

Señales de audio para TV móvil, iPad e iPod

Thomas Lund
TC Electronic A/S
Risskov, Denmark

Resumen - Durante cinco años, el autor de este artículo ha estudiado de forma sistemática las capacidades audio de los dispositivos para TV móvil y Pods de Apple, Nokia, Samsung y Sony Ericsson. Este artículo es la primera presentación ante el público de una parte de las pruebas e investigaciones acerca de lo que un usuario de dispositivos móviles es capaz de oír, y lo que no. Tomando en consideración los resultados de las pruebas y los criterios de percepción, obtenemos una guía para el manejo óptimo de los programas para dispositivos móviles. Además, este artículo presenta una ruta de audio totalmente transparente y desvinculada de códecs de HDTV para múltiples plataformas personales, alcanzando la meta propuesta sin necesidad de un "procesado salchicha". Las técnicas descritas aquí van encaminadas a obtener una alta calidad de audio, y se basan completamente en standards abiertos y en una baja carga de trabajo.

INTRODUCCIÓN

Entre los amantes del audio, el sonido digital se ha granjeado una mala fama por su uso continuado de reducción de datos y por sus continuas guerras por el volumen percibido que han provocado gran número de damnificados en el mundo de la música, el broadcast y el cine.

La mayoría de nuestra herencia musical de los últimos 20 años es patética y su sonido es incluso peor cuando uno la hace sonar en un sistema de reproducción de alta calidad [1-4]. Desafortunadamente, podemos decir que la música es ya un caso perdido porque hoy por hoy no podemos encontrar ninguna sesión de grabación o master lineal que no contenga algún tipo de compresión de datos. Ojala nos volviéramos a encontrar con un periodo musical parecido a lo que nos ofrecieron los Beatles, Kathleen Ferrier, Dylan o Pink Floyd; pero eso no lo sabremos sin echar una mirada retrospectiva dentro de unos años.

Otro campo en el que es aplicada sin ningún pudor la hiper compresión es en los anuncios para broadcast. No obstante, dada su corta vida útil no creo que en el futuro nos vayamos a lamentar por no haberlos podido escuchar correctamente en su momento.

Pero lo que resulta casi más preocupante es que el mundo del cine también ha entrado en una auténtica espiral de la muerte en cuanto a la gestión de niveles de picos. A pesar de los sistemas de reproducción estandarizados, gracias al SMPTE, Dolby y THX, la ganancia de reproducción está siendo reducida sistemáticamente durante el proceso de mezcla de películas, así como en los teatros y cines, anulando de esa forma las grandes ventajas de una escucha bien calibrada. Una reciente encuesta sobre el nivel

de reproducción en cines de Dinamarca dio un valor medio de "4.75" en la arbitraria escala Dolby. La reducción de ganancia en 8-10 dB sobre la media es el triste signo de los tiempos.

En la base del problema está la medición del nivel de picos. Tanto los más novatos como los mayores expertos se dejan engañar por un instrumento que debería ser usado *exclusivamente* para evitar la saturación. Y lo que es peor aún, la forma más barata de implementar medidores en el audio digital se basa en la detección de muestreos de picos, razón por la que los técnicos y editores siguen mirando en su software ProTools, MediaComposer, Logic, Final Cut, Premiere etc. Hay que recalcar que los medidores de muestreos de picos son famosos por no ser ni siquiera fiables a la hora de detectar señales que vayan a provocar una sobremodulación [1-3].

En todos los campos anteriores, en la producción audio deberíamos eliminar esa costumbre de hacer uso de una inadecuada medición y normalización de nivel de picos. De igual forma, en los estudios de grabación se debería tapar los medidores de muestreos de picos con esparadrapo para no verlos.

Este artículo trata sobre cómo conseguir una señal audio con buen nivel de sonido y distribuida de forma fiable en plataformas móviles, no de las formas de trabajar para una producción con los mínimos requisitos posibles. Lo que hace que sea posible una transmisión predecible es un nuevo standard de broadcast centrado en el volumen percibido y no en el nivel de picos. La piedra angular para todo esto es el ITU-R BS.1770-3 que funciona perfectamente sea cual sea el género y plataforma, e independientemente de si se emplea audio lineal o una amplia gama de códecs de reducción de datos en parte de la cadena de señal.

EL STANDARD

Dado que el ITU-R BS.1770 fue revisado en 2011, en él ya ha sido incluido un importante sistema de medición que permite una discriminación fiable entre la señal principal y el ruido de fondo. La normalización del audio para broadcast con base en el nivel de volumen percibido de la señal principal (mezzoforte) ofrece una serie de ventajas importantes:

- Una medición óptima con cualquier tipo de señal audio.
- Un solo nivel objetivo => transparente a la hora de la producción.
- Un solo nivel objetivo => simple, funciona bien sin metadatos.
- Medición de fácil aplicación: Inicio y parada automática.
- Standard abierto. Ninguna molesta patente de por medio.

- Más margen o headroom que una medición basada en voz hablada.

Puede encontrar detalles sobre los cinco primeros aspectos en [5, 6]. La última revisión de este standard al momento de publicar este artículo es la BS.1770-3, que en la actualidad es la referencia mundial [7].

Relación (o razón) de picos con respecto al volumen percibido, PLR

El disponer de una definición transparente del volumen percibido del programa no es la única virtud del BS.1770. Este standard también especifica una técnica para librarse de la sobremodulación, por medio de la observación del nivel de "picos real", una técnica mucho más fiable en comparación con la de muestreos de picos.

Los programas o las pistas musicales pueden ser evaluadas gracias a esto por su relación (o razón) de picos con respecto al volumen percibido, o más bien por su relación (o razón) de picos reales con respecto al volumen percibido, abreviado como "PLR". Esto es una medición de lo exigente que será un programa en cuanto a margen o headroom para la ruta de señal emitida-.

