para asegurar que no van a aparecer factores incontrolados que influyan en la prueba de escucha, de forma que no surjan ambigüedades. Como ejemplo puede indicarse que si la secuencia real de los fragmentos de audio es idéntica para todos los participantes en una prueba de escucha, no podrá asegurarse que los juicios emitidos por dichos participantes sean debidos a dicha secuencia y no a los distintos niveles de degradación presentados. De acuerdo con ello, las condiciones de prueba deben disponerse de manera que revelen los efectos de los factores independientes y sólo de esos factores.

En situaciones en las que cabe esperar que las posibles degradaciones y otras características se distribuyan de forma homogénea a lo largo de toda la prueba de escucha, puede aplicarse una aleatorización verdadera a la presentación de las condiciones de prueba.

Cuando puede suponerse una falta de homogeneidad, este hecho debe tenerse en cuenta al presentar las condiciones de prueba. Por ejemplo, cuando el material que va a evaluarse presenta diversos niveles de dificultad, el orden de presentación de los estímulos debe distribuirse aleatoriamente, tanto en el transcurso de una sesión como entre sesiones.

De forma similar, las pruebas de escucha deben diseñarse de manera que no se sobrecargue excesivamente a los participantes hasta el punto de disminuir la exactitud de sus apreciaciones. Salvo en los casos en que la relación entre la imagen y el sonido es importante, conviene que la evaluación de los sistemas de audio se lleve a cabo sin acompañamiento de imágenes.

Una consideración importante es la inclusión de las condiciones de control adecuadas. Normalmente, dichas condiciones consisten en la inclusión de materiales de audio sin degradar, presentadas de forma imprevisible a los participantes. La diferencia entre los juicios emitidos en respuesta a estos estímulos de control y los que responden a los estímulos potencialmente degradados es lo que permite llegar a la conclusión de que las apreciaciones constituyen una evaluación real de las degradaciones.

Posteriormente se discutirán algunas de estas consideraciones. Debe entenderse que los temas relativos al diseño, ejecución y análisis estadístico del experimento son complejos y que en una Recomendación como éste sólo pueden ofrecerse pautas generales. Se recomienda que al principio del proceso de planificación de las pruebas de escucha se cuente con la presencia de profesionales experimentados en análisis estadístico y diseño de experimentos o, al menos, se les consulte.

### **3 Selección de los oyentes**

#### **3.1 Oyentes expertos**

Es importante que los datos de las pruebas de escucha para evaluar pequeñas degradaciones en los sistemas de audio procedan exclusivamente de participantes con experiencia en detectar dichas pequeñas degradaciones. Cuanto mayor sea la calidad alcanzada en los sistemas que deben someterse a prueba, más importante será contar con oyentes expertos.

#### **3.2 Criterios para la selección de los participantes**

Los resultados de las pruebas subjetivas en sistemas de sonido con pequeñas degradaciones llevadas a cabo mediante un grupo seleccionado de oyentes no se pueden extrapolar, en principio, al público en general. Normalmente, el objetivo es investigar si un grupo de oyentes expertos, sometidos a unas ciertas condiciones, pueden percibir degradaciones relativamente ligeras y a la vez realizar una estimación cuantitativa de las degradaciones introducidas. La naturaleza exigente del procedimiento de prueba tiene por objeto resaltar los problemas que pueden aparecer durante el amplio periodo de exposición a las distintas condiciones que surgen en la vida real una vez que se ha presentado el sistema al usuario.

A veces hay una razón para introducir una técnica de rechazo antes (selección previa) o después (selección posterior) de la prueba real. En algunos casos pueden utilizarse ambos tipos de rechazos. Se entiende por eliminación un proceso en el que se omiten todas las apreciaciones expresadas por un participante en particular.

Cualquier tipo de técnica de rechazo que no se haya analizado y aplicado de forma minuciosa puede dar lugar a resultados sesgados. Por consiguiente, es muy importante que siempre que se eliminen algunos datos, el informe de prueba describa claramente el criterio aplicado de forma que el lector pueda llegar a sus propias conclusiones.

##### **3.2.1 Selección previa de los participantes**

Los procedimientos de selección previa incluyen métodos tales como pruebas audiométricas, selección de participantes basándose en su experiencia anterior y desempeño en pruebas anteriores, y eliminación de participantes en base a análisis estadísticos de pruebas previas. El procedimiento de adiestramiento puede utilizarse como método para realizar la selección previa.

El argumento principal para introducir una técnica de selección previa es el de aumentar la eficacia de la prueba de escucha. Sin embargo, ello debe también contrapesarse con el riesgo de limitar excesivamente la representatividad de los resultados.

##### **3.2.2 Selección posterior de los participantes**

Los métodos de selección posterior pueden dividirse aproximadamente en dos tipos; uno se basa en las incoherencias respecto al resultado medio y el otro en la capacidad del participante para realizar identificaciones correctas. El primer tipo nunca es justificable. Siempre que se lleve a cabo una prueba de escucha subjetiva con el método de prueba recomendado en esta Recomendación, se dispone automáticamente de la información necesaria para el segundo tipo de selección posterior. En el Apéndice 1 se propone un método estadístico para aplicar este segundo método.

Los métodos se utilizan fundamentalmente para eliminar participantes que no son capaces de realizar las discriminaciones adecuadas. La aplicación de un método de selección posterior puede aclarar las tendencias en el resultado de una prueba. Sin embargo, teniendo en cuenta las sensibilidades de la «variabilidad de los participantes» a los diferentes estímulos, deben tomarse todas las precauciones necesarias.

### **3.3 Número de oyentes**

El número adecuado de oyentes puede estimarse calculando la varianza y determinando la resolución necesaria del experimento.

Cuando las condiciones de una prueba de escucha están rígidamente determinadas por los aspectos técnicos y de comportamiento de los participantes, la experiencia ha demostrado que a menudo bastan los datos procedentes de 20 de ellos para extraer las conclusiones adecuadas de la prueba. Si puede realizarse un análisis de los mismos a medida que se lleva a cabo la prueba, no es necesario seguir procesando los resultados de más participantes cuando se ha alcanzado un nivel apropiado estadísticamente significativo para extraer las conclusiones adecuadas de la prueba.

Si se espera que algunos de los sistemas sean prácticamente transparentes, es necesario contar con un número mayor de participantes para garantizar que una cantidad suficientemente amplia de éstos supere la prueba de selección posterior.

Si por cualquier razón no puede lograrse un control estricto del experimento, se necesita un número más elevado de participantes para obtener la resolución necesaria.

El número de oyentes no se basa exclusivamente en consideraciones sobre la resolución deseada. Los resultados obtenidos con el tipo de experimento considerado en la presente Recomendación son, en principio, válidos únicamente para dicho grupo de oyentes expertos implicados realmente en la prueba. Por consiguiente, si se aumenta el número de oyentes, los resultados pueden aplicarse a un grupo más general de oyentes expertos y, por tanto, pueden considerarse a veces más convincentes. También puede que sea necesario aumentar el número de oyentes para tener en cuenta la probabilidad de que los participantes varíen su sensibilidad a distintos estímulos.

#### 4 Método de prueba

Para las evaluaciones subjetivas de sistemas que producen pequeñas degradaciones, es necesario seleccionar un método adecuado. Un método especialmente sensible, estable y que permite detectar con exactitud pequeñas degradaciones es el de «triple estímulo doblemente ciego con referencia oculta». Por consiguiente, es el que se utilizará para este tipo de prueba.

En la forma más adecuada y sensible de este método, sólo actúa un participante cada vez y lo hace seleccionando a discreción uno de entre tres estímulos («A», «B» y «C»). La referencia conocida siempre es el estímulo «A». La referencia oculta y el objeto son «B» y «C», asignados de manera «aleatoria», dependiendo del experimento.

Se solicita al participante que evalúe las degradaciones en «B» comparadas con las de «A», y las de «C» comparadas con las de «A», de acuerdo con la escala continua de degradación de cinco notas. Uno de los estímulos, «B» o «C», debe ser indistinguible del estímulo «A», el otro puede presentar degradaciones. Toda diferencia percibida entre la referencia y los otros estímulos debe interpretarse como una degradación.

Una vez que el participante haya finalizado la apreciación de un experimento, en el método preferido, debe ser posible proceder directamente al siguiente experimento. Estos pasos pueden repetirse hasta que el participante haya realizado una evaluación. De esta forma, el procedimiento de prueba establece su propio ritmo.

La escala de apreciaciones puede considerarse continua con «referencias» obtenidas de la escala de degradación de cinco notas del UIT-R que aparece en la Recomendación UIT-R BS.1284 indicada en el Cuadro 1.

CUADRO 1

Degradación	Nota
Imperceptible	5,0
Perceptible, pero no molesta	4,0
Ligeramente molesta	3,0
Molesta	2,0
Muy molesta	1,0

NOTA 1 – Se ha visto que la utilización de puntos de referencia intermedios predefinidos puede introducir sesgos [Poulton, 1992]. Es posible utilizar las escalas numéricas sin descripciones de puntos de referencia. En tales casos, habrá que indicar la orientación prevista de las escalas. Esto puede contribuir a superar los problemas de traducción que se plantean en las comparaciones de las pruebas efectuadas en diferentes idiomas.

En caso de que no se utilicen puntos de referencia intermedios, resulta esencial normalizar los resultados correspondientes a cada sujeto con respecto a la media y a la desviación típica. La siguiente ecuación puede aplicarse para lograr dicha normalización, reteniéndose al mismo tiempo la escala original.

$$Z_i = \frac{(x_i - x_{si})}{s_{si}} \cdot s_s + x_s$$

donde:

- $Z_i$ : resultado normalizado
- $x_i$ : nota del participante  $i$
- $x_{si}$ : nota media para el participante  $i$  en la sesión  $s$
- $x_s$ : nota media de todos los participantes en la sesión  $s$
- $s_s$ : desviación típica para todos los participantes en la sesión  $s$
- $s_{si}$ : desviación típica para el participante  $i$  en la sesión  $s$ .

El uso de escalas sin punto de referencia intermedio impide también la interpretación de los resultados en términos absolutos.

Se recomienda utilizar la escala con una resolución de un número decimal.

El método de prueba consta de dos partes: fase de adiestramiento o de familiarización y fase de apreciación.

#### 4.1 Fase de adiestramiento o de familiarización

Antes de realizar la apreciación formal, debe permitirse a los participantes familiarizarse con los dispositivos y el entorno de prueba y con el proceso y las escalas de apreciación, así como los métodos de utilización. Los participantes deben familiarizarse también con los elementos estudiados. Para las pruebas más sensibles se les debe presentar todo el material que deberán evaluar posteriormente en las sesiones de apreciación formal. Durante el proceso de familiarización o adiestramiento, debe agruparse preferentemente a los participantes (por ejemplo, de tres en tres) de manera que puedan intercambiar opiniones libremente y discutir entre sí sobre las degradaciones que detecten.

En el Apéndice 3 aparece, como ejemplo, un modelo de «Instrucciones para los oyentes». Estas instrucciones incluyen una descripción de la técnica de presentación de estímulos denominada «Triple estímulo doblemente ciego con referencia oculta». Si se lleva a cabo de manera correcta el proceso de familiarización puede transformar a algunos participantes con habilidad acústica inicialmente baja en expertos a efectos de la prueba. Al finalizar dicho proceso, los participantes deben haber adquirido un conocimiento preciso de la escala que se utilizará en la fase de apreciación formal que seguirá al proceso de familiarización o adiestramiento.

#### 4.2 Fase de apreciación

Al iniciar la primera sesión de apreciación formal del día debe hacerse a cada uno de los participantes una presentación oral de las instrucciones de prueba, completada preferentemente con un texto escrito. Inmediatamente antes de comenzar la apreciación formal pueden presentarse varias comparaciones a título ilustrativo.

