

UIT-R

Sector de Radiocomunicaciones de la UIT

Recomendación UIT-R BS.2132-0
(10/2019)

**Método para la evaluación subjetiva de la
calidad de las diferencias audibles de los
sistemas sonoros utilizando múltiples
estímulos sin una referencia dada**

Serie BS
Servicio de radiodifusión
(sonora)



Prólogo

El Sector de Radiocomunicaciones tiene como cometido garantizar la utilización racional, equitativa, eficaz y económica del espectro de frecuencias radioeléctricas por todos los servicios de radiocomunicaciones, incluidos los servicios por satélite, y realizar, sin limitación de gamas de frecuencias, estudios que sirvan de base para la adopción de las Recomendaciones UIT-R.

Las Conferencias Mundiales y Regionales de Radiocomunicaciones y las Asambleas de Radiocomunicaciones, con la colaboración de las Comisiones de Estudio, cumplen las funciones reglamentarias y políticas del Sector de Radiocomunicaciones.

Política sobre Derechos de Propiedad Intelectual (IPR)

La política del UIT-R sobre Derechos de Propiedad Intelectual se describe en la Política Común de Patentes UIT-T/UIT-R/ISO/CEI a la que se hace referencia en la Resolución UIT-R 1. Los formularios que deben utilizarse en la declaración sobre patentes y utilización de patentes por los titulares de las mismas figuran en la dirección web <http://www.itu.int/ITU-R/go/patents/es>, donde también aparecen las Directrices para la implementación de la Política Común de Patentes UIT-T/UIT-R/ISO/CEI y la base de datos sobre información de patentes del UIT-R sobre este asunto.

Series de las Recomendaciones UIT-R

(También disponible en línea en <http://www.itu.int/publ/R-REC/es>)

Series	Título
BO	Distribución por satélite
BR	Registro para producción, archivo y reproducción; películas en televisión
BS	Servicio de radiodifusión (sonora)
BT	Servicio de radiodifusión (televisión)
F	Servicio fijo
M	Servicios móviles, de radiodeterminación, de aficionados y otros servicios por satélite conexos
P	Propagación de las ondas radioeléctricas
RA	Radio astronomía
RS	Sistemas de detección a distancia
S	Servicio fijo por satélite
SA	Aplicaciones espaciales y meteorología
SF	Compartición de frecuencias y coordinación entre los sistemas del servicio fijo por satélite y del servicio fijo
SM	Gestión del espectro
SNG	Periodismo electrónico por satélite
TF	Emisiones de frecuencias patrón y señales horarias
V	Vocabulario y cuestiones afines

Nota: Esta Recomendación UIT-R fue aprobada en inglés conforme al procedimiento detallado en la Resolución UIT-R 1.

Publicación electrónica
Ginebra, 2020

© UIT 2020

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

RECOMENDACIÓN UIT-R BS.2132-0

Método para la evaluación subjetiva de la calidad de las diferencias audibles de los sistemas sonoros utilizando múltiples estímulos sin una referencia dada

(2019)

Alcance

En esta Recomendación se describe un método que utiliza múltiples estímulos sin una referencia dada para la evaluación subjetiva de la calidad de las diferencias audibles de los sistemas de audio. Este método refleja muchos aspectos del método MUSHRA especificado en la Recomendación UIT-R BS.1534 pero, a diferencia de la Recomendación UIT-R BS.1534, amplía la evaluación de los sistemas para incluir condiciones en las que no se dispone de referencias y patrones ocultos conocidos.

Palabras clave

Prueba de escucha, calidad de audio, sistemas de sonido avanzados, evaluación subjetiva, evaluación percibida

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que en las Recomendaciones UIT-R y UIT-T existen muchas metodologías de pruebas subjetivas para evaluar la calidad subjetiva de los sistemas de audio, vídeo y voz;
- b) que la utilización de métodos subjetivos de ensayo normalizados es importante para el intercambio, la compatibilidad y la correcta evaluación de los datos de prueba;
- c) que se procura utilizar métodos de ensayo normalizados para la evaluación de sistemas de sonido avanzados;
- d) que en algunas aplicaciones no se dispone de señales de referencia o éstas no son apropiadas, de manera que no puede realizarse la evaluación de la calidad subjetiva de los sistemas de sonido en relación con una señal conocida y, en cambio, debe hacerse sin una referencia;
- e) que el proceso de producción de programas requiere el uso de sistemas tecnológicos para crear señales de audio y expresar una intención creativa, y que en esos casos puede haber condiciones en las que no se disponga de una señal de referencia o un comportamiento de sistema objetivo;
- f) que la introducción de sistemas de sonido avanzados, según se describe en la Recomendación UIT R BS.2051, proporciona nuevas herramientas para la expresión creativa en la producción y requiere nuevos métodos de evaluación subjetiva, incluidos los métodos para la asociación de los atributos de percepción a la calidad de audio percibida general,

recomienda

que se utilicen los procedimientos de prueba y evaluación que figuran en el Anexo 1 a la presente Recomendación para la evaluación subjetiva de las diferencias auditivas de los sistemas de audio, cuando no se disponga de una señal de referencia o de una referencia de sistema apropiadas.

Anexo 1

1 Introducción

Las pruebas de escucha subjetiva son una forma fiable de medir la calidad de percepción de los sistemas de audio. Existen métodos debidamente descritos y probados para evaluar la calidad de audio en un contexto de radiodifusión cuando se comparan los sistemas con una referencia conocida y no degradada, tanto a niveles de calidad altos como intermedios. La Recomendación UIT-R BS.1116 – Métodos para la evaluación subjetiva de pequeñas degradaciones en los sistemas de audio incluyendo los sistemas de sonido multicanal, está destinada a evaluar sistemas de audio de alta calidad con pequeñas degradaciones respecto de una señal de referencia determinada, y la Recomendación UIT-R BS.1534 – Método para la evaluación subjetiva del nivel de calidad intermedia de los sistemas de audio, está destinada a evaluar sistemas de audio a un nivel intermedio, apropiado para las aplicaciones de radiodifusión, pero claramente diferente respecto de una señal de referencia. Cabe señalar que el desarrollo de estos dos métodos estuvo motivado en gran medida por la necesidad de evaluar los efectos de los sistemas de codificación de audio de baja velocidad binaria.

En algunas aplicaciones no se dispone de una señal de referencia o ésta no es apropiada, por lo que la evaluación de la calidad subjetiva de los sistemas no puede realizarse en términos de fidelidad a una referencia. En la Recomendación UIT-R BS.1284 – Métodos generales para la evaluación subjetiva de la calidad de sonido, sólo se describen métodos dedicados a la gama de audio de alta calidad o que no ofrecen una valoración absoluta de la calidad de audio.

En esta Recomendación se describe un método que utiliza múltiples estímulos para la evaluación subjetiva de la calidad de las diferencias audibles de los sistemas sonoros en aplicaciones para las que no existe una referencia.

