

## RECOMENDACIÓN UIT-R BS.1770-1

**Algoritmos para medir la sonoridad de los programas radiofónicos  
y el nivel de cresta de audio real**

(Cuestión UIT-R 2/6)

(2006-2007)

**Cometido**

Esta Recomendación especifica los algoritmos de medición de audio destinados a determinar la sonoridad subjetiva de los programas y el nivel verdadero de cresta de la señal.

La Asamblea de Radiocomunicaciones de la UIT,

*considerando*

- a) que las modernas técnicas de transmisión sonora digital ofrecen una gama dinámica extremadamente amplia;
- b) que las modernas técnicas digitales de producción y transmisión del sonido proporcionan una combinación de formatos monofónicos, estereofónicos y multicanal, y que los programas radiofónicos se producen en todos esos formatos;
- c) que los oyentes desean que la sonoridad subjetiva de los programas de audio sea uniforme para los diferentes tipos de fuentes y programas;
- d) que hay muchos métodos de medición de los niveles de audio, pero que los métodos existentes utilizados en la producción de programas no dan ninguna indicación de la sonoridad subjetiva;
- e) que para el intercambio de programas es fundamental disponer de un algoritmo único recomendado a fin de realizar una estimación objetiva de la sonoridad subjetiva;
- f) que los futuros algoritmos complejos basados en modelos psicoacústicos pueden proporcionar mejores medidas objetivas de la sonoridad para una gran variedad de programas de audio;
- g) que los medios digitales se sobrecargan abruptamente y que esa sobrecarga debe evitarse, aun cuando sea momentánea,

*considerando también*

- h) que los niveles de cresta de la señal pueden aumentar debido a procesos habituales, como el filtrado o la reducción de la velocidad binaria;
- j) que las tecnologías de medición existentes no reflejan el nivel de cresta real de una señal digital ya que ese valor puede producirse en un lugar situado entre las muestras;
- k) que el estado de procesamiento de la señal digital hace que sea práctico aplicar un algoritmo que evalúe con precisión el nivel verdadero de cresta real de una señal;
- l) que la utilización de un algoritmo que proporcione el valor de cresta real permitirá obtener una indicación más precisa del margen que existe entre el nivel de cresta de una señal de audio digital y el nivel de recorte,

*recomienda*

- 1 que cuando sea necesario realizar una medición objetiva de la sonoridad de un canal o de un programa de audio para facilitar la distribución y el intercambio de programas, se utilice el algoritmo que se especifica en el Anexo 1;
- 2 que los métodos utilizados en la producción y postproducción de programas para indicar su sonoridad se basen en el algoritmo que se especifica en el Anexo 1;
- 3 que cuando se requiera una indicación del nivel de cresta real de una señal de audio digital, el método de medición se base en las directrices que figuran en el Anexo 2, o en un método cuyos resultados sean similares o mejores,

NOTA 1 – Los usuarios deben tener presente que el valor de la sonoridad obtenido es una estimación de la sonoridad subjetiva, por lo que implica un cierto grado de discrepancias que dependen de los oyentes, del equipo de audio y de las condiciones de escucha,

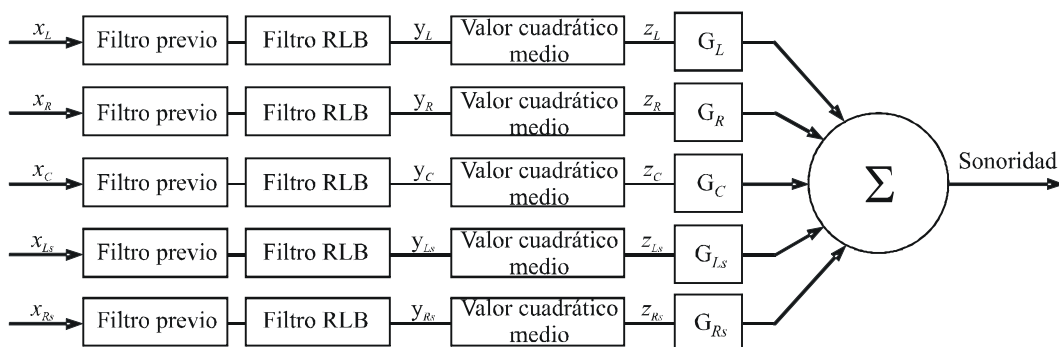
*recomienda también*

- 1 que se prosigan los trabajos a fin de ampliar el algoritmo especificado en el Anexo 1 para poder proporcionar indicaciones acerca de la sonoridad a corto plazo;
- 2 que se considere la posibilidad de tener que actualizar esta Recomendación si aparecen nuevos algoritmos relativos a la sonoridad cuyos resultados mejoren significativamente a los del algoritmo especificado en el Anexo 1.

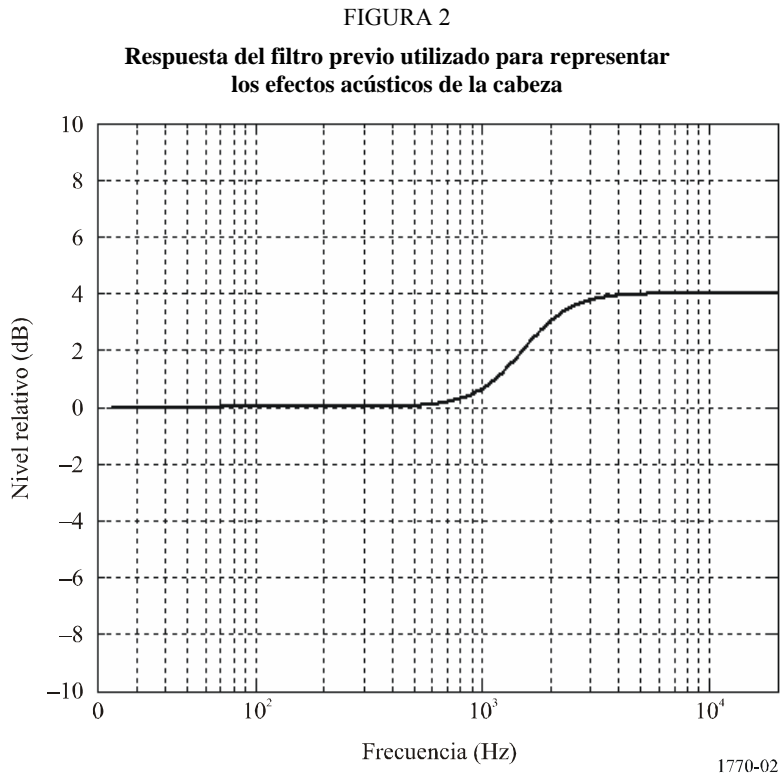
**Anexo 1****Especificación del algoritmo de medición de la sonoridad objetiva multicanal**

El presente Anexo especifica el algoritmo de medición de la sonoridad multicanal. En el diagrama de bloques de la Fig. 1 pueden observarse los diversos componentes del algoritmo. Para facilitar su descripción, se han previsto etiquetas en los diferentes puntos a lo largo del trayecto del flujo de las señales. El diagrama muestra las entradas de los cinco canales principales (izquierdo, central, derecho, sonido ambiente izquierdo y sonido ambiente derecho), lo que permite controlar los programas que contienen entre uno y cinco canales. En el caso de programas con menos de cinco canales, algunas de esas entradas no se utilizarán. En las mediciones no se incluye el canal de efectos de baja frecuencia (LFE, *low frequency effect*).

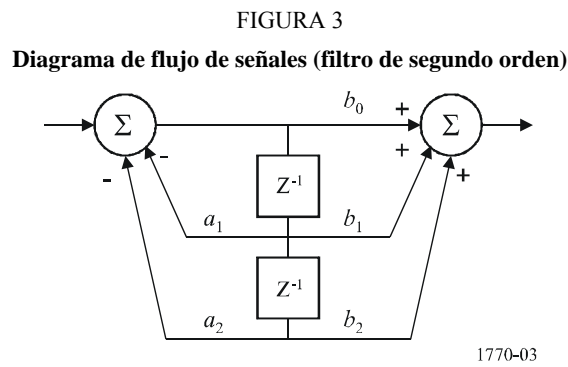
FIGURA 1

**Diagrama de bloques del algoritmo de medición de la sonoridad multicanal**

La primera etapa del algoritmo efectúa un prefiltrado de la señal antes de la medición de  $Leq(RLB)$ , como se ilustra en la Fig. 2. El prefiltrado representa los efectos acústicos de la cabeza, que se ha modelado como una esfera rígida.



