

EVALUACIÓN SUBJETIVA DE LA CALIDAD DEL SONIDO MEDIANTE TEST DE PAREJAS

PACS: 43.66.Lj

Ferrer Contreras, Miguel; de Diego Antón, María; Piñero Sipán, Gema; González Salvador, Alberto

Departamento de Comunicaciones, Universidad Politécnica de Valencia

Camino de Vera s/n,
46022 Valencia. España.

Tel: 963 877 300

Fax: 963 877 309

e-mail: mifercon@teleco.upv.es; mdediego@dcom.upv.es; gpinyero@dcom.upv.es;

agonzal@dcom.upv.es

ABSTRACT

The subjective sensation that the sound of a manufacturer object provides at the customers is a very important fact that carrier weight to the total satisfaction by the product and this does tools and models for sound quality have been needed (car industry, for example). The subjective view of the man is requested for this and the used of the method of paired comparisons is enough extended for to get this objective. This method allow to offer a ranking of the sound quality after a jury test (with adequate number of subject no necessarily experts) about the tested sounds is done. In this work a software for designing, running and analysing test according to the named method is presented. The capability of the software like helper for make test of pair and how it work is showed. The different results we can obtain of the analysis (repeatability and consistency of the subject, concordance of the different subjective characteristic analysed, and others parameters about the sounds like probabilities of choice, merits values,...) are proposed. Finally, an example of subjective evaluation about several engine sounds of diesel automobile under different rules of work is presented.

RESUMEN

La sensación subjetiva que la acústica de un producto industrial proporciona en un usuario final es uno de los factores más importantes que contribuyen a la satisfacción general del cliente por el producto. Por eso, sectores como el de la automoción están necesitando de herramientas y modelos capaces de evaluar la calidad del sonido, para lo que resulta imprescindible la opinión subjetiva humana. Uno de los métodos más usados para la emisión de juicios subjetivos sobre diferentes sonidos es el "Método de comparación por parejas", que permite valorar un conjunto de señales a partir de los resultados obtenidos tras la realización de diferentes tests por un número suficiente de jueces, no necesariamente expertos. En este trabajo se explica como se realizan estos ensayos y se presenta una herramienta software capaz de automatizar la mayor parte de las acciones necesarias para el diseño de tests que se ajusten al modo de operar de dicho método, así como facilitar la realización de los tests y proporcionar los resultados del análisis: tanto de los jueces (repetitividad, consistencia,...), como de los sonidos (probabilidad de selección, valoración,...), como de las diversas características subjetivas (concordancia,...) que se pretendan analizar en el experimento. Como ejemplo de funcionamiento del software se

presentan los resultados de la evaluación subjetiva de distintas señales de motor de automóvil diesel de las mismas rpm, bajo distintas condiciones de funcionamiento.

1. INTRODUCCIÓN

Cuando se quiere evaluar de forma subjetiva cualquier tipo de evento, tarde o temprano tendremos que acudir a las opiniones personales que diferentes sujetos (normalmente expertos) emiten sobre dicho evento. Traducido a la evaluación subjetiva de la calidad de los sonidos, tendríamos que buscar un jurado de expertos en el tema, que evaluaran los sonidos deseados según criterios preestablecidos. Esto conlleva enormes inconvenientes como: la elección correcta de los miembros del jurado, la dificultad que significa emitir juicios absolutos, consensuar las diferentes opiniones de cada juez, etc. Sin embargo, mediante el método de comparación de parejas, eliminamos la mayoría de estos inconvenientes, ya que los jueces sólo tienen que emitir juicios comparativos entre parejas de eventos (sonidos en este caso).

Un estudio de la calidad del sonido realizado mediante el método de comparaciones de parejas [1], consiste básicamente en presentar a cada juez los diferentes sonidos involucrados en dicho estudio, para que éste lo evalúe, pero con la particularidad de que dichos sonidos siempre son presentados de dos en dos, y los juicios emitidos por el juez, han de ser relativos entre cada pareja presentada. La información recogida de todos los jueces, permite clasificar los diferentes sonidos y emitir calificaciones tanto cualitativas como cuantitativas de los mismos.

