

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

ABSTRACT

In this article, it is presented subjective methodologies for sound quality. Differences between objective and subjective measurements and how critic listening solve artefacts detection through subjective measurement are explained. Some listening tests standards and methodologies are referred to. An appropriate selection of rooms criterial and source material for evaluations are including. Additionally, evaluation of perceptual codecs, speech codecs and speech intelligibility are mentioned in this info compilation. Also, statistical evaluation is clarified. Further, an example of a testing procedure and evaluation form is showed.

INTRODUCCIÓN

En esencia, casi todos los esfuerzos en ingeniería de audio comparten el mismo objetivo, el cual es el de proveer una óptima calidad de sonido para las condiciones presentadas. De igual manera, formalmente o no, cada ingeniero de audio escucha una señal de audio para determinar si cumple o no con los criterios de diseño. Comparando las complicaciones que presenta el diseño y desarrollo de un dispositivo, la verificación de calidad de sonido podría parecer ser una tarea relativamente pequeña. Pero, por supuesto, el problema es mucho más complejo de lo que parece. El hecho es que es difícil definir calidad sonora y más aún juzgarla.

Puede argumentarse que el significado de un "buen sonido" no es bien entendido, sin embargo, ¿Cómo determinamos que un dispositivo suena bien?, ¿Quién determina esto?, ¿Cómo podemos notar el hecho de que un altavoz de \$ 10 puede sonar relativamente bien, pero evidentemente este suena inferior a uno de \$ 100?, ¿Cuándo debemos elegir calidad sonora?, ¿Qué es preferible? Por ejemplo, ¿Es un sonido "cálido" mejor que un sonido "crujiente"?, ¿Cómo suena un sonido "cálido"?, ¿Cómo suena un sonido "crujiente"?, ¿Es igual un buen sonido para hip hop que para música clásica?, ¿Es lo mismo escuchar para cualquier persona en cualquier parte del mundo? Preguntas como estas son el principal tema de discusión en la evaluación subjetiva de la calidad sonora.

EVALUACIÓN SUBJETIVA VERSUS MEDICIONES OBJETIVAS

Las mediciones objetivas comparan resultados con puntos de referencia confiables para valorar el rendimiento. En muchos proyectos de ingeniería, las mediciones objetivas cuantifican completamente el rendimiento de un dispositivo. Sin embargo, la señal de salida de la mayoría de dispositivos de audio es entregada finalmente al oído humano y resulta difícil o imposible determinar a través de una medición objetiva si una señal de audio va a ser agradable de escuchar. Es posible argumentar que las mediciones objetivas de audio pueden proveer únicamente modelamiento limitado de calidad de sonido. Este dilema forma las bases para la evaluación subjetiva diseñada para examinar la calidad estética en señales de audio.

Existe una división entre el valor subjetivo de la calidad sonora, y mediciones objetivas. Las mediciones son de vital importancia, pero con dispositivos de audio y acústica, es el oído quien decide. En muchos casos, las personas son quienes realizan análisis impecables. Por ejemplo, la sonoridad es un factor sensorial y no una medición literal. En una investigación de sonoridad, los sujetos a prueba son presentados con varios sonidos, y a cada uno se le pide que compare la

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

sonoridad de un sonido "A" con un sonido "B". Los datos son sujetos a análisis estadísticos, y la dependencia de la sonoridad sobre la medición física de nivel sonoro es evaluada. Si la evaluación es conducida apropiadamente y hay suficientes oyentes participando, los resultados van a ser confiables. En este sentido, nosotros descubrimos, por ejemplo, que no hay una relación lineal entre nivel de sonido y sonoridad, altura y frecuencia, o timbre y calidad de sonido.

Es deseable como diseñadores correlacionar impresiones subjetivas con parámetros de diseño objetivos. Esto permite a diseñadores conocer donde existen limitaciones en la fidelidad de audio y por lo tanto concentrarse en donde se puede hacer mejoras. De la misma manera, por ejemplo, este conocimiento podría permitir la tasa de bits de un código de compresión y descompresión (códec) sea baja mientras conocemos los efectos que estos tienen sobre la fidelidad. La correlación entre las impresiones de los oyentes y el objetivo significa que medir el fenómeno es un problema difícil. Las correlaciones no siempre son conocidas. Una forma de correlacionar impresiones subjetivas con datos objetivos es con investigaciones, en particular a través de una escucha crítica. Con el tiempo, es posible que emerjan algunos patrones que puedan proveer correlaciones.

No hay un instrumento para medir directamente cualidades tales como *calidez* o *brillantez*. Sin embargo, en algunos casos, términos subjetivos pueden ser relacionados para mediciones objetivas. Comparar el sonido directo y las reflexiones tempranas, las cuales son integradas por el oído, para el sonido reverberante entero, es una forma directa de medición de un sonido impulsivo de una pistola o de otra fuente. Este es un ejemplo de como una evaluación puede ser correlacionada para una medición objetiva.

En cualquier evaluación subjetiva, la terminología tiene que ser rigurosamente definida y entendida por la persona encargada de evaluar (administrador o encuestador) y las personas a evaluar (oyentes o encuestados). Esto provee palabras que los oyentes pueden usar en su evaluación, además brinda uniformidad en el lenguaje usado por una amplia audiencia de oyentes, y esto brinda un punto de partida en la calificación objetiva de los comentarios subjetivos. Cualquier confusión en la terminología podría comprometer la integridad de la comunicación entre los oyentes y el administrador. Se puede considerar las siguientes palabras descriptivas, las cuales son aplicadas a salas de conciertos acústicas: Calidez, definición, reverberación, sonoridad, claridad, brillantez, vivo(a), seco(a), resonante, intimo(a), mixto(a). Estos términos deben ser definidos usando vocabulario que sea familiar para el oyente.

ESCUCHA CRITICA

Desde el punto subjetivo, la mejor forma de evaluar una señal de audio es escuchándola cuidadosamente, y usando una gran cantidad de oyentes. Este tipo de escucha crítica, cuando se analizada apropiadamente, es el método preferido para la evaluación de calidad de sonido subjetivo. La escucha debe ser ciega (la fuente se desconoce) y usando oyentes expertos, y además debería presentarse un análisis estadístico apropiado para entregar resultados estadísticos confiables. La escucha crítica puede tardar mucho tiempo y por lo tanto es costosa. En algunos casos, se discierne distorsiones solo después de repetidas pruebas de audición.

Cualquier dispositivo de audio manifiesta sus potenciales anomalías que podrían ser contadas en una prueba de escucha crítica. Identificando varios defectos comunes, los oyentes pueden evaluar

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

de manera más cercana el rendimiento. Los defectos deben ser definidos, y en algunos casos, son especificados en la evaluación. Generalmente es útil diseñar una evaluación de escucha acorde al tipo de dispositivo de audio que va a ser evaluado. Por ejemplo, para la fidelidad de audio de dispositivos se puede considerarse estas cuatro categorías:

Gran Deterioro. - Las diferencias de calidad de sonido son claramente audibles para un oído no entrenado. Por ejemplo, dos sistemas de altavoces idénticos, uno normal y otro con el twitter inhabilitado o averiado, podría constituir un gran deterioro.

Deterioro Medio. - Las diferencias de calidad de sonido son audibles para un oído no entrenado, pero puede requerir más que una escucha casual. La habilidad para cambiar fácilmente de ida y vuelta y comparar directamente dos fuentes, comparaciones A-B, hace estos deterioros evidentes. Por ejemplo, un par estéreo de altavoces con un altavoz invertido la polaridad a través de sus cables, esto podría constituir un deterioro medio.

Deterioro Pequeño. - Las diferencias de calidad de sonido son audibles para muchos oídos, sin embargo, puede ser necesario algo de entrenamiento y práctica. Por ejemplo, la diferencia de fidelidad entre un archivo de música codificado a 128 Kbps, y 256 Kbps, podría revelar pequeños deterioros en el archivo de 128 Kbps. Los deterioros podrían ser únicos y no familiares para el oyente, entonces estos son más difíciles de identificar y toma más tiempo detectar.