La música pop/rock y los anuncios de hoy en día por lo general tienen los valores PLR más bajos, lo que resulta un signo claro de un uso intensivo de los efectos de compresión y limitación en el proceso de producción. En un reciente estudio fueron comparados los valores PLR a lo largo del tiempo de las canciones más famosas en Estados Unidos, Reino Unido y Alemania, (vea Fig 1). Este estudio revela un claro pico cuando se produjo la introducción del CD a mitad de la década de los 80 y una continua bajada desde entonces [8]. Algunas canciones hoy en día tienen un valor PLR de menos de 8 dB. En el lado opuesto de la escala, la música clásica y el audio de las películas tienen los valores PLR más altos, algunas veces por encima de 20 dB.

FIG 1. PLR MEDIO PARA LAS 7.488 CANCIONES MÁS FAMOSAS EN ESTADOS UNIDOS, REINO UNIDO Y ALEMANIA, DE 1973 A 2011

Margen o headroom

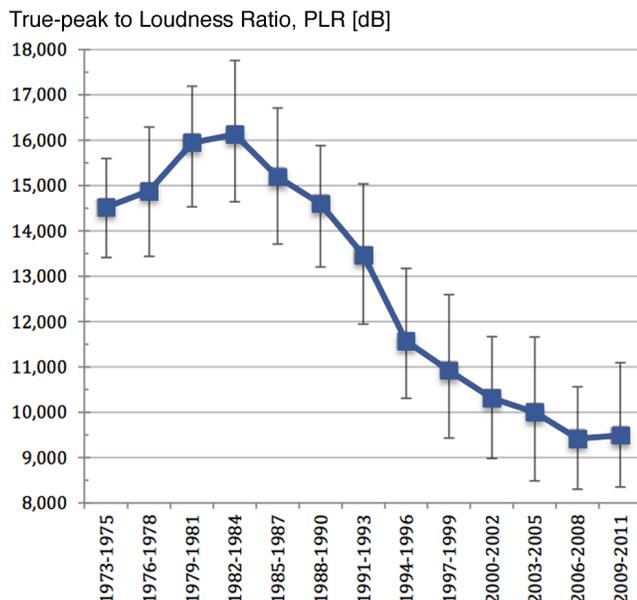
Para una determinada ruta de señal, se conoce como margen o headroom a la relación (o razón) entre el nivel de picos máximo que maneja y el nivel operativo RMS normal. Cualquiera de las partes de una ruta de señal supone en sí misma un posible cuello de botella en cuanto a este margen, y toda la cadena de señal se ve limitada por su eslabón más débil.

El método de normalización usado en una emisora produce un efecto importante sobre la cantidad de este margen disponible y, por tanto, en la forma en la que programas con valores PLR elevados pueden ser emitidos sin saturación o procesado. Este fenómeno aparece perfectamente ilustrado en la Fig 2, que es un ejemplo típico [6, 9].

La barra marcada como "1" corresponde a un programa con un rango de volumen percibido de 20 LU normalizado usando un anclaje para voz hablada, como el descrito en ATSC [10]. Un efecto colateral de este sistema pensado para cine es una pérdida muy marcada de margen en el broadcast, lo que produce una saturación o limitación. Ese mismo programa aparece en la barra marcada como "2", pero esta vez normalizado usando la antigua medición BS.1770 sin puerta de ruidos. La mayor parte del margen sigue siendo usado por el rango de volumen percibido, es decir, por partes del programa con mayor volumen que su punto de normalización. Finalmente, el programa también ha sido normalizado usando BS.1770-3, y allí ha desaparecido una menor cantidad del rango de volumen percibido. En la barra "3" pude ver cómo el mismo programa no necesita ningún tipo de procesado simplemente usando un sistema de normalización más inteligente [5, 6, 7, 9].-

Por tanto aquí consideramos el "margen" de una plataforma más o menos como la relación entre el nivel de picos real máximo y el nivel de destino basado en el BS.1770-3. Este último hace referencia al volumen percibido de programa usado; es decir -24 LUFS en la mayor parte del Mundo, donde el margen en DTV es por tanto de 22 dB (-2 dBTP en relación a -24 LUFS).

Nota: En este artículo usamos los valores "LUFS" compatibles con ISO en lugar de "LKFS". No obstante, los dos son idénticos, dado que -24 LUFS es exactamente lo mismo que -24 LKFS. Alguna gente piensa que los valores "U" con respecto a los "K" se diferencian por el hecho de indicar una medición con puerta de ruidos o no, pero no es así. Solo existe un único standard BS.1770, en la actualidad el -3, y emplea una puerta relativa de -10 LU.



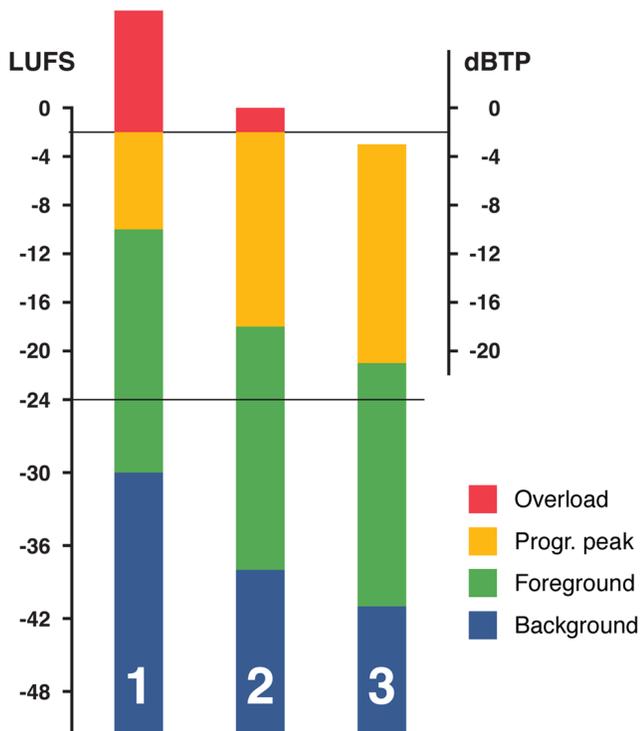


FIG 2. MÉTODO DE MARGEN VS. NORMALIZACIÓN
1: VOZ HABLADA, 2: BS.1770-1, 3: BS.1770-3

Dado que el balanceado inter-género y el margen es mejor con normalización BS.1770-3 que con otras alternativas, ha sido una decepción que la ITU no haya decidido deshacerse de la ambigua y limitada por patentes normalización de voz hablada en favor de la recientemente revisada ITU-R BS.1864 para el intercambio internacional de programas. En vez de eso son reconocidos los dos métodos, la norma BS.1770-3 y la de voz hablada. Una verdadera lástima en cuanto a las posibilidades de transparencia y de la señal audio en general.