Como la memoria auditiva a largo y medio plazo no es fiable, el procedimiento de prueba debe basarse exclusivamente en la memoria a corto plazo. Para ello, lo más adecuado es utilizar un método de conmutación casi instantánea (véase la Nota 1) junto con un sistema de triple estímulo, como el que se describe en el Apéndice 3. Este tipo de conmutación exige una alineación en el tiempo muy exacta entre los estímulos.

NOTA 1 – La conmutación instantánea exacta puede dar lugar a efectos secundarios si las ondas de estímulos sucesivos no son idénticas. Por ejemplo, es preferible utilizar una conmutación casi instantánea en la que la desconexión/cambio/conexión dure unos 40 ms.

En las evaluaciones más críticas, los participantes deben actuar de forma individual. Sólo de esa manera tendrán completa libertad para conmutar entre los estímulos en el método del triple estímulo. Esta libertad es esencial para que el participante pueda utilizar completamente su propio albedrío al realizar comparaciones detalladas entre los estímulos de cada experimento.

Es preferible que los participantes puedan conmutar entre los estímulos sin ayuda visual, de forma que, si lo desean, puedan mantener los ojos cerrados para concentrarse mejor en condiciones de distracción mínima. No deben aparecer fenómenos audibles (por ejemplo «clicks») en el sistema de conmutación, puesto que podrían interferir seriamente en el proceso de evaluación.

Una sesión de apreciación no debe durar más de 20 ó 30 min, si bien el carácter de autocontrol del ritmo de los experimentos señalado en este documento dará lugar a una variación en la duración de la prueba según los participantes. La experiencia sugiere que no deben programarse más de 10 ó 15 experimentos por sesión para lograr la duración de sesión deseada. La fatiga de los participantes puede convertirse en un factor fundamental que interferiría seriamente en la validez de sus juicios. Para evitar esta circunstancia, entre sesiones sucesivas de cada participante deben preverse periodos de descanso de una duración al menos igual a la de la sesión.

## 5 Atributos

A continuación se indican los atributos específicos de las evaluaciones monofónicas, estereofónicas y multicanal. Es conveniente evaluar en cada caso el atributo «calidad de audio básica». Los experimentadores pueden definir y evaluar otros atributos.

Un problema que puede aparecer con participantes que intenten evaluar más de un atributo en cada experimento es el de una sobrecarga en las respuestas. Si los participantes se encuentran sobrecargados o confundidos, intentando responder múltiples cuestiones sobre un estímulo determinado, ello puede dar lugar a apreciaciones no fiables para todas las cuestiones.

### 5.1 Sistema monofónico

*Calidad de audio básica:*

- Este atributo sencillo y general se utiliza para juzgar una o todas las diferencias detectadas entre la referencia y el objeto.

## 5.2 Sistema estereofónico de dos canales

### *Calidad de audio básica:*

- Este atributo sencillo y general se utiliza para juzgar una o todas las diferencias detectadas entre la referencia y el objeto.

Puede tener interés el siguiente atributo adicional:

### *Calidad de la imagen estereofónica:*

- Este atributo está relacionado con las diferencias entre la referencia y el objeto en términos de emplazamientos de la imagen sonora y sensaciones de profundidad y realidad del elemento de audio.

Aunque algunos estudios han demostrado que puede degradarse la calidad de la imagen estereofónica, no se han hecho todavía suficientes investigaciones como para justificar la utilización de un índice de calidad de imagen estereofónica distinto del de la calidad audio básica.

NOTA 1 – Hasta 1993 la mayoría de los estudios sobre evaluación subjetiva de pequeñas degradaciones en sistemas estereofónicos de dos canales utilizaban exclusivamente el atributo de calidad de audiobásica. Por consiguiente, el atributo de calidad de imagen estereofónica estaba incluido, implícitamente o explícitamente, en la calidad de audio básica como atributo global en dichos estudios.

## 5.3 Sistema estereofónico multicanal

### *Calidad de audio básica:*

- Este atributo sencillo y general se utiliza para juzgar una o todas las diferencias detectadas entre la referencia y el objeto.

Pueden tener interés los siguientes atributos adicionales:

### *Calidad de imagen frontal:*

- Este atributo se relaciona con el emplazamiento de las fuentes de sonido frontales. Incluye la calidad de imagen estereofónica y las pérdidas de definición.

### *Impresión de la calidad panorámica:*

- Este atributo está relacionado con la impresión espacial, el ambiente o los efectos panorámicos direccionales especiales.

## 6 Material de programa

Sólo se utiliza material crítico para poner de relieve las diferencias entre los sistemas sometidos a prueba. Se entiende por material crítico aquel que somete a un esfuerzo a los sistemas bajo prueba. No hay un material de programa «adecuado» de forma universal que pueda utilizarse para evaluar todos los sistemas bajo todas las condiciones. En consecuencia, debe encontrarse para cada sistema probado en cada experimento, el material de programa crítico apropiado. La búsqueda de buen material normalmente lleva tiempo; sin embargo, si no se proporciona el material crítico conveniente a cada sistema, los experimentos no pondrán de relieve las diferencias entre los sistemas y no podrán sacarse conclusiones al respecto.

Antes de aceptar como válido un resultado «nulo», debe demostrarse desde un punto de vista empírico y estadístico que la incapacidad para encontrar diferencias entre los sistemas no se debe a la insensibilidad del experimento por una elección inadecuada del material de audio o a cualquier otro aspecto incorrecto del experimento. En el caso extremo en que algunos o todos los sistemas sean completamente transparentes, puede que sea necesario programar pruebas especiales con notas de referencia bajas o medias a efectos explícitos de examinar la pericia de los participantes (véase el Apéndice 1).

Estas referencias deben conocerse (por ejemplo, basándose en investigaciones anteriores) para que puedan ser detectadas por los oyentes expertos, pero no así por los oyentes inexpertos. Estas referencias se introducen como elementos de la prueba que han de verificarse no solamente en lo que concierne a la experiencia del oyente sino también a la sensibilidad con respecto a los demás aspectos de la situación experimental.

Si estas referencias, sea que se encuentren incorporadas de manera impredecible en el contexto de los elementos aparentemente transparentes o en otros elementos de una prueba independiente, son identificados correctamente por todos los oyentes en un método de prueba normalizado (véase el § 3 de este Anexo), aplicando las consideraciones estadísticas esbozadas en el Apéndice 1, esto podrá aducirse como prueba de que la experiencia del oyente resulta aceptable y de que los demás aspectos de la situación experimental no plantean problemas de sensibilidad. Por consiguiente, en este caso los resultados en relación con la transparencia aparente de estos oyentes son prueba de una «transparencia auténtica» en lo que concierne a los elementos o sistemas respecto de los cuales los oyentes no puedan distinguir entre versiones codificadas y no codificadas.

Por otra parte, si estas referencias no permiten dicha identificación correcta por parte de ningún oyente, esto será indicio de que los oyentes carecen de la experiencia necesaria o de que existen defectos de sensibilidad en la situación, o ambas cosas a la vez. En este caso, no podrá interpretarse adecuadamente la transparencia aparente de los sistemas y el experimentador habrá de repetir la prueba con nuevos oyentes para reemplazar a los que no superaron esta prueba adicional e introducir otros cambios que intensifiquen la sensibilidad experimental.

En la búsqueda del material crítico, debe permitirse cualquier estímulo que pueda considerarse como posible material de difusión. No deben incluirse las señales sintéticas específicamente diseñadas para irrumpir en un sistema concreto. El contenido artístico o intelectual de una secuencia de programa no debe ser tan interesante ni tan aburrido o desagradable como para distraer la atención del participante e impedir que se centre en la detección de degradaciones. Debe tenerse en cuenta la frecuencia de aparición esperada de cada tipo de material de programa en difusiones reales. No obstante, debe entenderse que la naturaleza del material difundido puede variar a lo largo del tiempo debido a los cambios futuros que se producirán en las preferencias y estilos musicales. En el futuro, los modelos de percepción objetivos pueden ayudar a seleccionar el material crítico.

Al seleccionar el material de programa, es importante definir con precisión los atributos que van a evaluarse. La responsabilidad de selección del material debe recaer en un grupo de participantes adiestrados con un conocimiento básico de las degradaciones que cabe esperar. El punto de partida debe basarse en una amplia gama de material, que puede ampliarse con grabaciones especiales.

Para preparar las cintas de prueba de comparación subjetiva, el citado grupo de participantes adiestrados debe ajustar subjetivamente la sonoridad de cada pasaje antes de grabarle en el medio de prueba, lo que permitirá el uso posterior de este medio con una ganancia fija establecida para todos los elementos de programa.

Por consiguiente, en todas las secuencias de prueba el grupo de participantes adiestrados debe reunirse y llegar a un acuerdo sobre los niveles sonoros relativos en los pasajes de prueba individuales. Además, los expertos deben alcanzar un consenso sobre el nivel de presión sonora absoluto reproducido para la secuencia en su conjunto con respecto al nivel de alineación.

Al principio de cada grabación debe incluirse una ráfaga de tono (por ejemplo, 1 kHz, 300 ms, -18 dBFS (FS: por completo)) de nivel de señal de alineación que permita el ajuste de su nivel de alineación de salida al nivel de alineación de entrada necesario para el canal de reproducción (véase el § 8.4.1). En el caso de material de prueba con grabación digital, el nivel de alineación debe corresponder a -18 dB con respecto al máximo nivel de codificación posible del sistema digital [UER, 1992]. La señal del programa radiofónico debe controlarse de forma que las amplitudes de las crestas sólo raramente rebasen la amplitud de cresta de la señal máxima permitida definida en la Recomendación UIT-R BS.645 (una onda sinusoidal 9 dB por encima del nivel de alineación). En esas condiciones un medidor de cresta de programa indicará los niveles que no rebasan el nivel de la señal máxima permitida. La ráfaga de tono puede ser útil también para la alineación temporal de referencia y los estímulos de la prueba.

El posible número de pasajes que pueden incluirse en una prueba varía, pero deberá ser igual para cada objeto. Una estimación razonable es de  $1,5 \times$  (número de objetos), con un valor mínimo de 5 pasajes. Los pasajes de audio tendrán una duración típica entre 10 y 25 s. Debido a la complejidad de la tarea, el objeto u objetos deben estar disponibles. Sólo puede realizarse una selección adecuada si se establece un calendario apropiado.

Para la evaluación de sistemas monofónicos y estereofónicos, sería conveniente seleccionar los pasajes de fuentes fácilmente accesibles de forma que las cintas de prueba preparadas pudieran verificarse fácilmente, si es necesario, teniendo en cuenta las fuentes originales. El disco compacto SQAM es un ejemplo de dicha fuente. No obstante, es más importante utilizar pasajes críticos verdaderos, aun cuando provengan de fuentes menos fácilmente accesibles.

La calidad de funcionamiento de un sistema multicanal en condiciones de reproducción bicanal debe comprobarse utilizando un mezclado de referencia. Aunque la utilización de un mezclado fijo puede considerarse restrictivo en algunas circunstancias, es sin duda alguna la opción más sensata para su utilización por los organismos de radiodifusión a largo plazo. Las ecuaciones para el mezclado de referencia son las siguientes (véase la Recomendación UIT-R BS.775):

$$L_0 = 1,00 L + 0,71 C + 0,71 L_s$$

$$R_0 = 1,00 R + 0,71 C + 0,71 R_s$$

La preselección de pasajes de prueba adecuados para la evaluación crítica de la calidad del mezclado de referencia bicanal debería basarse en la reproducción de material de programa con un mezclado bicanal.

## 7 Dispositivos de reproducción

### 7.1 Consideraciones generales

Deben elegirse auriculares o altavoces de control de referencia con objeto de que todas las señales radiofónicas u otras señales de prueba puedan reproducirse de forma óptima; es decir, deben proporcionar sonido neutro para cualquier tipo de reproducción y deben poder utilizarse en evaluaciones monofónicas así como en evaluaciones de sistemas de sonido estereofónico de dos o más canales.