El método utiliza el enfoque de presentación de estímulos múltiples empleado en la Recomendación UIT-R BS.1534 como base para la comparación de los sistemas de sonido objeto de la prueba. Se pide al Evaluador que proporcione calificaciones para cada sistema objeto de prueba en términos de:

- 1) Calidad subjetiva global del sonido.
- 2) Calificaciones de atributos (conjuntos predefinidos de atributos seleccionados).

Las calificaciones de calidad subjetiva global del sonido se obtienen utilizando la escala continua de calidad definida en la Recomendación UIT-R BS.1534.

Dado que la fase de calificación de atributos, descrita en las Recomendaciones UIT-R BS.1116, UIT-R BS.1284, UIT-T P.835 y UIT-T P.806, es optativa, se seleccionan previamente para cada experimento varios atributos de calidad de sonido pertinentes, de preferencia entre léxicos existentes y validados. Los evaluadores califican estos atributos en escalas lineales de 100 puntos.

Gracias al análisis estadístico de estos dos tipos de datos es posible deducir:

- La calidad subjetiva global relativa del sonido de cada sistema de sonido.
- Opcionalmente, las características perceptivas de determinados atributos de cada sistema de sonido.
- Opcionalmente, la ponderación relativa de las distintas características perceptivas con respecto a la calidad percibida de los sistemas sometidos a prueba.

2 Terminología

Calidad subjetiva global – El atributo único que capta todos los aspectos de la calidad del sonido que se está evaluando, esto es, la «calidad de audio básica» definida en la Recomendación UIT-R BS.1284. En esta Recomendación se utiliza el término «calidad subjetiva global» para evitar una posible confusión con la «calidad de audio básica» definida en la Recomendación UIT-R BS.1116.

La principal diferencia entre la «calidad de audio básica» y la «calidad subjetiva global» es la diferencia en el proceso de evaluación de la calidad, no los diferentes atributos percibidos que se resumen con estos dos términos de calidad globales. Para la «calidad de audio básica», la evaluación se realiza comparando dos o más estímulos entre sí, definiéndose uno de ellos como la referencia (por ejemplo juzgar la diferencia cuantitativa entre una versión comprimida de un elemento de audio comparada con el original sin comprimir). Por su parte, la «calidad subjetiva global» es el juicio cuantitativo comparado únicamente con una referencia interna, esto es, la expectativa del oyente sin ninguna referencia externa (por ejemplo las reproducciones binaurales).

Variable controlada – Las variables que están controladas en el experimento para permitir un diseño estructurado y controlado del experimento. También se conocen como variables independientes ya que su valor es independiente de las demás variables experimentales.

Respuesta variable – Las variables para las que los evaluadores ofrecen una respuesta, calificando el estímulo percibido en una escala determinada. También se conocen como variables dependientes, ya que sus valores dependen de otras variables experimentales, por ejemplo las variables independientes/controladas.

Condición – Un conjunto de valores de las variables controladas utilizadas en la evaluación.

Ensayo (*Trial*) – Un paso en el proceso de evaluación en el que los sistemas objeto de la evaluación (o un subconjunto de los mismos) se presentan en una determinada condición y los evaluadores les atribuyen una calificación en términos de una variable de respuesta.

Repetición (*Replicate*) – Una condición de prueba repetida en que la(s) misma(s) variable(s) de respuesta es/son calificada(s) por un determinado Evaluador con los mismos valores de variable(s) controlada(s).

Descriptivo – Describir de manera objetiva e imparcial.

Atributo – Una característica especificada de calidad percibida que puede valorarse utilizando una escala de calificación. La calidad subjetiva global percibida puede considerar la valoración de múltiples atributos.

Elemento de programa – Un elemento de material de audio utilizado en la evaluación en combinación con otras variables controladas.

Estímulo – Una presentación individual de un elemento de programa a través de un sistema con arreglo a un conjunto de valores de variable controlada.

Léxico – Un conjunto de atributos de percepción descriptivos con nombres, definiciones y escalas de calificación claros.

3 Prácticas generales

Se utilizan muchas estrategias de investigación diferentes para reunir información fiable en el ámbito del interés científico. En la evaluación subjetiva de la calidad de los sistemas de audio se utilizarán métodos experimentales formales para garantizar la fiabilidad de los resultados y su interpretación. La reunión de datos fiables a partir de experimentos subjetivos requiere el control y la manipulación de las condiciones experimentales de manera que, cuando se presente a los evaluadores de manera

controlada, el experimento produzca datos de alta calidad. Se necesitan un diseño y una planificación experimental cuidadosos para garantizar que se reduzcan al mínimo los factores incontrolados, que pueden añadir ruido no deseado al experimento. Por ejemplo, si todas las condiciones de un experimento se presentan a todos los evaluadores en una secuencia fija e idéntica, ello daría lugar a un efecto de sesgo en el orden de presentación que no puede eliminarse de los datos y su interpretación. Una mejor práctica a este respecto es asegurar que las condiciones se presenten a cada evaluador en un orden aleatorio o utilizando un diseño equilibrado para reducir al mínimo cualquier posible efecto de sesgo de orden. El procedimiento de prueba recomendado se presenta en detalle en el § 5.

Para garantizar la mejor calidad de los datos en esos experimentos, es importante tener en cuenta algunos de los siguientes detalles, que forman parte de la presente Recomendación.

Hay que emplear evaluadores experimentados, ya que suelen proporcionar datos de alta calidad. La selección preliminar y definitiva de los evaluadores experimentados se describen en la Recomendación UIT-R BS.1534. Para poder evaluar de manera significativa la calidad de funcionamiento de los sistemas que se están evaluando, también es importante emplear elementos críticos del programa y seleccionar los atributos de percepción que mejor diferencian los sistemas que se están probando y contribuyen a la calidad percibida de la experiencia.

Para el diseño del experimento, el experimentador debe planificar cuidadosamente la duración del mismo. Incluir un número y una gama suficientes de elementos de prueba críticos dará una visión más general de la calidad de funcionamiento de los sistemas de sonido objeto de estudio. También puede ser conveniente comparar muchos sistemas de sonido. Esos objetivos son habituales, pero también conllevan una penalización en términos de tiempo y costo, además del riesgo de sobrecargar a los evaluadores. Los métodos para facilitar la planificación de los recursos (incluida la estimación de la duración del ensayo) figuran en el Adjunto informativo 1 del Anexo 1 de la presente Recomendación.

A fin de que los experimentos puedan proseguirse fielmente o ser repetidos más tarde o en un lugar alternativo, el informe de la prueba no sólo debe incluir los resultados, sino todos los detalles del experimento. Las directrices para la presentación de informes se describen en las Recomendaciones UIT-R BS.1116 y UIT-R BS.1534.

4 Parámetros experimentales

En esta sección se definen los parámetros experimentales clave para permitir el diseño estructurado de experimentos controlados. Éstos se dividen en dos categorías principales, a saber, las variables experimentales controladas y las variables de respuesta.