PERCEPCIÓN Y TRANSITORIOS

El concepto de sensación se puede dividir en dos partes; la recepción y la percepción. El sonido se define como la energía mecánica transmitida a través de un medio, por lo general el aire, y debe quedar dentro de un intervalo de frecuencias concreto para que el oído humano lo reconozca como un estímulo. De la misma forma que nuestros ojos solo pueden detectar un pequeño rango de frecuencias de las ondas electromagnéticas con las que se encuentran, y su pequeño punto central (la mácula) escasamente puede cubrir con precisión un margen del tamaño aproximado con el que vemos la luna en el cielo.

No obstante, el ancho de banda real de nuestros sentidos es mucho mayor que el de nuestra consciencia. Por este motivo nuestro cerebro siempre está aplicando prioridades entre nuestro oído, vista, tacto, gusto, olfato y otros tipos de evaluaciones somáticas hasta reducir toda la información a un total de 30-40 bits por segundo, vea Fig 6 [11].

Desde el punto de vista fisiológico, los datos de los distintos sentidos llegan al tallo cerebral, que es el centro

encargado de asignar prioridades y de la correlación intermodal. El procesado temporal más agudo que tenemos se realiza en esta región, y se trata de la comparación I/D de los oídos usada para la localización y posicionamiento. Es importante tener en cuenta que el procesado de las sensaciones lleva su tiempo y que el cambio de prioridad de un sentido a otro conlleva incluso más tiempo aún. Unos 400-500 ms para ser preciso, vea Fig 3. En la ilustración siguiente puede ver los revolucionarios descubrimientos de Libet y que fueron cuestionados durante años [12, 13]. Observe cómo la sensación de alarma asociada con determinados sentidos compensa esta latencia.

The Latency of Sensation

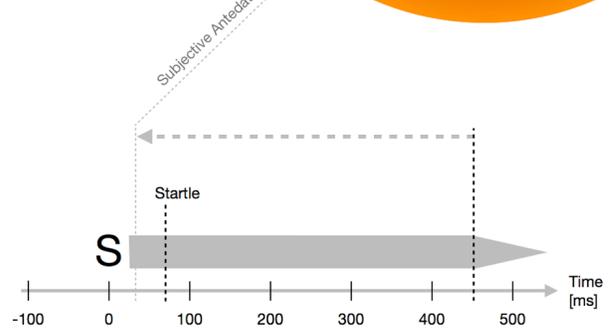


FIG 3. LA SENSACIÓN TARDA UN TIEMPO Y USA ANTICIPACIÓN
S: ESTÍMULO. CÍRCULO NARANJA: ACTUACIÓN CONSCIENTE

La Fig 3 justifica el porqué es importante la medición de volumen percibido momentáneo del ITU-R [19] y EBU [5] a 400 ms (vea Fig 5), y porqué no tiene tanta importancia definir un intervalo de tiempo menor para la medición del volumen percibido.

La Fig 4 muestra el efecto de cuello de botella a nivel de la percepción que se produce en el tallo cerebral. La ruta de audición primaria es rápida y no usa muchas sinapsis, por lo que siempre reaccionamos a cualquier alarma acústica rápidamente (70-100 ms). El oído es el verdadero rey en cuanto a los sentidos temporales, y si le asociamos una señal de alarma, hace que podamos reaccionar a cualquier amenaza rápidamente desde que nacemos, a mayor velocidad que para ganar un record del mundo de velocidad.

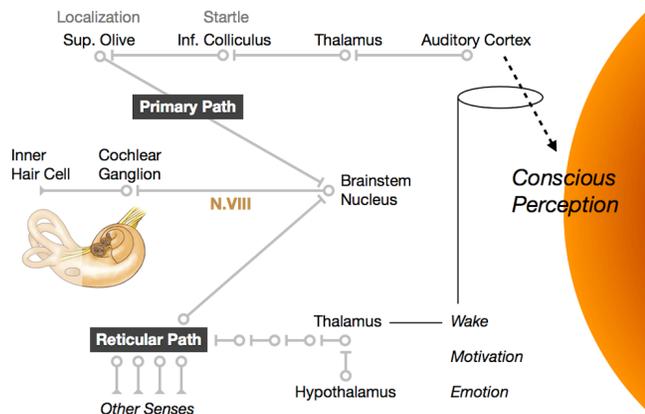


FIG 4. RUTAS DE ESCUCHA PRIMARIA Y RETICULAR
N.VIII: NERVIU CRANEAL DE ESCUCHA

Sin embargo, la ruta reticular con su "panel de control" final, el tálamo, decide a qué sentido prestamos atención y le dedica un mayor tiempo (Fig 3). Esta es la razón por la que la combinación de conducción y teléfonos móviles es un cóctel mortal.

Evidentemente somos capaces de escuchar sonidos con una duración inferior a los 400 ms, pero no debemos confundir esos sonidos con la sensación general de volumen percibido que todos los humanos compartimos. Los fonemas, las partículas mínimas del lenguaje, son mucho más cortas, pero al contrario que ocurre con el volumen percibido, para detectarlas y reconocerlas necesitamos aprendizaje [14]. La inteligibilidad, claridad y el volumen percibido no son lo mismo.

Los transitorios juegan un papel muy importante para los dos primeros aspectos tanto en la voz hablada como en la música. Los transitorios tienen incluso más importancia cuando el emisor está en un entorno silencioso y el oyente en uno ruidoso. En esos casos el reflejo Lombard no sirve para hacer que la voz hablada sea más limpia y menos dependiente de los transitorios [15, 16]. En tales casos, la restricción del PLR por medio de una limitación de transitorios puede ser negativa para la claridad de la voz hablada y para su inteligibilidad.