Algunas deficiencias en la calidad se perciben más claramente en el caso de reproducción por auricular, sin embargo otras se detectan con más facilidad si se utilizan altavoces. Por consiguiente, sería necesario determinar el tipo adecuado de dispositivo de reproducción mediante pruebas previas subjetivas.

Debe utilizarse la reproducción por altavoces especialmente en casos en que las deficiencias afecten a las características de la imagen sonora estereofónica.

Para evaluar los sistemas de sonido estereofónicos de dos canales, puede que sea necesario utilizar tanto altavoces como auriculares estereofónicos. Para evaluar sistemas de sonido monofónicos, puede utilizarse un altavoz central y/o auriculares.

La elección de unos u otros para experimentos individuales o grupos de experimentos permitirá establecer una correlación entre la audibilidad de un efecto y el transductor utilizado, siendo menor el número de participantes. De forma alternativa, si los participantes pueden conmutar a discreción entre altavoces y auriculares, no será posible determinar dicha correlación entre la audibilidad de un efecto y el transductor utilizado.

Para evaluar los sistemas de sonido multicanal con o sin acompañamiento de imágenes, deben utilizarse altavoces, si se desean estudiar las influencias en todos los canales de reproducción activados simultáneamente.

En todos los casos, deben adaptarse acústicamente cada altavoz a las gamas de frecuencia pertinentes, de forma que las diferencias de timbre inherentes a dichos altavoces se reduzcan al mínimo.

### 7.2 Altavoz de control de referencia

#### 7.2.1 Consideraciones generales

El «altavoz de control de referencia» es un equipo de escucha en estudio de alta calidad que comprende una unidad integrada de sistemas de altavoces alojada en un bastidor específicamente dimensionado, con una ecualización especial, amplificadores de potencia de alta calidad y filtros de separación.

Las características electroacústicas deben cumplir los siguientes requisitos mínimos, medidos en condiciones de campo libre. La referencia para los valores del nivel sonoro absoluto es una distancia de medición de 1 m al centro acústico, a menos que se especifique otra cosa.

#### 7.2.2 Requisitos electroacústicos

##### 7.2.2.1 Amplitud en función de la respuesta en frecuencia

Para la preselección de los altavoces, la curva de respuesta en frecuencia en la gama de 40 Hz a 16 kHz medida en bandas de un tercio de octava y utilizando ruido rosa en los ejes principales (ángulo respecto al eje = 0°), debe encontrarse preferentemente dentro de una franja de tolerancia de 4 dB. Las curvas de respuesta en frecuencia medidas para ángulos respecto al eje de  $\pm 10^\circ$  no deben diferir de la respuesta en frecuencia en el eje principal más de 3 dB y para ángulos respecto al eje de  $\pm 30^\circ$  (únicamente en el plano horizontal), en más de 4 dB.

Deben adaptarse las respuestas en frecuencia de los distintos altavoces. A ser posibles las diferencias no deben rebasar el valor de 1,0 dB en la gama de frecuencias de al menos 250 Hz a 2 kHz.

NOTA 1 – La curva de respuesta de la sala de prueba indicada en el § 8.3.4 describe la característica en frecuencia del campo sonoro en la sala de escucha.

### 7.2.2.2 Índice de directividad

El índice de directividad,  $C$ , medido con ruido en banda de un tercio de octava, en la gama de frecuencias entre 500 Hz y 10 kHz debe encontrarse dentro de los siguientes límites:

$$6 \text{ dB} \leq C \leq 12 \text{ dB}$$

El índice de directividad debe aumentar suavemente con la frecuencia.

### 7.2.2.3 Distorsión no lineal

Se aplica al altavoz una señal de entrada de tensión constante que produce un nivel de presión sonora medio de 90 dB. Respecto a dicho nivel de presión sonora, ningún componente de distorsión armónica en la gama de frecuencias fundamental de 40 Hz a 16 kHz, deberá rebasar los siguientes valores:

$$\begin{array}{ll} -30 \text{ dB (3\%)} & \text{para } f < 250 \text{ Hz} \\ -40 \text{ dB (1\%)} & \text{para } f \geq 250 \text{ Hz} \end{array}$$

### 7.2.2.4 Fidelidad de los transitorios

El tiempo de caída a un nivel de  $1/e$  (aproximadamente 0,37) del nivel original (únicamente en el eje principal) medido en un osciloscopio, debe ser:

$$t_s < 5 / f$$

siendo  $f$  la frecuencia.

Ello significa que el tiempo de caída de una ráfaga de tono sinusoidal no debe superar cinco veces el periodo de la onda sinusoidal correspondiente.

### 7.2.2.5 Retardo de tiempo

Las diferencias de retardo de tiempo entre los canales de un sistema estereofónico o multicanal no deben superar los 100  $\mu$ s.

NOTA 1 – Esto no incluye el retardo de tiempo entre el altavoz y la posición de escucha.

En el caso de sistemas con acompañamiento de imagen, el retardo de tiempo total del altavoz de control de referencia en combinación con el sistema o sistemas sometidos a prueba no debe rebasar los límites establecidos en la Recomendación UIT-R BS.775.

### 7.2.2.6 Gama dinámica

El nivel sonoro de funcionamiento máximo que puede producir un altavoz durante un periodo de tiempo de, al menos, 10 min sin que sufra daños térmicos o mecánicos y sin sobrecargar los circuitos activados, medido con un programa que simula una señal de ruido (de acuerdo con la Publicación 268-1c de la Comisión Electrotécnica Internacional (CEI)), debe ser:

$$L_{ef \text{ máx}} > 108 \text{ dB}$$

La medición se realiza con un medidor de nivel sonoro ajustado a respuesta plana y valor eficaz (lento).

El nivel de ruido acústico equivalente generado por un solo altavoz de control de referencia y un amplificador asociado, referido a una distancia de 1 m desde el centro acústico (véase la Nota 1) debe ser:

$$L_{\text{ruido}} < 10 \text{ dBA}$$

NOTA 1 – El centro acústico es el punto de referencia a efectos de medición. Normalmente corresponde al punto medio geométrico de la superficie que radia las frecuencias más elevadas del altavoz. Este punto debe venir indicado por el fabricante.

### 7.3 Auriculares de control de referencia

#### 7.3.1 Consideraciones generales

Los auriculares de control de referencia son equipos de escucha en estudio de alta calidad ecualizados para dar una respuesta en campo difuso.

#### 7.3.2 Requisitos electroacústicos

##### 7.3.2.1 Respuesta en frecuencia

La Recomendación UIT-R BS.708 especifica la respuesta en frecuencia de los auriculares de control con campo difuso en estudios.

##### 7.3.2.2 Retardo de tiempo

Las diferencias de retardo de tiempo entre los canales de un sistema estereofónico no deben rebasar los 20  $\mu$ s.

En el caso de sistemas con acompañamiento de imagen, el retardo de tiempo total de los auriculares de control de referencia en combinación con el sistema o sistemas sometidos a prueba no debe rebasar los límites establecidos en la Recomendación UIT-R BS.775.

## 8 Condiciones de escucha

### 8.1 Consideraciones generales

El término «Condiciones de escucha» describe los requisitos acústicos complejos de un campo sonoro de referencia que afecta a un oyente situado en el punto de escucha de referencia de una sala de escucha, para el sonido reproducido por los altavoces. Incluye:

- las características acústicas de la sala de escucha,
- la disposición de los altavoces en la sala de escucha,
- el emplazamiento del punto o zona de escucha de referencia,

que establecen las características del campo sonoro resultante en dicho punto o zona.

Como la tecnología actual no permite aún realizar una descripción completa del campo sonoro de referencia, que sólo puede efectuarse mediante parámetros acústicos, se indican algunos requisitos geométricos y de acústica de salas para una sala de escucha de referencia, a fin de asegurar la viabilidad de las condiciones de escucha descritas.

### 8.2 Sala de escucha de referencia

#### 8.2.1 Consideraciones generales

Para realizar las pruebas subjetivas en el caso de reproducción por altavoces deben observarse los siguientes requisitos. Se describen los requisitos mínimos para una sala de escucha de referencia.

En el caso de reproducción por auriculares únicamente, la sala de escucha debe satisfacer al menos el requisito de nivel de ruido de fondo.

#### 8.2.2 Propiedades geométricas

Los valores siguientes describen las dimensiones efectivas adecuadas de una sala de escucha de referencia. Si la sala de prueba no puede tener estas dimensiones, deben satisfacerse al menos los requisitos sobre condiciones de campo sonoro y disposición de los altavoces mencionados en los puntos siguientes.

##### 8.2.2.1 Tamaño de la sala (superficie)

- Para reproducción monofónica o estereofónica de dos canales: 20 a 60 m<sup>2</sup>.
- Para reproducción estereofónica multicanal: 30 a 70 m<sup>2</sup>.

NOTA 1 – Las salas más pequeñas imponen limitaciones en cuanto al número máximo de oyentes que pueden acoger simultáneamente.

##### 8.2.2.2 Forma de la sala

La sala debe ser simétrica respecto al plano vertical en la perpendicular media de la base estereofónica. El suelo debe tener la forma de un rectángulo o un trapecio, preferentemente.

**8.2.2.3 Proporciones de la sala**

Las dimensiones deben observar las siguientes relaciones para asegurar una distribución razonablemente uniforme de los tonos propios en baja frecuencia de la sala:

$$1,1 w / h \leq l / h \leq 4,5 w / h - 4$$

siendo:

- $l$ : longitud
- $w$ : anchura
- $h$ : altura.

Además, se cumplirán las condiciones  $l / h < 3$  y  $w / h < 3$ .

**8.2.3 Propiedades acústicas de la sala**

**8.2.3.1 Tiempo de reverberación**

El valor medio del tiempo reverberación,  $T_m$ , medido en la gama de frecuencias 200 Hz a 4 kHz debe ser:

$$T_m = 0,25 (V / V_0)^{1/3} \quad \text{s}$$

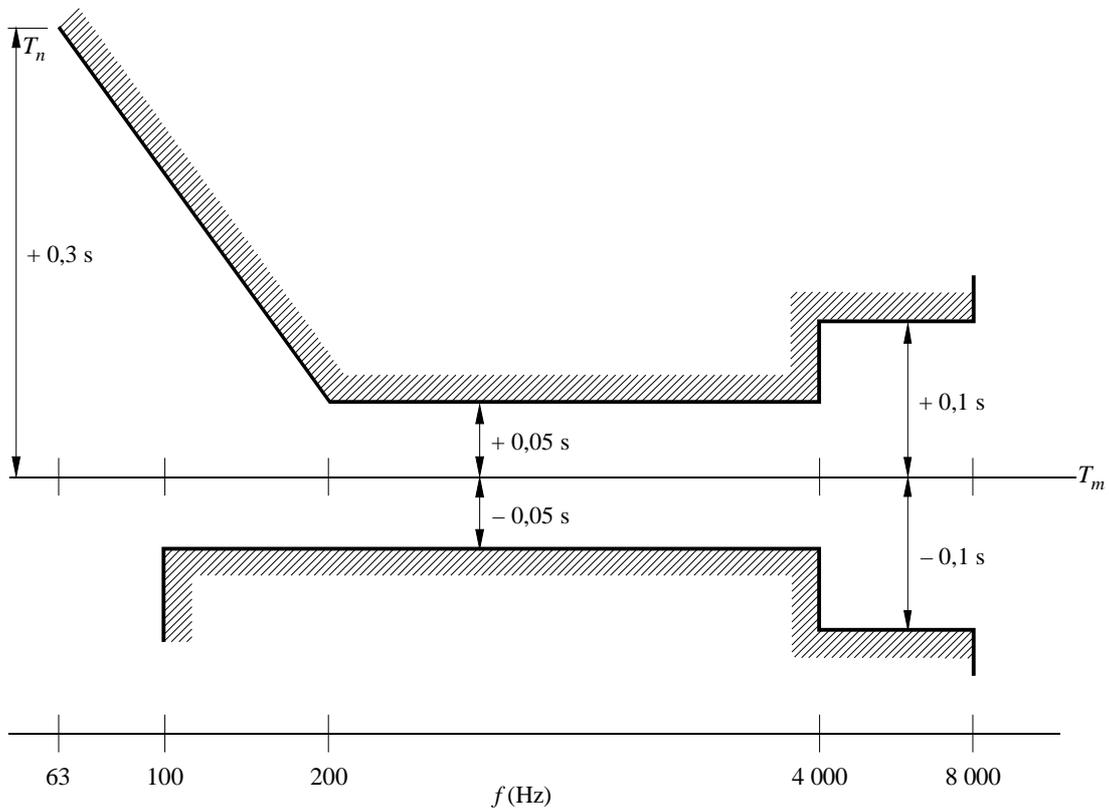
siendo:

- $V$ : volumen de la sala
- $V_0$ : volumen de referencia de 100 m<sup>3</sup>.