4.1 Variables experimentales controladas

Las variables experimentales controladas (o variables independientes) se utilizan para definir los parámetros que se controlan dentro del experimento, lo que permite un diseño estructurado y controlado del experimento que conducirá a un análisis estadístico exhaustivo. Típicamente, se definen variables controladas para parámetros tales como los sistemas que se están evaluando, el material del programa, los evaluadores y las réplicas. Para cada variable controlada, el experimentador debe definir el número de niveles. Por ejemplo, la inclusión de 10 elementos de programa diferentes en un ensayo se corresponde con 10 niveles de la variable elemento de programa. El número de niveles se utiliza luego en el diseño del experimento y en el posterior análisis estadístico.

4.1.1 Sistemas sometidos a evaluación

En estos experimentos, el experimentador se interesa por el estudio de la calidad de la percepción de la tecnología o el sistema que se está evaluando.

El número de sistemas bajo evaluación debe situarse entre 5 y 9, sobre la base de la Ley de Miller (Miller, G.A., 1956) para minimizar el error en la calificación del evaluador. Cuando el número de sistemas de sonido que se desea evaluar exceda de 9, se pueden encontrar más orientaciones en el § 5.1.

Siempre que sea posible, el experimentador debe incluir uno o más sistemas de calidad bien entendida para permitir que los resultados para los sistemas sometidos a prueba se consideren en su contexto.

4.1.2 Material del programa

La selección del material de prueba debe seguir los procedimientos descritos en las Recomendaciones UIT-R BS.1116 y UIT-R BS.1534. Si bien no existe un material de programa universalmente adecuado que pueda utilizarse para evaluar todos los sistemas en todas las condiciones, debe buscarse explícitamente material de programa esencial. La búsqueda del material adecuado suele llevar tiempo; no obstante, a menos que se encuentre material esencial para cada sistema, los experimentos no conseguirán poner de manifiesto las diferencias entre sistemas y no permitirán alcanzar conclusiones.

4.1.3 Evaluadores

Se recomienda recurrir a evaluadores experimentados para garantizar la calidad de los datos de prueba recopilados. Estos evaluadores deben contar con experiencia en la audición del sonido con mentalidad crítica. Dichos evaluadores obtendrán resultados más fiables y de forma más rápida que los no experimentados. También es importante señalar que la mayoría de los evaluadores no experimentados tienden a verse más influidos por los distintos tipos de efectos perturbadores tras una exposición frecuente a los mismos. Un evaluador experimentado se escoge por su capacidad para realizar una prueba de escucha. Esta capacidad puede cualificarse y cuantificarse en términos de la fiabilidad y capacidad de discriminación del evaluador durante una prueba sobre la base de una serie de evaluaciones, tal y como se indica a continuación:

- 1) Discriminación: Una medida de la capacidad de percibir las diferencias entre los elementos de prueba.
- 2) Fiabilidad: Una medida de la similitud de las calificaciones repetidas del mismo elemento de prueba.

Sólo deberán incluirse en el análisis final de los datos los evaluadores considerados evaluadores experimentados para cualquiera de las pruebas (véase la Recomendación UIT-R BS.1116). Se basan en, al menos, una calificación repetida por cada evaluador y permiten cualificar y cuantificar la experiencia del evaluador con un experimento. Los métodos se aplican como parte del proceso de preselección de los evaluadores en un experimento piloto o, preferiblemente, como proceso de preselección y de postselección (utilizando las respuestas de la evaluación principal). Un experimento piloto es a menudo una evaluación a menor escala asociada a un experimento principal e incluye un conjunto representativo e muestras de prueba a evaluar dentro del experimento principal. A fin de evaluar la experiencia del oyente, el experimento piloto deberá comprender un subconjunto pertinente de estímulos de prueba, representativos de toda la gama de estímulos y efectos perturbadores que se evaluarán durante el experimento principal.

Nominalmente, se han de emplear unos 20 (de preferencia más) evaluadores experimentados.

4.1.4 Repetición

Una manera de evaluar la calidad de los datos reunidos y el desempeño de los evaluadores es pedir juicios repetidos sobre una condición por cada evaluador. Se sugiere que se repitan al menos dos muestras para comprobar el rendimiento del evaluador (véase la sección anterior), sin aumentar excesivamente el tamaño de un experimento.

4.1.5 Variables experimentales controladas adicionales

Puede haber ocasiones en que el experimento requiera variables experimentales controladas adicionales. Resulta bastante normal y aceptable, y éstas pueden añadirse de manera similar y estructurada a las definidas en el § 4.1. El experimentador debe ser consciente de que, al aumentar el número de variables controladas, aumentará el tamaño y la duración del experimento.

4.1.6 Variables de respuesta

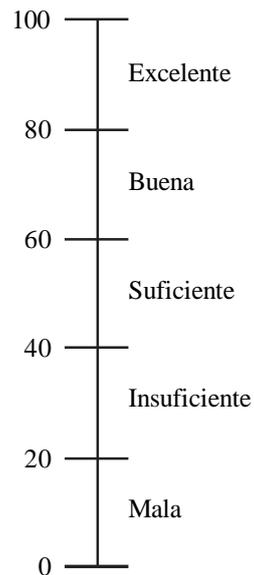
Para cada condición, se pide a los evaluadores que den su evaluación utilizando variables de respuesta. Se utilizarán dos tipos diferentes de variables de respuesta, que se describen a continuación, con su dimensionalidad asociada

- Calidad subjetiva global (por sistema).
- Calificaciones de atributos (optativas, por sistema).