El filtro previo está definido por el filtro de la Fig. 3, con los coeficientes que se especifican en el Cuadro 1.



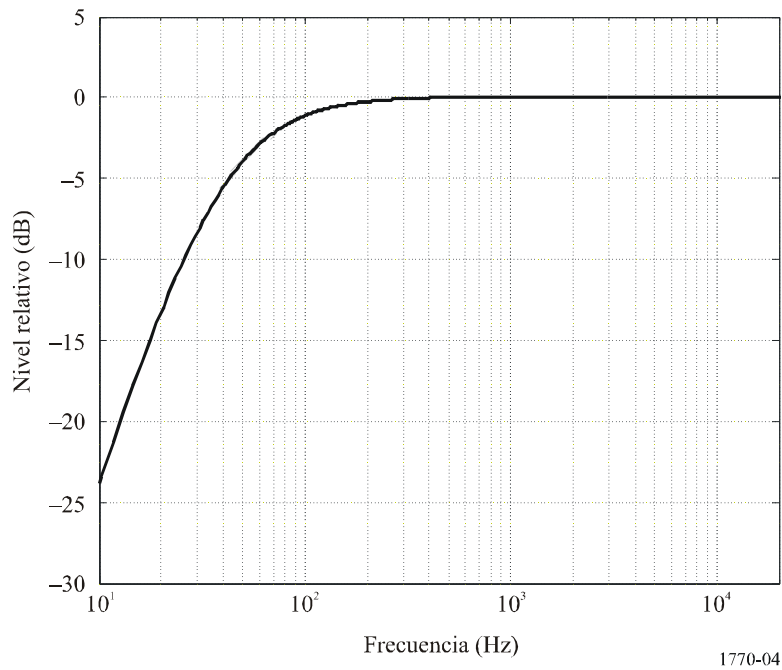
CUADRO 1  
**Coefficientes del filtro para el filtro previo a fin de modelar una cabeza esférica**

		$b_0$	1,53512485958697
$a_1$	-1,69065929318241	$b_1$	-2,69169618940638
$a_2$	0,73248077421585	$b_2$	1,19839281085285

Esos coeficientes de filtro son para una velocidad de muestreo de 48 kHz. Con otras velocidades de muestreo será necesario utilizar coeficientes diferentes, que se escogerán de manera que proporcionen la misma respuesta en frecuencia a 48 kHz que la del filtro especificado. Debido a la precisión interna del hardware disponible, puede que sea necesario cuantificar los valores de esos coeficientes. Las pruebas han demostrado que los resultados del algoritmo no son sensibles a las pequeñas variaciones de esos coeficientes.

La segunda etapa del algoritmo aplica la curva de ponderación RLB, que consiste en un filtro paso alto único (véase la Fig. 4).

FIGURA 4  
Curva de ponderación RLB



La curva de ponderación RLB se especifica como un filtro de segundo orden (véase la Fig. 3), con los coeficientes que se especifican en el Cuadro 2.

CUADRO 2

**Coefficientes del filtro para la curva de ponderación RLB**

		$b_0$	1,0
$a_1$	-1,99004745483398	$b_1$	-2,0
$a_2$	0,99007225036621	$b_2$	1,0

Esos coeficientes del filtro son para una velocidad de muestreo de 48 kHz. Con otras velocidades de muestreo será necesario utilizar coeficientes diferentes, que se escogerán de manera que proporcionen la misma respuesta en frecuencia a 48 kHz que la del filtro especificado.

Una vez aplicado el filtro previo y realizado el filtrado RLB, se mide la energía cuadrática media en el intervalo de medición  $T$  como:

$$z_i = \frac{1}{T} \int_0^T y_i^2 dt, \tag{1}$$

donde  $y_i$  es la señal de entrada filtrada dos veces, por el filtro previo, para modelar los efectos de la cabeza, y por la curva de ponderación RLB. ( $i = L, R, C, Ls, Rs, N$  donde  $N$  es el número de canales).

Una vez calculado para cada canal el nivel cuadrático medio ponderado,  $z_i$ , la etapa final consiste en sumar los  $N$  canales de la siguiente manera:

$$\text{Sonoridad} = -0,691 + 10 \log_{10} \sum_i^N G_i \cdot z_i \quad \text{LFKS} \tag{2}$$

La ponderación de frecuencia en esta medida, que es la concatenación del prefiltrado y la ponderación RLB, se designa como ponderación K. El resultado numérico del valor de la sonoridad que se calcula mediante la ecuación (2) debería ir seguido por la designación LKFS (*Loudness, K weighted, relative to nominal full scale*) que significa, sonoridad, ponderada en K, en relación con la plena escala nominal. La unidad LKFS es equivalente a un decibelio, en el sentido de que un aumento en el nivel de una señal en 1 dB hará que la lectura de la sonoridad aumente en 1 LKFS.

Si se aplica una onda sinusoidal de 1 kHz a 0 dB FS a la entrada del canal izquierdo, central o derecho, la sonoridad indicada será igual a -3,01 LKFS.

En el Cuadro 3 figuran los coeficientes de ponderación para los diferentes canales.

CUADRO 3

**Coefficientes de ponderación para los canales de audio individuales**

Canal	Ponderación, $G_i$
Izquierdo ( $G_L$ )	1 (0 dB)
Derecho ( $G_R$ )	1 (0 dB)
Central ( $G_C$ )	1 (0 dB)
Sonido ambiente izquierdo ( $G_{Ls}$ )	1,41 (~ +1,5 dB)
Sonido ambiente derecho ( $G_{Rs}$ )	1,41 (~ +1,5 dB)

Cabe observar que si bien este algoritmo ha demostrado ser eficaz para los programas de audio que tienen un contenido de radiodifusión típico, de manera general no es adecuado para evaluar la sonoridad subjetiva de tonos puros.

## Apéndice 1 al Anexo 1

### Descripción y desarrollo del algoritmo de medición multicanal

En este Apéndice se describe un algoritmo recientemente elaborado para medir de manera objetiva la sonoridad percibida de las señales de audio. Puede utilizarse para medir con precisión la sonoridad de señales monofónicas, estereofónicas y multicanal. Una de las ventajas fundamentales del algoritmo propuesto es su sencillez, lo que permite su aplicación a muy bajo coste. En este Apéndice se describen también los resultados de las pruebas subjetivas formales realizadas para crear un banco de datos subjetivos que se han de utilizar para evaluar los resultados del algoritmo.

#### 1 Introducción

Hay numerosas aplicaciones en las que es necesario medir y controlar la sonoridad percibida de las señales de audio, como por ejemplo, las aplicaciones de radiodifusión y televisión, en que la naturaleza y el contenido del material de audio cambian con frecuencia. En esas aplicaciones, el contenido de audio puede alternar continuamente entre música, palabra y efectos sonoros, o alguna combinación de ellos. Esos cambios en el contenido del material de los programas pueden traer consigo cambios significativos de la sonoridad subjetiva. Además, a menudo se aplican a las señales diversas formas de procesamientos dinámicos que pueden producir un efecto importante en la sonoridad percibida de la señal. Evidentemente, el problema de la sonoridad subjetiva reviste también gran importancia para la industria de la música en la que habitualmente se utiliza procesamiento dinámico para maximizar la sonoridad percibida de una grabación.

Desde hace algunos años, el Grupo de Trabajo 6P de Radiocomunicaciones lleva a cabo esfuerzos a fin de determinar un medio objetivo de medición de la sonoridad percibida de material de programas típicos para aplicaciones de radiodifusión. En la primera etapa de su labor el UIT-R examinó los algoritmos para la sonoridad monofónica objetiva exclusivamente, y comprobó que una medición de la media cuadrática ponderada,  $Leq(RLB)$ , permitía obtener los mejores resultados en el caso de las señales monofónicas [Soulodre, 2004].

Es evidente que para las aplicaciones de radiodifusión, se necesita un medidor de sonoridad que pueda funcionar con señales monofónicas, estereofónicas y multicanal. El presente documento propone un nuevo algoritmo de medición de la sonoridad que funciona satisfactoriamente con estos tipos de señales de audio. Este algoritmo se basa en una extensión directa del algoritmo  $Leq(RLB)$ . Además, el nuevo algoritmo multicanal conserva la reducida complejidad informática del algoritmo monofónico  $Leq(RLB)$ .