Micro deterioro o deterioro muy pequeño. - Las diferencias de calidad de sonido son sutiles y requieren paciencia escuchando por parte de oyentes entrenados y con experiencia. En muchos casos, las diferencias no son audibles bajo condiciones normales de escucha, con música a niveles normales. Podría ser necesario amplificar la música, o usar señales de prueba a un nivel bajo como tonos puros y silencio con "dither". Por ejemplo, ligeras distorsiones audibles sobre una onda sinusoidal de 1 kHz con dither a -90 dBFS podría constituir un micro deterioro.

Cuando escuchamos grandes deterioros como en el ejemplo de los altavoces, la evaluación de calidad de sonido puede depender de mediciones objetivas y de condiciones subjetivas para establecer diferencias y encontrar causas para los defectos. Sin embargo, cuando comparamos equipos de alta fidelidad como convertidores de alta fidelidad de digital a analógico, los deterioros pequeños son considerablemente más difíciles de cuantificar. Por lo tanto, es importante tener metodologías para identificar, categorizar y describir diferencias sutiles.

El Desarrollo de estas metodologías requiere la capacitación de oyentes críticos, evaluaciones en curso, y discusiones entre diseñadores y oyentes para cerrar el ciclo. Además, es necesario encontrar sistemáticamente umbrales de audición para varios defectos. Esta búsqueda introduce la audición subjetiva de los defectos y umbrales de audición.

Cuando las diferencias entre dispositivos son pequeñas, uno de los enfoques es estudiar el residuo (diferencia de señal) entre ellas. Las salidas análogas podrían ser aplicadas a un convertor análogo-digital de alta calidad; una señal es invertida; las señales son alineadas en el tiempo con precisión; las señales son sumadas (restadas por la inversión). Luego la señal resultante puede ser estudiada. Por ejemplo; si un dispositivo añade ruido, será solo el ruido el que se escuche como resultado. Dos señales idénticas producirán un archivo nulo. El proceso de preparación de una señal residual es mostrado en la Figura 1.

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

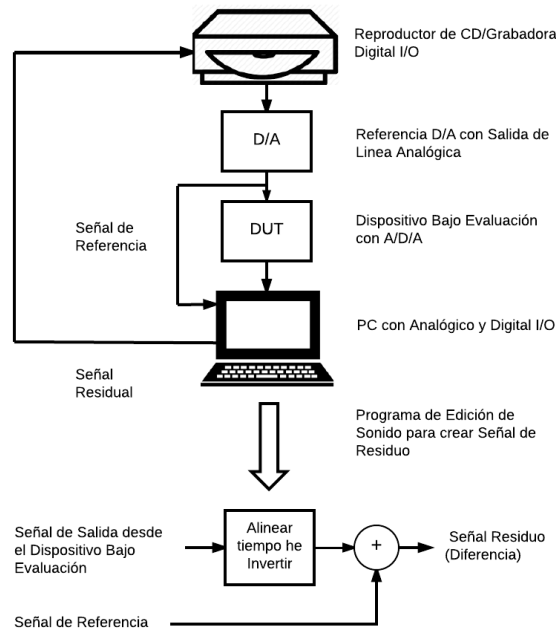


Figura 1 Método usado para crear una señal residual.

Se prefieren oyentes experimentados a los no experimentados porque oyentes expertos están más familiarizados con distorsiones sutiles y peculiares. Un oyente experto es más confiable detectando detalles y deterioros que no son notados por un oyente casual. Ha sido estadísticamente estimado que un oyente con experiencia puede reemplazar siete sin experiencia. Los oyentes en cualquier evaluación deben escuchar distorsiones que el dispositivo bajo evaluación pueda mostrar. Por ejemplo, oyentes podrían comenzar su preparación con ejemplos de tasas de bits bajas o señales izquierda menos derecha, o señales residuales con distorsiones expuestas entonces se familiarizan con las *firmas* que sonoramente pueden dejar los dispositivos. Cuando una referencia es usada, esta debe ser de la más alta calidad posible. Por ejemplo, una grabación a 16 bits podría no ser una referencia adecuada cuando se evalúan dispositivos de alta calidad porque algunos dispositivos podrían superar el rendimiento la calidad de referencia, es decir, un dispositivo que admita 24 bits no será evaluado con una referencia de 16 bits.

Algunas evaluaciones de oyentes podrían ser conducidas usando auriculares de alta calidad; condiciones externas como características de un altavoz; acústica de la habitación, y posición del altavoz y el oyente son eliminados. Los auriculares pueden permitir escucha crítica de detalles audibles sutiles. En algunos casos, auriculares al aire libre (ecualizados para dar una respuesta plana en campo difuso) son recomendados porque presentan una reproducción de sonido más natural; sin embargo, es necesaria una habitación tranquila. Auriculares cerrados pueden ser usados para minimizar la audibilidad del ruido ambiente. Otras evaluaciones usan altavoces. Por ejemplo, estas son necesarias para evaluaciones de dispositivos multicanal. Cuando son usados altavoces, la acústica del lugar juega un papel importante en la evaluación. Una sala de escucha adecuada debe proveer una acústica adecuada y además proveer un ruido de ambiente bajo incluyendo aislamiento

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

acústico. En algunos casos, las habitaciones son diseñadas y construidas de acuerdo a criterios estándar. Los cuartos de escucha son descritos más adelante en este artículo.

Para dar conclusiones útiles, los resultados de una evaluación de escucha deben ser sujetos a un análisis estadístico. Se debe tener cuidado para generar un análisis válido. El número de oyentes, el número de pruebas, el intervalo de confianza y otras variables pueden afectar la validez de las conclusiones. En muchos casos, varias evaluaciones de escucha, diseñadas desde diferentes perspectivas, pueden ser empleadas y analizadas para determinar completamente la calidad de un dispositivo. Los análisis estadísticos son descritos más adelante en este artículo.

EVALUACIONES DE ESCUCHA, METODOLOGÍAS Y ESTÁNDARES

Numerosas metodologías y estándares han sido desarrollados para las pruebas de escucha. Estas pueden ser seguidas de cerca o usadas como guías prácticas para desarrollar otras evaluaciones. En cualquier caso, una evaluación de escucha debe contener material de origen, flujo de señal, evaluadores, metodología y análisis. Normalmente la asunción es que la calidad de sonido es igual a la transparencia. Una señal bajo evaluación es transparente cuando no hay diferencia audible entre esta y la referencia. Bajo este umbral, nosotros podemos medir la distancia perceptual a partir de la transparencia y describir deficiencias. La calidad de sonido es por lo tanto la distancia subjetiva entre una señal evaluada, y una referencia conocida. Cualquier diferencia es una distorsión.

Una evaluación ideal de escucha debería satisfacer al menos cinco criterios:

1. Ser consistentemente repetible con diferentes tiempos, lugares, y oyentes.
2. Considerar solo parámetro(s) audible(s) seleccionado(s).
3. Revelar la magnitud de diferencias con un puntaje.
4. Informar comentarios de los oyentes.
5. Entregar resultados estadísticamente válidos.

Algunas evaluaciones de escucha pueden establecer si un dispositivo es perceptualmente transparente; es decir, si oyentes expertos pueden notar la diferencia entre el archivo original y el procesado, usando señales de prueba y una variedad de música. En una prueba *ABX*, se presenta las fuentes conocidas *A* y *B*, y una fuente desconocida *X* que puede ser cualquiera *A* o *B*; la asignación es realizada pseudo-aleatoriamente para cada prueba. El oyente debe identificar si *X* ha sido asignado a *A* o *B*. La respuesta a la pregunta de si el oyente puede notar la diferencia entre *A* y *B*. La prueba *ABX* no puede ser usada para concluir que no hay diferencia, por el contrario, esta puede mostrar que una diferencia es escuchada. Pequeñas muestras de canciones (quizás entre 15 y 20 segundos) pueden ser presentados para identificar distorsiones. Esto es útil para analizar individualmente a los sujetos de la prueba *ABX*, e informar el número de sujetos que escuchen la diferencia. Los softwares *ABXTeser* para usuarios Mac y *PLugin ABX Comparator 2.0.4* del reproductor *foobar2000* en Windows pueden ser utilizados para este fin, sin embargo, estos no son los únicos existen un sinnúmero de aplicaciones en la web que pueden ser consideradas para pruebas de escucha.