De forma similar, la limitación de picos en la música tiende a desajustar el balance entre sonido directo y reverberante en favor de este último, lo que nuevamente puede ser negativo para la claridad. Sin embargo, en el caso de la música la historia tiene una complicación añadida. Al igual que ocurre con los fonemas en el lenguaje hablado, tenemos que aprender a reconocer los transitorios musicales para que podamos apreciarlos. O nos hemos familiarizado con los instrumentos acústicos reales o hemos aprendido cómo es su sonido ayudados de un sistema de reproducción mínimamente decente.

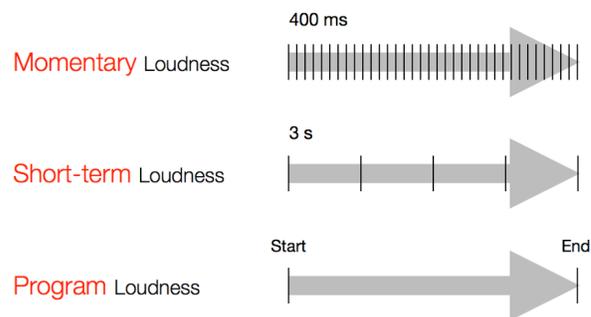


FIG 5. VOLUMEN PERCIBIDO EN TRES ESCALAS DE TIEMPO
POR ITU-R BS.1770, BS.1771 Y EBU R128

El cada vez menor valor PLR en la música pop no es algo bueno para las nuevas generaciones. Incluso aunque un niño pudiese conseguir unos altavoces de gama alta, no tendría prácticamente ninguna posibilidad de aprender cómo suenan los transitorios porque, de partida, no estarían en la fuente. Por este motivo, muchos niños no llegan a oír transitorios. Hemos comprobado esto en los empleados más jóvenes de TC. Con unos buenos altavoces, la mayoría no podía detectar diferencias en una pista musical tras escucharla con un PLR de 6, 9 o incluso 12 dB menos. Es como si fuese un idioma que nunca hubiesen aprendido.

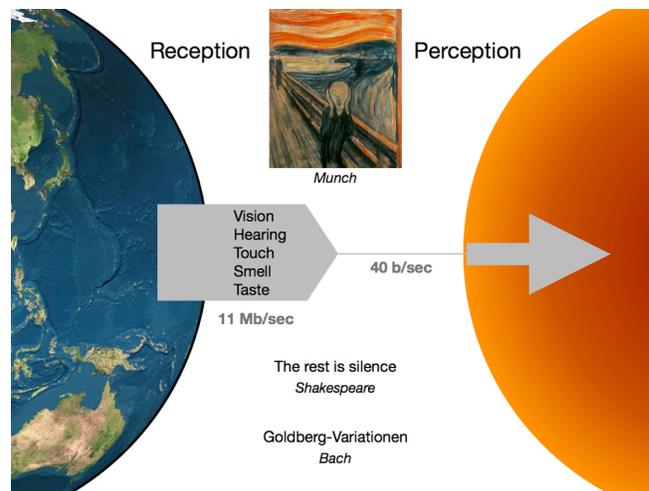


FIG 6. ANCHO DE BANDA DE LA CONSCIENCIA
EL ARTE NOS PERMITE LIBERARNOS DE UNA REALIDAD DE 40 BITS/SEG

Pero el que los niños no hayan aprendido a apreciar los transitorios, no es motivo para eliminarlos. Shakespeare no se dedicó a hacer tebeos solo porque algunos no supiesen leer. El arte es nuestra oportunidad de experimentar algo más que 40 bits/segundo cuando los estímulos se desarrollan en nuestra mente, vea Fig 6. Como profesionales, tenemos la obligación de asegurar que el arte basado en señales audio mantenga todo su potencial.

Rango de volumen percibido, LRA

Al contrario que lo que ocurre con el PLR, cualquiera es capaz de detectar el rango de volumen percibido, que es una medición estadística de la variación de volumen percibido de una pista o programa [9, 17].

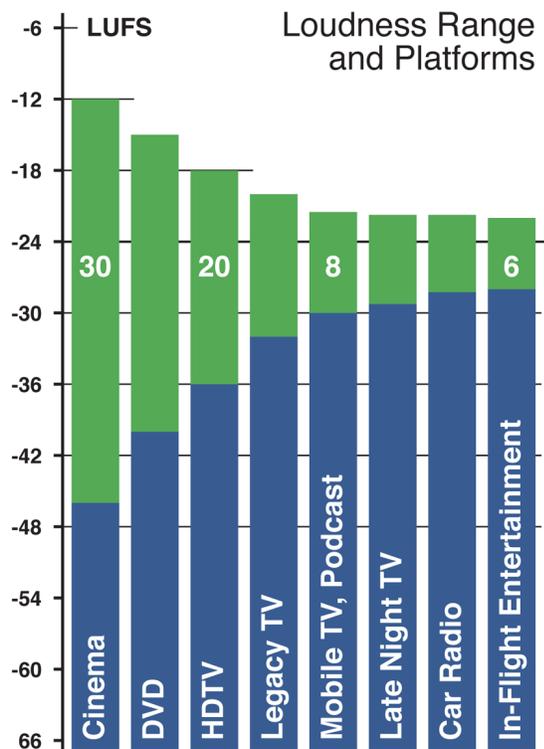


FIG 7. RANGO SUGERIDO DE VOLUMEN PERCIBIDO MÁXIMO Y APLICACIONES DE DESTINO
EL LRA PARA TV MÓVIL DEBERÍA SER MUY SUPERIOR A 8 LU

El rango de volumen percibido no es simétrico alrededor del objetivo. Por ejemplo, en el contenido típico de broadcast con un LRA de 8 LU, normalizado a -24 LUFS, la mayor parte del rango de volumen percibido probablemente procederá de fuentes con un valor inferior a -24 LUFS, como puede ver en la Fig 7. Las partes esenciales del programa pueden tener un nivel de volumen percibido a corto plazo alrededor de los -30 LUFS, que - como vimos en la Tabla 2 - se supone que producirá un SPL de alrededor de 72 dB (78.6-6 dB con ruido rosa) en un iPhone 5 con la ganancia al máximo.