4.1.7 Calidad subjetiva global del sonido

FIGURA 1

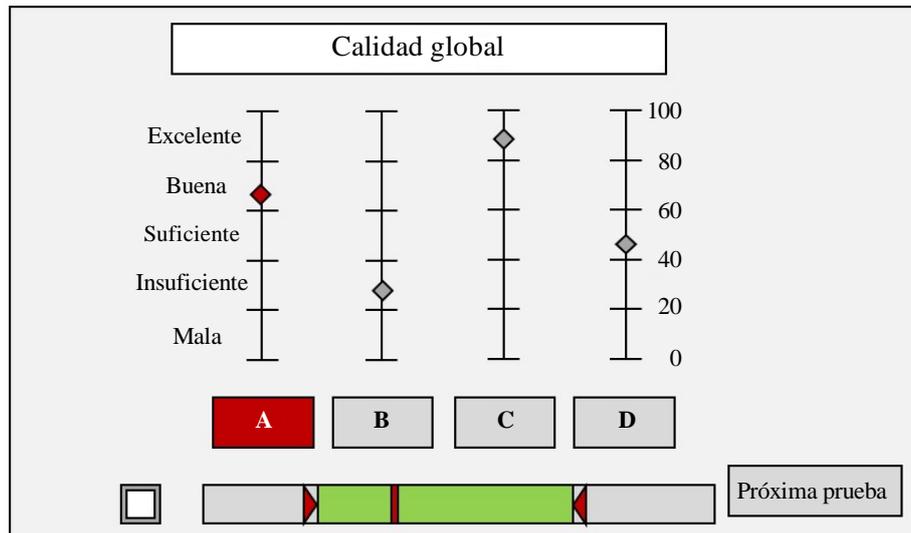
Escala de calidad continua



BS.2132-01

Inicialmente se pide a los evaluadores que proporcionen su calificación de calidad subjetiva de sonido global utilizando la escala de calidad continua (CQS)¹. Se pide a los evaluadores que valoren la calidad de sonido subjetiva global de cada presentación y que proporcionen su calificación en la CQS. La CQS consiste en una escala lineal de 100 puntos (típicamente >10 cm) que se divide en cinco intervalos iguales con los adjetivos como se ilustra en la Fig. 1. Se presentan múltiples sistemas en un solo ensayo con un elemento de programa común, cada uno con su propia escala de calificación, como se muestra en la Fig. 2.

FIGURA 2
Ejemplo de interfaz de usuario gráfica de prueba de la calidad subjetiva de sonido global



BS.2132-02

4.1.8 Calificaciones de atributos

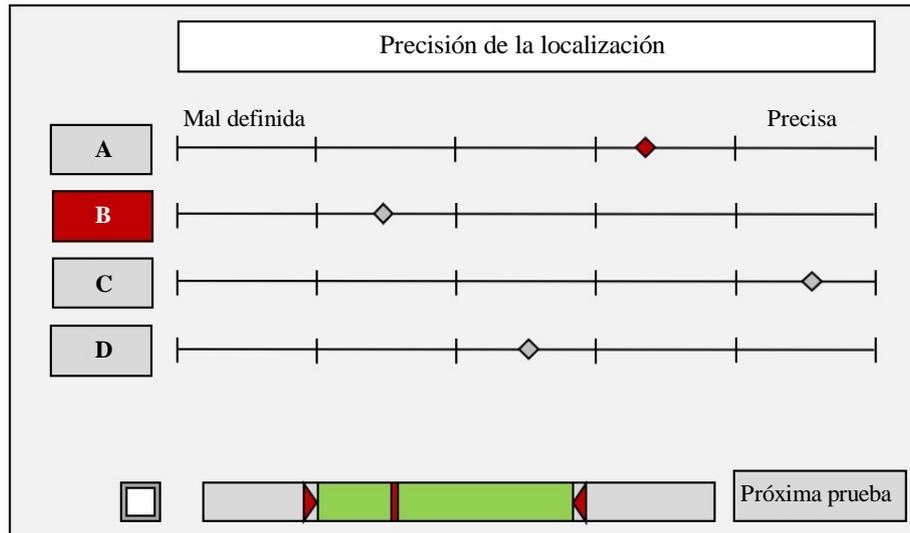
Tras la calificación de la calidad subjetiva de sonido global, se pide además a los evaluadores que califiquen combinaciones de elementos de sistema – programa en términos de cada atributo descriptivo. De nuevo, se presentan sistemas múltiples en un ensayo único con un elemento de programa común, cada uno de ellos con su propia escala de calificación.

Las escalas de atributos son típicamente continuas de 100 puntos.

En la Figura 3 se muestra un ejemplo de interfaz para la calificación de un atributo, donde el atributo y su definición son la precisión de la localización.

¹ Esta escala se utiliza también para evaluar la calidad de la imagen (Recomendación UIT-R BT.500 – Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión) y Recomendación UIT-R BS.1534 – Método para la evaluación subjetiva del nivel de calidad intermedia de los sistemas de audio.

FIGURA 3

Ejemplo de interfaz de interfaz para la calificación de un atributo (*precisión de la localización*)

BS.2132-03

5 Protocolo de prueba

5.1 Diseño del experimento

El experimento debe ser diseñado cuidadosamente para asegurar que arroje datos de alta calidad, al tiempo que se minimizan las fuentes de efectos aleatorios o incontrolados. El diseño también resulta beneficioso para estimar la magnitud y duración del experimento, así como para proporcionar la estructura para el análisis estadístico. El diseño consta de dos aspectos clave, a saber, el diseño del tratamiento y el diseño de la asignación de estímulos, como se describe a continuación.

5.1.1 Diseño del tratamiento

El diseño del tratamiento especifica qué variables controladas se utilizarán dentro del experimento, excluyendo (la variable del evaluador).

Para los experimentos de tamaño moderado, se recomienda un diseño factorial completo en el que se evalúen todas las combinaciones posibles de los niveles de las variables controladas. Para un experimento factorial completo, el número de condiciones se obtiene multiplicando el número de niveles dentro de cada variable independiente.

5.1.2 Diseño de la asignación de estímulos

El diseño de la asignación de estímulos define cómo deben presentarse las condiciones a cada evaluador.

Se recomienda un diseño «intrasujetos», en el que se presentan todas las condiciones a cada evaluador. Se controla el orden de la presentación – típicamente de manera aleatoria – para reducir al mínimo los efectos los sesgos sistemáticos. Es conveniente una presentación totalmente equilibrada.

5.1.3 Subdivisión de grandes experimentos

En ciertas situaciones, el tamaño del experimento puede llegar a ser demasiado grande y engorroso utilizando un diseño factorial intrasujetos completo. Tales casos pueden ocurrir cuando se han de evaluar muchos sistemas de sonido o cuando la duración total de la prueba por evaluador se hace irrazonablemente larga.

En esos casos se puede considerar un diseño más avanzado del experimento.

En la presente sección sólo se ilustran los enfoques que pueden considerarse para tratar esos casos. Sin embargo, el experimentador interesado debe consultar la bibliografía sobre el diseño del experimento para obtener la mejor orientación.

A continuación se facilitan ejemplos de dos soluciones diferentes para tratar experimentos de mayor tamaño.

5.1.3.1 Diseño de bloques separados

En el § 4.1.1 se recomienda limitar el número máximo de sistemas objeto de evaluación dentro del intervalo 5-9. Cuando deban evaluarse muchos sistemas de sonido, puede considerarse un diseño de bloques separados. Cuando, por ejemplo, 14 sistemas de sonido requieren evaluación, la evaluación global podría consistir en dos ensayos de siete sistemas de sonido. Para controlar cualquier efecto de sesgo relacionado con la presentación dividida, la asignación de los sistemas de sonido a cada ensayo debe ser aleatoria. Sin embargo, esta asignación al azar no debería afectar al número total de condiciones presentadas a cada evaluador o en el experimento general. Véase el § 5.1.1.

Cuando se utilice el diseño de bloques separados, es importante que se incluya el factor de definición de bloques como parte del análisis. Idealmente, la definición de los bloques no debería ser un factor estadísticamente significativo.