Otras evaluaciones de escucha pueden ser usadas para estimar el margen, o cuanto la señal puede ser degradada antes de perder transparencia. Otras evaluaciones son designadas para calcular la

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

transparencia relativa. Esta es una tarea más difícil. Si dos dispositivos presentan ruido apreciable, solo el oído humano puede determinar cuál dispositivo es preferible. Además, diferentes oyentes podrían tener diferentes preferencias en esta elección del menor de dos males. Por ejemplo, un oyente podría estar más preocupado por la reducción del ancho de banda mientras que otro está más molesto por el ruido.

La escucha crítica debe usar el método doble ciego en el cual ni el evaluado ni el evaluador conocen la identidad de la selección. Por ejemplo, en una evaluación "triple estímulo A-B-C, referencia oculta, doble ciego" el oyente es presentado con una señal conocida no procesada A, y dos señales B y C desconocidas. Cada estímulo es grabado por quizás unos 10 a 15 segundos de duración. Una de las señales conocidas es idéntica a la referencia conocida y la otra es la señal codificada bajo análisis. Las asignaciones son hechas al azar y cambiadas para cada ensayo. Los oyentes deben asignar un puntaje para las dos señales desconocidas, clasificándola con la referencia conocida. El oyente puede presentar cualquiera de los estímulos, con repetidas escuchas. Los ensayos son repetidos con diferentes estímulos. Auriculares o altavoces pueden usarse. El nivel de reproducción debe ser arreglado en un ensayo particular para resultados consistentes. La escala mostrada en la Figura 2 puede ser usada para puntuar. Estos cinco puntos de escala de deterioro fueron ideados por el Comité Consultivo de Radio Internacional (CCRI) y es a menudo usado para evaluaciones subjetivas. Paneles de oyentes expertos clasifican los deterioros que escuchan sobre unos 41 puntos en escala continua en categorías desde 5.0 (transparente) a 1.0 (deterioros muy molestos)

Grado Absoluto	5.0	Imperceptible	0.0	Grado de Diferencia
	4.9	Perceptible pero no molesta	-0.1	
	4.8		-0.2	
	4.7		-0.3	
	4.6		-0.4	
	4.5		-0.5	
	4.4		-0.6	
	4.3		-0.7	
	4.2		-0.8	
	4.1		-0.9	
	4.0		-1.0	
	3.9	Ligeramente molesto	-1.1	
	3.8		-1.2	
	3.7		-1.3	
	3.6		-1.4	
	3.5		-1.5	
	3.4		-1.6	
	3.3		-1.7	
	3.2		-1.8	
	3.1		-1.9	
3.0	-2.0			
2.9	Molesto	-2.1		
2.8		-2.2		
2.7		-2.3		
2.6		-2.4		
2.5		-2.5		
2.4		-2.6		
2.3		-2.7		
2.2		-2.8		
2.1		-2.9		
2.0		-3.0		

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

	1.9		-3.1	
	1.8		-3.2	
	1.7		-3.3	
	1.6		-3.4	
	1.5	Muy molesto	-3.5	
	1.4		-3.6	
	1.3		-3.7	
	1.2		-3.8	
	1.1		-3.9	
	1.0		-4.0	

Figura 2 Escala de deterioro del CCIR.

La señal seleccionada por los oyentes como la referencia escondida esta dado un puntaje por defecto de 5.0. Restando la puntuación dada a la referencia oculta actual la puntuación dada a la señal alterada codificada se obtiene el grado de diferencia subjetiva (GDS). Por ejemplo, material no comprimido, original, puede recibir un puntaje promedio de 4.8 en la escala. Si un dispositivo obtiene un puntaje promedio de 4.8, el GDS es 0 y se dice que dispositivo es transparente (sujeto a análisis estadístico). Un menor puntaje de GDS (por ejemplo 2.6) evalúa que tan lejos de la transparencia un dispositivo se encuentra. Numerosas técnicas de análisis estadísticas pueden ser usadas. Quizás 50 oyentes son necesarios para unos buenos resultados estadísticos.

MUSHRA (MUltiple Stimulus with Hidden Reference and Anchors o en español Estímulos múltiples con Referencia Escondida y Anclas) es un método evaluación usado cuando existen degradaciones. Este método usa una referencia escondida y una o más anclas escondidas; un ancla es un estímulo con una limitación audible conocida. Por ejemplo, una de las anclas es una señal codificada de paso bajo. Una escala continua con cinco divisiones es usada para calificar el estímulo: excelente, buena, pobre y mala. MUSRA es especificada en ITU-R BS.1534. Otros asuntos en evaluación de sonido son descritos en ITU-T P.800, P.810, y P0830; ITU-R BS.562-3, BS.644.1, BS.1284, BS.1285, y BS.1286, entre otros.

En particular, evaluaciones subjetivas de escucha pueden ser conducidas usando la ITU-R Recomendaciones BS1116-1. Esta metodología selecciona direcciones de materiales de audio, rendimiento de reproducción del sistema, escala de calificación y métodos de análisis de datos.

SALAS DE ESCUCHA ESTANDAR

Aunque los auriculares permiten una escucha muy precisa, dispositivos como altavoces deben por supuesto ser probados en una sala. Cualquier sala afecta enormemente las características del sonido que perciben los oyentes. Para normalizar los efectos acústicos de la sala, varios diseños de salas han sido desarrollados.

Además de la metodología de la prueba de audición, la recomendación estándar BS.1116-1 de la ITU-R describe una referencia de la sala de escucha. El estándar BS.1116-1 recomienda un piso de área de 20 a 60 m², para reproducciones estéreo y monoaurales y un área de 30 70 m² para reproducciones multicanal. Para distribuciones de ondas estacionarias en bajas frecuencias, los estándares recomiendan que las proporciones de la sala cumplan con los siguientes criterios:

$$1.1 \frac{w}{h} \leq \frac{l}{h} \leq 4.5 \frac{w}{h} - 4$$

Ecuación 1

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

$$\frac{w}{h} < 3$$

Ecuación 2

$$\frac{l}{h} < 3$$

Ecuación 3

Donde:

l, w, h , son las dimensiones de la sala, longitud, ancho y altura.

El nivel de presión sonora de un nivel de 1/3 de octava, sobre un rango de 50 a 16000 Hz, medidos desde la posición de escucha con ruido rosa es definido por una curva de respuesta estándar. El tiempo de reverberación en una sala promedio es determinado por:

$$T_{ave} = 0.25 \left(\frac{V}{V_0} \right)^{1/3}$$

Ecuación 4

Donde:

T_{ave} es el tiempo de reverberación promedio.

V es el volumen de la sala de escucha.

V_0 es un volumen de referencia de 100m³.

Este tiempo de reverberación es además especificado para ser relativamente constante en el rango de frecuencias de 200 a 4000 Hz, y sigue variaciones entre 63 y 8000 Hz. Reflexiones tempranas en el rango de 1000 a 8000 Hz que llegan a la posición del oyente con 15 ms deben ser atenuadas por al menos 10 dB con respecto al sonido directo desde los altavoces. Es recomendable que el nivel de ruido de fondo no supere la relación de ISO de NR10, con NR15 como un límite máximo.