La Fig 7 le muestra un rango de volumen percibido que le debería servir de objetivo cuando vaya a mezclar programas normales con una plataforma concreta en mente. Esta medición ha demostrado ser muy útil durante la producción al ayudar a establecer fácilmente valores esperados, que es la finalidad principal de las guías EBU [5]. Incluso la música clásica o las series para HDTV no deberían tener un LRA que superase los 20 LU, si bien no hay ningún límite concreto.

Al contrario de lo que ocurre con el LRA, un valor PLR alto por lo general no es un problema para el usuario de un pod. Por ejemplo, un programa con un PLR de 15 dB está bien, siempre y cuando su rango de volumen percibido no sea superior a 7 ó 8 LU. Por ejemplo, escuche "New Frontier" de Donald Fagen. Tiene un valor PLR elevado (18 dB), pero con un LRA de 6 LU no necesita ningún tipo de procesado para sonar muy bien en un iPhone - o en un par de increíbles altavoces [4].

ESTUDIO EN APPLE ITUNES

En la convención AES de 2009 en New York, el autor tomó parte de un panel en el que se trataba sobre las guerras en cuanto al volumen percibido en la música, contribuyendo con un estudio sobre la función de normalización incluida en iTunes, llamada Sound Check [18]. Los datos demostraron que, en general, esta función Sound Check es positiva, siendo capaz de ajustar la ganancia en pistas desiguales en un playlist basado en el volumen percibido. Las canciones antiguas y modernas pueden cohabitar sin ajustes de ganancia, aunque la normalización no está basada en BS.1770 sino en un algoritmo de Apple.

En general, este Sound Check demostró dar unos resultados bastante aceptables, adaptando la mayoría de las pistas +/- 2LU con respecto a donde habrían quedado con BS.1770-3. Esto es una mejora inmensa si lo comparamos con la escucha sin el Sound Check activado, ya que entonces el volumen percibido inter-pistas puede llegar a desviaciones de +/- 10 LU o incluso más.

Observamos que el *nivel objetivo medio* al usar Sound Check es -16.2 LUFS en una escala BS.1770-3; lo que es muy razonable si tenemos en cuenta la Fig 1. Partiendo de la base de la gran atención que pone esta empresa en ofrecer una señal audio detallada, presumiblemente este es el nivel al que está optimizada la estructura de ganancia de los dispositivos Apple.

También podemos ver de qué forma esa normalización positiva en líneas generales queda anulada en caso de que el nivel de picos pase por encima de 0 dBFS, por lo que las pistas con un valor PLR superior a 16 dB serían normalizadas para sonar a un nivel inferior a -16.2 LUFS.

Una de las primeras razones por las que escribí este artículo también tiene relación con iTunes: Durante sus viajes, el autor de este artículo suele escuchar los podcast de BBC Radio 4, en particular el programa "In Our Time" en el que se trata de temas históricos, culturales y científicos de una forma muy amena. Y hubo un programa concreto que trataba sobre Benjamin Franklin que no pude llegar a escuchar con el suficiente volumen percibido. Muchas partes se perdieron en el ruido de fondo del vuelo.

Midiéndolo con posterioridad, la combinación de un valor LRA relativamente alto y un volumen percibido de programa bastante bajo, vea Fig 8, solo era una parte del problema. No obstante, el Sound Check hubiera sido capaz de realzar ese podcast en 7 dB (de -23.3 a -16.2 LUFS) si ese PLR no hubiese sido ya excesivo. El programa, simplemente, quedó registrado a un nivel bajo y su escucha durante un vuelo fue imposible.

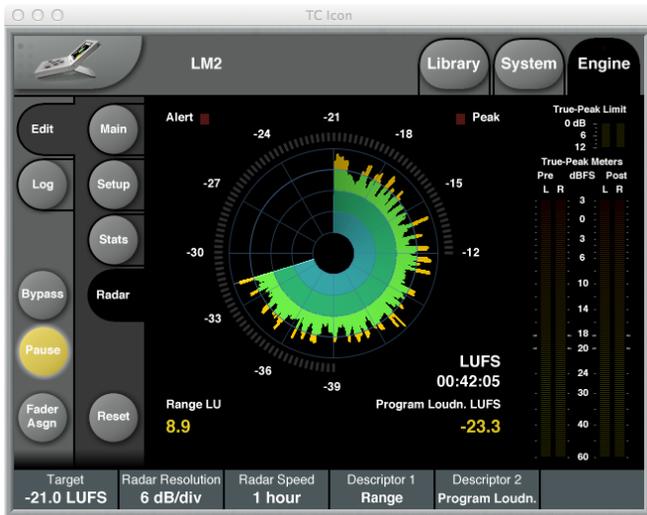


FIG 8. TRAZADO DE VOLUMEN PERCIBIDO DEL PODCAST DE LA BBC QUE DIÓ PIE A ESTE ARTÍCULO. PARTE DEL PROGRAMA ES DEMASIADO SILENCIOSO PARA UN IPOD

PRUEBAS CON TV MÓVIL E IPOD

Desde el punto de vista del audio, los dispositivos móviles funcionan bajo unas condiciones que no podemos llamar ideales: Sus dimensiones limitan la cantidad de voltaje y corriente (es decir potencia) que pueden enviar a unos auriculares; el entorno de escucha suele ser ruidoso y la calidad final depende en gran medida del precio.

Hemos hecho pruebas con iPods, iPads y Smartphones con una serie de parámetros audio. Utilizamos a Otto, un maniquí con cabeza y torso con micrófonos de condensador internos muy precisos, monitores de campo cercano, monitores principales y un software de medición de SPL y volumen percibido calibrado. Puede ver todo este montaje en Fig 9.



FIG 9. OTTO EN EL ESTUDIO ESCUCHANDO CON EARPODS DE APPLE

Otto fue calibrado usando ruido rosa a -23.0 LUFS con respecto a los monitores de campo cercano colocados a 240 cm del torso. Esto dio como resultado un SPL de 77.0 dB C por canal, 80.0 dB en ambos, en la posición de Otto. Los micrófonos binaurales daban señal a un analizador usando el Leq integrado con sumatorio de potencia de los canales como BS.1770. El punto 80 dB SPL fue usado como

referencia para la realización de los test siguientes de los dispositivos móviles y auriculares, y repetimos la calibración dos veces al día.