El método de comparación de parejas, resulta por tanto, mucho más sencillo a la vez que robusto para la evaluación subjetiva de sonidos, sobre todo cuando no se dispone de un grupo de expertos que puedan formar parte del jurado. Ejemplos de aplicación de dicho método para el estudio y generación de modelos de calidad del sonido pueden contemplarse en [3] y [4].

Si además, disponemos de algún tipo de herramienta software capaz de automatizar todas las tareas que involucra realizar un test de parejas (sobre todo la toma de datos y su análisis posterior), podríamos reducir notablemente el tiempo de postprocesado así como presentar al jurado un entorno más amigable y cómodo a la hora de realizar el test (circunstancia que favorece la fiabilidad de los resultados emitidos por los jueces).

TESTSOUND es una herramienta software desarrollada para facilitar y automatizar todo el proceso, desde su diseño hasta su análisis, pasando naturalmente por la ejecución de tests de comparación de parejas aplicados al estudio de la calidad del sonido. En este documento se introducen las posibilidades que dicha herramienta incorpora y se presenta un ejemplo de utilización que ilustra la validez del mismo.

2. ESTUDIO DE LA CALIDAD DEL SONIDO MEDIANTE EL MÉTODO DE COMPARACIÓN DE PAREJAS

Realizar un test de comparación de parejas para el estudio de la calidad del sonido (como el propuesto en [2]) puede resultar un trabajo largo y laborioso tanto en la propia ejecución del test como en el procesado de los datos obtenidos del mismo. Antes de que el jurado realice el test se debe proceder al diseño del mismo, que consiste en decidir los sonidos a evaluar, las comparaciones de parejas a realizar entre ellos (normalmente todos con todos menos consigo mismo), el orden en el cual serán presentadas dichas comparaciones y las características de los sonidos sobre las que los jueces deben emitir su opinión. Con estas premisas, el test quedaría diseñado, y estaría listo para ser puesto a disposición del jurado en la realización propiamente dicha del test. Si no se tienen los medios adecuados para automatizar dicha ejecución, este proceso, que es lo verdaderamente importante del test, podría resultar difícil y confuso para los jueces. El jurado dejará reflejadas sus opiniones sobre algún tipo de plantilla (ya sea de papel o informática) para que puedan ser tratadas posteriormente con el objeto de emitir las conclusiones finales del test. El trabajo de postprocesado también resultaría bastante laborioso si no se pudiera automatizar, ya que recoger las opiniones del jurado y tratarlas de forma manual supone un gran esfuerzo no exento de cometer errores. Una vez realizado este procesado de los datos, debemos tener la suficiente información como para emitir juicios fidedignos de la calidad del sonido.

Si disponemos de los sonidos a evaluar en algún formato informático, podríamos usar TESTSOUND para todo el proceso antes descrito, con la ventaja y la comodidad que supone automatizar e integrar dicho proceso en una única herramienta software.

3. USO DE TESTSOUND PARA EL DISEÑO DE TEST DE PAREJAS

Como ya se ha comentado, el diseño del test es el primer paso para la evaluación de la calidad del sonido mediante el método de comparación de parejas.

Una vez decididos que sonidos son los que formarán parte del test, tan sólo hay que interactuar con el programa de forma totalmente intuitiva ya que paso a paso nos irá guiando y pidiendo los datos que necesita para el diseño del test.

Lo primero que aparecerá en el diseño es una pantalla como la de la figura 1, donde tendremos que introducir el número de sonidos (llamados señales) que aparecerán en el test así como el número de comparaciones de las que consta. Después asociaremos secuencialmente a cada señal el archivo (en formato wav) que contiene a cada uno de los sonidos, de forma análoga a la apertura de cualquier archivo en cualquier aplicación Windows.