La especificación IEC 60268-13 (originalmente IEC268-13) describe una sala de escucha tipo residencial para evaluación de altavoces. Las especificaciones 60268-13 recomienda para el piso un área de 25 a 40 m² para reproducciones estéreo y monoaurales y un área de 30 a 45m² para reproducciones multicanales.

Para distribuir espacialmente frecuencias estacionarias bajas en la sala, las especificaciones recomiendan tres criterios para las relaciones de las dimensiones de las salas.

$$\frac{w}{h} \leq \frac{l}{h} \leq 4.5 \left(\frac{w}{h} \right) - 4$$

Ecuación 5

$$\frac{w}{h} < 3$$

Ecuación 6

$$\frac{l}{h} < 3$$

Ecuación 7

Donde:

l, w, h , son las dimensiones de la sala, longitud, ancho y altura.

El tiempo de reverberación (medido de acuerdo a los estándares de la ISO 3382 en 1/3 de bandas de octava con la sala desocupada) está especificado para caer dentro de un rango de 0.3 a 0.6s en el rango de frecuencias de 200 a 4000 Hz. Alternativamente, el tiempo promedio de reverberación

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

debería ser de 0.4s y está dentro de la curva de frecuencia dada en el estándar. El nivel de ruido de ambiente no debería exceder los NR15 (20-25 dBA).

Los estándares de la EBU 3276 especifican a una sala de escucha con un área mayor a 40m² y un volumen no menor que 300m³. Las relaciones de dimensión y tiempo de reverberación de la sala siguen las especificaciones BS.1116-1. Además, para mitigar efectos no deseados de resonancias modales, las proporciones de dimensión debería diferir en más de +-5 por ciento. La respuesta de la sala medida a 1/3 de octava con ruido rosa sigue una curva estándar.

FUENTE DEL MATERIAL DE AUDIO

Para revelar distorsiones es importante usar material de audio que estrese el dispositivo bajo análisis. Sobre todo, porque diferentes dispositivos responden diferentemente, una variedad de material es necesario, incluyendo material que estrese cada dispositivo específicamente. Además, los canales deben ser bien grabados para evitar introducir distorsiones propias. Generalmente, música con transiente, tonos complejos, contenido armónico abundante alrededor de la región más sensible del oído, 1 a 5 kHz, es útil. Particularmente ejemplos desafiantes incluyendo castañuelas, triangulo, clave, pandereta, dialogo, trompeta y el bajo.

La selección de pistas de prueba es generalmente una elección altamente personal. En efecto, cuando es posible los oyentes deberían ser alentados a utilizar su propia selección de música. Un aspecto importante de la escucha crítica es el familiarizarse con las pistas de prueba. Esto solo sucede después de escuchar una pista sobre muchos sistemas de reproducción, en muchas salas, donde uno gana entendimiento del contenido exacto de un sonido grabado. Solo entonces la pista puede usarse como referencia para evaluar un dispositivo. La Tabla 1 indica algunos ejemplos de pistas de audio que han sido útiles en evaluaciones de escucha.

Tabla 1 Ejemplos de pistas musicales usadas para escucha subjetiva.

Pink Floyd, Dark Side of the Moon, "Speak to Me"
Pink Floyd, Delicate Sound of Thunder, "On the turning Away"
KT Tunstall, Eye to the Telescope, "Black Horse and the Cherry Tree"
KT Tunstall, Eye to the Telescope, "Suddenly I see"
Five for Fighting, The Battle for Everything, "100 Years"
Jennifer Warnes, Famous Blue Raincoat, "First We Take Manhattan"
Dire Straits, The very Best of Dire Straits, "Sultans of Swing"
Dire Straits, On every Street, "Calling Elvis"
Dire Straits, Brothers in Arms, "Brothers in Arms"
Madonna, Ray of Light, "Candy Perfume Girl"
Santana, Supernatural, "Migra"
Blur, 13, Tender
Gomez, Liquid Skin, "Hangover"
The Mavericks, Trampoline, "Dream River"
Allison Krauss, New Favorite, "New Favorite"
Steely Dan, Gaucho, "Babylon Sisters"

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

Algunas veces es útil resaltar aspectos particulares de una pista musical para un oyente. Por ejemplo, sobre la pista de Charly García, "No Voy en Tren " una guía de escucha podría ser:

- Frescura de los instrumentos.
- Respuesta en transientes de la batería.
- Dicción e intimidad de la línea vocal.
- Sibilancia natural.
- Guitarra líder suave.
- Voces internas como piano, cuerdas y sintetizadores.
- Reverberación sobre las voces.
- Mezcla brillante con claridad.
- Compresión natural.
- Equilibrio de mezcla a niveles bajos.
- Sobre todo alta fidelidad.

Sobre la pista de Pink Floyd "On the Turning Away," algunos apuntes musicales son:

- Suave respuesta en el bajo.
- Voces claras y con presencia.
- Reverberación audible, eco y efectos de doblaje en las voces.
- Guitarra acústica limpia con detalles del instrumento.
- Teclados limpios.
- Pandereta crujiente.
- Bombo fuerte.
- Tambores grandes.
- Guitarra rítmica y melódica prominente.

Cuando se provee guías de escucha como estas, como se ha señalado es necesario que los oyentes entiendan el significado de los términos que han sido empleados. Diferencias en el uso podrían claramente invalidar los resultados de cualquier sesión de escucha.

Como una alternativa para el contenido musical, a veces es útil emplear contenido no musical. Este contenido es seleccionado debido a características acústicas específicas. Porque estas pistas son generalmente menos complicadas que la música, y quizás de corta duración, y podría ser más fácil de usar. Algunos ejemplos incluyen: colas de reverberación, ondas sinodales, ruido rosa correlacionado y no correlacionado, discursos de hombres y mujeres, castañuelas, y escalas cromáticas de piano muestreadas.

EVALUACIÓN DE CÓDECS PERCEPTUALES

Los códecs perceptuales ofrecen oportunidades únicas y desafían a las pruebas subjetivas. Los códecs perceptuales son diseñados para aprovechar las debilidades en el sistema de audición humano. Y para evitar detección de su transparencia degradada. A velocidades de bits muy bajas, un códec no puede ser designado para ser transparente, entonces evaluaciones de escucha son necesarias para optimizar la calidad de sonido minimizando distorsiones audibles.

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

Cuando un códec no es transparente, distorsiones tales como cambios de timbre, ráfagas de ruido, sonido de ambiente granular, desplazamiento de imágenes estéreo, y el ruido espacialmente enmascarado puede ser usado para identificar la "*firma*" del dispositivo. La reducción del ancho de banda es además es fácilmente evidente, pero una constante reducción del ancho de banda es menos notable que un ancho de banda continuamente cambiante. Cambios en contenidos con altas frecuencias en sucesivos bloques de transformación pueden crear distorsiones audibles.

La voz es a menudo una señal de prueba difícil porque este códec requiere de una alta resolución en tiempo y frecuencia. Con bajas velocidades de bits o largas ventanas de transformación, los códecs de voz pueden asumir una cualidad reverberante extraña. Una grabación envolvente o estéreo de una audiencia aplaudiendo a veces revela errores de codificación espacial. Códecs de sub banda pueden ser ruido de cuantificación no enmascarado que aparece como una ráfaga de ruido en un bloque procesado.

En códecs de transformación, los errores son reconstruidos como funciones básicas (por ejemplo, una impresión coseno) de una transformación de un códec. Un códec con una larga longitud de bloque puede exponer una ráfaga de ruido con pre-eco justo después de una transigente, o podría tener un sonido de campanilleo o un ataque suavizado. La transformación de códecs distorsionados tienden a ser más audibles en altas frecuencias. Cambios en bits asignados a altas frecuencias pueden resultar en un sonido giratorio debido a cambios en el timbre de altas frecuencias.