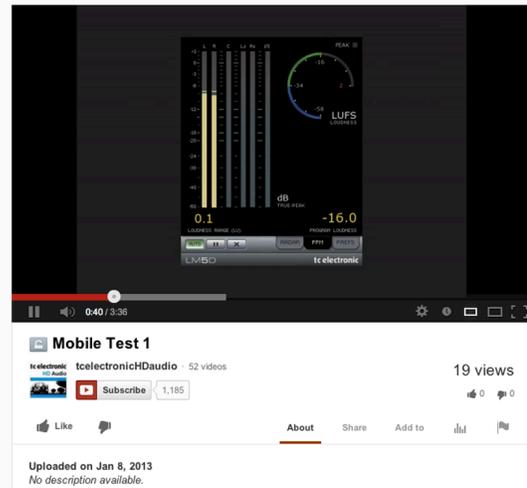


FIG 10. VIDEO DE YOUTUBE CON RUIDO ROSA A -16.0 LUFS

Probamos los dispositivos móviles con una serie de señales que iban desde la voz hablada y música al ruido rosa y los tonos base. Para cada una de las señales de prueba podíamos ver su valor de picos real y de volumen percibido de programa desde la pantalla del DUT mientras eran reproducidas, vea Fig 10. Los videos de prueba fueron cargados con clips de YouTube con audio con reducción de datos AAC. En el caso de las unidades Apple, repetimos los mismos test usando ficheros QuickTime con audio lineal. Para el estudio SPL que nos concierne aquí, no hubo diferencias significativas entre YouTube y el QuickTime lineal.

El objetivo principal era predecir el SPL que uno puede obtener para un volumen percibido de programa determinado usando una plataforma personal y al escuchar con auriculares. En varias ocasiones en el mundo real, el autor se quedó sin ganancia de reproducción en su iPod al escuchar podcasts.

RESULTADOS DEL TEST

Cada uno de los dispositivos móviles fue probado con su ganancia de reproducción al máximo, generando de esa forma una señal medible en los oídos de Otto a través de unos auriculares. Esto suministró datos SPL cuando un nivel de volumen percibido de programa determinado era reproducido a través de cada par de auriculares. Los DUT fueron verificados usando auriculares AKG K240S (55 ohmios) como una referencia común, y también con los auriculares que vienen originalmente con cada una de las unidades, indicados en la tabla como "standard".

Escogimos los AKG por su sonido y por ser un auténtico standard industrial como auriculares semi-abiertos y con una impedancia que no supondría ninguna limitación en cuanto a corriente de salida.

Plataforma	Auriculares	PN -24	PN -16
Apple iPod Nano G2	AKG K240S	75.9	83.9
Apple iPod Nano G3	AKG K240S	76.1	84.0
Apple iPad	AKG K240S	77.4	85.3
Apple iPhone 5	AKG K240S	71.8	79.6
Nokia Lumia 920	AKG K240S	66.3	74.3
Sony Eric's Xperia	AKG K240S	70.2	78.1
Samsung Gal S3	AKG K240S	69.2	77.2
Samsung Gal IIS	AKG K240S	69.9	77.8

TABLA 1. DB SPL AL REPRODUCIR RUIDO ROSA A -24 LUFS Ó -16 LUFS A TRAVÉS DE UNOS AURICULARES DE REFERENCIA

Plataforma	Auriculares	PN -24	PN -16
Apple iPod Nano G2	Apple antiguo	86.0	94.0
Apple iPod Nano G2	Apple nuevo	86.8	94.8
Apple iPod Nano G3	Apple antiguo	86.4	94.4
Apple iPod Nano G3	Apple nuevo	82.8	90.7
Apple iPad	Apple antiguo	86.0	93.9
Apple iPad	Apple nuevo	87.0	95.0
Apple iPhone 5	Apple nuevo	78.6	86.6
Nokia Lumia 920	Standard	86.2	94.2
Samsung Gal S3	Standard	81.8	89.8
Samsung Gal IIS	Standard	82.5	90.5

TABLA 2. LO MISMO QUE LA TABLA 1, PERO CON LOS AURICULARES INCLUIDOS CON CADA DISPOSITIVO. EN EL CASO DE APPLE, HEMOS INDICADO SI SE TRATA DEL MODELO NUEVO O DEL ANTIGUO DE AURICULARES
VERIFICACIÓN SUBJETIVA Y NOTAS

Para probar los programas basados en voz hablada y música en condiciones reales, el autor también los escuchó a través de los dispositivos usando sus auriculares standard y los auriculares de referencia AKG K240S. De cara a calcular el nivel de normalización, usamos con los programas un medidor, tal como mostramos en *Fig 10*.

Basándonos en una amplia variedad de programas, solo unos de los auriculares intraurales verificados demostraron tener una respuesta espectral decente y una buena imagen sonora, los nuevos "EarPods" standard de Apple. Por lo general, los auriculares de tipo intraural bloquean el ruido exterior incluso menos que los AKG semi-abiertos.

Con los nuevos Apple EarPods, la escucha de música pop/rock dentro de un coche o un tren hace necesario un SPL de 80-83 dB como mínimo para una buena experiencia. Cuando escuchamos una voz hablada limpia en las mismas condiciones, se puede suponer que las palabras se perderán por debajo de 78 dB SPL aproximadamente. En el futuro,

probablemente se considerará que no será suficiente un valor por debajo de 78 dB SPL.

Aunque no es el objetivo principal de este estudio, la pérdida de capacidad auditiva como resultado de un uso excesivo de un Pod debe ser algo a tener muy en cuenta: Otto también escuchó las señales a través de unos auriculares "Beats by dr. dre". Al emitir la señal por un dispositivo Apple, el SPL superó los 100 dB a -16 LUFS!

INTERPRETACIÓN DE LOS RESULTADOS

Las *Tablas 1* y *2* dan una estimación del SPL que experimentaría un usuario de dispositivos móviles cuando subiese al máximo la ganancia ("volumen"), al reproducir una señal audio a -24 LUFS ó -16 LUFS.

