

RECOMENDACIÓN UIT-R BS.1284*

Métodos para la evaluación subjetiva de la calidad de sonido – Requisitos generales

(Cuestión UIT-R 85/10)

(1997)

La Asamblea de Radiocomunicaciones de la UIT,

considerando

- a) que la introducción de nuevos tipos de procesamiento de la señal de sonido, tales como la codificación digital y la reducción de la velocidad binaria, de nuevos tipos de señales de televisión que utilizan componentes multiplexados en el tiempo y de nuevos servicios como la televisión mejorada y la televisión de alta definición (TVAD), pueden hacer necesario introducir métodos nuevos o modificaciones en la realización de evaluaciones subjetivas de la calidad de sonido;
- b) que estas técnicas entrañan sus propias degradaciones de la señal;
- c) que las pruebas subjetivas de escucha permiten evaluar el grado de la molestia ocasionada al oyente por una degradación de la señal deseada durante su transmisión entre la fuente y el oyente;
- d) que pueden aplicarse muchos métodos distintos de prueba subjetiva;
- e) que, en particular, en otras Cuestiones UIT-R se insta urgentemente a la elaboración de directrices para la evaluación subjetiva;
- f) que es muy deseable normalizar los métodos de prueba subjetiva y la interpretación de los resultados, para que puedan establecerse las mejores comparaciones posibles entre los resultados obtenidos en diferentes momentos y/o lugares;
- g) que es muy deseable que las escalas de apreciación que se utilicen para describir la calidad subjetiva del sonido permitan la aplicación de métodos de procesamiento estadístico más coherentes, independientemente del idioma en que se expresen las opiniones;
- h) que sería deseable contar con una única escala de evaluación para los programas de radiodifusión sonora y de televisión;
- j) que las propiedades geométricas y acústicas de las salas de control y las salas de escucha pueden influir considerablemente en la audición, por lo que deberían especificarse ajustadamente las condiciones de escucha,

recomienda

1 que los procedimientos de prueba y evaluación señalados en el Anexo 1 a la presente Recomendación se utilicen para la evaluación subjetiva de la calidad del sonido reproducido.

* La Comisión de Estudio 6 de Radiocomunicaciones efectuó modificaciones de redacción en esta Recomendación en 2002 de conformidad con la Resolución UIT-R 44.

ANEXO 1

1 Consideraciones generales

Este Anexo se divide en los siguientes párrafos, en los que se detallan los requisitos referentes a los diferentes 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 Tratamiento estadístico de los datos
- 10 Presentación de los resultados
- 11 Contenido de los informes de prueba
- 12 Referencias Bibliográficas
- 13 Bibliografía

La idea es que esta Recomendación sirva de guía a la evaluación general de la calidad del sonido. La Recomendación se basa en 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. No obstante, los requisitos de esta última Recomendación son estrictos, ya que se concibieron pensando en la evaluación de pequeñas degradaciones. Las evaluaciones más generales suelen entrañar diferencias mayores y, en consecuencia, no suelen obligar a controlar tan ajustadamente los parámetros de prueba. La Recomendación UIT-R BS.1116 contiene un glosario de términos, algunos de los cuales se utilizan en esta Recomendación.

En la Recomendación UIT-R BS.1283 – Evaluación subjetiva de la calidad de sonido – Guía para las Recomendaciones existentes, se citan otras Recomendaciones de la UIT que pueden ser relevantes en ciertos casos especiales.

2 Diseño del experimento

Al diseñar las pruebas, habría que tener en cuenta las consideraciones del § 2 de la Recomendación UIT-R BS.1116. No obstante, debido al hecho de que las degradaciones sometidas a prueba pueden no ser pequeñas, no siempre es indispensable recurrir a una referencia. En caso de utilizar una referencia, no hay necesidad de que la misma no esté degradada en sentido absoluto.

En general, se requieren conocimientos especializados de estadística para diseñar la prueba. Esto incluye la determinación del número de observaciones requeridas, los métodos estadísticos para analizar los datos y la interpretación correcta de los resultados del análisis estadístico, lo que entraña, a su vez, la verificación de la validez de las hipótesis del modelo.

3 Selección de los oyentes

Resulta preferible recurrir a oyentes no expertos. Se ha señalado que los oyentes no expertos pueden ser representativos de la población general y que es posible que los expertos sean demasiado críticos. En todo caso, los oyentes no expertos se convertirán a la larga en expertos si se familiariza a éstos con los aparatos. Por consiguiente, las pruebas con expertos indican más adecuada y rápidamente cuáles serán los resultados probables a largo plazo. En caso de duda, debe investigarse la relación entre las opiniones de los expertos y no expertos.

En general, debería adiestrarse a los oyentes para familiarizarlos con los procedimientos, los materiales y los entornos de prueba.

4 Método de prueba

4.1 Escalas de apreciación

Se recomienda recurrir a las siguientes escalas de cinco notas para la evaluación subjetiva de la calidad de sonido o la degradación. La naturaleza y objetivo de las pruebas determinarán cuál de las dos escalas es la más apropiada.