Con un tipo de códec, por ejemplo, con códecs MP3, el MPEG estándar indica que todas las complicaciones de decodificadores deberían ejecutarse idénticamente y sonar lo mismo. Estos son codificadores es más probable que introduzcan diferencias sonoras. Sin embargo, algunos decodificadores podrían no implementar apropiadamente el MPEG estándar y por lo tanto no son compilados. Por ejemplo, no podrían soportar la intensidad de codificación estéreo. Codificadores MP3, por ejemplo, pueden diferir significativamente en el rendimiento del audio. Otro factor es el método de codificación de estéreo conjunto, y como este es optimizado para un número de canales particulares, ancho de banda y tasa de bits. Muchos códecs tienen un rango de tasa de bits óptimo; la calidad no mejora significativamente por encima de esas tasas, y la calidad puede decrecer drásticamente por debajo de estas. Un entendimiento de las propiedades de diseño de dispositivos como códecs puede ayudar a enfocar el alcance una prueba de escucha y, por lo tanto, mejora su eficacia.

EVALUACIÓN DE CÓDECS DE VOZ

El rendimiento de los sistemas de voz codificados como los usados en teléfonos celulares puede ser valorado a través de evaluaciones de escucha subjetivas. Por ejemplo, la puntuación media de la opinión (Mean Opinion Score o MOS) es a menudo usada. En esta escala de cinco puntos, una puntuación de 4.0 o mayor denota una alta calidad o un rendimiento casi transparente. Puntuaciones entre 3.5 y 4.0 son a menudo permisibles para sistemas de comunicación por voz. Puntuaciones bajo 3.0 denotan una baja calidad. Por ejemplo, una señal puede ser altamente inteligible, pero la voz puede sonar sintética o poco natural. La escala de calidad de circuito de mérito (CM) puede ser usada para calificar el rendimiento subjetivo. El promedio de puntuaciones CM obtenidas de oyentes entrega una respuesta de opinión media. El rango de puntuaciones CM esta mostrada en la siguiente tabla.

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

Tabla 2 Rango de puntuaciones CM para rendimiento subjetivo de códecs de voz.

CM 5	Excelente	Voz perfectamente comprensible.
CM 4	Buena	Voz fácilmente comprensible .
CM3	Justo	Voz comprensible con un pequeño esfuerzo, repeticiones son ocasionalmente necesarias.
CM2	Pobre	Voz comprensible solo con considerable esfuerzo, repeticiones frecuentes son necesarias.
CM1	Insatisfactorio	Voz no comprensible.

La prueba de diagnóstico de rima (Diagnostic Rhyme Test o DRT) es un método subjetivo estandarizado usado para evaluar inteligibilidad del habla. En esta prueba de la ANSI (American National Standard Institute), los oyentes son presentados con pares de palabras. Ellos deben escoger que palabra que perciben. Las palabras son diferentes solo en sus consonantes líderes y las parejas son seleccionadas de manera que seis atributos de inteligibilidad de la palabra son evaluados en estado presencia o ausencia de esta. La prueba de diagnóstico de medida de aceptabilidad (Diagnostic Acceptability Measure o DAM) también se utilizada para evaluar rendimiento de códec de voz.

Históricamente, los sistemas telefónicos proveen un ancho de banda estrecho desde aproximadamente 300 a 3400 Hz. Es generalmente aceptado el corte de frecuencias bajas, aunque cause un sonido "delgado" debido a la falta de frecuencias graves y en consecuencia decrezca la calidad de sonido, no disminuye mucho la inteligibilidad de la palabra, esto ha sido reportado ya que la frecuencia de corte superior de 3400 Hz permite que el 97% de los sonidos de voz sean entendidos. El reconocimiento de oraciones puede alcanzar el 99% de inteligibilidad debido a la ayuda del contexto. Los nuevos anchos de banda para los sistemas telefónicos de voz proveen una respuesta desde 50 Hz a 7000 Hz. Aunque la progreso en mediciones de inteligibilidad es ligero, estos sistemas proveen un sonido de voz más natural con claridad en la calidad del sonido.

EVALUACIÓN DE INTELIGIBILIDAD DE HABLA EN SALAS

La inteligibilidad del habla es la más alta prioridad de diseño para cualquier espacio acústico destinado para la palabra hablada. Este es el caso de muchos lugares de culto, auditorios, y teatros. En salas donde la amplificación no es usada, los elementos acústicos de diseño de la sala deben ser cuidadosamente considerados para proveer un algo nivel de inteligibilidad de la palabra. Los sistemas de sonido son a menudo usados para superar limitaciones acústicas, y para proveer inteligibilidad en lugares muy grandes.

La inteligibilidad de la palabra en una sala es a menudo estimada usando mediciones basadas en sujetos, es decir, usando experimentación en vivo. Un interlocutor lee una lista de palabras y frases,

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

y los oyentes en la sala escriben lo que ellos escucharon, la lista incluye ejemplos de sonidos importantes de habla. Una evaluación puede usar entre 200 y 1000 palabras. Por ejemplo, algunas palabras en español usadas en una prueba de inteligibilidad son: casa, mano, moto, torta, luna, mesa, perro, bota, globo, plato, cama, nene, silla, gallina, hamaca, jabón, zapatilla, pelota, caballo gato, vaca, helado, boca, gordo, vaso, entre otras.

Mientras más alto el porcentaje de palabras y frases correctamente entendidas, más alta la inteligibilidad de la palabra. En algunos casos, dificultad de escucha es medida. Cuando el nivel de la palabra es el mismo que el del ruido, la inteligibilidad puede ser alta, pero los oyentes pueden todavía tener dificultad en entender que está siendo dicho, y es requerirá una atención considerable. Cuando el nivel de la palabra es aumentado por 5 o 10 dB sobre el nivel del ruido, la inteligibilidad no mejora mucho, pero oyentes informan mucha menor dificultad al escuchar.

EVALUACIÓN ESTADÍSTICA DE LAS EVALUACIONES DE ESCUCHA

Para que las interpretaciones de resultados de evaluaciones de escucha sean significativas deben ser consideradas cuidadosamente. Por ejemplo, en una prueba ABX, si un oyente identifica la referencia en 12 de 16 pruebas, ¿Ha tenido una diferencia audible notoria? El análisis estadístico es usado para interpretar los resultados. Debido a que la evaluación es una muestra, nosotros definimos nuestros resultados en términos de probabilidad. Mientras más grande la muestra más confiable el resultado. La preocupación central es el significado de los resultados. Si los resultados son significativos, son debido a diferencias audibles. De lo contrario son debidos a una casualidad. En una evaluación ABX, una puntuación correcta de 8 de 16 veces indican que el oyente no ha escuchado diferencias; el resultado lo pudo obtener adivinando. Una puntuación de 12/16 podría indicar una diferencia audible, pero podría ser obtenida al azar. Para evaluar esto, nosotros podemos definir una hipótesis nula H_0 que el resultado fue obtenido al azar y una hipótesis alternativa H_1 que se debe a una diferencia audible. El nivel de importancia α es la probabilidad de que la puntuación sea obtenida al azar. El criterio de importancia α' es el umbral escogido para α que va a ser aceptado. Si α es menor o igual a α' luego nosotros aceptamos que la probabilidad es suficientemente elevada para aceptar la hipótesis que el puntaje es debido a diferencias audibles. La selección de α' es arbitraria pero un valor de 0.05 es a menudo usado. Esto puede ser analizado con distribución binomial usando la ecuación:

$$z = \frac{c - 0.5 - np_1}{[np_1(1-p_1)]^{1/2}} \quad \text{Ecuación 8}$$

Donde:

z es la desviación estándar normal

c es el número de respuestas correctas

n es el tamaño de la muestra

p_1 es la proporción de respuestas correctas en una población debido al azar.