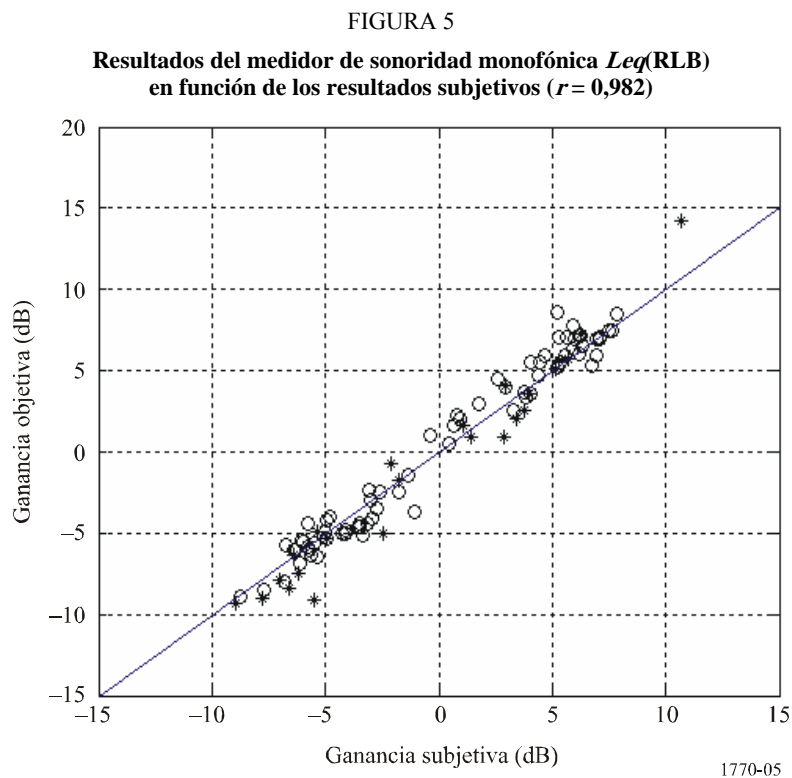
#### 2 Antecedentes

En la primera etapa del estudio del UIT-R, se elaboró un método de prueba subjetivo para determinar la percepción de sonoridad de los materiales típicos de programas monofónicos [Soulodre, 2004]. Las pruebas subjetivas se llevaron a cabo en cinco lugares diferentes del mundo, a fin de crear un banco de datos subjetivos destinado a evaluar los resultados de los algoritmos potenciales de medición de la sonoridad. Se procedió a ajustar la sonoridad de diversas secuencias de audio monofónicas con una secuencia de referencia. Para las secuencias de audio se utilizó material de radiodifusión real (televisión y radio).

Conjuntamente con esas pruebas, siete proponentes diferentes presentaron diez medidores/algoritmos de sonoridad monofónica desarrollados comercialmente para su evaluación por el Laboratorio de Percepción Sonora del Centro de Investigaciones en Comunicaciones (*Communications Research Centre*), Canadá.

Además, Soulodre contribuyó con dos algoritmos adicionales básicos de sonoridad para su utilización como base de resultados [Soulodre, 2004]. Ambas mediciones objetivas consistían en una simple función de ponderación de frecuencia, seguida de un bloque de medición de la media cuadrática. Una de las dos mediciones, *Leq*(RLB), utiliza una curva de ponderación de frecuencia de paso alto denominada curva-B de baja frecuencia revisada (RLB).

La otra medida, *Leq*, es simplemente una medición del valor cuadrático medio sin ponderación.



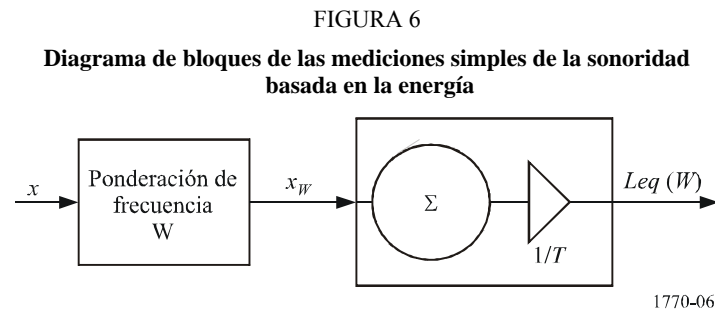
La Fig. 5 muestra los resultados del estudio inicial del UIT-R relativo al medidor de sonoridad *Leq*(RLB). El eje horizontal indica la sonoridad subjetiva relativa obtenida del banco de datos subjetivos, mientras que el eje vertical indica la sonoridad prevista por la medición de *Leq*(RLB). Cada punto del gráfico representa el resultado de una de las secuencias de audio de la prueba. Los círculos representan secuencias de audio basadas en la voz, mientras que los asteriscos corresponden a secuencias no basadas en la voz. Puede observarse que los puntos están muy agrupados en torno de la diagonal, lo que indica que los resultados del medidor *Leq*(RLB) son muy buenos.

Se observó que de todos los medidores evaluados, el que arrojaba mejores resultados era el *Leq*(RLB) (aunque estadísticamente los resultados de algunos medidores basados en la psicoacústica eran igualmente buenos). Se observó que los resultados del *Leq* eran casi tan buenos como los de RLB. Esas observaciones indican que para un material de radiodifusión monofónica típico, una simple medición de la sonoridad basada en la energía es, comparada con las mediciones más complejas que pueden comprender modelos perceptivos detallados, igualmente buena.

### 3 Diseño del algoritmo $Leq(RLB)$

El algoritmo de sonoridad  $Leq(RLB)$  se diseñó específicamente para ser muy sencillo. En la Fig. 6 se ilustra un diagrama de bloque del algoritmo  $Leq(RLB)$ , que consiste en un filtro paso alto seguido de un dispositivo para promediar la energía a lo largo del tiempo. La salida del filtro se introduce en un bloque de procesamiento que suma la energía y calcula el promedio a lo largo del tiempo.

El objeto del filtro es proporcionar una cierta ponderación perceptiva del contenido espectral de la señal. Una de las ventajas de utilizar esta estructura básica para medir la sonoridad, radica en que todos los procesamientos se pueden llevar a cabo con simples bloques de dominio del tiempo, y recurriendo poco a la informática.



El algoritmo  $Leq(RLB)$  que se ilustra en la Fig. 6, es simplemente una versión con frecuencia ponderada de una medición de *nivel de sonido equivalente* ( $Leq$ ), que se define de la siguiente manera:

$$Leq(W) = 10 \log_{10} \left[ \frac{1}{T} \int_0^T \frac{x_W^2}{x_{Ref}^2} dt \right] \quad \text{dB} \quad (3)$$

donde:

- $x_W$ : señal a la salida del filtro de ponderación
- $x_{Ref}$ : nivel de referencia
- $T$ : longitud de la secuencia de audio

El símbolo  $W$  en  $Leq(W)$  representa la ponderación de frecuencia, que en este caso era la curva-B de baja frecuencia revisada (RLB).

### 4 Pruebas subjetivas

A fin de evaluar las posibles mediciones de sonoridad multicanal es necesario realizar pruebas subjetivas formales destinadas a crear un banco de datos subjetivos. A continuación, los posibles algoritmos de medición de la sonoridad se pueden evaluar según su capacidad para predecir los resultados de las pruebas subjetivas. El banco de datos proporciona clasificaciones de la sonoridad percibida para una amplia diversidad de materiales de programas monofónicos, estereofónicos y multicanal. Para las pruebas, se utilizó material de programas perteneciente a emisiones reales de radio y televisión de todo el mundo, y también se utilizaron CD y DVD. En las secuencias se incluyó música, películas de televisión y de cine, acontecimientos deportivos, radiodifusión de noticias, efectos sonoros y publicidad. En esas secuencias se incluyeron asimismo partes habladas en varios idiomas.

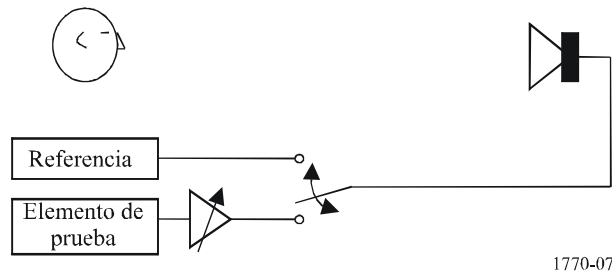


#### 4.1 Realización de la prueba subjetiva

Las pruebas subjetivas consisten en establecer una correspondencia de sonoridad. Las personas escuchan una amplia gama de material de programas típicos y ajustan el nivel de cada elemento de prueba hasta que la sonoridad percibida corresponde a la señal de referencia (véase la Fig. 7).

La señal de referencia se reproduce siempre a un nivel de 60 dBA, valor determinado por Benjamin para que sea un nivel típico de escucha de los televidentes de los hogares reales [Benjamin, 2004].