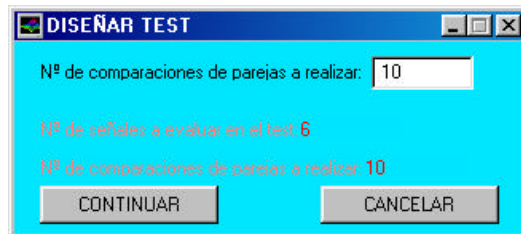


Figura 1: Ventana inicial para el diseño de test de sonido

Llegado a este punto, faltaría determinar qué parejas de sonidos presentar en el test y en qué orden, lo que se puede realizar fácilmente con el programa gracias a diversas herramientas como la capacidad de reordenar aleatoriamente las parejas seleccionadas o especificar manualmente cuál es dicho orden.

Por último, para que el diseño quede completado, faltaría determinar la parte que interactúa con el jurado, es decir, qué características de los sonidos deseamos que los jueces evalúen. TESTSOUND permite que el juez pueda evaluar hasta 14 características en el mismo test de las cuales trece pueden ser determinadas en el diseño del mismo como queda reflejado en la figura 2 (quedando siempre fija en cada diseño la característica preferencia).

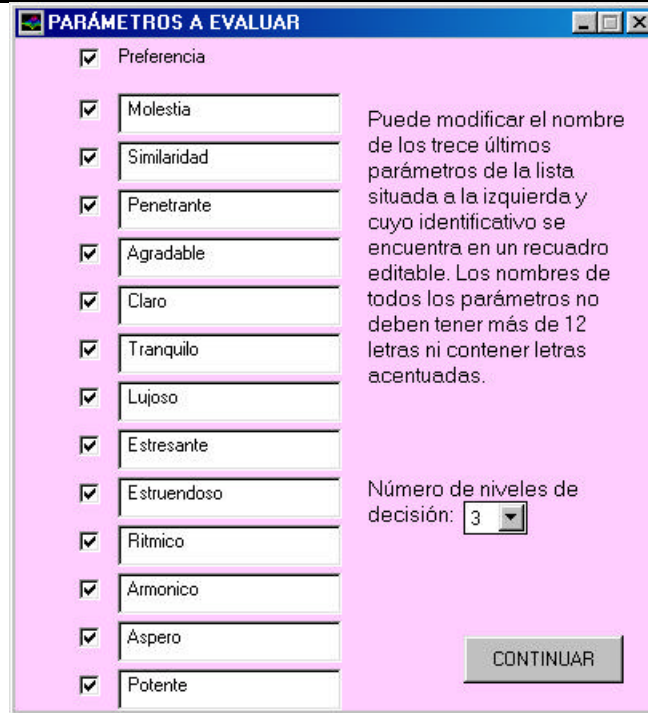


figura 2: Ventana para la selección de las características de los sonidos a analizar en el test.

Toda esta información quedaría guardada en un archivo de diseño (con extensión tst) que posteriormente puede ser invocado por el programa para ser ejecutado o analizado.

4. USO DE TESTSOUND PARA LA EJECUCIÓN DE TEST DE PAREJAS

A la hora de la ejecución de los test TESTSOUND pretende ser práctico a la par que cómodo para los jueces. La ejecución del test deberá realizarla cada juez delante de un ordenador con tarjeta de sonido a través de la cual podrá escuchar los diferentes sonidos a evaluar. Para cada comparación, los jueces tendrán en la pantalla del ordenador una ventana como la de la figura 3. Durante las comparaciones, los jueces podrán escuchar alternativamente los sonidos a evaluar mientras unos indicadores parpadean con la misma frecuencia con que las señales son reproducidas para diferenciarlas entre si. Este proceso continúa hasta que el juez ha emitido todos los juicios que se le requieren en el test para cada una de las comparaciones, lo que tiene que realizar de una forma sencilla seleccionando sobre la casilla que mejor representa su opinión (ver figura 3).