5.1.3.2 Diseño entre sujetos

Si el número de condiciones que debe evaluar cada evaluador es elevado, esto puede dar lugar a un número irrazonablemente grande de sesiones de escucha, especialmente cuando la duración total de la prueba puede superar las cuatro horas por evaluador.

Un enfoque para tratar esta cuestión es emplear un diseño entre sujetos (o entre grupos). Se trata de un método genérico por el que se presentan diferentes condiciones a los distintos evaluadores o grupos de evaluadores. Una forma sencilla de reducir el número de condiciones presentadas a cada evaluador (o grupo de evaluadores) sería asignar a cada evaluador (o grupo) un subconjunto diferente de los elementos del programa. Esto debe hacerse con cuidado, a fin de asegurar que el número total de condiciones presentadas esté equilibrado entre cada evaluador (o grupo de evaluadores).

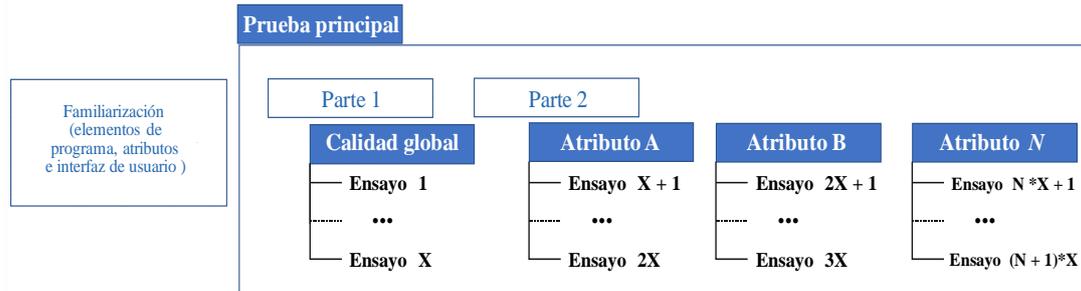
Al emplear tales diseños, es importante que la prueba incluya los factores de agrupación como parte del análisis. Lo ideal sería que el agrupamiento no fuera un factor estadísticamente significativo.

5.2 Administración de la prueba

La prueba se administrará en dos etapas, además de una familiarización que irá seguida de la prueba propiamente dicha cuando se utilicen los atributos. El orden en que se presentan los diferentes elementos a los evaluadores se ilustra en la Fig. 4. En la figura se supone que se evaluarán N atributos, utilizando M elementos del programa. Habrá un total de x ensayos para las clasificaciones generales de la calidad subjetiva de sonido global e y ensayos para las clasificaciones de los atributos.

FIGURA 4

Diagrama de flujo general de la estructura de prueba global, incluida la familiarización y la prueba principal

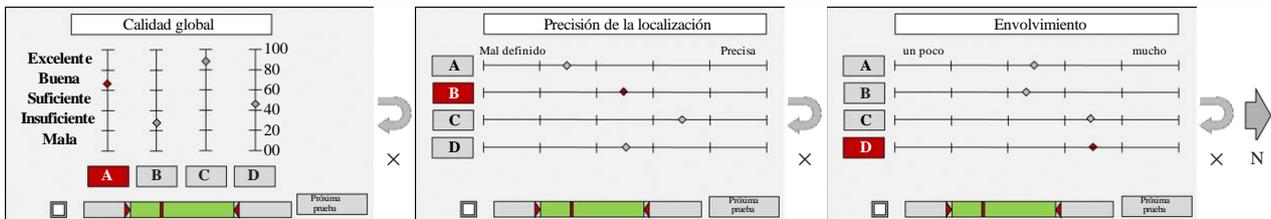


BS.2132-04

Para garantizar la recopilación de datos de alta calidad, los evaluadores deben familiarizarse con el protocolo de la prueba, las interfaces de usuario, los elementos del programa, así como con los atributos de percepción. Además, debe permitirse a cada evaluador escuchar los estímulos de la prueba, revisar los atributos y probar la interfaz de usuario. La familiarización puede incluir la escucha libre y a ciegas de los estímulos con un subconjunto de la prueba con el único propósito de familiarizarse. En la figura 5 se muestra el proceso preferido antes descrito. El evaluador evalúa los estímulos individuales para un atributo utilizando las interfaces individuales en un ensayo.

FIGURA 5

Proceso de evaluación para estímulos individuales utilizando interfaces individuales



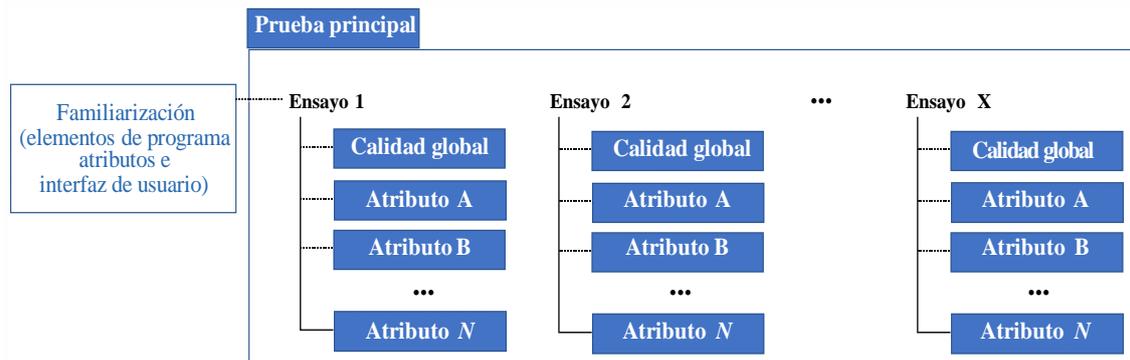
BS.2132-05

5.2.1 Procedimiento optativo

En la Figura 6 se muestra un flujo optativo de la prueba general. Pueden presentarse al evaluador interfaces individuales para cada atributo, o una interfaz combinada para múltiples atributos. En la Figura 7 se muestra un ejemplo de interfaz gráfica de usuario (GUI) en el que un evaluador evalúa cada estímulo utilizando interfaces individuales para cada atributo. En la Figura 8 se muestra un ejemplo de GUI en el que un evaluador evalúa múltiples atributos en una interfaz combinada.