En la Fig. 1 se indican las tolerancias que se aplican a  $T_m$  en la gama de frecuencias 63 Hz (véase la Nota 1) a 8 kHz.

NOTA 1 – Es difícil medir valores reducidos del tiempo de reverberación a bajas frecuencias.

FIGURA 1  
Límites de tolerancia para el tiempo de reverberación,  
con respecto al valor medio,  $T_m$



### **8.3 Condiciones del campo sonoro de referencia**

#### **8.3.1 Consideraciones generales**

Las características del campo sonoro en la zona de escucha adquieren gran importancia en la percepción subjetiva o la evaluación de la calidad de fenómenos sonoros y su reproducibilidad en otros lugares o salas de escucha. Estas características resultan de la interacción del altavoz o altavoces y la sala de escucha, y toman como referencia la disposición de escucha utilizada (véase el § 8.5).

Actualmente pueden describirse las siguientes características.

#### **8.3.2 Sonido directo**

##### **8.3.2.1 Respuesta en frecuencia del altavoz de control**

La respuesta en frecuencia del altavoz o altavoces, medida en condiciones de campo libre, debe cumplir los requisitos indicados en el § 7.2.2.

#### **8.3.3 Sonido reflejado**

##### **8.3.3.1 Reflexiones primarias**

Las reflexiones primarias ocasionadas por las superficies fronterizas de la sala de escucha que llegan al área de escucha durante un intervalo de tiempo de hasta 15 ms después del sonido directo deberían atenuarse en la gama 1-8 kHz en al menos 10 dB con respecto al sonido directo.

##### **8.3.3.2 Energía residual**

Además de los requisitos especificados para la reverberación y las reflexiones primarias (véase el § 8.2.3), es necesario evitar otras anomalías significativas en el campo sonoro tales como los ecos vibratorios, las coloraciones tonales, etc.

##### **8.3.3.3 Tiempo de reverberación**

(Véase el § 8.2.3.1.)

#### **8.3.4 Campo sonoro estacionario**

##### **8.3.4.1 Curva de respuesta de la sala de prueba**

Las curvas de respuesta de la sala de prueba se definen como las respuestas en frecuencia de un tercio de octava de los niveles de presión sonora producidos por el altavoz o a cada uno de los altavoces de control en la posición de escucha de referencia, utilizando ruido rosa en la gama de frecuencias 50 Hz-16 kHz. Las curvas de respuesta de la sala de prueba medida deben encontrarse dentro de los límites de tolerancia representados en la Fig. 2.

Las diferencias entre las curvas de respuesta de la sala de prueba producidas por cada uno de los altavoces frontales (estereofónico o multicanal) en el punto de escucha de referencia no debe rebasar el valor de 2 dB en toda la gama de frecuencias.

##### **8.3.4.2 Ruido de fondo**

El ruido de fondo continuo (producido por un sistema de aire acondicionado, un equipo interno u otras fuentes exteriores), medido en la zona de escucha a una altura de 1,2 m por encima del suelo no debe rebasar, preferentemente, los valores indicados en la curva NR 10 (véanse las Figs. 3 y 4).

Bajo ninguna circunstancia el nivel de ruido debe rebasar los valores indicados en la curva NR 15.

El ruido de fondo no debe ser de naturaleza tonal, cíclica o impulsiva perceptible.

### **8.4 Nivel de escucha**

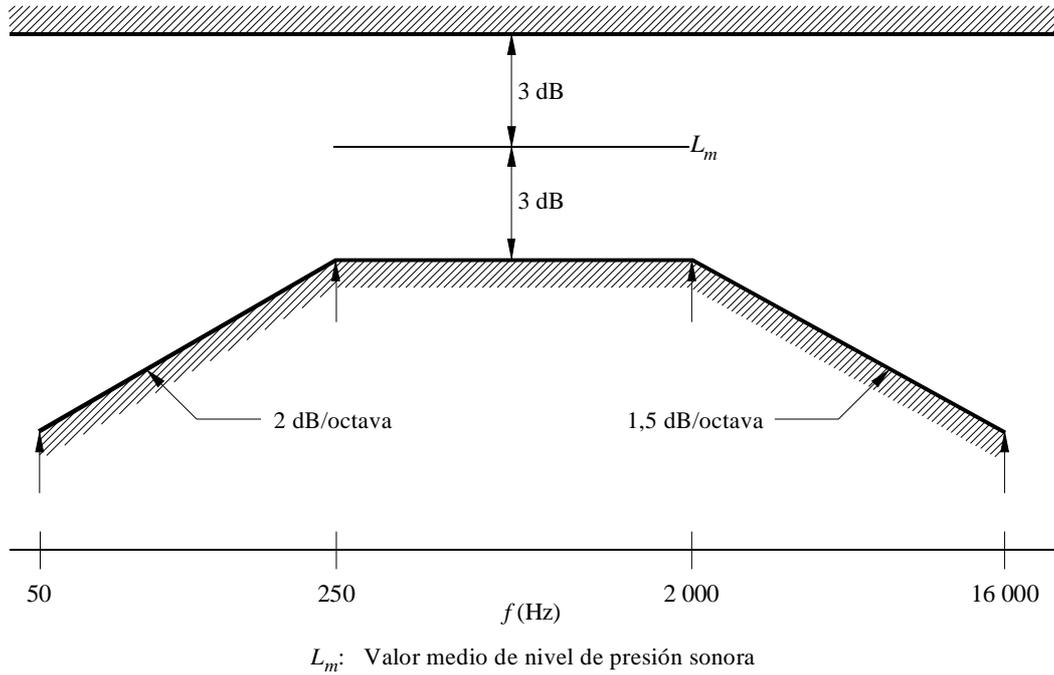
#### **8.4.1 Reproducción por alta voz**

##### **8.4.1.1 Nivel de presión sonora de funcionamiento (nivel de escucha de referencia)**

El nivel de escucha de referencia se define como el nivel de escucha preferido, producido por una señal de medición determinada en el punto de escucha de referencia. Caracteriza la ganancia acústica del canal de reproducción para asegurar que el mismo pasaje de audio produce el mismo nivel de presión sonora en distintas salas de escucha.

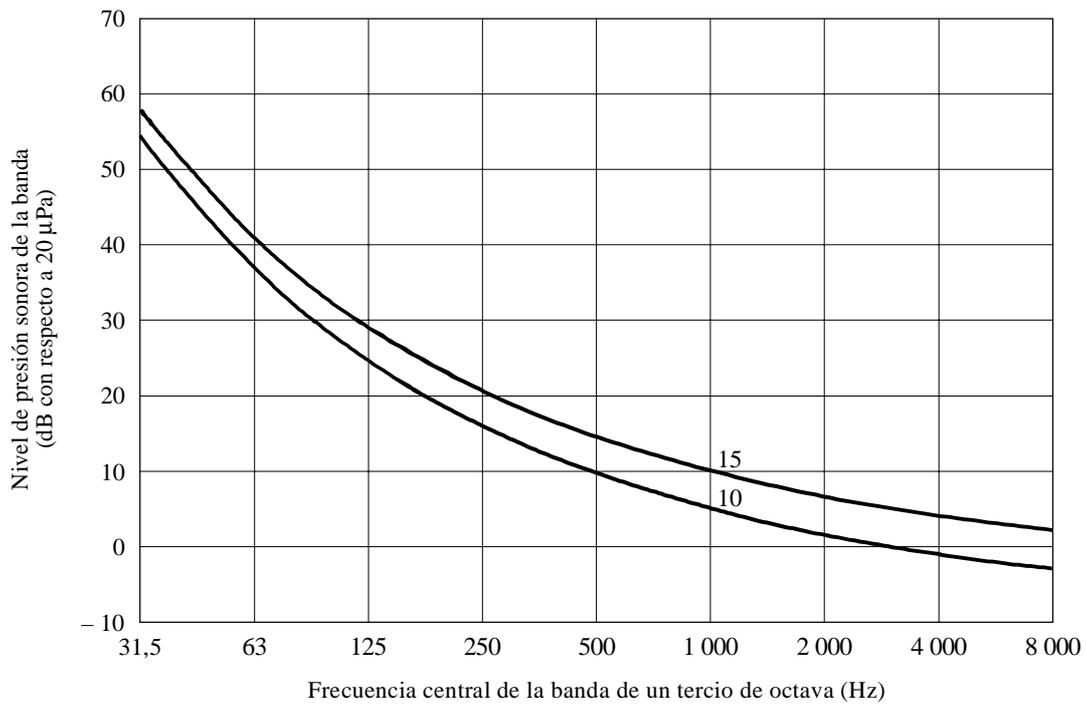
El nivel de alineación de cada uno de los altavoces en una disposición de escucha debe establecerse mediante ruido rosa.

FIGURA 2  
Límites de tolerancia para la curva de respuesta de la sala de prueba



1116-02

FIGURA 3  
Límites del nivel de ruido de fondo en bandas de un tercio de octava.  
Curvas de índice de ruido, basadas en las antiguas curvas NR de la ISO, Recomendación ISO R1996 (1972)

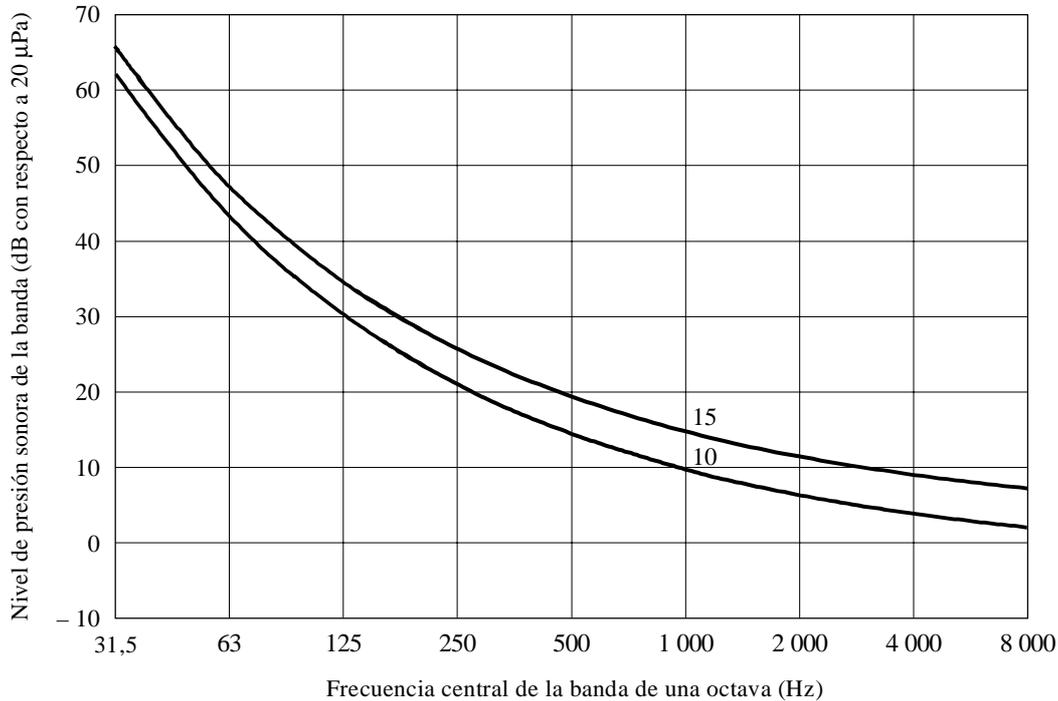


Curvas de índice de ruido NR 10 (recomendada) y NR 15 (máxima)

1116-03

FIGURA 4

**Límites del nivel de ruido de fondo en bandas de una octava.  
Curvas de índice de ruido, basadas en las antiguas curvas NR  
de la ISO, Recomendación ISO R1996 (1972)**



Curvas de índice de ruido NR 10 (recomendada) y NR 15 (máxima)

1116-04

Para una señal de medición con una tensión eficaz igual al «nivel de señal de alineación» (0 dBµ0s de acuerdo con la Recomendación UIT-R BS.645; -18 dB por debajo del nivel de corte de una grabación en cinta digital de acuerdo con [UER, 1992]), aplicada consecutivamente a la entrada de cada canal de reproducción (es decir, un amplificador de potencia y su altavoz asociado), la ganancia del amplificador deberá ajustarse para obtener un nivel de presión sonora de referencia (con ponderación CEI/A, lento) de:

$$L_{ref} = 85 - 10 \log n \pm 0,25 \quad \text{dBA}$$

siendo  $n$  el número de canales de reproducción en el montaje total.