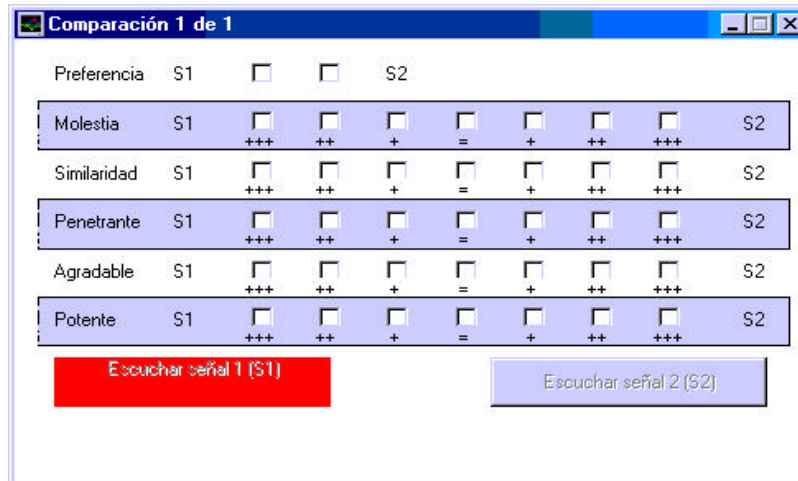


Figura 3: Aspecto de la ventana que visualizará cada juez en la realización del test

5. USO DE TESTSOUND PARA EL ANÁLISIS DE LOS TESTS PREVIAMENTE REALIZADOS

TESTSOUND realiza de forma automática el análisis de los resultados obtenidos de las opiniones de los jueces. A continuación exponemos la información útil que se puede obtener de este análisis.

5.1 Sobre los jueces:

Podemos medir la aptitud de cada juez (si el diseño del test está bien realizado) mediante los parámetros siguientes:

Repetitividad: Da una idea de las veces que un mismo juez mantiene la misma opinión en las diferentes ocasiones que se le plantea una misma comparación. Lo normal en los test de parejas es que una misma comparación aparezca dos veces intercambiando el orden de aparición de los sonidos, por lo que en estos casos podemos dar una medida de la repetitividad de una forma clara, como el tanto por ciento de las dobles comparaciones en las que se mantuvo la misma opinión [1], [2].

Consistencia: Nos indica el grado de coherencia de las opiniones de cada juez advirtiéndonos sobre si mantiene los mismos criterios a la hora de evaluar o éstos son cambiantes [1],[2]. Con los datos de repetitividad y consistencia podemos descartar o seleccionar las opiniones de los jueces que no cumplan unos requisitos mínimos, para los análisis posteriores.

5.2 Sobre las características de los sonidos:

Concordancia: en el diseño del test debemos decidir qué características de los sonidos deseamos evaluar. La concordancia de cada una de las características nos da una idea de si estas características han sido entendidas de la misma forma por los jueces o ha habido discrepancias. En definitiva, lo que nos viene a decir es si son significativas o la excesiva variedad de opiniones las convierten en irrelevantes [1].

5.3 Sobre los sonidos a evaluar:

Quizás el objetivo más importante del test sea el conseguir información de los sonidos que componen el mismo. Sobre dichos sonidos podemos conocer el grado de preferencia o la opinión que le merece al jurado (o a cualquier grupo seleccionado de jueces) cada sonido, para cada una de las características que interesan en el test. Estas calificaciones de los sonidos son tanto cualitativas como cuantitativas, pudiendo obtener 'notas' de cada sonido para cada una de las características. Estas valoraciones de los sonidos se calculan a partir de los índices de preferencia de cada uno de los sonidos (que ya de por si se pueden considerar como una valoración puesto que expresan el número de veces que un determinado evento es preferido a los demás teniendo en cuenta las comparaciones donde dicho evento aparece). A partir de estos índices de preferencia y mediante transformaciones matemáticas podemos calcular diferentes valoraciones de los sonidos. La valoración usada en nuestros programas esta basada en el modelo de Bradley-Terry [1], que se fundamenta en que la valoración subjetiva de una cierta característica es una variable aleatoria con una distribución de probabilidad determinada (doblemente exponencial).