FIGURA 7  
Metodología de la prueba subjetiva

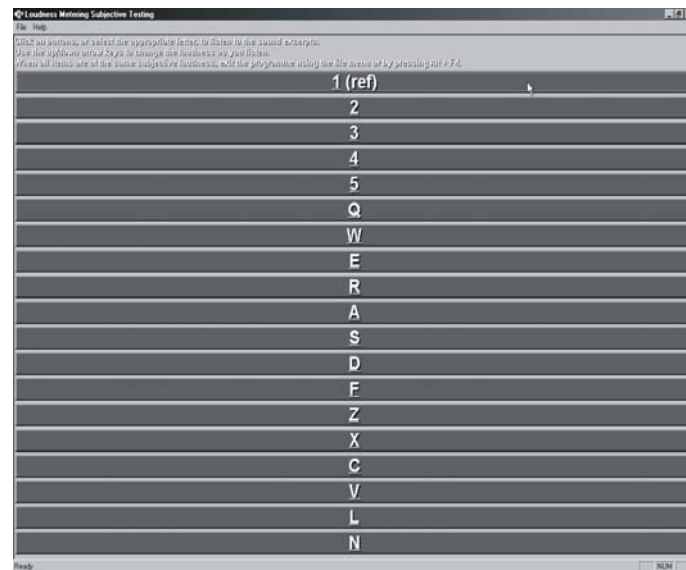


Mediante un sistema de pruebas subjetivas multicanal basado en programas informáticos, desarrollado y financiado por la Australian Broadcasting Corporation, los oyentes pueden cambiar instantáneamente de elemento de prueba y ajustar el nivel (sonoridad) de cada uno de ellos. En la Fig. 8 puede verse una pantalla de captura de los software de prueba. El nivel de los elementos de prueba se puede ajustar por pasos de 0,25 dB. Si se selecciona el botón de «1», se accede a la señal de referencia, cuyo nivel se mantiene fijo.

Mediante el teclado del ordenador, el sujeto selecciona un elemento de prueba determinado y ajusta su nivel hasta que la sonoridad corresponda a la señal de referencia. Los sujetos pueden pasar instantáneamente de uno cualquiera de los elementos de prueba a otro, seleccionando la tecla apropiada. Las secuencias se repiten ininterrumpidamente (en bucle) durante las pruebas. El programa informático registra las ganancias establecidas para cada elemento de prueba según lo decide el sujeto. Por consiguiente, las pruebas subjetivas producen un conjunto de valores de ganancia (dB) necesario para hacer corresponder la sonoridad de cada secuencia de prueba con la secuencia de referencia. Ello permite determinar directamente la sonoridad relativa de cada elemento de prueba.

FIGURA 8

## Interfaces de usuario del sistema de prueba subjetiva



1770-08

Antes de pasar a las pruebas ciegas formales, cada sujeto sigue una sesión de formación durante la cual se familiariza con los programas informáticos de prueba y con lo que debe hacer en el experimento. Como muchos de los elementos de prueba contienen una mezcla de voz y otros sonidos (música, ruidos de fondo, etc.), se instruye particularmente a los sujetos para que hagan corresponder la sonoridad de la señal global y no sólo el componente vocal de las señales.

Durante las pruebas ciegas formales, los elementos de prueba se presentan a cada sujeto en orden aleatorio. Por consiguiente, el orden de presentación de los elementos de prueba nunca será el mismo para dos sujetos distintos. La razón de ello es eliminar toda posibilidad de distorsión del resultado debido a un efecto de orden.

#### 4.2 El banco de datos subjetivo

El banco de datos subjetivo utilizado para evaluar los resultados del algoritmo propuesto consiste en tres conjuntos de datos separados, creados a partir de tres pruebas subjetivas independientes llevadas a cabo durante varios años.

El primer conjunto de datos es el resultado del estudio original del UIT-R en que los sujetos hacían corresponder la sonoridad percibida de 96 secuencias monofónicas de audio. Para crear este conjunto de datos, se llevaron a cabo pruebas subjetivas en cinco lugares diferentes del mundo con un total de 97 oyentes. Un panel de tres personas formado por miembros del GT 6P SRG3 de Radiocomunicaciones seleccionaron la secuencia de prueba, así como los elementos de referencia. La señal de referencia de este experimento consistió en la voz de mujeres de habla inglesa. Las secuencias se escuchaban a través de un único altavoz situado directamente frente al oyente.

Como resultado del estudio monofónico original del UIT-R, algunos de los proponentes de algoritmos argumentaron que la gama y el tipo de señales que se utilizaron para las pruebas subjetivas no eran suficientemente amplios. Más adelante, afirmaron que por esa razón el algoritmo simple basado en la energía  $Leq(RLB)$  había superado a todos los demás algoritmos.

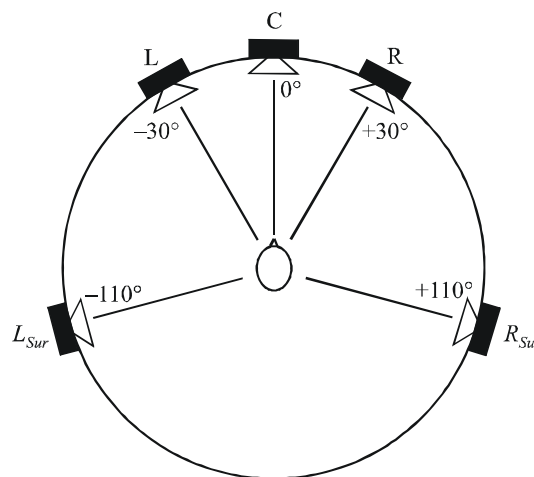
A fin de hacer frente a esta preocupación, se solicitó a los proponentes que presentasen nuevas secuencias de audio para realizar una nueva ronda de pruebas subjetivas. Se les alentó a que realizaran las secuencias monofásicas que consideraban con mayores posibilidades de desafiar al algoritmo  $Leq(RLB)$ . Sólo dos de los proponentes presentaron nuevas secuencias.

Utilizando esas nuevas secuencias, se llevaron a cabo pruebas subjetivas formales en el Laboratorio de Percepción Sonora del Centro de Investigaciones en Comunicaciones, Canadá. Un total de 20 personas hicieron clasificaciones de la sonoridad para 96 secuencias monofónicas. Para las pruebas se utilizó la misma metodología subjetiva que para la creación del primer banco de datos, y también se utilizó la misma señal de referencia. Los resultados de esas pruebas formaron el segundo conjunto de datos del banco de datos subjetivos.

El tercer conjunto de datos consistió en clasificaciones de sonoridad correspondientes a 144 secuencias de audio. Las secuencias de prueba consistieron en 48 elementos monofónicos, 48 elementos estereofónicos y 48 elementos multicanal. Además, la mitad de los elementos monofónicos se reprodujo a través del canal central (mono), mientras que la otra mitad de esos elementos se reprodujo a través de los altavoces izquierdo y derecho (mono dual). Se procedió de esta manera para representar las dos formas diferentes en que se puede escuchar una señal monofónica. Para esta prueba, la señal de referencia consistió en la voz de una mujer de habla inglesa con ambiente estereofónico y un nivel bajo de música de fondo. En esta prueba, en que se utilizó la configuración de altavoces especificada en la Recomendación UIT-R BS.775 y que se describe en la Fig. 9, participaron en total 20 personas.

FIGURA 9

Configuración de altavoces utilizada para el tercer conjunto de datos



1770-09

Los dos primeros conjuntos de datos se limitaron a secuencias de prueba monofónicas y, por consiguiente, las imágenes no intervinieron. En el tercer conjunto de datos, que también comprendió secuencias estereofónicas y multicanal, las imágenes fueron una consideración importante que debía tenerse en cuenta. Se estimó que las imágenes y el ambiente dentro de una secuencia podían tener un efecto significativo en la sonoridad percibida de la secuencia. Por consiguiente, se escogieron secuencias estereofónicas y multicanal a fin de incluir una amplia gama de estilos de imágenes (por ejemplo, panorámico central con respecto a izquierdo y derecho no panorámico, fuentes en el frente con respecto a fuentes todo alrededor) y de variar diferentes ambientes (por ejemplo, seco frente a reverberante).

El hecho de que las personas tenían que hacer corresponder simultáneamente la sonoridad de señales monofónicas, dual monofónicas, estereofónicas y multicanal significa que la dificultad de esta prueba es por su naturaleza mayor que la de las precedentes, que se limitaban a señales monofónicas. La dificultad aumentó debido a la diversidad de estilos de imágenes y a la variación

de ambiente. Hubo una cierta inquietud pues se temió que a raíz de esos factores las personas pudiesen sentirse abrumadas por la prueba. Afortunadamente, las pruebas preliminares mostraron que la tarea era razonable y que las 20 personas eran realmente capaces de llegar a resultados coherentes.