Con un puntaje de 12/16, $z = 1.75$. La distribución binomial por lo tanto produce un nivel importante de 0.038. La probabilidad de obtener un puntaje alto de 12/16 al azar (y no por diferencias audibles)

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

es 3.8%. En otras palabras, hay un 3.8% de probabilidad que el oyente no note la diferencia. Sin embargo, ya que α es menor a α' ($0.038 < 0.05$) nosotros concluimos que el resultado es significativo y que hay una diferencia audible, al menos de acuerdo a como nosotros seleccionemos nuestro criterio de importancia. Si α' es seleccionado para ser 0.01, luego el mismo resultado de 12/16 no sería relevante y nosotros podríamos concluir que el puntaje es obtenido al azar.

Nosotros podemos definir parámetros que caractericen el riesgo que estamos aceptando mal una hipótesis. Un error de riesgo Tipo 1 (a menudo conocido como α') es el riesgo de rechazar la hipótesis nula cuando esta es en realidad verdadera. Este valor es determinado por el criterio de importancia; si $\alpha'=0.05$ entonces nosotros podemos estar mal el 5% de las veces asumiendo resultados significativos. El error de riesgo tipo 2 (β) define el riesgo de aceptar una hipótesis nula cuando esta es falsa. El riesgo Tipo 2 es basado en el tamaño de la muestra, el valor α' , en el valor de puntuación por oportunidad, y el efecto del tamaño o la más pequeña puntuación sería significativa. Estos valores pueden ser usados para calcular el tamaño de la muestra usando la ecuación:

$$n = \left\{ \frac{z_1[p_1(1-p_1)]^{1/2} + z_2[p_2(1-p_2)]^{1/2}}{p_2 - p_1} \right\}^2 \quad \text{Ecuación 9}$$

Donde:

n es el tamaño de la muestra.

p_1 es la proporción de respuestas correctas en una población debido al azar.

p_2 es el tamaño del efecto: proporción hipotética de respuestas correctas en una población debido a diferencias audibles.

z_1 es el valor de la distribución binomial correspondiente al error de riesgo tipo 1.

z_2 es el valor de la distribución binomial correspondiente al error de riesgo tipo 2.

Por ejemplo, en una evaluación ABX, si el riesgo tipo 1 es 0.05, el riesgo tipo 2 es 0.10, y el tamaño del efecto es 0.70, entonces el tamaño de la muestra podría ser 50 pruebas; cuanto menor sea el tamaño de la muestra, es decir, el número de pruebas mayor será el riesgo de error. Por ejemplo, si son conducidos 32 pruebas, $\alpha'=0.05$, y el tamaño del efecto es 0.70. Para alcanzar un resultado estadísticamente significativo, un puntaje de 22/32 es necesario.

El análisis de distribución binomial provee buenos resultados cuando un gran número de muestras están disponibles. Otros tipos de análisis estadísticos como la teoría de detección de señal puede ser aplicada a ABX y otras evaluaciones subjetivas. El análisis estadístico puede parecer impresionante, pero estos resultados no pueden validar una evaluación que es inherentemente defectuosa.

EJEMPLO DE UN PROCEDIMIENTO Y UN FORMULARIO DE EVALUACIÓN

Una evaluación de escucha, preparada por Ken Pohlmann, usa comparación por parejas para evaluar pequeñas degradaciones. El análisis compara un dispositivo bajo análisis (DUT) con una referencia,

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

o un DUT contra otro DUT. En este ejemplo, son utilizados auriculares para la reproducción. Un ejemplo de evaluación del flujo de señal es mostrado en la figura 3. Este análisis cumple 3 requisitos:

1. Determina si un oyente puede oír una diferencia entre el par de fuentes, si es así, el oyente es examinado.
2. Pregunta al oyente examinado por comentarios subjetivos.
3. También pregunta al oyente examinado calificar la fuente usando una escala numérica.

La evaluación está en 2 partes. El formulario presentado a los oyentes es mostrado en la figura 4. La parte uno es una prueba *ABX*; esta determina si el oyente puede escuchar una diferencia entre las dos fuentes, el análisis es doble-ciego, usando una evaluación *ABX*. Por ejemplo, si un oyente tiene 12 respuestas correctas de 16, nosotros estamos 95% confiados que el escucho una diferencia (se dice que el oyente va a ser examinado). Nosotros podemos mejorar la confianza incrementando el umbral (por ejemplo, 13 o 14 respuestas correctas), o incrementando el número de pruebas (por ejemplo, 20,50,100). Si un oyente no puede escuchar ninguna diferencia, ellos no son tomados en cuenta. Su opinión no es útil.

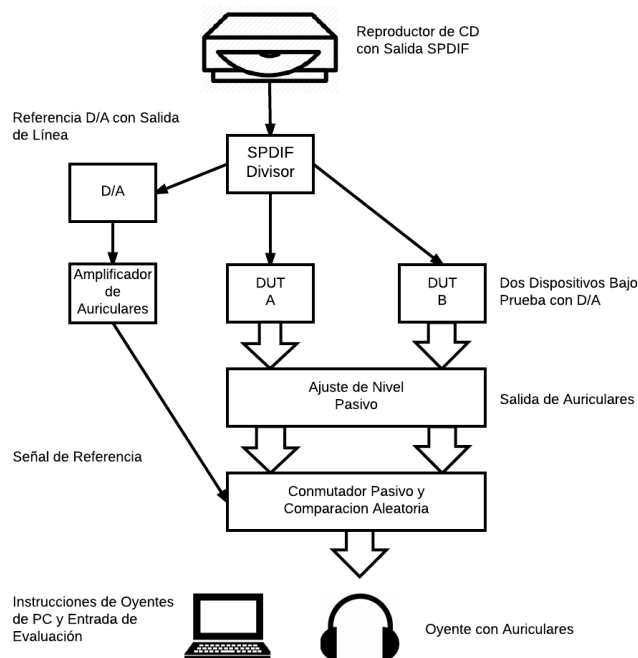


Figura 3 Flujo de señal usado para una evaluación de escucha subjetiva.

La parte dos es una evaluación no ciega. El oyente examinado escribe comentarios describiendo su evaluación subjetiva. El formulario señala múltiples áreas a escuchar (balance tonal, rango dinámico, etc.). Además, se pregunta al oyente por una puntuación numérica.

El criterio en la parte dos cubre un amplio rango de parámetros de escucha. Estos son seleccionados por escucha de auriculares y son omitidos algunos parámetros tradicionales asociados con los efectos de altavoces en salas. Los criterios son comprensivos, pero en muchos casos el oyente puede

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

concentrarse en unos pocos criterios (como el bajo nivel de detalle o distorsiones audibles) y no tiene ningún comentario significativo en otras áreas.

En la parte dos, el par comparado puede ser un *DUT* y una referencia, o dos *DUTs*. En el primer caso, el oyente no tiene que escribir comentarios acerca de la referencia, en el segundo, el o la oyente debe escribir comentarios sobre ambos *DUTs*. La razón para comentar es que ayuda a los diseñadores a entender que está escuchando el oyente, y posiblemente objetando algo.

En la parte dos, tanto a la referencia o el *DUT A* se otorga arbitrariamente una puntuación de 0. La otra fuente se califica en una escala de ± 10 para este criterio. Esto permite para los *DUTs* que suenan peor, o mejor que la referencia. Como se ha señalado, dos *DUTs* pueden ser puntuadas como una; esto es esencialmente un ahorro de tiempo. Si es posible, es recomendable que cada *DUT* sea evaluado y puntuado con la referencia. Esto provee una puntuación estable.

En algunas otras evaluaciones, esencialmente, la parte uno y la parte dos son fusionadas. El oyente debe evaluar y puntuar a ciegas. Esto no es recomendado. Sobrecarga a los oyentes y puede disminuir la calidad de los comentarios y las puntuaciones. Además, conduce a problemas prácticos. Si es revelado que un oyente no escuchó una diferencia en un ensayo, entonces estos comentarios pueden ser omitidos. Pero si el oyente no escuchó una diferencia en un ensayo, ¿pueden ser confiables los resultados? qué pasa si, sobre ese ensayo en particular, ¿el oyente estuvo bien por esa ocasión? Es mejor tener dos partes, en las cuales sea determinado con anterioridad, con un error estadístico, que el oyente está escuchando una diferencia y por lo tanto sus comentarios son válidos.