FIGURA 6

Diagrama de flujo optativo de la estructura de prueba global, incluida la familiarización y la prueba principal



BS.2132-06

FIGURA 7

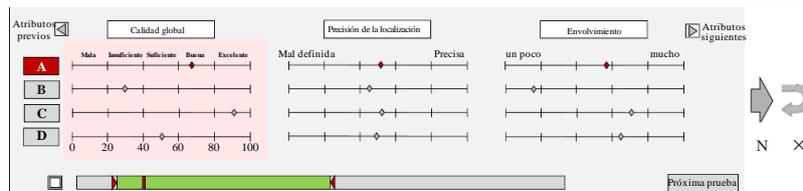
Proceso de evaluación de cada atributo para los mismos estímulos usando interfaces individuales



BS.2132-07

FIGURA 8

Proceso de evaluación de múltiples atributos para los mismos estímulos utilizando una interfaz combinada



BS.2132-07

5.3 Instrucciones al evaluador

Para asegurarse de que los evaluadores sean plenamente conscientes de la tarea que deben realizar, se les debe proporcionar instrucciones tanto por escrito como verbalmente antes del experimento. Las instrucciones deben sensibilizar a los evaluadores respecto de la tarea que ha de realizarse sin sesgos excesivos, y presentarles el protocolo de la prueba, las variables de respuesta (calidad subjetiva de sonido global, definiciones de atributos) y la forma en que se van a utilizar con la GUI de la prueba. En el anexo 2 se dan ejemplos de dichas instrucciones.

6 Entorno de prueba

6.1 Entorno de escucha

Las pruebas se realizarán en una sala de audición que sea conforme con las disposiciones de la Recomendación UIT-R BS.1116.

6.2 Dispositivos de reproducción

Pueden utilizarse auriculares o altavoces para la prueba; sin embargo, no se permite el uso de ambos en una misma sesión de prueba. Todos los evaluadores deben utilizar el mismo tipo de transductor. Deben utilizarse altavoces o auriculares de control de referencia, según se especifica en la Recomendación UIT-R BS.1116.

La disposición y los requisitos de los altavoces, así como las ubicaciones de escucha del evaluador deben configurarse de preferencia con los requisitos que figuran en la Recomendación UIT-R BS.1116.

Las disposiciones de altavoces deberían seleccionarse de preferencia entre las definidas en la Recomendación UIT-R BS.2051.

6.3 Calibración

Para garantizar resultados repetibles y reproducibles, es importante tener cuidado al configurar el equipo que va a utilizarse en la prueba, y el material de prueba que se utilizará.

Sonoridad relativa de los elementos

La sonoridad relativa de los distintos elementos no debe relacionarse en modo alguno con las mediciones de la sonoridad especificadas en la Recomendación UIT-R BS.1770. Se ajustarán breves extractos de audio al nivel de sonoridad deseado. Se conservará la diferencia entre un elemento alto (*fortissimo*) y otro bajo (*pianissimo*), que puede ser, por ejemplo, de 15dB. Hay que ajustar la sonoridad relativa de cada elemento, de preferencia por parte de un grupo de evaluadores experimentados. Para mantener esta diferencia en el nivel de reproducción, es importante probar adecuadamente los distintos elementos.

Sonoridad relativa de los estímulos

Pequeñas diferencias de sonoridad tienden a introducir un sesgo a favor del estímulo más fuerte. Tales diferencias deben ser eliminadas entre los diferentes estímulos de un elemento. Se utilizará la medición objetiva (en lugar de subjetiva) de la sonoridad especificada en la Recomendación UIT-R BS.1770. Si no es posible utilizar una medición objetiva, la sonoridad de cada extracto deberá ser ajustada subjetivamente por un grupo de evaluadores cualificados antes de su inclusión en una prueba.

Sincronización de elementos

La conmutación instantánea entre estímulos que son versiones del mismo elemento del programa procesadas de manera diferente no debe resultar en un cambio temporal perceptible. Para obtener detalles sobre la presentación de los estímulos, remítase a la Recomendación UIT-R BS.1534.

6.3.1 Calibración del sistema de reproducción

Para las pruebas realizadas con altavoces, la disposición de los altavoces utilizada debe ser preferentemente una de las que figuran en las Recomendaciones UIT-R BS.775 y UIT-R BS.2051. Alternativamente, se debe utilizar la notación definida en la Recomendación apropiada para describir la disposición utilizada en la prueba.

El nivel del sistema de reproducción se ajustará según lo descrito en la Recomendación UIT-R BS.1116.

Los detalles de la calibración de los altavoces de sonidos graves (*subwoofers*) y los sistemas de gestión de bajos quedan fuera del alcance de este documento. El resultado de la gestión de graves debe ser que la respuesta de frecuencia de una combinación de altavoces y subwoofers individuales sea plana (dentro de la tolerancia especificada en la Recomendación UIT-R BS.1116).

Se ha observado a partir de secuencias de prueba previas 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 afectará a la audibilidad de algunos de los efectos perturbadores a evaluar. Por consiguiente, si los sujetos ajustan la ganancia del sistema, ello debe anotarse en los resultados de la prueba.

Las diferencias de retardo de tiempo entre los canales de un sistema estereofónico o multicanal no deben superar los 20 μ s para los auriculares, o 100 μ s para los altavoces, según se especifica en la Recomendación UIT-R BS.1116.

En el caso de sistemas con acompañamiento de imagen, el retardo de tiempo total de los auriculares o 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.

NOTA – La medición de parámetros acústicos de sistemas de sonido avanzados puede resultar considerablemente más compleja que en el caso de los anteriores sistemas de audio multicanales. Al proceder a las mediciones, hay que ser cuidadosos con la selección del micrófono de medición y la orientación del mismo (véase el Informe UIT-R BS.2419). En la Recomendación UIT-R BS.1116 se facilita información adicional sobre los requisitos electroacústicos y las características de respuesta de la sala de operaciones.

7 Análisis estadístico

Se lleva a cabo el análisis estadístico de los datos de calificación del evaluador para proporcionar una visión de la calidad subjetiva de los sistemas que se están evaluando. Las calificaciones medias se calculan para dar indicaciones de esta calidad de funcionamiento y se utiliza la estimación de la varianza de los datos para indicar la fiabilidad de las diferencias de calidad de funcionamiento del sistema observadas.

Cuando se reúnen los datos, cada evaluador proporciona las calificaciones de los atributos de los sistemas objeto de la evaluación. Estos sistemas se prueban con diferentes elementos del programa. El evaluador califica cada sistema respecto de una lista de atributos de percepción utilizando escalas predefinidas. Para cada elemento del programa, el evaluador califica cada atributo en el mismo conjunto exacto de escalas de atributos. El evaluador también califica cada combinación de elementos del sistema y del programa en función de la calidad subjetiva general.

Para cada elemento del programa, los evaluadores deben proporcionar sus calificaciones de atributos de cada sistema, así como las calificaciones subjetivas generales de la calidad del sonido.

En la Recomendación UIT-R BS.1534 se proporcionan detalles para el análisis estadístico de los datos de calidad subjetiva global del sonido y de los datos para cada atributo descriptivo.

Además, el análisis de los datos sensoriales obtenidos al aplicar esta Recomendación arroja conclusiones comparables a las del análisis de los datos sensoriales obtenidos mediante un análisis descriptivo cuantitativo más clásico. Esos análisis incluyen el Análisis de Varianza (ANOVA) realizado para cada atributo sensorial, así como el análisis multivarianza (como el uso del Análisis de los Componentes Principales, PCA).