## 5 Diseño del algoritmo de sonoridad multicanal

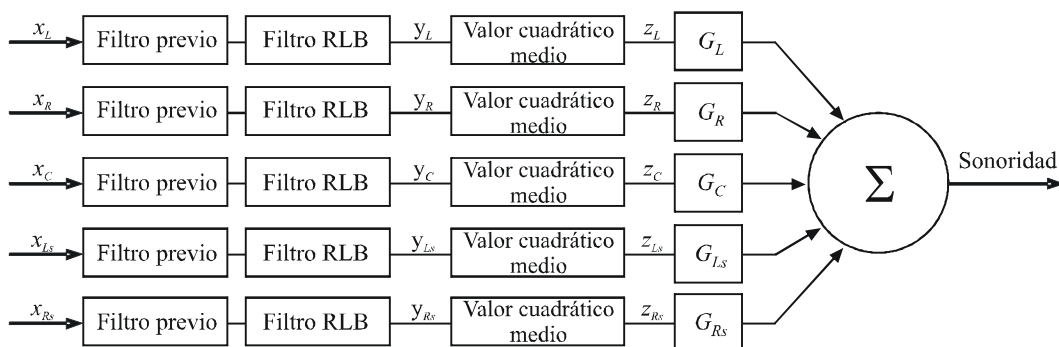
Como ya se indicó, el algoritmo *Leq*(RLB) se diseñó para operar con señales monofónicas, y en un estudio precedente se demostró que para esos casos sus resultados eran satisfactorios. El diseño de un algoritmo de sonoridad multicanal trae aparejados ciertos retos adicionales. Un requisito fundamental para que un algoritmo multicanal sea satisfactorio es que debe funcionar igualmente bien tanto para las señales monofónicas como para las dual monofónicas y las estereofónicas. Es decir, esos formatos deben considerarse como casos especiales de una señal multicanal (aunque los casos son sumamente comunes).

En el presente estudio suponemos que las señales multicanal se ajustan a la configuración normalizada de canales 5.1 de la Recomendación UIT-R BS.775. No se ha hecho ningún esfuerzo para representar el canal LFE.

Con el medidor de sonoridad multicanal, la sonoridad de cada canal de audio individual se mide de manera independiente mediante el algoritmo monofónico *Leq*(RLB) (véase la Fig. 10). No obstante, antes de efectuar la medición con *Leq*(RLB) se aplica un prefiltrado a cada canal.

FIGURA 10

Diagrama de bloques del medidor de sonoridad multicanal propuesto



1770-10

El objeto del prefiltrado es eliminar los efectos acústicos que se producen en la cabeza debido a las señales entrantes. Aquí, la cabeza ha sido modelada como una esfera rígida. El mismo prefiltrado se aplica a cada canal. Posteriormente, los valores de sonoridad resultantes se ponderan ( $G_i$ ) según el ángulo de llegada de la señal, y se suman (en el dominio lineal) para obtener una medición compuesta de la sonoridad. Las ponderaciones se efectúan para permitir que el agente perciba los sonidos que le llegan por detrás con mayor intensidad que los sonidos que le llegan de frente.

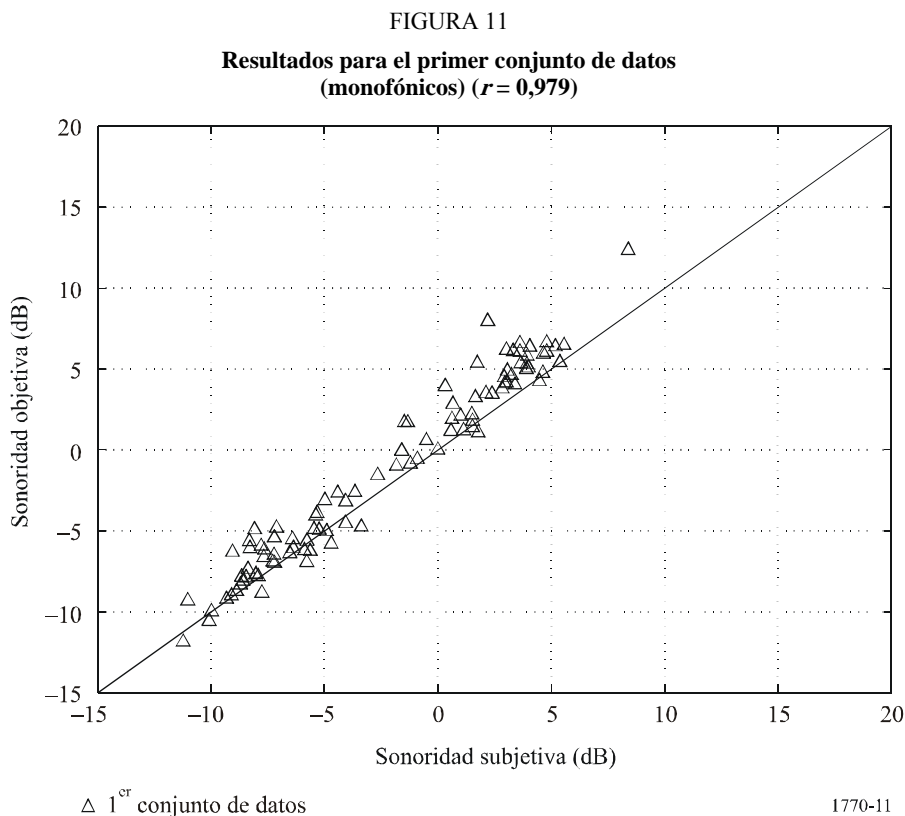
Una ventaja fundamental del algoritmo de sonoridad multicanal propuesto es su sencillez. Es un algoritmo constituido enteramente de bloques sumamente básicos de procesamiento de la señal que se puede aplicar en el dominio temporal a equipos pocos costosos. Otra ventaja clave de este algoritmo es su escalabilidad. Como el procesamiento aplicado a cada canal es idéntico, resulta sumamente sencillo realizar un medidor que pueda tener en cuenta cualquier número de canales, desde 1 hasta  $N$ . Además, puesto que las contribuciones de los canales individuales se suman como

valores de sonoridad, más que a nivel de la señal, el algoritmo no depende de la fase ni de la correlación entre canales. Ello hace que la medición de la sonoridad propuesta sea mucho más genérica y sólida.

## 6 Evaluación del algoritmo multicanal

Las 336 secuencias de audio utilizadas en los tres conjuntos de datos se procesaron mediante el algoritmo multicanal propuesto y se registraron las clasificaciones de sonoridad previstas. Como resultado de este proceso, se pudieron evaluar los resultados globales del algoritmo en base a la concordancia entre las clasificaciones previstas y las clasificaciones subjetivas reales obtenidas con las pruebas subjetivas formales.

Las Figs. 11, 12 y 13 muestran los resultados del medidor de sonoridad propuesto para los tres conjuntos de datos. En cada figura, el eje horizontal indica la sonoridad subjetiva de cada secuencia de audio en el conjunto de datos. El eje vertical indica la sonoridad objetiva prevista por el medidor de sonoridad propuesto. Cada punto del gráfico representa el resultado correspondiente a una secuencia de audio individual. Cabe observar que un algoritmo objetivo perfecto daría como resultado que todos los puntos de datos estarían situados sobre la diagonal de pendiente 1, que pasa por el origen (como lo ilustran las figuras).