Calidad		Degradación	
5	Excelente	5	Imperceptible
4	Buena	4	Perceptible, pero no molesta
3	Adecuada	3	Ligeramente molesta
2	Insuficiente	2	Molesta
1	Mala	1	Muy molesta

Para realizar las pruebas de comparación, puede aplicarse un método basado en la siguiente escala de comparación de siete notas o un método basado en diferencias numéricas en el que se utilicen las escalas de cinco notas mencionadas. En general, estos dos métodos no son equivalentes y es posible que arrojen resultados distintos.

(Resulta esencial indicar claramente la dirección prevista de la comparación.)

Comparación	
3	Mucho mejor
2	Mejor
1	Ligeramente mejor
0	Igual
-1	Ligeramente peor
-2	Peor
-3	Mucho peor

NOTA 1 – Las escalas deben considerarse como si fueran continuas y a este respecto se recomienda una resolución de un decimal.

NOTA 2 – Se ha visto que la utilización de puntos de referencia intermedios predefinidos puede introducir sesgos. Es posible utilizar las escalas numéricas sin descripciones de puntos de referencia. En esos casos, hay que indicar la orientación prevista de las escalas. Esto puede contribuir a allanar los obstáculos de traducción planteados cuando se comparan los resultados de las pruebas escritas en diferentes idiomas.

Si no se utilizan puntos de anclaje intermedios, es esencial que los resultados para cada sujeto se normalicen con respecto a la desviación media y típica. Cabe aplicar la ecuación (1) para realizar dicha normalización, reteniendo al mismo tiempo la escala original.

$$Z_i = \frac{(x_i - x_{si})}{s_{si}} \cdot s_s + x_s \quad (1)$$

donde:

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

4.2 Procedimiento de prueba

Las pruebas pueden consistir en presentaciones únicas, comparaciones de dos sistemas (uno de los cuales puede ser la referencia) o comparaciones múltiples, con o sin referencias. Las presentaciones pueden repetirse, si así se requiere. Estos procedimientos de prueba deberían utilizarse conjuntamente con las escalas de apreciación del § 4.1.

Las limitaciones a corto plazo de la memoria humana pueden exigir que los extractos de los programas no duren más de 15 a 20 s; éstos pueden ser muy cortos (unos cuantos segundos) para ciertas pruebas. En caso de que la secuencia consista en una pieza musical, la frase no debería parecer interrumpirse. Si la secuencia de prueba no está bajo el control del participante, es necesario proporcionarle una clara indicación sobre la presentación en curso.

Ninguna sesión debería durar más de 15 a 20 min sin interrupción. Si hay necesidad de que las sesiones sean consecutivas, deberían separarse por pausas de al menos la misma duración.

El aparato de conmutación no debería introducir disturbio audible alguno.

En caso de aplicar las pruebas a cada oyente por separado, es muy conveniente que éstos controlen la conmutación entre los estímulos, según se describe en la Recomendación UIT-R BS.1116.

5 Atributos

Dependiendo de los objetivos de la prueba, pueden utilizarse diferentes números y tipos de atributos para describir la calidad percibida.

Deben definirse claramente los atributos que se utilicen.

5.1 Calidad de sonido básica

El atributo de calidad de audio básica abarca todos los aspectos de la calidad del sonido que se evalúe, e incluye de manera no exhaustiva lo siguiente: timbre, transparencia, imagen estereofónica, presentación espacial, reverberación, ecos, distorsiones de armónicos, ruido de cuantificación, «pops», «clics» y ruido de fondo. Para evaluar las pequeñas degradaciones, en la Recomendación UIT-R BS.1116 se define de forma distinta el atributo de calidad de audio básica.

5.2 Atributos que especifican detalladamente la calidad del sonido estereofónico de dos canales y multicanal

En los § 5.2 y 5.3 de la Recomendación UIT-R BS.1116 se presenta una lista de atributos.

5.3 Atributos que especifican la relación entre el sonido y la imagen de acompañamiento

En el § 5 de la Recomendación UIT-R BS.1286 – Métodos para la evaluación subjetiva de los sistemas de audio con acompañamiento de imagen, se indica una lista de atributos.

5.4 Atributos principales para especificar detalladamente la evaluación absoluta de la calidad del sonido

La Unión Europea de Radiodifusión, en su documento técnico 3286 «Méthodes d'évaluation subjective de la qualité de programmes radiophoniques», proporciona una lista de atributos.

5.5 Atributos que especifican detalladamente la calidad del sonido transmitido/codificado digitalmente

En el Anexo 2 se proporciona una lista de los principales atributos.

6 Material de programa

Dependiendo de cuál sea el objetivo preciso de las pruebas y, en particular, de la categoría del sistema de transmisión o reproducción de programas de radiodifusión sonora que se esté probando, el material de prueba puede seleccionarse deliberadamente atendiendo a su comportamiento muy crítico con respecto a las deficiencias introducidas por el sistema que se esté probando. En otros casos, puede utilizarse material menos crítico.

En el § 6 de la Recomendación UIT-R BS.1116 se exponen detalladamente los factores relacionados con el material de programa de pruebas crítico y su selección con diferentes propósitos.