8 Referencias

- [1] Recomendación UIT-R BS.2051 – Sistemas de sonido avanzados para la producción de programas
- [2] Recomendación UIT-R BS.775 – Sistema de sonido estereofónico multicanal con y sin acompañamiento de imagen
- [3] Recomendación UIT-R BS.645 – Señales de prueba y métodos de medida para los enlaces radiofónicos internacionales

- [4] Recomendación UIT-R BS.1116 – Métodos para la evaluación subjetiva de pequeñas degradaciones en los sistemas de audio
- [5] Recomendación UIT-R BS.1534 – Método para la evaluación subjetiva del nivel de calidad intermedia de los sistemas de audio
- [6] Recomendación UIT-R BS.1770 – Algoritmos para medir la sonoridad de los programas radiofónicos y el nivel de cresta de audio real
- [7] Recomendación UIT-R BS.1864 – Prácticas operativas en materia de sonoridad en el intercambio internacional de programas de televisión digital
- [8] Miller, G.A. (1956), *The magical number seven, plus or minus two: Some limits on our capacity for processing information*. *Psychological Review*. 63 (2): 81–9
- [9] Recomendación UIT-R BT.500-13 – Metodología para la evaluación subjetiva de la calidad de las imágenes de televisión
- [10] Recomendación UIT-T P.835 – Metodología de prueba subjetiva para evaluar los sistemas de comunicación vocal que utilizan un algoritmo de cancelación de ruido
- [11] Recomendación UIT-T P.806 – Metodología de prueba de la calidad subjetiva con numerosas escalas de puntuación
- [12] Informe UIT-R BS.2419 – Efecto de la directividad del micrófono en relación con la calibración y la equalización del nivel de los sistemas avanzados de sonido

Adjunto 1 al Anexo 1 (informativo)

Herramienta Excel para estimar la duración de la prueba de un experimento

La herramienta Excel que se facilita más abajo sirve para estimar la duración de la prueba de un experimento de evaluación subjetiva del sonido a efectos de la planificación de recursos. En el Cuadro 1 se muestra un ejemplo de un experimento. Los campos en marrón son los campos de ingresos. Los campos en verde son los campos de resultados.



Copy of
R15-WP6C-190715-T

CUADRO 1

Copia de pantalla del cuadro Excel facilitado para estimar la duración de la prueba de un experimento de evaluación subjetiva del sonido

Variables experimentales controladas (variables independientes)					
Variable	Etiqueta	Descripción	Núm. de niveles	Cálculo del núm. de niveles	Grados de Libertad (DOF)
1	SYSTEM	Núm. de sistemas objeto de la prueba	7	7	6
2	PROGRAMME	Núm. de elementos de programa	3	3	2
3	REPLICATE	Núm. de presentaciones o repeticiones	1	1	0
4	ASSESSOR	Núm. de evaluadores	20	20	19
Total			31		27

Parámetros de la prueba			
Parámetro	Núm.	Unidades	Comentarios
Núm. de condiciones experimentales controladas (total)	21		Excluyendo a los evaluadores
Núm. de condiciones de la prueba (por repetición)	21		Excluyendo a los evaluadores
Núm. total de variables independientes	420		Incluyendo a los evaluadores
Núm. de bloques	1		1 = diseño intrasujetos >1 = diseño entre sujetos
Núm. de elementos de PROGRAMA por bloque	3		Debe ser entero ≥2
Núm. de calificaciones por condición	20		
Núm. total de calificaciones por evaluador	147		
Tiempo medio de calificación estimado por condición	20	s	
Duración estimada total por evaluador	0.8	hora	
Duración de la sesión	2	hora	máx. < 2 horas incl. descansos
Núm. de sesiones estimado por evaluador	1	sesiones	
Núm. total de sesiones	20	sesiones	
Núm. de puntos de datos por variable de respuesta	420		
Núm. total de puntos de datos en el experimento	2940		

Variables de respuesta (variables dependientes)			
Variable	Etiqueta	Descripción	Núm. de niveles
Total		Calidad subjetiva global	1
Descriptiva		Envolvimiento, anchura de la fuente, etc.	6

**Adjunto 2
al Anexo 1
(informativo)**

Ejemplo de instrucciones a los evaluadores

A2.1 Instrucciones generales

Estimado oyente:

Bienvenido a esta prueba, en la que escuchará y evaluará diferentes sistemas de sonido con una serie de elementos de programa. La prueba emplea el método de «Estímulos Múltiples sin una Referencia Dada».

Dispone de unas dos horas en total para la prueba, que incluye una fase de familiarización seguida de la prueba principal. Durante la fase de familiarización, se familiarizará con los elementos del programa, las interfaces de usuario, así como los atributos utilizados para evaluar cada sistema de sonido. Después de la familiarización, realizará una prueba en dos partes.

La Parte 1 consiste en la calificación de sistemas de sonido en términos de *calidad subjetiva global*.

Calidad subjetiva global – El atributo único que capta todos los aspectos de la calidad del sonido que se está evaluando.

La Parte 2 consiste en calificar cada uno de los sistemas de sonido en términos de los atributos siguientes:

- Profundidad del escenario.
- Envolvimiento.
- Inmersión.
- Precisión de la localización.
- Brillantez.
- Distorsión.

A continuación se facilitan las definiciones de los atributos, que se explicarán antes de la prueba.

Al realizar cada etapa de la prueba, por favor, escuche atentamente los elementos del programa tomándose el tiempo necesario para explorar y evaluar cada elemento.

Si tiene alguna pregunta o necesita más aclaraciones en relación con el protocolo de la prueba, por favor pregunte, preferiblemente durante o después de la fase de familiarización.

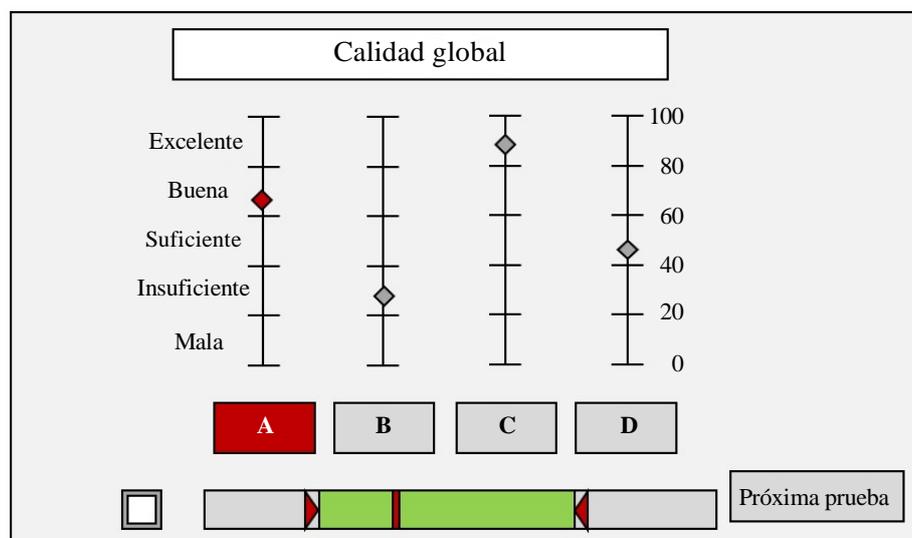
Se recomienda que haga una pausa cada 20-30 minutos para estirar las piernas y descansar brevemente.