En esta evaluación, el oyente usa un archivo que contiene pistas musicales y no musicales. El oyente está libre de seleccionar cualquier pista para la prueba *ABX* y para evaluar cualquier criterio. Algunas otras evaluaciones asignan pistas específicas para un criterio específico. Pero los oyentes son muy individualistas en como escuchan. Asignando pistas específicas se limita innecesariamente sus habilidades de escuchar detalles. Sin embargo, cuando escriben sus comentarios, los oyentes deberían proveer comentarios sobre cuales pistas están escuchado, en este caso es importante para análisis subsecuentes.

Idealmente, una evaluación de audición de preferencias como esta podría usar de 10 a 20 oyentes expertos; este consenso debería proveer resultados certeros. En la práctica, esto es difícil de lograr. Un enfoque todavía válido es usar 1 o 2 oyentes expertos en una evaluación doble ciega. Para facilitar esto, la evaluación elimina el sesgo del oyente, verifica que diferencias están siendo escuchadas, y por lo tanto se puede confiar en el puntaje de unos cuantos oyentes expertos.

Después de una escucha preliminar, se preparará un archivo de compilación utilizando estas (y otras) pistas. El formulario entregado a los oyentes también provee definiciones de algunos términos comunes para escuchas subjetivas. Como se ha señalado a menos que el significado de los términos de la evaluación sea bien comprendido, los resultados subjetivos no serán significativos.

Figura 4 Ejemplo de un Formulario Subministrado a los Oyentes que Realizan una Evaluación en una Prueba de Audición de Pequeños Deterioros.

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

Esta sección es un ejemplo de un formulario de evaluación proporcionado a oyentes rindiendo una evaluación subjetiva.

Nombre del Oyente:
Fuente de Audio A:
Fuente de Audio B:
Auriculares Usados:
Fecha:

Evaluación de Escucha Parte 1

Instrucciones para el Oyente: Use el comparador *ABX* para comparar dos fuentes de audio, usando auriculares. Una Fuente es *A*, y la otra es *B*. En cada prueba, *A* o *B* es asignada aleatoriamente para *X*. Su tarea es escuchar *A*, *B*, y *X* y determinar si *A* se encuentra asignada a *X*, o si *B* se encuentra asignada a *X*. Ud. puede escuchar a *A*, *B* y *X* tantas veces como se requieran. Ud. puede usar las pistas de audio que desee. Ud. puede escribir cualquier comentario, así como también que pista de música es la más útil si esto le ayuda a completar la evaluación; estos comentarios no van a ser calificados. Luego escriba si *X* es *A* o *B* para cada prueba. Por favor complete 16 pruebas usando esté formulario:

Número de Prueba	Comentarios	Notas musicales	¿X está asignada a A o B?
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
Su puntuación para esta evaluación: /16			

MÉTODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

El comparador *ABX* revelará las respuestas correctas en el modo "Respuestas". Si Ud. puntuó correctamente 12 o más respuestas correctas, ¡felicitaciones! Ud. probablemente está escuchando diferencias entre *A* y *B*. Ud. puede proceder a la parte dos de la evaluación. Si su puntuación es menor a 12, Ud. puede no estar escuchando ninguna diferencia. Ud. puede reiniciar el comparador *ABX* e intentar de nuevo, escuchando más de cerca a pequeños detalles. La evaluación *ABX* podría automáticamente crear nuevas tareas.

Evaluación de Escucha - Parte Dos

Instrucciones para el Oyente: Use el comparador *ABX* en el modo "*respuestas*" para seleccionar entre *A* y *B*. Solo hay un intento, las asignaciones de *A* y *B* no cambian, y Ud. verá que fuente está escuchando. En la parte dos, Ud. no está siendo calificado. Por el contrario, Ud. está puntuando la calidad del sonido de las fuentes *A* y *B*. Su tarea es escuchar *A* y *B* tantas veces como desee. Ud. puede usar cualquier pista de música que Ud. quiera.

La fuente *A* podría ser un dispositivo bajo análisis, o una referencia conocida. Ud. está comparando con la fuente *B*. Por favor escuche los diferentes Criterios de Sonido, y escriba comentarios y anote las pistas musicales para las 2 fuentes. Las definiciones de los términos de Criterios de Sonido están listados a continuación. La fuente *A* o la referencia siempre tiene un puntaje de 0 (cero o neutral). Compare la fuente *B* con este 0, que es un punto de referencia, y luego puntúe la calidad del sonido de *B*. Su puntuación será un entero entre -10 (muy malo) y +10 (muy bueno) Por ejemplo, si la fidelidad de *B* es solo un poco mejor que *A*, entonces su puntaje de fidelidad será +1. Por favor use la siguiente escala de puntuación:

-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
Muy Malo			Algo Malo				Igual				Algo Bueno				Muy Bueno					

Antes de comenzar la evaluación, es importante que Ud. entienda el significado de los términos de criterios de Calidad de Sonido (como balance tonal, rango dinámico, etc.). Los términos son definidos en el formulario. Estos términos podrían tener un significado que es diferente del significado que Ud. está acostumbrado. Es importante que Ud. use los termino como están definidos aquí.

Fuente A o Referencia			Fuente B	
Criterio de sonido	Comentarios y Anotaciones Musicales	Puntaje	Comentarios y Anotaciones Musicales	Puntaje ± 10
Balance Tonal		0		
Rango Dinámico		0		
Claridad de Bajos		0		
Claridad en Rango Medio		0		
Claridad en Agudos		0		
Imagen		0		
Espacialidad		0		
Detalles a Bajo Nivel		0		
Distorsiones Audibles		0		
Fidelidad Total		0		
			Puntaje Fuente B	

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

Definiciones de Criterios de Sonido

Balance Tonal. (Balance Espectral, Balance de Octava). Un dispositivo de reproducción que es tonalmente balanceado tiene una respuesta de frecuencia consistente que idealmente es plana a través del ancho de banda audible. Ningún área de frecuencia es dominante, o carente. Un sonido que tiene un balance tonal pobre podría tener pico o pendientes en la respuesta de frecuencia, y esto podría exagerar u oscurecer ciertas notas o cambiar el timbre. Bajas frecuencias deberían mezclarse naturalmente con altas frecuencias. Un balance correcto además requiere una extensión de frecuencias bajas y altas buenas, en el fondo y en el tope del espectro.

Rango Dinámico. El sonido debería tener un viaje natural a través del espectro de volumen desde lo más suave a lo más fuerte, sin tensión. A altos niveles, el sonido no debería endurecer o comprimir el sonido con una disminución en el rango dinámico registrado. No debería haber realces espectrales a diferentes niveles de volumen. No debería haber bombeo dinámico, como un rápido decrecimiento en frecuencias medias cuando una nota de bajo es pulsada. A niveles máximos de volumen, cuando suena un pasaje musical estridente, el sonido debería mantenerse claro, sin recortes y libre de distorsión excesiva en todas las frecuencias. A altos niveles, escuchar frecuencias bajas que estén más duras o nítidas en el timbre; por ejemplo, el sonido de un bombo puede cambiar. Escuchar cambios en el espectro dinámico, cualquier diferencia en timbre o balance espectral que cambia de volumen suave a fuerte. A altos niveles, esto podría ser causado por compresión en uno o varios rangos de frecuencias. A bajos niveles, el sonido debería mantener un balance espectral musicalmente apropiado. Esto no está relacionado con la compensación de volumen; la compensación de volumen está relacionada con el volumen de reproducción, no a la sonoridad y la dinámica del material musical. El máximo nivel de volumen debería ser adecuadamente estridente para una reproducción satisfactoria.