NOTA 1 – Esta hipótesis, de canal de ganancias iguales, puede no ser la adecuada en lo que concierne a ciertos materiales de fuente.

(En secuencias de prueba previas se ha llegado a la conclusión de que los oyentes individuales pueden preferir niveles de escucha absoluta distintos. Si bien no se trata de una opción preferible, no siempre es posible impedir que los participantes tengan tal grado de flexibilidad. Actualmente no se sabe si este hecho afecta la audibilidad de algunos de los estímulos a evaluar. Por consiguiente, si los participantes ajustan la ganancia del sistema, debe tenerse en cuenta esta circunstancia en los resultados de la prueba.)

#### 8.4.2 Reproducción por auricular

El nivel debe ajustarse de forma que se logre una sonoridad igual al campo sonoro de referencia producido por los altavoces. Para determinar esa igualdad en la sonoridad el participante debe situarse en el punto de escucha de referencia.

### 8.5 Disposiciones de escucha

#### 8.5.1 Consideraciones generales

La disposición de escucha describe el emplazamiento de los altavoces y de los lugares de escucha (zona de escucha) en la sala de escucha.

Normalmente las pruebas de escucha se llevarán a cabo en las posiciones de referencia y en otras recomendadas. Sin embargo, también es necesario evaluar todos los efectos debidos a escuchas realizadas fuera del centro de referencia. Por esa razón se incluyen las posiciones de escucha «en caso más desfavorable».

**8.5.1.1 Altura y orientación de los altavoces de control**

La altura de todos los altavoces de control, medida en el centro acústico del altavoz, debe ser de aproximadamente 1,2 m por encima del nivel del suelo. Ello representa la altura del oído de un oyente sentado. La orientación de los altavoces debe ser tal que sus ejes de referencia pasen a través de la posición de referencia a una altura de 1,2 m.

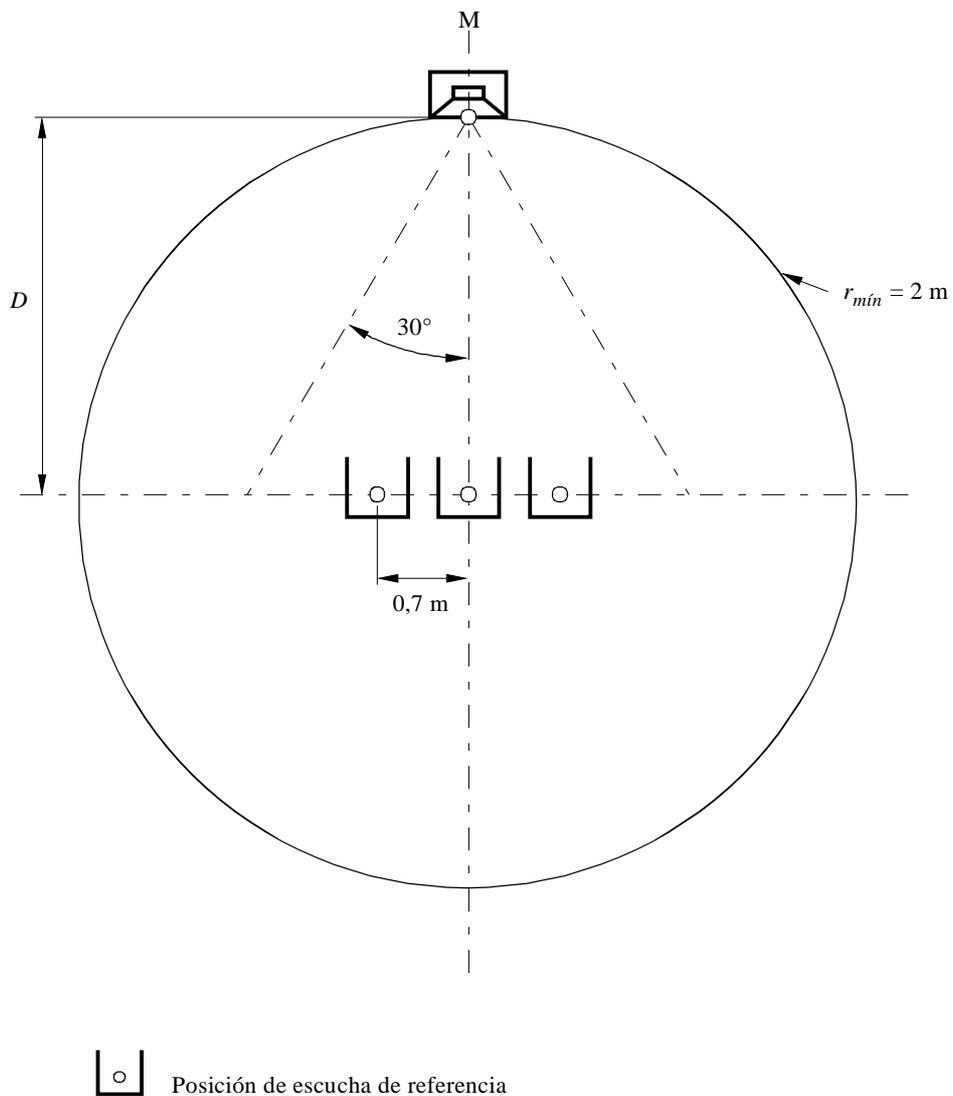
**8.5.1.2 Distancia a las paredes**

En el caso de altavoces no adosados a las paredes, la distancia del centro acústico de un altavoz a las superficies reflectantes circundantes debe ser al menos de 1 m.

**8.5.2 Reproducción monofónica (Fig. 5)**

Para la reproducción de señales monofónicas, debe utilizarse un solo altavoz. La distancia de escucha mínima debe ser de 2 m y todas las posiciones de escucha deben encontrarse dentro de un ángulo de  $\pm 30^\circ$  con respecto al eje del altavoz.

FIGURA 5  
**Disposición de la prueba de escucha de referencia con el altavoz M y zona de escucha permitida para sistemas de sonido monofónico**



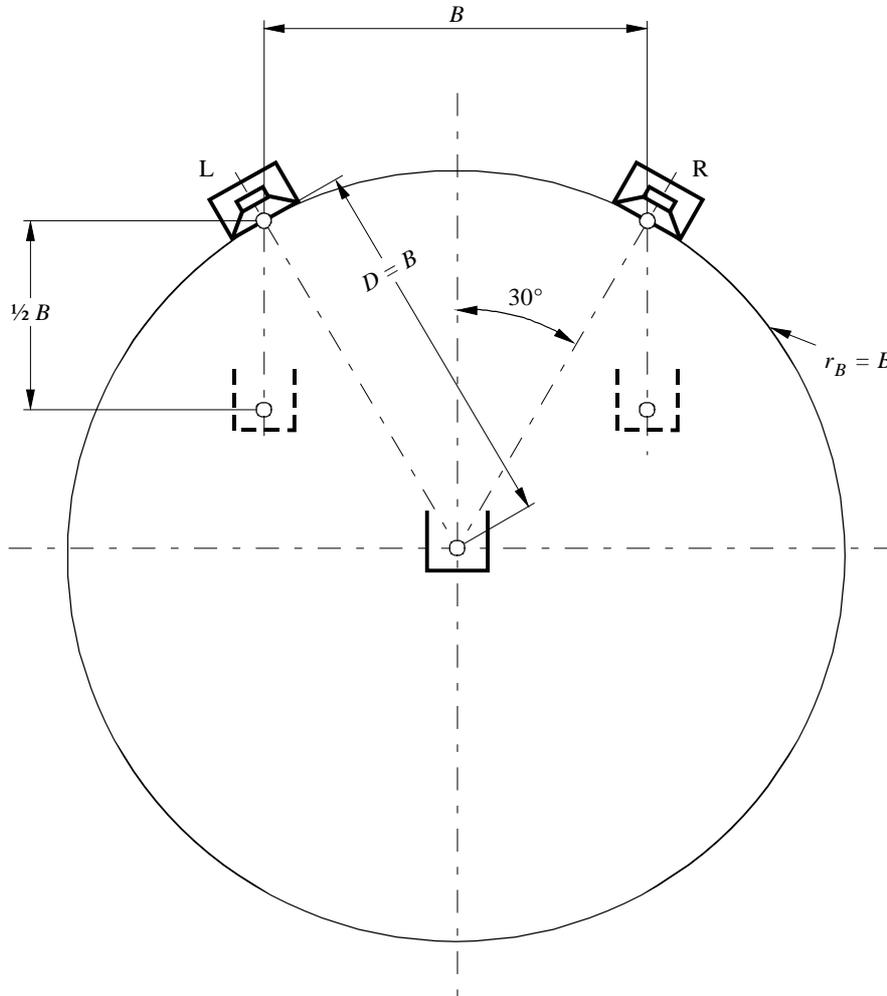
$D$ : Distancia de escucha

1116-05

## 8.5.3 Reproducción estereofónica de dos canales (Fig. 6)

FIGURA 6

Disposición de la prueba de escucha con los altavoces L y R para sistemas de sonido estereofónico con pequeñas degradaciones



 Posición de escucha de referencia

 Posiciones de escucha en caso más desfavorable

$B$ : Distancia entre altavoces  
 $D$ : Distancia de escucha

1116-06

8.5.3.1 Distancia entre altavoces,  $B$ 

Los límites preferidos son  $B = 2$  a  $3$  m. Pueden aceptarse valores de  $B$  de hasta  $4$  m en salas adecuadamente diseñadas.

8.5.3.2 Distancia de escucha,  $D$  (distancia entre el altavoz y el oyente)

Los límites de la distancia de escucha son  $D = 2$  a  $1,7 B$  (m).