A2.1.1 Calificación de la calidad global

Se le pedirá que evalúe la calidad global de cada muestra de sonido en términos de la *calidad subjetiva global* utilizando una escala de calidad continua de 0-100 puntos, según se muestra en la Fig. 9. Por favor, escuche atentamente cada muestra y con la frecuencia que sea necesaria y, cuando esté listo, dé sus calificaciones. Una vez que haya calificado todas las muestras, pulse «siguiente» para comenzar la siguiente prueba.

FIGURA 9

Ejemplo de interfaz de usuario gráfica de prueba de la calidad subjetiva de sonido global



BS.2132-09

A2.1.2 Atributos descriptivos

Para cada prueba, se le pedirá que evalúe la calidad del sonido de cada uno de los sistemas respecto de uno de los atributos (ver Cuadro 2).

CUADRO 2

Etiquetas y definiciones de los atributos

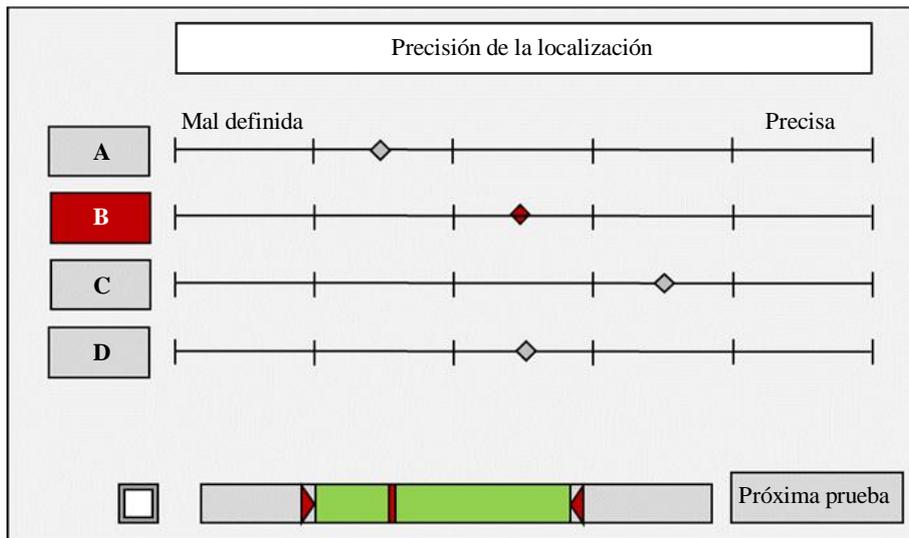
Atributo	Definición	Ejemplo	Etiqueta inferior	Etiqueta superior
Profundidad del escenario	La percepción de la profundidad de la imagen sonora. Tiene en cuenta tanto la profundidad general del escenario como la distancia relativa de las fuentes de sonido individuales.		plana	profunda
Envolvimiento	¿Está rodeado por el sonido reproducido y da una sensación de espacio alrededor o rodeándolo? La sensación de estar rodeado por el sonido.	En qué medida	poco	mucho
Inmersión	Extensión percibida de una fuente de sonido en dirección vertical. Sensación de ser barrido para sentirse rodeado o completamente cubierto.	¿Hasta qué punto el sonido aparece por encima y por debajo de usted? ¿Hasta qué punto el sonido le cubre verticalmente?	poca	muchas
Precisión de la localización	¿En qué medida están bien colocados y separados los instrumentos y voces individuales en la imagen de sonido espacial? ¿Con cuánta precisión están ubicadas las fuentes de sonido individuales en la sala? Si las fuentes de sonido individuales están inadvertidamente dispersas o extendidas, la precisión es baja. ¿Pueden los instrumentos y voces individuales ubicarse y distinguirse claramente en la imagen espacial de sonido? ¿Se han ubicado de manera precisa las fuentes de sonido individuales en la sala?		imprecisa	precisa
Brillantez	Extensión de agudos o de altas frecuencia: <ul style="list-style-type: none"> – Poca: como si escucharas la música a través de una puerta, apagada, borrosa o tediosa. – Mucha: ligereza, pureza y claridad, con espacio para los instrumentos. Claridad en las frecuencias superiores sin ser agudas o chillonas y sin distorsión. 		poca	muchas

CUADRO 2 (fin)

Atributo	Definición	Ejemplo	Etiqueta inferior	Etiqueta superior
Distorsión	Sonidos adicionales y no deseados que añaden efectos perturbadores al sonido.	Pueden aparecer en forma de efectos perturbadores temporales o de timbre que pueden producir un sonido «agudo», «áspero» o «roto», o un timbre temporal, por ejemplo.	poca	mucha

FIGURA 10

Interfaz de usuario de pruebas de atributos descriptivos



BS.2132-10

Cuando esté listo, comience la fase de familiarización.

Le damos las gracias por su participación.

**Adjunto 3
al Anexo 1
(informativo)**

Ejemplos de casos de utilización para la evaluación subjetiva y la elaboración de perfiles descriptivos de la calidad de sonido de los sistemas de audio sin una referencia dada

La introducción de sistemas de sonido avanzados, tal y como se describe en la Recomendación UIT-R BS.2051, ofrece herramientas para la expresión creativa en la producción. La evaluación de estos sistemas puede incluir condiciones en las que no se dispone de referencias y patrones ocultos conocidos. Los métodos descritos en esta Recomendación son aplicables en estas condiciones.

Además, pueden existir otros casos de uso en los que los experimentadores podrían beneficiarse del uso de esta metodología para la caracterización de la calidad subjetiva de sus sistemas o señales.

Entre los ejemplos de casos de uso aplicables a este método figuran las evaluaciones subjetivas de la calidad de:

- El comportamiento de reproductores de sistemas de producción de sonido avanzados cuando no se dispone de referencias para la intención del productor o éstas no son apropiadas.
 - La reproducción de un programa de sonido avanzado con distintas configuraciones de altavoces por un mismo reproductor de producción.
 - Sistemas para la reproducción de teatro a domicilio de contenidos de sistemas de sonido avanzados, cuando no existe ningún sistema que proporcione a priori un objetivo conocido de mejor calidad.
 - algoritmos de mezclado ascendente o descendente.
 - Sistemas de micrófonos para la grabación y producción de sonido espacial.
 - Procesadores de reverberación para la producción de sonido espacial.
 - Sistemas de procesamiento dinámico multibandas y ajustes para la distribución de radio.
 - Técnicas de restauración de sonido.
-