Todos los resultados obtenidos del análisis los podemos presentar tanto de forma numérica como gráfica, resultando esta última opción bastante interesante puesto que permite visualizar gráficas comparativas tanto del análisis global con todos los sonidos, como comparativos entre cualquier par de sonidos.

6. EJEMPLO DE RESULTADOS OBTENIDOS MEDIANTE TESTSOUND

En este apartado se presentan los resultados obtenidos tras la realización de un test de parejas con la herramienta descrita, de modo que pueda validarse el correcto funcionamiento de dicha

herramienta, así como contrastar los resultados que se pueden obtener con otras metodologías de evaluación subjetiva. El test se realizó sobre cinco señales de motor de automóvil a 3000 rpm y con diferentes cargas, relacionadas con el volumen de combustible inyectado por cilindro y ciclo. Se trataba de discernir cómo afecta el volumen de inyección en la calidad del sonido de motor. Para ello se realizó un test con un jurado compuesto por 16 personas que debían que evaluar 20 comparaciones entre las 5 señales (todas con todas) para cinco características de los sonidos (preferencia, molestia, tranquilo, agradable y estresante). Cada juez invirtió unos 15 minutos de duración media para la realización del test y los resultados de repetitividad y consistencia fueron en su mayoría aceptables (mayores del 90%). Sobre los sonidos, el programa nos ofrece una puntuación de cada uno de ellos según el modelo de Bradley-Terry, para cada característica, lo que queda reflejado en la tabla 1 y en la figura 4.

Inyección	Preferencia	Agradable	Tranquilo	Estresante	Molesto
0 mm ³	1.9316	2.4377	3.1686	-2.4322	-2.3903
12 mm ³	1.1109	1.4713	1.4837	-1.4760	-1.3220
28 mm ³	0.2340	0.1833	0.3909	-0.6696	-0.7066
48 mm ³	-1.9295	-2.5896	-2.7262	2.6795	2.8721
68 mm ³	-1.3469	-1.5026	-1.5026	1.8984	1.5467

Tabla 1: Valoración de las señales según el modelo de Bradley -Terry.

Una de las propiedades que posee el modelo de Bradley-Terry consiste en que la suma de todas las valoraciones de una determinada característica de los sonidos es igual a cero [1] (por eso aparecen valoraciones negativas), pero lo importante de dichas valoraciones reside en que son relativas entre las señales que aparecen en el test y mediante sencillas aplicaciones lineales podríamos desplazarlas a cualquier otro margen de notas (si interesa). En la figura 4, también aparece representado el índice de preferencia (en tanto por ciento) comprobándose como existe una total correspondencia entre ambas valoraciones. Es decir, que los sonidos mejor valorados mediante el modelo de Bradley-Terry, se corresponden como era de esperar con los que mas veces han sido elegidos.

Los resultados ofrecidos por el programa nos permiten extraer varias conclusiones sobre este estudio particular. Por ejemplo se observa que a mayor volumen de inyección, peor es la valoración obtenida por los jueces para las características positivas de la calidad del sonido (agradable y tranquilo), aunque parece que hay un cierto nivel de volumen (48mm³) para el cual el empeoramiento de la calidad ya no es tan notable. También podemos observar que existe cierta correlación entre las puntuaciones obtenidas por preferencia, agradable y tranquilo y también entre las de molesto y estresante como se podía esperar.