Claridad de Bajo. (definición, diferenciación). El sonido tiene bastante claridad cuando Ud. Puede escuchar detalles musicales en la región del bajo. Instrumentos como la batería y bajo deberían tener una calidad de sonido natural que es sólido y apretado sin ablandamiento. Las transientes tienen un ataque limpio con un golpe de bajo preciso. Cuando el sonido del bajo tiene una claridad pobre, podría ser espeso, estruendoso, débil o delgado.

Claridad en Rango Medio. (Rango Medio, Definición, Distinción). El sonido tiene mucha claridad cuando puede escuchar detalles musicales en la región media. Voces e instrumentos deberían tener una calidad de sonido natural y suave. Las voces deberían ser inteligibles. Las transientes deberían ser limpias con un ataque ligero. Cuando los sonidos del rango medio tienen una claridad pobre, las voces podrían sonar sibilantes, roncadas o nasales. Los instrumentos podrían sonar espesos o ruidosos. Ud. podría escuchar cambios en el espectro como una coloración, y zumbido, o un sonido cornetoso. Podría ser borroso en frecuencias de rango medio causadas por intermodulación de bajas frecuencias. Podría haber intermodulación cuando muchas voces, como cuando un coro cantando. Podría haber un sonido brillante duro en el rango de frecuencias medias altas. Escuchar en cualquier momento distorsión como un efecto peine que puede cambiar en timbre o un sonido hueco.

Claridad en Agudos. (claridad en brillos, definición, distinción). El sonido tiene mucha claridad cuando Ud. Puede escuchar detalles en la región de frecuencias altas. Los instrumentos deberían tener un sonido limpio y natural con una sensación de "aire" alrededor de estos. Los transientes deberían ser crujientes, limpios, y delineados. Cuando un sonido agudo es pobre, los instrumentos podrían sonar ruidosos, filudos, demasiado brillantes, demasiado aburridos, o estruendosos. Podría ser un sonido de brillo duro en altas frecuencias.

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

Imagen. Cuando se reproduce un sistema estéreo, debería ser una propagación consistente de imágenes a través del panorama estéreo. El panorama debería alcanzar continuamente de un canal al otro sin brechas. Las imágenes deberían ser generalmente estables y localizadas precisamente a través del panorama, y no muy difuso. Las imágenes deberían ser diferentes y distintamente separadas como voces individuales; escuchar el "espacio en blanco entre instrumentos." Debería haber una profundidad exacta adelante y atrás en una grabación; para lograr esto, debe reproducir exactamente la proporción de sonido directo, para sonidos reverberantes. Una imagen pobre podría resultar en la localización de un instrumento en más de un lugar, o frecuencias altas y bajas viniendo de diferentes lugares. La localización debería no cambiar con cambios de tono, volumen o timbre. Movimientos de cabeza de un oyente o cambios de asiento no debería cambiar mucho la imagen. Sonido localizado para un subwoofer es otro ejemplo de una imagen pobre. Un sonido monoaural habría de tener una imagen en el centro, pero no otras, esta imagen amplia es muy pobre.

Espacialidad (ambiente). Un sonido con espacialidad estéreo debería generalmente ser abierto, con amplitud y anchura y profundidad de adelante hacia atrás. Tiene plenitud; llena simétricamente el espacio musical y tiene la presencia espacial y la acústica geométrica de una sala natural de música interpretada en un escenario. La vivacidad y presencia debería envolverle espacialmente dando una sensación de que "Ud. está ahí". Más que escuchar un ambiente reverberante, Ud. está en el ambiente. El ambiente debería ser difuso y viniendo de todas las direcciones, y no localizado. El ambiente debería tener balance espectral; la reverberación no debe ser demasiado tenue, o demasiado estruendosa, el espacio es consistente con el espacio de la grabación; Ud. debería estar dispuesto a discernir diferencias en el tamaño y la respuesta acústica de diferentes ambientes grabados. Una espacialidad que no es espaciosa podría ser cerrada o estrecha, esta podría ser plana o con falta de presencia. Una espacialidad pobre podría tener distorsiones de fase extrañas un sonido desfasado con efectos direccionales inconsistentes, y cambios espectrales resultantes desde pequeños cambios en la posición de la cabeza del oyente.

Detalles a Bajo Nivel. Sonidos muy suaves deberían ser reproducidos limpiamente. Sonidos ambientales naturales, ruido, y el tono de la sala deberían ser todos audibles no oscurecidos o enmascarados por sonidos introducidos. Las colas de reverberación deberían decaer suave y limpiamente al silencio. El dispositivo debería no añadir ningún otro sonido suave, sin linealidades, ruidos de granulación u otros ruidos. Escuchar cuidadosamente cualquier ruido que sea dependiente en la señal de audio.

Distorsiones Audibles. (Otros sonidos extraños). Un dispositivo reproductor debería ser limpio y solo transmitir los sonidos en la grabación original. Idénticamente, no debería añadir sonidos para nada, especialmente sonidos que son independientes de la señal de audio. Sonidos extraños deberían incluir ruido de fondo, otros ruidos, *hum*, *buzz*, *clicks*, *pops* y otras distorsiones. Asegúrese que algunos sonidos extraños que escuche no estén en la grabación original. No deberían ser añadidos sonidos o cambios de timbres especialmente en altas frecuencias.

Fidelidad Total. Este criterio resume sus impresiones y preferencias. Un dispositivo con buena fidelidad es transparente, ninguna adición o substracción desde el sonido originalmente grabado. El sonido debería tener relación preestablecida con el dispositivo reproductor. Esto mide la exactitud general y la amabilidad del sonido.

CONCLUSIONES

En un futuro, es posible que mediciones objetivas sean desarrolladas serán evaluadas comprensivamente en el rendimiento estético de una señal de audio. Tales mediciones modelarían

METODOS SUBJETIVOS PARA LA EVALUACIÓN DE CALIDAD DE SONIDO

el rendimiento procesado por el oído humano, eliminando la necesidad de oyentes. Parámetros como claridad, plenitud, imagen, y espacialidad serían cuantificados, o quizás descartados en favor de mejores mediciones. Quizás el sistema podría permitir que ingenieros calibren el sistema para proveer resultados estéticos como por ejemplo un sonido cálido. Sin embargo, hasta que mediciones objetivas sean desarrolladas, escuchas subjetivas proveerán un valioso aporte en el desarrollo de dispositivos de audio. A pesar que consume tiempo, es apropiado que oyentes formen las señales que serán apreciadas por sus pares.

BIBLIOGRAFÍA

- Ballou, G. (2015). *Handbook for sound engineers* (Fifth ed.). Burlington: Focal Press.
- D'Onofrio Julieta, N. A. (Julio de 2012). Verificación de la inteligibilidad del habla con audífonos en pacientes adultos: Listas de palabras fonéticamente balanceadas del Dr. Tato y cols. vs. Listas de palabras P.I.P.-C 25. Argentina.
- European Broadcasting Union . (Mayo de 1998). Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic. págs. 1-17.
- Lachlan. (18 de Febrero de 2014). ABX Testing: 128k AAC vs Lossless. Obtenido de <https://www.youtube.com/watch?v=UrfX-g8auc8>
- Miyara, F. (s.f.). El ruido y la inteligibilidad de la palabra.
- Obregon Gaviria, D., & Sierra Gallon, J. S. (2017). EVALUACIÓN OBJETIVA Y SUBJETIVA DE PARÁMETROS DE AUDIO RELACIONADOS CON LA CALIDAD EN PRODUCCIONES MUSICALES FINALIZADAS. Medellin, Colombia.
- Pawlowski, P. (2017, Septiembre 09). *foobar2000*. Retrieved from foobar2000: http://www.foobar2000.org/components/view/foo_abx
- Union Internacional de Telecomunicaciones. (2015). Métodos para la evaluación subjetiva de pequeñas degradaciones en los sistemas de audio . *UIT-R*, 1-33.