Con los auriculares de referencia y cada uno de los niveles de destino, el SPL de reproducción baja en un rango de 11 dB. Por lo general, los dispositivos Apple son mejores a la hora de dar señal a los auriculares de referencia AKG que los de otras marcas (*Tabla 1*). Con programas a -16 LUFS, el nivel SPL es suficientemente alto para escuchar música y una voz hablada limpia en la mayoría de condiciones posibles, mientras que dispositivos de otras marcas dieron unos niveles más suaves.

Usando los auriculares de cada dispositivo, se observó que en general se obtenía un volumen suficiente cuando la señal audio era reproducida a -24 LUFS. No obstante, si dejamos de lado los auriculares Apple, el resto de auriculares incluidos con los dispositivos suenan tan mal que habitualmente los usuarios se comprarán otros, añadiendo pues otra variable a la ecuación.

Los antiguos Apple iPods e iPads escasamente dan 1 dB más de volumen con los nuevos auriculares que con los antiguos, lo cual contrasta con los nuevos iPhones e iPods que dan un nivel menor cuando se usan los nuevos Apple Earbuds (*Tabla 2*). La razón para que exista esta diferencia puede ser técnica, pero es más probable que se trate de un intento de Apple de mantener el SPL bajo control.

LÍNEAS GENERALES PARA TV MÓVIL Y PODCAST

Las emisoras deberían meditar con calma acerca de sus necesidades inmediatas y futuras a la hora de decidir la mejor estrategia global para la gestión de emisiones para dispositivos móviles y Podcast. Los procesos deben ser automáticos, transparentes, con buen sonido y flexibles.

Automáticos

Resulta una pérdida de un tiempo muy valioso el tener que preparar el contenido para más de una plataforma, es decir HD. La transcodificación para TV móvil con un nivel de destino entre -18 y -14 LUFS debe producirse de forma automática. La *Fig 11* usa -16 LUFS como objetivo.

Transparente

A partir de la producción, debería ser algo fácil comprobar cómo va a sonar un programa en una plataforma concreta, por lo que no tendría que haber ningún punto oscuro o ambiguo en la transcodificación; es decir, si un programa es normalizado a nivel de voz hablada o a su volumen percibido de programa, o si los valores de metadatos son los correctos o no.

Con buen sonido

Para optimizar la calidad audio, el nivel de destino no debería ser incrementado más de lo necesario. -16 LUFS para el caso de TV móvil es una opción razonable. Algunos programas necesitan ciertas restricciones en LRA y PLR, pero debe evitar el "procesado salchicha" para que no desaparezca la distinción entre ruido de fondo y señal principal, incluso al trabajar con TV móvil.

Flexible

Nadie puede conocer con precisión cuáles serán las necesidades del oyente del mañana, pero lo que sabemos es que vivimos en un mundo dinámico. Tenga cuidado de no centrarse únicamente en un códec de reducción de datos concreto, porque el Ogg Vorbis o la codificación sin pérdida de señal puede que se conviertan en la mejor opción del mañana.

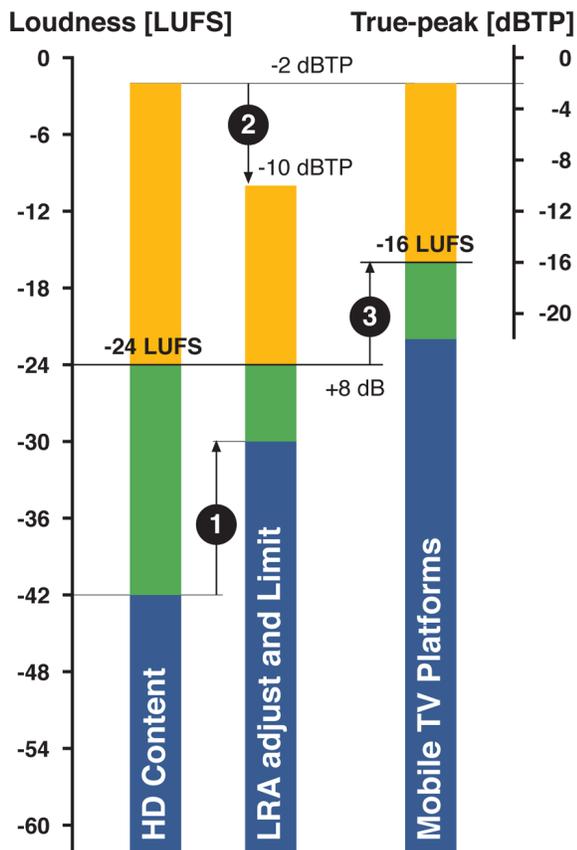


FIG 11. CONVERSIÓN AUTOMÁTICA DE HD A MÓVIL E POD TRES PASOS FÁCILES CUANDO LOS PROGRAMAS QUEDAN EN -24 LUFS

Si sumamos todo eso, en la Fig 11 puede ver una codificación inter-plataforma ideal y sencilla: El uso de BS.1770-3 para normalizar el contenido a -24 LUFS. Siga los números de la ilustración para -

- 1) Aumentar el nivel de los sonidos de nivel bajo que quedarían ahogados en un dispositivo móvil.
- 2) Limitar los picos a -10 dBTP.
- 3) Añadir un desfase de ganancia estática de, por ejemplo, +8 dB.

Conclusión

Hemos creado un caso con la intención de no buscar el mínimo común denominador para la señal audio en Pod y TV móvil. Para los usuarios de TVs planas y altavoces del tamaño de encendedores, las plataformas personales y sus auriculares son lo más cerca que pueden estar hoy en día de una experiencia audio decente. Sin embargo, hay dos cosas que pueden hacer que incluso eso no sea posible: La hiper-compresión de la señal fuente y/o una fuerte reducción de datos en las plataformas a través de las cuales el usuario escucha la señal.

En este artículo hemos presentado un estudio del nivel SPL de distintos dispositivos iPod y TV móviles, con diferencias entre los distintos fabricantes de unos 11 dB. Si a eso unimos también los distintos auriculares, la variación llega a los 20 dB. El SPL en TV móvil también ha sido enlazado al mejor método de normalización audio en emisiones broadcast, la medición del volumen percibido de programa de ITU-R BS.1770-3, con lo que, al menos, el nivel queda bien definido en la transmisión.