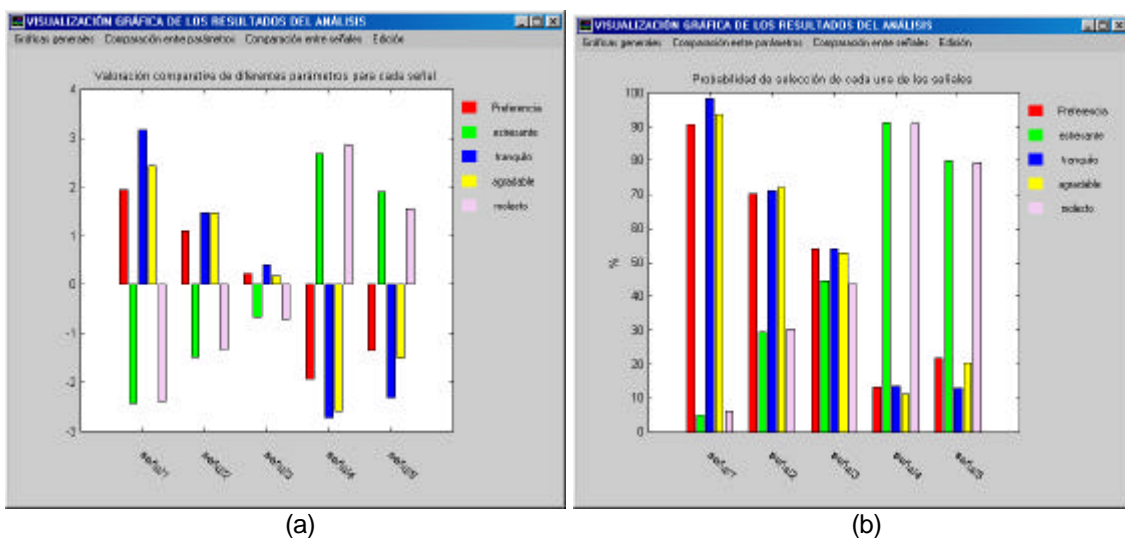


Figura 4: Comparativa gráfica de las valoraciones de todas las características para todas las señales tanto según el modelo de Bradley-Terry (a) como los índices de preferencia(b).

A partir de este experimento, se puede extraer la conclusión de que cuanto menor es el volumen de inyección a 3000 rpm, mejor valorado es el ruido que produce el mismo y que existe una cierta carga límite a partir de la cual el sonido parece no empeorar.

Para reforzar esta opinión, la Tabla 2 muestra las puntuaciones de los mismos sonidos emitidas por un jurado de expertos completamente independiente del anterior. Se comprueba que son totalmente compatibles con las obtenidas mediante el test de parejas.

Señal	Nota 1	Nota 2	Nota 3	Nota 4	Nota media
12 mm ³	7	6.6	8	6	6.875
28 mm ³	6.5	6.5	6.5	6.5	6.5
48 mm ³	6	6.5	6	6	6.125
68 mm ³	7	7	6	5.5	6.375

Tabla 2: Puntuaciones proporcionadas por un grupo de expertos.

Puede observarse como, aunque las calificaciones son muy parecidas (poca diferencia entre los sonidos), la nota media obtenida tiende a empeorar a medida que el volumen de inyección aumenta, a excepción de los dos últimos sonidos cuya valoración parece intercambiarse. Este hecho también sucede con los resultados del test. Incluso en un caso difícil como éste donde las puntuaciones resultan tan parecidas, los resultados obtenidos mediante TESTSOUND resultan completamente compatibles con los emitidos por un jurado de expertos.

7. REFERENCIAS

- [1] H.A. David, "The method of paired comparisons", . Oxford University Press, New York, 1988.
- [2] N.C. Otto Wakefield, "A subjective evaluation and analysis of automotive starter sounds" , Noise Control Engineering Journal, Vol 41 N° 3, pp. 337-382, noviembre-diciembre 1993.
- [3] M. Hussain, J. Gölles, A. Ronacher, H. Schiffbänker," Statistical evaluation of an annoyance index for engine noise recordings" , SAE 911080.
- [4] H. Schiffbänker, F.K. Brandl, G.E. Thier, "Development and application of an evaluation technique to assess the subjective character of engine noise" , SAE 911081.
- [5] E. Zwicker, H. Fastl, "Psychoacoustics-fact and models", Springer-Verlag, 2ª Edición.