**8.5.3.3 Posiciones de escucha**

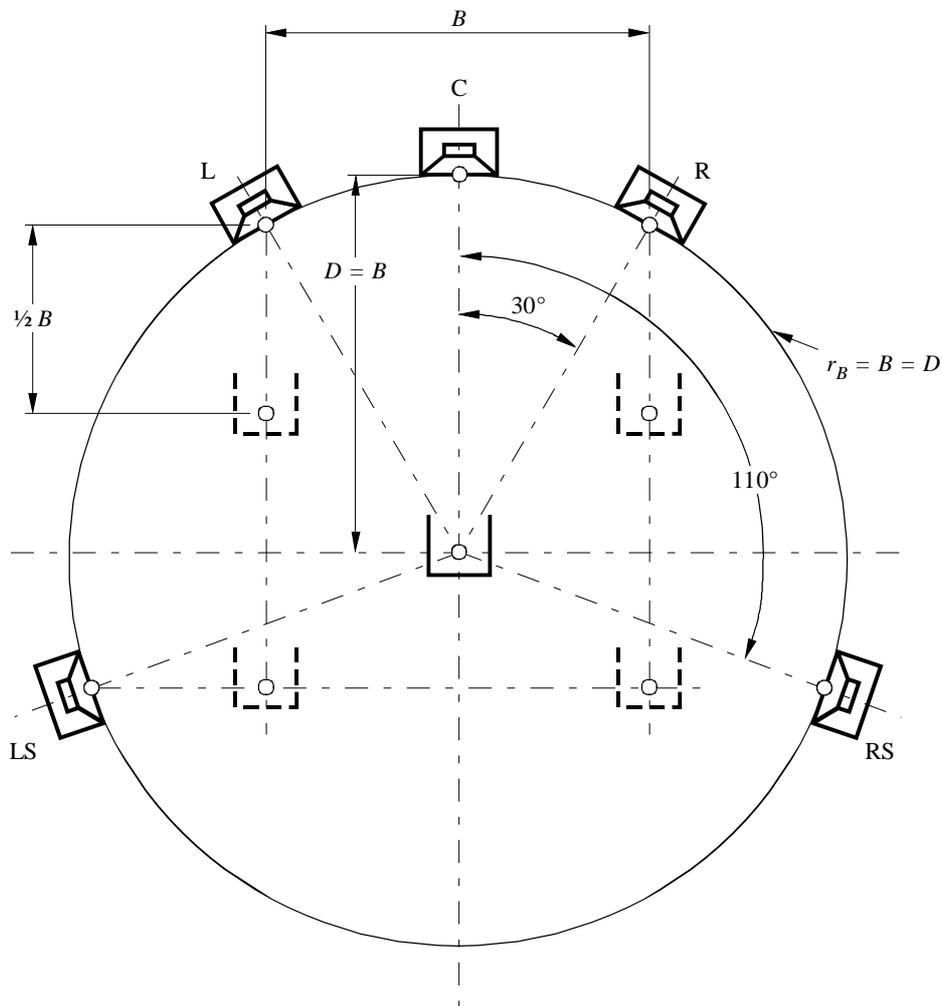
El llamado punto de escucha de referencia se define por el ángulo de escucha de  $60^\circ$ .

La zona de escucha recomendada no debe rebasar el radio de 0,7 m alrededor del punto de escucha de referencia. En la Fig. 6 también se muestran posiciones de escucha adicionales correspondientes al «caso más desfavorable».

**8.5.4 Reproducción estereofónica multicanal (Fig. 7)**

La disposición de escucha debe corresponder en principio al esquema de sonido multicanal 3/2, especificado en la Fig. 1 de la Recomendación UIT-R BS.775 – «Disposición de los altavoces de referencia L/C/R y LS/RS».

FIGURA 7  
**Disposición de la prueba de escucha con los altavoces L/C/R y LS/RS para sistemas de sonido multicanal con pequeñas degradaciones**



 Posición de escucha de referencia

 Posiciones de escucha en caso más desfavorable

$B$ : Distancia entre altavoces  
 $D$ : Distancia de escucha

#### 8.5.4.1 Distancia entre altavoces

Los límites preferidos son  $B = 2$  a 3 m. Pueden aceptarse valores de  $B$  de hasta 5 m en salas adecuadamente diseñadas.

#### 8.5.4.2 Distancia de escucha y ángulo básico

La distancia de escucha de referencia debe ser  $B$  y, por consiguiente, el ángulo básico de referencia debe ser de  $60^\circ$ .

#### 8.5.4.3 Posiciones de escucha

El denominado punto de escucha de referencia se define por el ángulo de escucha de  $60^\circ$ , como se ha indicado anteriormente. En la Fig. 7 también se representan posiciones de escucha adicionales correspondientes al «caso más desfavorable».

## 9 Análisis estadístico

El objeto fundamental del análisis estadístico de los resultados de prueba es identificar con exactitud la calidad de funcionamiento media de cada uno de los sistemas sometidos a prueba y la fiabilidad de cualquier diferencia entre los valores de calidad de funcionamiento media. Este último aspecto obliga a efectuar una estimación de la variabilidad o varianza de los resultados.

Si se han realizado las pruebas de acuerdo con los procedimientos indicados en otros puntos de la presente Recomendación, es probable que la escala presente intervalos regulares; es decir, todos los pasos de la escala de apreciaciones tendrán aproximadamente el mismo tamaño. Sin embargo, la propiedad de la escala obtenida no impide ni exige la utilización de un método estadístico en particular.

Siempre que se cumplan razonablemente los supuestos estadísticos paramétricos básicos, este método es el más sensible y potente y por consiguiente es el que se recomienda. Solamente en el caso de que en importantes propiedades de los datos se detecten desviaciones apreciables con respecto a las hipótesis subyacentes en el modelo de análisis de varianza (ANOVA), habría que considerar otros métodos de análisis (por ejemplo, los no paramétricos). Específicamente, se aconseja aplicar como primera etapa un análisis primario o del modelo ANOVA. Posteriormente, pueden emplearse otros métodos (tales como el de prueba- $t$ , Neuman-Keuls, Scheffe, etc.) que utilizan estimaciones de la varianza proporcionadas por el ANOVA para estudiar más detalladamente dónde se encuentran los efectos globales más significativos (de haberlos) que revela el ANOVA.

A menudo, una hipótesis específica puede ser validada por varios métodos estadísticos distintos. Evidentemente, la base para llegar a una decisión puede reforzarse si una hipótesis particular también se cumple para una validación con un método estadístico alternativo. Por consiguiente, se sugiere aplicar un análisis de datos suplementario (tales como el Wilcoxon, etc.).

También es importante considerar en algún instante los aspectos psicométricos, que indudablemente tienen cierta influencia sobre el tipo de conclusiones significativas que pueden obtenerse a partir de una escala no física.

Hay que señalar que, a menos de que pueda demostrarse que la escala de apreciación es lineal, sólo pueden hacerse comparaciones entre las diferentes apreciaciones basándose en el orden de las notas.

## 10 Presentación de los resultados de los análisis estadísticos

### 10.1 Consideraciones generales

La presentación debe realizarse de forma que tanto los lectores expertos como los inexpertos puedan evaluar la información correspondiente. En principio, todo lector desea conocer los resultados globales del experimento, preferiblemente en forma gráfica. Tal presentación puede esforzarse con información cuantitativa más precisa, si bien los análisis numéricos detallados deben aparecer en Apéndices.

## 10.2 Notas absolutas

La presentación de las notas medias absolutas, para el objeto y la referencia oculta por separado, puede proporcionar una impresión inicial bastante acertada de los datos.

Sin embargo, debe tenerse en cuenta que esto no constituye una base adecuada para realizar un análisis estadístico detallado, debido al hecho de que cuando se utiliza el método de prueba recomendado en este texto, los participantes saben explícitamente que una de las fuentes en la comparación por pares es idéntica a la referencia. En consecuencia, las observaciones no son independientes y el análisis estadístico de estas notas absolutas no aporta una información significativa y, por lo tanto, no debe realizarse.

## 10.3 Notas distintas

La diferencia entre las notas otorgadas a la referencia oculta y al objeto es el punto de partida adecuado para efectuar los análisis estadísticos. Una representación gráfica revela claramente las distancias reales a la transparencia, que normalmente tienen gran interés.

## 10.4 Nivel de significación e intervalo de confianza

El informe de la prueba debería proporcionar al lector información acerca de la naturaleza inherentemente estadística de todos los datos subjetivos. Habría que explicitar los niveles de significación, así como otros detalles acerca de los métodos y resultados estadísticos que contribuyan a dar una idea más clara al lector. Dichos detalles podrían incluir los intervalos de confianza o las barras de error en los gráficos.

Evidentemente, no existe un nivel de significación «correcto». Sin embargo, tradicionalmente se elige el valor de 0,05. En principio, es posible utilizar una prueba de una cola o de dos colas según las hipótesis que se estén verificando.

## 11 Contenido de los informes de prueba

Los informes de prueba deben indicar, de la manera más clara posible, los motivos fundamentales del estudio, los métodos utilizados y las conclusiones extraídas. Deben presentarse detalles suficientes como para que, en principio, una persona con ciertos conocimientos pueda repetir el estudio a fin de verificar de forma empírica los resultados. Un lector informado debe ser capaz de entender e interpretar los detalles más importantes de la prueba así como las razones fundamentales para el estudio, los métodos de diseño del experimento y su ejecución y los análisis y sus conclusiones.

Debe prestarse especial atención a los puntos siguientes:

- la especificación y selección de los participantes y pasajes;
- los detalles físicos de los equipos y del entorno de escucha, incluidas las dimensiones de la sala y las características acústicas, los tipos y emplazamiento de los transductores y la especificación de los equipos eléctricos;
- el diseño del experimento, el adiestramiento, las instrucciones, las secuencias y procedimientos de prueba y la generación de datos;
- el procesamiento de los datos, incluyendo los detalles de las estadísticas inferenciales descriptivas y analíticas;
- las bases detalladas de todas las conclusiones extraídas.

## REFERENCIAS BIBLIOGRÁFICAS

POULTON, E.C. [1992] Bias in quantifying judgments. Lawrence Erlbaum Associates, Hillsdale, Estados Unidos de América.

UER [1992] Recomendación R-68. Niveau d'alignement dans les équipements numériques de production du son ainsi que dans les magnétoscopes numériques. Unión Europea de Radiodifusión, Ginebra, Suiza.

## Consideraciones estadísticas de la selección posterior de los participantes

### 1 Evaluación de la experiencia de los participantes

El método del triple estímulo doblemente ciego con referencia oculta proporciona dos notas en cada experimento y hace posible que cada participante, de forma individual, compare directamente estas dos notas y pueda realizarse un examen de estas comparaciones para todos los experimentos de dicho individuo. En cada experimento puede calcularse la diferencia algebraica entre las dos apreciaciones siempre, evidentemente, restando en el mismo sentido. Supóngase que se resta la nota para la referencia oculta de la nota para el objeto.

Si el participante no ha podido, de forma general, identificar correctamente la referencia oculta con respecto al objeto, la media de todas las notas distintas de dicho participante en la prueba de escucha debe ser cero o próxima a dicho valor, puesto que existirán notas positivas y negativas que tienden a equilibrarse unas a otras al calcular la media. Si el participante ha sido capaz de detectar correctamente la referencia oculta del objeto, la media de las notas diferentes se desviará de cero en sentido negativo, puesto que la mayoría de las notas serán negativas en vez de positivas.

Los datos tenidos de este modo se someten a una prueba- $t$  unilateral, para evaluar la probabilidad de que la media de la distribución para cada participante sea 0. Si esta hipótesis nula se rechaza para un determinado participante, puede concluirse que los datos correspondientes al mismo dimanar de una distribución con una media superior a 0 en el sentido negativo, a un determinado nivel de confianza. Cabe concluir, por consiguiente, que cada participante respecto del cual esto no sea cierto ha demostrado que, en general, no estaba simplemente conjeturando; en ese caso podrá decirse que estos participantes han demostrado una experiencia lo suficientemente amplia como para justificar la inclusión de sus datos en el análisis final de los resultados experimentales. Los datos de los demás participantes, es decir, aquellos que, en general, se encontraban conjeturando en base a este criterio estadístico, pueden descartarse del análisis ulterior.