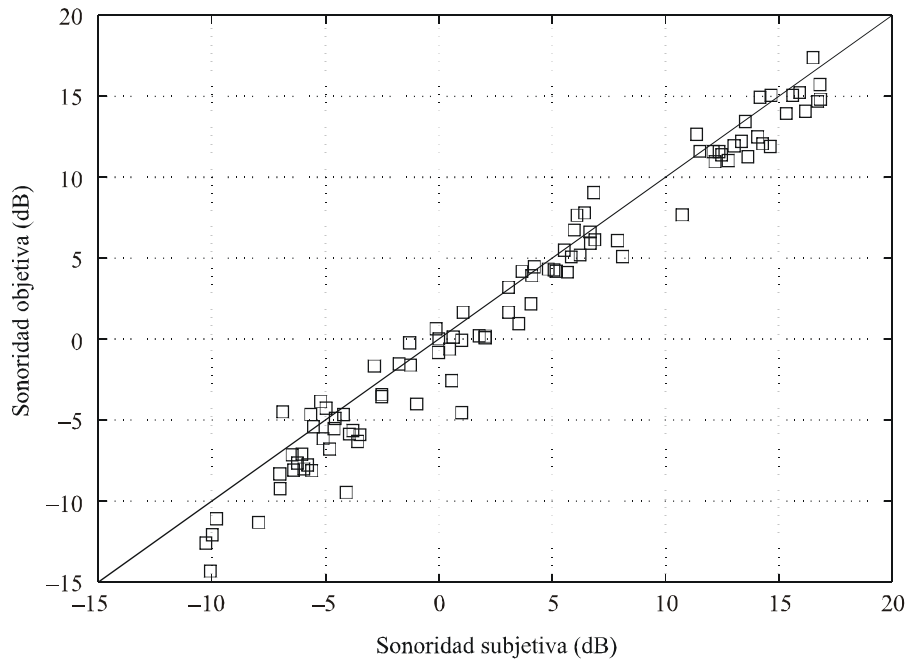


Puede observarse en la Fig. 11 que el algoritmo de sonoridad multicanal propuesto prevé con mucha precisión los resultados del primer conjunto de datos (monofónico). La correlación entre las clasificaciones de sonoridad subjetiva y las mediciones de sonoridad objetiva es de  $r = 0,979$ .

En la Fig. 12, puede verse que la correlación entre las clasificaciones de sonoridad subjetiva y las mediciones de sonoridad objetiva para el segundo conjunto de datos es también muy buena ( $r = 0,985$ ). Es interesante señalar que aproximadamente la mitad de las secuencias de este conjunto de datos eran musicales.

FIGURA 12

Resultados para el segundo conjunto de datos  
(monofónicos) ( $r = 0,985$ )

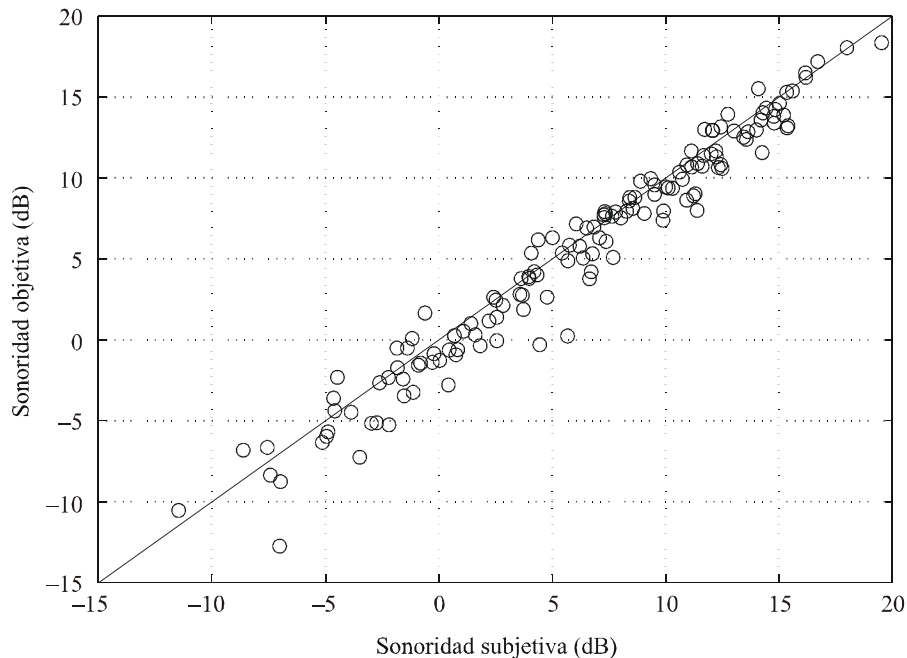


□ 2<sup>o</sup> conjunto de datos

1770-12

FIGURA 13

Resultados para el tercer conjunto de datos  
(monofónicos, estereofónicos y multicanal) ( $r = 0,980$ )



○ 3<sup>er</sup> conjunto de datos

1770-13

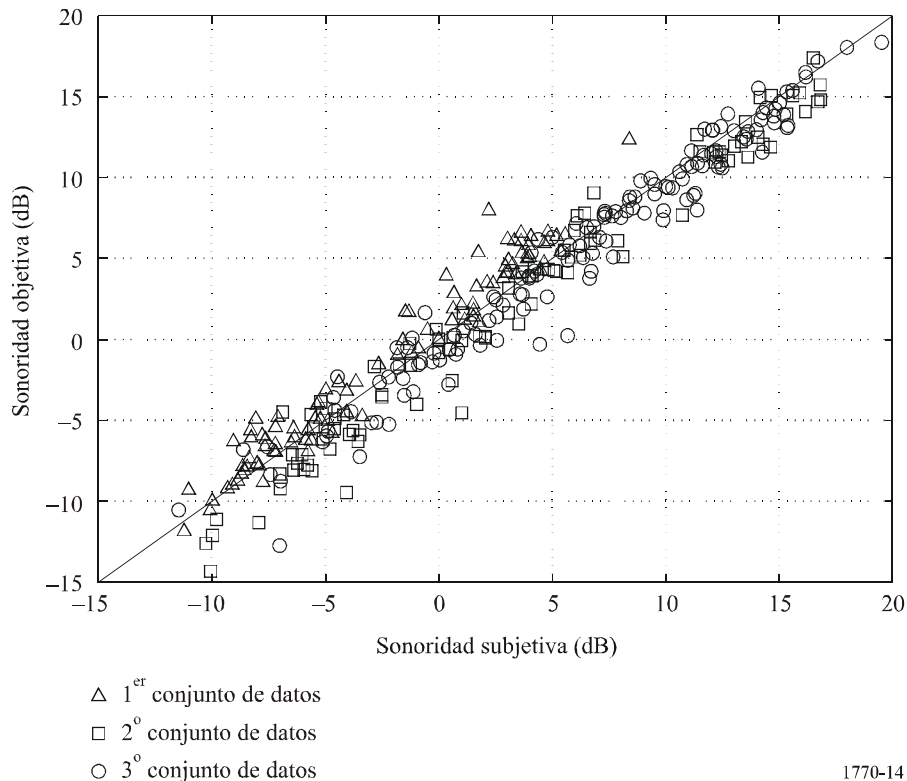
La Fig. 13 ilustra los resultados correspondientes al tercer conjunto de datos, que comprende señales monofónicas, dual monofónicas, estereofónicas y multicanal. Los resultados para el caso multicanal que aparecen en las Figs. 13 y 14 corresponden a algoritmos especificados, pero con las ponderaciones del canal de sonido ambiente puestas a 4 dB (propuesta original) en lugar de 1,5 dB

(especificación final). Se ha verificado que el cambio de 4 dB a 1,5 dB no tiene ningún efecto significativo en los resultados. Una vez más, se comprueba que los resultados del algoritmo son muy buenos, con una correlación de  $r = 0,980$ .

Es útil examinar los resultados del algoritmo para la totalidad de las 336 secuencias de audio que forman el banco de datos subjetivos. Por consiguiente, la Fig. 14 combina los resultados de los tres conjuntos de datos. Puede observarse que los resultados son sumamente buenos para todo el banco de datos subjetivos, con una correlación global de  $r = 0,977$ .

FIGURA 14

**Resultados combinados para los tres conjuntos de datos ( $r = 0,977$ )**



1770-14

Los resultados de esta evaluación indican que el algoritmo de medición de la sonoridad multicanal, basado en la medición de sonoridad  $Leq$ (RLB), funciona muy bien para las 336 secuencias del banco de datos subjetivos. Este banco de datos proporciona una amplia gama de material de programas, entre los que figuran música, películas de televisión y de cine, acontecimientos deportivos, radiodifusión de noticias, efectos sonoros y publicidad. También se han incluido en las secuencias partes de voz en varios idiomas. Además, los resultados demuestran que el medidor de sonoridad propuesto funciona perfectamente con señales monofónicas, dual monofónicas, estereofónicas y multicanal.

## Referencias Bibliográficas

- SOULODRE, G.A. [mayo de 2004] Evaluation of objective loudness meters. 116th Convention of the Audio Engineering Society, Berlin, Preprint 6161.
- BENJAMIN, E. [octubre de 2004] Preferred listening levels and acceptance windows for dialog reproduction in the domestic environment. 117th Convention of the Audio Engineering Society, San Francisco, Preprint 6233.