Al usar unos auriculares standard, los dispositivos de Apple no solo demostraron ser ζ que tenían mejor sonido, sino también los que ofrecían la mayor ganancia. La ganancia de los auriculares resulta un elemento crucial de la calidad de sonido para que los programas y pistas musicales no tengan que ser hiper-comprimidos antes de su transmisión. Si todos los Pods y TV móviles fuesen de Apple, no sería necesario un nivel destino superior a los -24 LUFS en broadcast móvil.

No obstante, si tenemos en consideración otros sistemas con menores capacidades, hemos justificado que un volumen percibido de programa de -16 LUFS es una opción válida que sigue permitiendo una reproducción de alta calidad sin un excesivo procesado dinámico. Hemos descrito un método automático, simple y transparente de convertir una señal audio con especificaciones HD a -24 LUFS a especificaciones móviles a -16 LUFS, basado completamente en standards abiertos.

El volumen percibido de programa y el rango de volumen percibido son parámetros importantes a la hora de preparar los programas para TV móvil, mientras que allí el nivel de picos tiene una importancia menor. No obstante, un margen muy bajo en la plataforma tiene un efecto adverso sobre la calidad del sonido. El nivel de picos real para plataformas móviles no tiene que verse limitado a -2 dBTP. Los códecs que producen pérdida de datos necesitan un límite de picos conservador únicamente si la medida es la de

muestreo de picos, o si el nivel queda cerca de la escala completa todo el tiempo. Este no es el caso cuando el nivel objetivo es ajustado a -16 LUFS. En consecuencia, el mejor sonido con el máximo margen para los transitorios viene a partir de una transmisión móvil que permita un nivel de picos real hasta los 0 dBTP, o al menos hasta los -0.5 dBTP. Parte del mito de un bajo nivel de picos deriva de la pérdida de margen del AC3 a la hora del "downmix", pero eso no debería dañar el sonido de la TV móvil. El margen de los transitorios también ayuda a la claridad e inteligibilidad de la voz hablada.

El evitar disponer de una ganancia de salida adecuada en las plataformas personales es algo negativo porque fuerza a que las fuentes audio sean aplastadas para hacer que el programa pueda ser escuchado. Desafortunadamente, el comité europeo de estándares técnicos, el CENELEC, no ha resultado de ayuda al obligar a una serie de restricciones en la cantidad de ganancia disponible. Esto no es una medida inteligente para reducir el SPL en Pods.

La solución más razonable y responsable, la revigorización del audio como uno de los posibles medios de transmisión del arte, debe usar alternativas a las "soluciones" tipo CENELEC y eliminar la actual dependencia de la reducción de datos con pérdida de señal. Con respecto a lo primero, animamos a aquellos fabricantes que lean este artículo a que consideren una normalización y control de ganancia integrados [20] en la siguiente generación de TV móviles, iPods, iPads y otros dispositivos no profesionales.

BIBLIOGRAFÍA

- [1] Nielsen, S.H. & Lund, T., "Level Control in Digital Mastering", *Paper of 107th AES Conv.*, New York, NY, 1999.
- [2] Nielsen, S.H. & Lund, T., "Overload in Signal Conversion", *Paper #11 of 23rd International AES Conference*, Copenhagen, Denmark, May 2003.
- [3] Lund, T., "Stop Counting Samples", *Paper of 121st AES Convention*, San Francisco, CA, 2006.
- [4] Serinus, J.V., "Winning the Loudness Wars", *Stereophile Magazine*, New York, NY, Nov 2012.
- [5] EBU, "EBU Technical Recommendation R128 - Loudness normalisation and permitted maximum level of audio signals", *European Broadcasting Union*, 2010.
- [6] Lund, T., "The CALM Act and Cross-platform Broadcast", *Paper NAB BE Conf.*, Las Vegas, NV, 2012.
- [7] ITU-R, "Rec. ITU-R BS.1770-3, Algorithms to measure audio programme loudness and true-peak audio level", *International Telecommunications Union*, 2012.
- [8] Ortner, R., "Je lauter desto bumm! - The Evolution of Loud", *Donau Universität, Krems, Austria*, 2012.
- [9] Skovenborg, E. & Lund, T., "Loudness Descriptors to Characterize Wide Loudness Range Material", *Paper of 127th AES Convention*, New York, NY, 2009.
- [10] ATSC, "Techniques for Establishing and Maintaining Audio Loudness for Digital Television", *Advanced Television Systems Committee*. Doc A/85, 2011.
- [11] Küpfmüller, K., "Nachrichtenverarbeitung im Menschen", *Springer*. Berlin, Germany, 1962.
- [12] Libet, B., "Subjective Referral of the Timing for a Conscious Sensory Experience", *Brain*, 102, Oxford, England, 1977.
- [13] Libet, B., "The Timing of Mental Events: Libet's Experimental Findings and Their Implications", *Consciousness and Cognition*, 11, 2002.
- [14] Tremblay, K. et al., "Central auditory system plasticity: Generalization to novel stimuli following listening training", *Journal of Acoustical Society of America*, 102(6), 1997.
- [15] Summers, W.V. et al., "Effects of noise on speech production: Acoustic and perceptual analyses", *Journal of Acoustical Society of America*, 84(3), 1988.
- [16] Lau, P., "The Lombard Effect as a Communicative Phenomenon", *UC Berkeley Report*, CA, 2008.
- [17] Skovenborg, E., "Loudness Range (LRA) – Design and Evaluation", *Paper of 132nd AES Convention*. Budapest, Hungary, 2012.
- [18] Lund, T. et al., "Loudness Wars", *Data from "tribunals"*. AES Conventions 127, 129, 131, 133. 2009-2012.
- [19] ITU-R, "Rec. ITU-R BS.1771-1, Requirements for loudness and true-peak indicating meters", *International Telecommunications Union*, 2012.
- [20] Camerer, F. et al., "Loudness Normalization: The Future of File-Based Playback", *Paper for the audio industry*. Vienna, Austria, 2012. <http://www.music-loudness.com>