Hay que recordar que las recomendaciones formuladas en este texto versan exclusivamente sobre pequeñas degradaciones. Si se demuestra, por cualquier razón que ello sea, que se ha incluido un gran número de «grandes» degradaciones en una prueba, y no así sólo «pequeñas» degradaciones, el método de postselección aplicado ingenuamente, según se describe anteriormente, puede llevar a obtener conclusiones erróneas o inadecuadas. En este contexto, por «gran» degradación se entiende una degradación que será relativamente fácil de detectar, incluso por oyentes «no expertos». Huelga decir que unas cuantas degradaciones auténticamente «pequeñas» (dificiles de detectar), incorporadas en un contexto en que la mayoría de las degradaciones son «grandes» (fáciles de detectar), tendrán poco peso en una prueba- $t$  como la descrita anteriormente. Por esta razón, los expertos que juzguen correctamente los elementos de pequeña degradación pueden ser indistinguibles en cuanto a su desempeño general de los no expertos que actúan a niveles de «conjetura» con respecto a estos elementos. Esto es así, debido a que en las evaluaciones de la prueba- $t$  el desempeño relativo a los elementos de pequeña degradación pueden diluirse en el ruido estadístico, ya que el mayor peso, en lo que respecta a la magnitud de  $t$ , vendría dado por los elementos de gran degradación.

Incluso en el caso de pruebas de «pequeñas degradaciones», es prácticamente inevitable detectar algunos elementos de gran degradación o fácil detección, aun cuando éstos no constituyan ni con mucho la mayoría de los elementos. Habida cuenta de lo anterior, se recomienda que, con el propósito exclusivo de efectuar pruebas- $t$  de selección posterior suficientemente rigurosas, todos los elementos «de fácil» detección o gran degradación se excluyan normalmente del procedimiento de la prueba- $t$  para evaluar la experiencia de los oyentes. Estos elementos, podrían ser todos los que han recibido notas medias bajas de los participantes, por ejemplo en un intervalo de  $-2,0$  a  $-4,0$ . Tratándose de dichos elementos, la mayoría de los sujetos distinguirán correctamente el objeto de la referencia oculta y la inclusión de dichos elementos en la prueba- $t$  obstaculizará en lugar de facilitar la evaluación de las diferencias de grado en la experiencia de los participante. Dejar elementos de gran degradación en los análisis de la prueba- $t$  conduciría a exagerar o sobrestimar la experiencia aparente de los participantes.

El caso opuesto ocurre cuando existe un número excesivo de elementos «auténticamente transparentes» (véase el § 5 de esta Recomendación). De ser así, habría que omitir los elementos aparentemente transparentes («demasiado difíciles») en las pruebas- $t$  de selección posterior. A continuación, los elementos especiales introducidos por su carácter crítico conocido tendrían más peso en las pruebas- $t$ , que es lo que se pretende. Los efectos para la prueba- $t$  resultantes de dejar en la misma los elementos aparentemente transparentes conducirían a subestimar la experiencia de los participantes.

En general, los elementos que son permanentemente «demasiado difíciles» o «demasiado fáciles» tienen un carácter no diferencial en lo que respecta a distinguir los expertos adecuados de los no idóneos.

La única ventaja de aplicar correctamente las pruebas- $t$  de selección posterior es que en nivel suficiente de experiencia para un experimento dado se evalúa sobre la base del desempeño de los participantes en dicho experimento. En una serie de experimentos en los que los mismos participantes intervienen en diferentes experiencias, basándose en los resultados de la selección posterior, puede llegarse a la conclusión de que, aunque todos los participantes hayan superado con éxito la preselección, algunos de ellos podrían tener la experiencia adecuada en relación con un subconjunto de los experimentos pero no así para todos ellos. Por consiguiente, en esos casos es posible aceptar o, en su caso, rechazar los datos de un determinado participante para una serie de resultados concretos de la prueba. Esto contribuye, en efecto, a refinar el concepto de «experiencia» más allá de lo que es posible si se recurre únicamente a la selección previa.

No obstante, habría que estar atentos a una cuestión. Un participante no suficientemente experto no puede aportar datos de calidad. En consecuencia, se justifica rechazar sus datos, aduciendo un grado insuficiente de experiencia determinado objetivamente mediante una selección previa rigurosa. Ahora bien, no es posible garantizar que los datos de un participante que haya superado adecuadamente la posterior selección en una prueba- $t$  sean necesariamente datos de calidad. Como ejemplo extremo de lo anterior, puede suceder que un participante distinga correctamente objetos de referencias ocultas en el 100% de las pruebas de un experimento y que, sin embargo, los datos revelen que dicho participante ha dado una nota de 1,0 a todos los objetos en todas las pruebas. Dicho de otro modo, el conjunto total de datos obtenidos a partir de ese participante pueden entrañar notas de discriminación de  $-4,0$  para todas las pruebas.

Suponiendo que a los demás participantes en el experimento corresponda una distribución «más usual» de notas en todas las pruebas, la pauta muy singular de respuesta del participante precitado (todas las notas de discriminación « $-4,0$ ») podrían llevar a la conclusión de rechazar tales datos. Sin embargo, excepto quizá un caso que supusiera claramente un desvío considerable, como el descrito aquí con propósitos de ilustración, resultaría muy difícil aplicar tales criterios post hoc a la aceptabilidad de los datos. Esto equivaldría a conformar deliberadamente los datos con arreglo a las preconcepciones del experimentador, en lugar de aceptar la evidencia empírica de los resultados reales.

Hay que señalar que NO deben utilizarse dichos métodos post hoc. Mientras el número total de participantes en un experimento sea el adecuado, incluso los datos de un participante experto que se desvíen en gran medida de la media tendrán un efecto de distorsión muy reducido en el conjunto total de datos. Es muy frecuente que se obtengan resultados significativos y duplicables después de realizar experimentos de sensibilidad, aun cuando en éstos hayan intervenido participantes atípicos pero expertos. Una vez concluido un experimento, si hay sospechas acerca del carácter fiable de los datos, la única solución consistirá en repetir íntegramente el experimento, utilizando un nuevo grupo de participantes y procurar corregir las insuficiencias sospechadas en los procedimientos experimentales aplicados anteriormente.

## 2 Evaluación ulterior de la experiencia de los oyentes

Al aumentar la calidad de los códecs que se han percibido como deficientes, quedarán inevitablemente menos oyentes con el suficiente grado de experiencia para discernir la calidad de los dispositivos de codificación restantes. Un oyente que tenga la experiencia suficiente para participar en una prueba que incluya aparatos relativamente «fácilmente audibles» puede no contar con experiencia suficiente en una prueba en que estos aparatos más audibles no se encuentren presentes. Por otra parte, si bien la nota  $t$  de un oyente puede indicar que su experiencia es suficiente para el experimento, considerado éste en su totalidad, es posible que dicha experiencia no baste para detectar diferencias entre la señal de referencia y una señal codificada de muy elevada calidad. En este caso, los datos del participante añadirían «ruido estadístico» a los datos totales, lo que encubriría las diferencias reales percibidas por otros participantes.

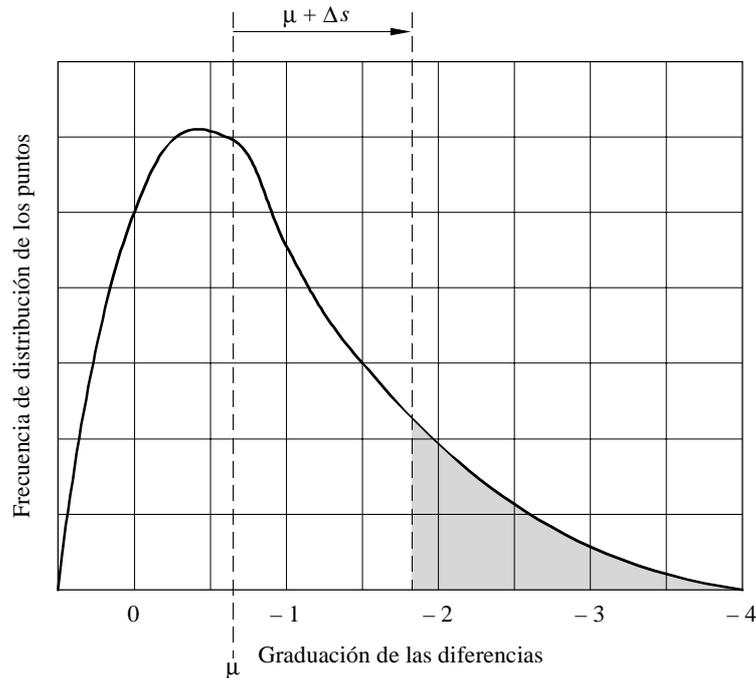
### APÉNDICE 2

#### AL ANEXO 1

## Evaluación del nivel de experiencia de un participante

En la actualidad, todos los datos de un sujeto en una determinada prueba se utilizan para evaluar su nota  $t$ . Los datos de todos los sujetos con notas  $t$  suficientemente altas se incluyen, a continuación, en el ANOVA.

FIGURA 8

Método para descartar puntos de datos con anterioridad a la prueba- $t$ 

1116-08

En la presente propuesta sugerimos que se realicen pruebas- $t$  sucesivas para subconjuntos de los datos de cada participante. En cada una de estas pruebas el criterio de evaluación del nivel de experiencia de un participante debería ser más estricto que en la anterior.

Habría que reevaluar el nivel de experiencia de un participante y en caso de que éste haya demostrado un nivel suficiente de experiencia, sus datos se incluirían en el ANOVA siguiente. En consecuencia, en cada prueba se hace más estricto el criterio de experiencia suficiente y el siguiente ANOVA se realiza con los datos de los participantes restantes. En lo que sigue se esbozan los criterios propuestos para evaluar la experiencia.

El procedimiento que debe seguirse se indica en el Fig. 8 para un conjunto hipotético de datos. En primer lugar, se calcula la media y la desviación típica para los datos del participante. Acto seguido, basándose en estas magnitudes estadísticas, se determinarán las correspondientes notas  $z$  (véase la Nota 1) correspondientes a los datos del participante. Todos los puntos de datos de un participante que sobrepasen un cierto criterio ( $\mu + \Delta 1 s$ ) se descartarán y se procederá a realizar una nueva prueba- $t$  en relación con los puntos de datos restantes. Como se indica en la Figura, los puntos de datos que sobrepasan  $\mu + \Delta 1 s$  (el área sombreada) se descartan y los puntos de datos restantes (el área no sombreada) se utilizan en la prueba- $t$  siguiente. En caso de que, para los puntos de datos restantes, la prueba- $t$  indique que el participante sigue contando con la experiencia suficiente, sus datos se incluirían en el ANOVA siguiente. Si el participante no puede demostrar una experiencia suficiente en la prueba- $t$ , sus datos se eliminarían en su totalidad de todos los ANOVAS posteriores. Este proceso se repetirá, aplicándose cada vez un criterio de experiencia más estricto,  $\mu + \Delta 2 s$ . El proceso se repetirá  $N$  veces con los criterios  $\mu + \Delta i s$  ( $i = 0, 1, \dots, N$ ). Se están investigando los valores adecuados de  $\Delta i s$  y  $N$ , basándose en los datos de los estudios realizados anteriormente en el Communications Research Center (Canadá).

NOTA 1 – La nota  $z$  representa la nota normalizada para una distribución con media cero y desviación típica 1. Se define como  $z = \frac{x - \mu}{s}$ , donde  $x$  es un punto de datos,  $\mu$  es la media de la muestra y  $s$  la desviación típica para la muestra:

$$s = \sqrt{\frac{N \sum x^2 - (\sum x)^2}{N(N - 1)}}$$

## APÉNDICE 3

## AL ANEXO 1

**Ejemplo de instrucciones a los participantes**

La terminología utilizada en estas instrucciones no se ajusta exactamente a las definiciones del glosario.

**1 Fase de adiestramiento o familiarización**

El objeto de la fase de adiestramiento es permitir a los oyentes familiarizarse con las posibles distorsiones y degradaciones producidas en los sistemas sometidos a prueba y poder identificarlas. Después del adiestramiento usted debe saber «qué es lo que tiene que escuchar». Esta tarde se le solicitará que evalúe a ciegas todo el material de audio que va a escuchar esta mañana. Durante esta fase también se familiarizará con el procedimiento de prueba.