## Anexo 2

### Directrices para una medición precisa del nivel «de cresta real»

En este Anexo se describe un algoritmo destinado a evaluar el nivel de cresta real dentro de una señal de audio digital MIC lineal de un solo canal. Para la discusión siguiente se supone que la velocidad de muestreo es de 48 kHz. El nivel de cresta real es el máximo valor (positivo o negativo) de la onda de la señal en el dominio de tiempo continuo, y puede ser superior al mayor valor de muestreo en el dominio del tiempo muestreado a 48 kHz. El algoritmo permite estimar la señal tal como es y, opcionalmente, cómo sería en el caso de que algún equipo situado más adelante suprimiese el componente de cc de la señal. Una preacentuación opcional moderada de alta frecuencia en el trayecto de la señal de medición de cresta puede permitir al algoritmo notificar, para señales de alta frecuencia, un nivel de cresta superior al que realmente existe. El objeto de ello es que los desplazamientos de fase de las subsiguientes etapas del procesamiento de la señal (como los filtros de Nyquist) puedan producir un crecimiento de las crestas de las señales de alta frecuencia; en algunas aplicaciones esta característica puede ser útil a fin de brindar una mejor protección contra un recorte que tenga lugar más adelante.

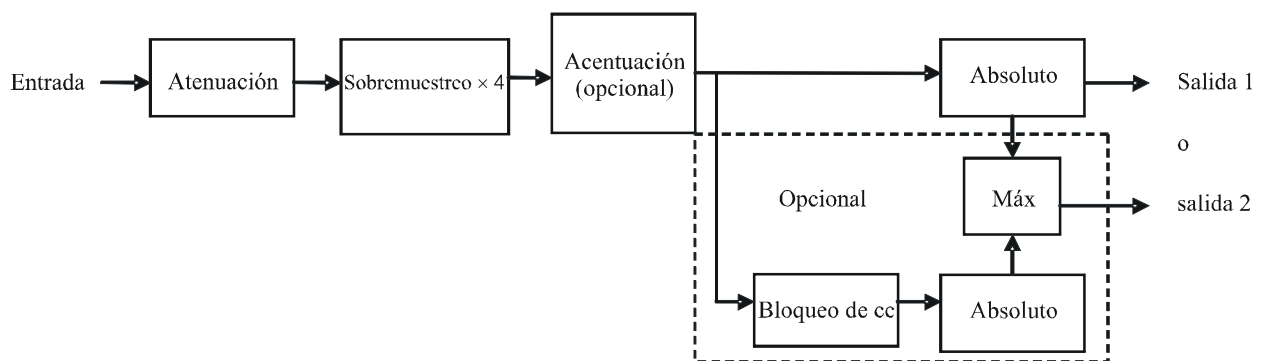
#### 1 Resumen

Las etapas del proceso son las siguientes:

- 1 Atenuación: atenuación de 12,04 dB
- 2 Sobremuestreo  $\times 4$
- 3 Acentuación: Filtro de estantería («shelving filter») de preacentuación con un cero a 14,1 kHz y un polo a 20 kHz (opcional)
- 4 Bloqueo de cc (opcional)
- 5 Absoluto: Valor absoluto
- 6 Max: Detección de valor más elevado (opcional, se incluye si está incluido el bloqueo de cc).

La detección del valor absoluto, tanto antes como después del bloqueo de cc, permite estimar el nivel de cresta de la señal en el punto actual de medición, así como el nivel de cresta si en algún dispositivo situado más adelante se suprime la componente de cc de la señal.

#### 2 Diagrama de bloques





### 3 Descripción detallada

La primera etapa consiste en imponer una atenuación de 12,04 dB (desviación de 2 bits). El propósito de esta operación es proporcionar un margen para el procesamiento de la señal subsiguiente utilizando una aritmética de números enteros. Esta etapa no es necesaria si los cálculos se efectúan con coma flotante.

El filtro de sobremuestreo  $\times 4$  aumenta la velocidad de muestreo de la señal, que pasa de 48 kHz a 192 kHz. Esta versión de la velocidad de muestreo más alta de la señal indica con mayor precisión la forma de onda real que se representa dentro de la señal. Se prefieren las velocidades de muestreo y las tasas de sobremuestreo más elevadas (véase el Apéndice 1 a este Anexo). Las señales entrantes con velocidades de muestreo más altas exigen proporcionalmente un sobremuestreo menor (por ejemplo, para una señal entrante con una velocidad de muestreo de 96 kHz, bastará con un sobremuestreo  $\times 2$ ).

El filtro opcional de estantería de preacentuación hace que el algoritmo indique un nivel de cresta más elevado para las componentes de la señal de frecuencia más alta. Esto puede lograrse al tener en cuenta que la dificultad de medir y controlar los valores de cresta de las componentes de la señal de frecuencia más alta es mayor debido a los efectos de dispersión (desplazamiento de fase) que se producen en los numerosos filtros de Nyquist que aparecen frecuentemente a lo largo de una cadena de señales de radiodifusión.

El filtro opcional de bloqueo de cc cubre el caso en que la señal es muy asimétrica o contiene algunos desplazamientos de cc. Además de permitir medir el valor de cresta de la señal actual (incluidos la asimetría y el desplazamiento de cc), la inclusión de esta sección opcional permite medir la señal tal como sería si algún dispositivo situado más adelante incluyera un filtro de bloqueo de cc.

El valor absoluto de las muestras se obtiene mediante la inversión de las muestras de valor negativo. En ese punto, la señal es unipolar, con valores negativos que han sido sustituidos por valores positivos de la misma magnitud. La Salida 1 es el tren de valores de salida si no se introduce el bloqueo de cc opcional.

Si se introduce, el bloqueo «MAX» selecciona la más amplia de cada una de las muestras procedentes de los dos trayectos de señal. En ese caso, la salida se toma de la Salida 2.

Los bloqueos de sistema subsiguientes (no figuran ni se especifican en el presente documento) permiten comparar los valores de muestra de salida con el nivel nominal de señal de cresta del 100% (1/4 de la escala total si se aplicaron a la entrada 12 dB de atenuación), lo que da una estimación del nivel de cresta real con respecto a la escala total digital.

Los sonómetros que están en consonancia con estas directrices y utilizan una frecuencia de muestreo sobremuestreada de por lo menos 192 kHz, deben indicar el resultado en unidades de dB TP. Esta designación expresa los decibelios en relación con la medición de cresta real en plena escala de 100%.

## Apéndice 1\* al Anexo 2

### Consideraciones para una medición precisa de las crestas de las señales de audio digitales

#### *Planteamiento del problema*

En los sistemas de audio digitales, los medidores de cresta registran a menudo los valores «de cresta de la muestra» y no los valores «de cresta reales».

Un medidor de cresta de la muestra en general funciona mediante la comparación del valor absoluto (rectificado) de cada muestra entrante con el valor obtenido en el medidor. Si la nueva muestra es superior, sustituye a la lectura actual; de lo contrario, la lectura actual se multiplica por una constante ligeramente inferior a la unidad para producir una descomposición logarítmica. Esos medidores están omnipresentes ya que son fáciles de realizar, pero no siempre registran los valores de cresta reales de la señal de audio.

Por consiguiente, la utilización de un medidor de cresta de la muestra cuando es importante medir con precisión las crestas de los programas, puede plantear problemas. Lamentablemente, la mayoría de los medidores de cresta digitales miden las crestas de la muestra, lo que en general no es evidente para el operador.

El problema se plantea ya que los valores de cresta reales de una señal muestreada en general aparecen entre las muestras y no precisamente en el instante muestreado. Por esa razón, el medidor de cresta de la muestra no los registra correctamente.

Ello trae aparejadas varias anomalías conocidas de los medidores de cresta de la muestra:

- *Lecturas de crestas contradictorias*: Se observa a menudo que si se reproduce repetidamente una grabación analógica en un sistema digital con un medidor de cresta de la muestra, en cada reproducción se obtienen lecturas bastante diferentes de las crestas de programa. Del mismo modo, si una grabación digital se reproduce de manera repetida a través de un convertidor de velocidad de muestreo antes de realizar la medición, las crestas registradas son, del mismo modo, diferentes en cada reproducción. Ello se debe a que en cada reproducción los instantes muestreados pueden caer en partes diferentes de la verdadera señal.
- *Sobrecargas inesperadas*: Como las señales muestreadas pueden contener sobrecargas, incluso cuando no tienen ninguna muestra a escala total digital (o próxima de ésta), la indicación de sobrecarga de un medidor de cresta de la muestra es poco fiable. Las sobrecargas pueden ocasionar recortes en los procesos subsiguientes, como por ejemplo en los convertidores *D/A* particulares o durante la conversión de la velocidad de muestreo, incluso si no han sido previamente registradas por el medidor de cresta de la muestra (y aún siendo inaudibles cuando se controlan en ese punto).
- *Sublectura y batido de tonos medidos*: Los tonos puros (como los tonos alineados) próximos de los factores enteros de la frecuencia de muestreo pueden producir sublecturas o lecturas que varían constantemente, incluso si la amplitud del tono es constante.