Siempre que la idea sea que el sistema transporte sonido de gran calidad, debería utilizarse un material de tipo crítico. Para garantizar la comparabilidad de los datos de prueba obtenidos en diferentes lugares y/o tiempos, deberían utilizarse las mismas secuencias de programa.

En todo caso, el contenido de una secuencia de programa no debería ser ni tan interesante ni tan desagradable o aburrido que distraiga al oyente.

7 Dispositivos de reproducción

7.1 Pruebas que no incluyen altavoces (o auriculares) como parte del sistema sometido a prueba

Deberían seguirse las recomendaciones del § 7 de la Recomendación UIT-R BS.1116. Hay que señalar, sin embargo, que el recurso a mediciones del nivel de presión sonora ponderadas «A» con una señal de banda ancha no permite necesariamente evaluar con exactitud la sonoridad subjetiva. Esto rige especialmente cuando el sistema de reproducción incluye ciertos componentes con diferentes anchuras de banda.

Puede resultar necesario utilizar métodos alternativos con el fin de garantizar disposiciones de ganancia correctas para todos los canales de reproducción.

7.2 Pruebas que incluyen altavoces (o auriculares) como parte del sistema sometido a prueba

Las pruebas en las cuales se incluyan dispositivos de reproducción en el sistema objeto de prueba, deberían diseñarse con arreglo a las especificaciones del sistema.

En las pruebas de comparación los sistemas deben ajustarse recíprocamente con exactitud en cuanto a la sonoridad.

8 Condiciones de escucha

A menos de que formen parte de la prueba, las condiciones de escucha deberían ser conformes con los requisitos estipulados en el § 8 de la Recomendación UIT-R BS.1116.

Hay que señalar que estos requisitos pueden ser demasiado rigurosos para ciertos tipos de pruebas.

9 Tratamiento estadístico de los datos

Los datos subjetivos deberían evaluarse para obtener los valores medios y los intervalos de confianza. Esta operación describirá los datos y, si la discriminación resultante es inadecuada para satisfacer los objetivos de la prueba, habría que proceder a un tratamiento ulterior. Cabe la posibilidad de utilizar los métodos del § 9 de la Recomendación UIT-R BS.1116. En general, será necesario contar con conocimientos especializados de estadística para analizar los datos.

Es posible mejorar el valor global de la prueba, si los datos se analizan una vez más para verificar los supuestos subyacentes de la prueba y evaluar la fiabilidad del participante.

10 Presentación de los resultados

En la medida de lo posible la presentación de los resultados debería estar en consonancia con el § 10 de la Recomendación UIT-R BS.1116.

11 Contenido de los informes de prueba

En la medida de lo posible habría que informar sobre todos los aspectos de las pruebas, con arreglo a lo descrito en la Recomendación UIT-R BS.1116, aun cuando no se traduzcan a la práctica o controlen todos los aspectos.

Por ejemplo, aunque no se imparta adiestramiento, el informe debería dar constancia de este hecho.

ANEXO 2

Las diferentes categorías de aparatos pueden entrañar técnicas de codificación o transmisión digitales.

Para evaluar las degradaciones de las señales de sonido ocasionadas por los procesos de codificación o transmisión digitales es posible utilizar una serie de categorías para analizar o clasificar el tipo de aparato.

<i>Categoría de aparatos</i>	<i>Explicación</i>
Defectos de cuantificación	Defectos relacionados con una insuficiente resolución digital, por ejemplo, distorsiones granulares y cambios no estacionarios en el nivel de ruido
Distorsión de la característica de frecuencia	Ausencia de frecuencias elevadas o bajas, exceso de frecuencias elevadas como las correspondientes a sonidos silbantes o seseos, efectos de los formantes y efectos de los filtros de peine
Distorsión de las características de ganancia	Cambios en el nivel (ganancia) o la gama dinámica de las señales de fuente, saltos de nivel (pasos)
Efectos de modulación periódicos	Variaciones periódicas de la amplitud de la señal, tales como gorjeos, bombeos o parpadeos
Efectos de modulación no periódicos	Efectos asociados con los transitorios, por ejemplo, chapoteos o ráfagas y deformación de los procesos transitorios
Distorsión no lineal	Distorsión no lineal de armónicos o inarmónicos y distorsiones por repliegue del espectro
Distorsión temporal	Pre y post-ecos, arrastre corto (pérdida de transparencia temporal de la señal de fuente) y asincronismo de las señales o canales
Sonido adicional (ruido)	Sonidos no esenciales que no guardan relación con el material de fuente, por ejemplo, clics, ruido, componentes tonales
Sonido faltante	Pérdida de los componentes sonoros del material de fuente, por ejemplo, ocasionados por fallos de enmascaramiento
Efecto de las correlaciones (diafonía)	Diafonía lineal o no lineal entre canales, fugas o correlación entre canales
Distorsión de la calidad de imagen espacial	Todos los aspectos, incluido ensanchamiento, movimiento, estabilidad de localización, equilibrio, exactitud de localización, cambios de la espaciosidad