Escuchará las versiones de referencia (original) y procesada de cada material de audio. En la pantalla de vídeo la versión de referencia se identificará con la letra «A» y la versión procesada de la señal y la «referencia oculta» por las letras «B» y «C». Puede conmutar libremente entre «A», «B» o «C» en cualquier instante durante la presentación. Con ello logrará una comparación precisa y detallada entre «A», «B» y «C». Debe evaluar las diferencias que encuentre entre «A» y «B» y entre «A» y «C». Las secuencias de audio durarán normalmente entre 10 y 25 s y puede escucharlas repetidamente todo el tiempo que desee. Durante el adiestramiento puede utilizar altavoces, auriculares o ambos. Dispone de tres horas para familiarizarse con todos los pasajes sobre los que deberá emitir un juicio formal en la fase de evaluación ciega que se celebrará esta tarde.

Durante las pruebas de la tarde se le solicitará que califique las presentaciones de acuerdo con la escala del Cuadro 2.

CUADRO 2

Degradación	Nota
Imperceptible	5,0
Perceptible, pero no molesta	4,0
Ligeramente, molesta	3,0
Molesta	2,0
Muy molesta	1,0

Debe indicarse a los participantes el significado de la escala resaltando que la escala se considera como una escala continua de intervalos iguales en la que se señalan unos puntos de referencia definidos para unos valores específicos.

Como cada experimento de la tarde contiene una referencia oculta (es decir, una réplica perfecta de la referencia), cabe esperar al menos una nota de 5,0 (pero sólo una (véase la Nota 1)) en cada uno de ellos. Si considera que «B» o «C» es mejor que la referencia, ello supone que ha encontrado una diferencia «perceptible, pero no molesta» y deberá otorgar una nota comprendida entre 4,0 y 4,9 de acuerdo con la diferencia detectada.

Durante la fase de adiestramiento debe considerar individualmente la forma de interpretar las degradaciones audibles en términos de la escala de notas; es importante, por consiguiente, que no discuta en ningún instante esta interpretación de tipo personal con el resto de los participantes.

NOTA 1 – Con el cambio recomendado se pretende obligar al participante a hacer una «mejor conjetura» con respecto al estímulo que representa el material codificado. Estimamos que ciertos participantes pueden detectar en la actualidad estímulos muy pequeños, pero que, debido a su enfoque conservador, darán dos notas de 5,0, en lugar de comprometerse. El cambio recomendado resolvería este problema.

## 2 Ejemplo del contenido de una etapa de adiestramiento

El adiestramiento básico debería impartirse a grupos de unos cuatro participantes en la mañana del primer día con una duración de hasta tres horas. Habría que enviar por adelantado instrucción a los participantes.

La sesión de adiestramiento debería incluir los siguientes puntos:

- Una breve introducción con los motivos y la finalidad de la prueba.
- Una grabación de los pasajes de la prueba seleccionados para que los participantes se familiaricen con la presentación del sonido y tomen conocimiento del material de programa que evaluarán posteriormente.
- Una breve explicación de los sistemas objeto de prueba y del tipo de degradaciones, así como una presentación oral de las categorías de degradación definidas por el panel de selección previa.
- Una demostración de las degradaciones, para lo cual se utilizarán algunos de los elementos más degradados.
- Una explicación del atributo que habrá que apreciar cuantitativamente.
- Una explicación de la escala de cinco notas para las degradaciones.
- El adiestramiento sobre conmutación y apreciación.

En los días siguientes a la prueba, habría que recordar a los participantes los puntos examinados en la sesión básica de adiestramiento. Esto podría incluir escuchar una vez más los elementos de la prueba, antes de pasar a las experiencias formales.

## 3 Fase de apreciación ciega

El objeto de la prueba ciega es evaluar el diverso material de audio que escuchará esta mañana durante la fase de adiestramiento.

En cada prueba oírás tres versiones de un material de audio determinado, denominadas «A», «B» y «C» en la pantalla de vídeo. «A» siempre es la versión de referencia (*original*) con la que se comparan y evalúan «B» y «C». «B» o «C» es la versión procesada y la otra es la referencia oculta (idéntica a la referencia). No se le va a decir cuál de ellas es la versión procesada y cuál la referencia oculta, de ahí la aplicación del término «ciego» para esta fase de apreciación. Podrá conmutar libremente entre «A», «B» o «C» en cualquier instante. Las secuencias de audio pueden repetirse indefinidamente hasta que usted esté seguro de sus evaluaciones. Cuando lo desee puede pasar al experimento siguiente, una vez que quede satisfecho con la evaluación del experimento en curso.

En cada experimento se solicita que evalúe la diferencia percibida (caso de existir) entre «B» y «A», por un lado, y la diferencia entre «C» y «A», por el otro, utilizando la escala de cinco notas indicada en el Cuadro 3. Por consiguiente, en cada experimento deben otorgarse dos notas, una para «B» y otra para «C». Cabe esperar al menos una nota de 5,0 (pero sólo una (véase la Nota 1 del § 1 de este Apéndice)) en cada experimento. Se le ruega que introduzca sus notas en el ordenador al final de cada experimento.

En vez de introducir las notas en el ordenador puede utilizarse una hoja de papel graduado.

Debe mostrarse a los participantes el Cuadro 3 y debe haber disponible una copia a lo largo de todas las sesiones de evaluación ciega.

Debe indicarse a los participantes el significado de la escala, poniendo de relieve que la escala se considera como una escala continua de intervalos iguales en la que se señalan unos puntos de referencia definidos para unos valores específicos.

CUADRO 3

Degradación	Nota
Imperceptible	5,0
Perceptible, pero no molesta	4,0
Ligeramente, molesta	3,0
Molesta	2,0
Muy molesta	1,0

## APÉNDICE 4

## AL ANEXO 1

**Evaluación subjetiva: Glosario**

A continuación se definen los términos utilizados en esta Recomendación. Véase también la Fig. 9 que ilustra la interrelación entre algunos de estos términos.

**Atributo**

Característica percibida de un elemento de audición de acuerdo con una definición determinada oral o escrita.

**Elemento**

Pasaje procesado por el sistema sometido a prueba.

**Emplazamiento**

Lugar donde se lleva a cabo la prueba de escucha. Puede ser simplemente el emplazamiento geográfico o la posición del participante en la sala de escucha. Puede ser uno de los factores en la prueba.

**Estímulo**

Combinación del objeto y de la referencia oculta o de la referencia y la totalidad o parte de un pasaje.

**Experimento**

Partes en que se divide una sesión; comienza con la presentación de un conjunto de estímulos y finaliza con su evaluación.

**Grupo de oyentes**

Conjunto de participantes que producen los datos de una prueba de escucha.

**Nota**

Expresión numérica de la magnitud de un atributo de acuerdo a una escala determinada.

**Objeto**

Sistema sometido a prueba, representado por un número de pasajes procesados por el sistema sometido a prueba.

**Participante**

Individuo que evalúa los estímulos en una prueba de escucha.

**Pasaje**

Muestra de una pieza de música, de conversación o de otro elemento sonoro, adecuada para evaluar las características individuales o los parámetros de la calidad del sonido en un sistema determinado sometido a prueba.

Los pasajes de prueba son normalmente grabaciones de sonido (CD, R-DAT u otros formatos fuente o de grabación).

**Prueba ciega**

Prueba en la que la única fuente de información que se da al participante en el experimento son los estímulos.

**Prueba doblemente ciega**

Prueba ciega en la que no hay posibilidad de que se produzcan interacciones incontroladas entre el experimentador y la prueba de escucha.

**Referencia**

Pasaje de prueba, reproducido sin el procesamiento por un objeto de prueba, utilizado como base de comparación para una prueba de degradación.

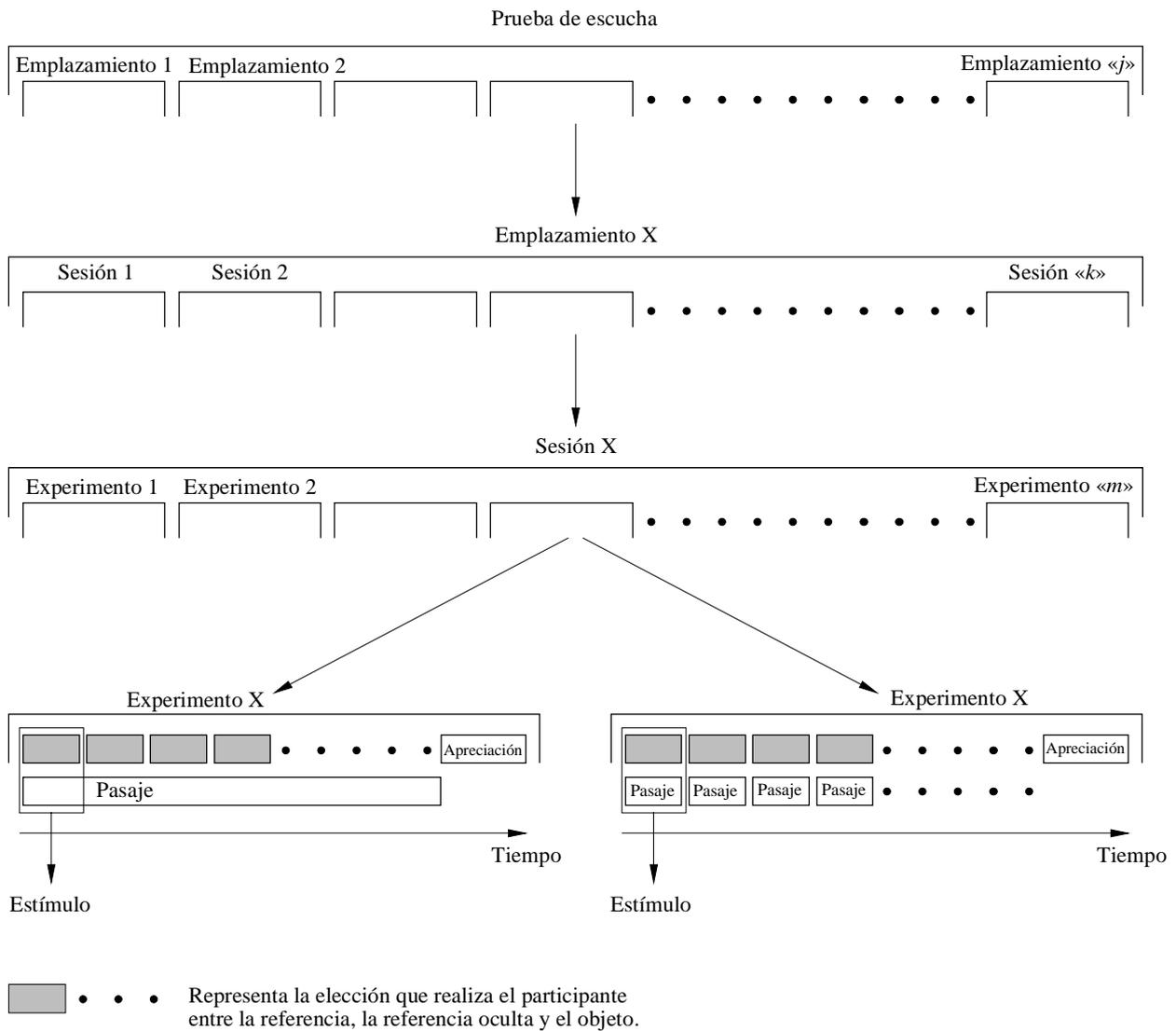
**Referencia oculta**

Referencia no identificada por el participante en la prueba.

**Sesión**

Conjunto total de experimentos que debe evaluar un participante o un grupo de escucha en un periodo continuado.

FIGURA 9  
Representación de la relación entre algunos de los términos utilizados en el glosario



Los dos experimentos mostrados son los puntos extremos de una gama de posibles disposiciones.