---

\* NOTA 1 – El siguiente texto informativo es una contribución del Grupo de Trabajo SC-02-01 sobre Normas de la AES, presentada a través del Relator del Grupo de Trabajo 6J de Radiocomunicaciones sobre medición de la sonoridad.

### ¿Gravedad del problema?

De manera general, cuanto más alta es la frecuencia de la señal de cresta de la muestra medida, mayores son las posibilidades de error.

Para tonos puros continuos es fácil demostrar, por ejemplo, una sublectura de 3 dB para un tono de fase errónea a un cuarto de la frecuencia de muestreo. La sublectura para un tono a la mitad de la frecuencia de muestreo puede ser casi infinita. No obstante, la mayoría de las señales de audio digitales no contienen energía significativa a esa frecuencia (ya que está en gran parte excluida por los filtros utilizados para evitar errores de cuantificación («anti-aliasing») en el punto de conversión D/A, y porque los sonidos «reales» en general no están dominados por altas frecuencias continuas).

Los tonos continuos que no se aproximan a los factores enteros bajos de la frecuencia de muestreo no producen sublecturas en los medidores de cresta de la muestra ya que la frecuencia de batido (la diferencia entre  $n \cdot f_{\text{tono}}$  y  $f_s$ ) es elevada en comparación con la inversa de la velocidad de disminución del medidor. En otras palabras, con frecuencia el instante de muestreo se encuentra suficientemente próximo de la cresta real del tono como para que el medidor no produzca sublectura.

No obstante, para los transitorios individuales, ese mecanismo no oculta las sublecturas, por lo que cuanto más alto es el contenido en frecuencias del transitorio, mayor será la sublectura potencial. Es normal que en sonidos «reales» ello suceda para los transitorios con contenidos significativos de alta frecuencia, y la sublectura de esos puede comúnmente ser de varios dB.

Como normalmente los sonidos reales tienen un espectro que disminuye a las frecuencias más altas, y debido a que esta circunstancia no cambia al aumentar la frecuencia de muestreo, la sublectura del medidor de cresta de la muestra es menos severa para frecuencias de muestreo originales más elevadas.

### ¿Solución?

A fin de medir el valor de cresta real de una señal muestreada es necesario «sobremuestrear» (o «aumentar las muestras») la señal, fundamentalmente mediante la recreación de la señal original entre las muestras existentes y, por consiguiente, aumentando la frecuencia de muestreo de la señal. Esta propuesta parece discutible; ¿cómo se puede recrear una información que ya parece perdida? De hecho, según la teoría del muestreo es posible hacerlo, ya que se sabe que la señal muestreada no contiene ninguna frecuencia por encima de la mitad de la frecuencia de muestreo original.

¿Qué tasa de sobremuestreo se necesita? Para saberlo, es necesario responder a las dos preguntas siguientes:

- ¿Cuál es el máximo error de sublectura aceptable?
- ¿Cuál es la relación entre la frecuencia más elevada que se ha de medir y la frecuencia de muestreo (la máxima «frecuencia normalizada»)?

Si conocemos esos criterios, es posible calcular la tasa de sobremuestreo que se necesita (incluso sin tener en cuenta el detalle de la realización del sobremuestreo) mediante un método sencillo de «papel milimetrado». Simplemente, puede considerarse que la sublectura resultará de un par de muestras a la velocidad sobremuestreada que se produce simétricamente a cada lado de la cresta de una senoide a nuestra máxima frecuencia normalizada. Este es el caso de sublectura más desfavorable.

Para: tasa de sobremuestreo,  $n$

máxima frecuencia normalizada,  $f_{\text{norm}}$

frecuencia de muestreo,  $f_s$

se observan que:

el periodo de muestreo para la velocidad sobremuestreada es  $1/n \cdot f_s$

el periodo de la máxima frecuencia normalizada es  $1/f_{norm} \cdot f_s$

de donde resulta:

el máximo valor de la sublectura (dB) es  $20 \cdot \log(\cos(2 \cdot \pi \cdot f_{norm} \cdot f_s / n \cdot f_s \cdot 2))$

(2 en el denominador, ya que puede perderse una cresta por un máximo de la mitad del periodo de sobremuestreo)

o:

máximo valor de la sublectura (dB) =  $20 \cdot \log(\cos(\pi \cdot f_{norm} / n))$

El cuadro que figura a continuación se elaboró utilizando la ecuación precedente, que probablemente abarca la gama de los valores de interés:

Tasa de sobremuestreo	Sublectura (dB) máxima $f_{norm} = 0,45$	Sublectura (dB) máxima $f_{norm} = 0,5$
4	0,554	0,688
8	0,136	0,169
10	0,087	0,108
12	0,060	0,075
14	0,044	0,055
16	0,034	0,042
32	0,008	0,010

#### *Cómo realizar un medidor de cresta real*

La operación de sobremuestreo se lleva a cabo introduciendo muestras de valor cero entre las muestras originales, a fin de generar un tren de datos a la velocidad de muestreo deseada y, posteriormente, aplicando un filtro de «interpolación» paso bajo para excluir las frecuencias superiores a la máxima frecuencia deseada  $f_{norm}$ . Si ahora se aplica un algoritmo de cresta de la muestra a la señal sobremuestreada, se obtiene un medidor de valor de cresta real con la sublectura máxima deseada.

Es interesante examinar la realización de este sobremuestreador. Normalmente se crea ese tipo de filtro paso bajo como un filtro de respuesta a impulso finita (FIR) simétrico. En los casos en que se utilicen filtros de esa clase aplicados a señales audio de alta calidad, por ejemplo, en convertidores D/A de sobremuestreo (aparatos antiguos) o en convertidores de muestra de cresta, será necesario calcular «numerosas derivaciones» a fin de mantener un rizado en la banda de paso muy bajo, y obtener una atenuación muy elevada en la banda rechazada y una banda de transición estrecha. Asimismo, puede ser necesario mantener una palabra de gran longitud a fin de preservar la gama dinámica y reducir al mínimo la distorsión.

No obstante, como no va a escucharse la salida del sobremuestreador, sino que sólo va a utilizarse para visualizar una lectura o controlar un gráfico de barras, probablemente no se necesitará la misma precisión. En la medida en que el rizado de la banda de paso, junto con la adición de componentes no esenciales de la banda rechazada, no degrade la precisión de la lectura más allá de los objetivos, pueden considerarse cumplidos dichos objetivos. Esto reduce considerablemente el número de «derivaciones» necesarias, a pesar de que puede que siga siendo necesario obtener una banda de transición estrecha en función del objetivo de máxima frecuencia normalizada. Del mismo

modo, puede que la longitud de la palabra sólo tenga que ser suficiente para garantizar la precisión del objetivo en el borde inferior del gráfico de barras, a menos que se requieran resultados numéricos precisos para las bajas amplitudes.

Por consiguiente, es posible que se pueda realizar adecuadamente un sobremuestreador apropiado (posiblemente para numerosos canales) en un procesador de señales digitales o en un circuito lógico programable por campo corrientes de bajo coste, o incluso en un procesador aún más modesto. Por otro lado, se han realizado medidores de sobremuestreo utilizando microplaquetas de sobremuestreo de alta precisión destinadas a los convertidores *D/A*. Si bien esto es un desperdicio de silicio y de energía, los dispositivos son de bajo coste y su lectura es inmediata.

La manera más simple de determinar el número de derivaciones y los coeficientes de derivación necesarios para la especificación de un medidor particular es utilizar un programa de diseño de un FIR recursivo, como Remez o Meteor.

También puede ser un requisito de un medidor de cresta excluir el efecto de toda entrada de cc, ya que tradicionalmente los medidores de audio tienen el bloqueo de la cc. Por otro lado, si interesan los valores de cresta reales de la señal para eliminar la sobrecarga, el contenido de cc debe mantenerse y medirse. De ser necesario, la exclusión de la cc se puede lograr con una potencia de cálculo baja mediante la inclusión de un filtro paso alto de respuesta a impulso infinita de orden bajo a la entrada del medidor.

A veces, es necesario medir la amplitud de la cresta de la señal después de aplicar algún tipo de filtro de ponderación a fin de acentuar los efectos de ciertas partes de la banda de frecuencias. La realización depende de la naturaleza del filtro de ponderación particular que se utilice.

---