



## ANÁLISIS DE TÉCNICAS DE GRABACIÓN ASEQUIBLES PARA PAISAJES SONOROS

*Roberto San Millán-Castillo<sup>1\*</sup>*  
*Ignacio López-Peñalver<sup>1</sup>*  
*Leticia Martínez-Cano<sup>2</sup>*  
*Eduardo Morgado<sup>1</sup>*  
*Luca Martino<sup>1</sup>*

<sup>1</sup>Universidad Rey Juan Carlos – EIF – Departamento de Teoría de la Señal y Comunicaciones, Camino del Molino, 5, Fuenlabrada (Comunidad de Madrid, España)

<sup>2</sup>Universidad de Castilla-La Mancha – ITct – Departamento de Ingeniería Eléctrica, Electrónica, Automática y Comunicaciones, Campus Universitario, s/n, Cuenca (Cuenca, España)

### RESUMEN

El estándar ISO 12913-2 proporciona información sobre la recogida de datos en investigaciones y estudios relacionados con paisajes sonoros. Uno de sus puntos claves es el ambiente acústico. Para ello, recomienda el uso de sistemas calibrados y normalizados de grabación binaural, concretamente cabezas artificiales. Estos sistemas ofrecen ventajas, pero también inconvenientes. Este trabajo evalúa el comportamiento de técnicas de grabación alternativas más accesibles: un teléfono móvil, una grabadora estéreo y unos auriculares binaurales. Los resultados arrojan reflexiones sobre el estándar y podrían favorecer un desarrollo más dinámico de la investigación en paisajes sonoros.

### ABSTRACT

ISO 12913-2 provides information about data collection in soundscape studies and research. To this end, the standard recommends calibrated and standardized binaural recording systems, particularly head and torso simulators. These systems provide benefits but also some drawbacks. This work assesses the performance of alternative and more affordable recording techniques: a smartphone, a stereo recorder, and some recording binaural headphones. The results lead to some considerations of the standard and may contribute to a more active development of the research in soundscapes.

**Palabras Clave**— Paisajes Sonoros, Sistemas de Grabación, binaural.

### 1. INTRODUCCIÓN

La distancia notable que existe entre la teoría y la práctica en la investigación de Paisajes Sonoros (PS) se pone de manifiesto en [1,2]. En lo que a instrumentación se refiere, se destaca el efecto de diferentes configuraciones de captación de PS. La norma ISO 12913-2 recomienda el registro de señales binaurales mediante simuladores de cabeza y torso (HATS). Sin embargo, en la comunidad científica surgen cuestiones relacionadas con las ventajas de utilizar sistemas de instrumentación complejos y costosos frente a otros más sencillos y asequibles.

El sonido binaural recrea una sensación próxima a la realidad en tres dimensiones, permitiendo una localización precisa de las fuentes sonoras implicadas. Sin embargo, la puesta en práctica es compleja. Uno de los principales problemas es la obtención de una respuesta de transferencia de la cabeza adecuada (i.e., Head Related Transfer Function, HRTF), y que recoja con fidelidad las diferencias de fase entre los dos oídos y la influencia de la cabeza y el torso humanos. Esta HRTF es única para cada persona y, además, no suele recoger posibles movimientos de la cabeza de los usuarios. Estas diferencias se pueden subrayar si el sistema de reproducción carece de un control exhaustivo de fase sobre las señales [3, 4]. Aun así, existen bases de datos de HRTFs que mejoran significativamente estos inconvenientes [5].

Por otro lado, los resultados de la investigación en [2], que compara las grabaciones de PS mediante grabación monoaural y binaural, sugieren que las primeras son

\* *Autor de contacto:* roberto.sanmillan@urjc.es

*Copyright:* ©2023 First author et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

suficientes para evaluar la mayoría de los descriptores de PS, incluyendo la impresión general, el confort acústico, lo agradable, lo molesto, lo memorable y la sonoridad. Eso sí, destaca que las grabaciones binaurales darán una percepción más consistente que la monoaural en escenarios más reverberantes; como ocurre en otros artículos que destacan esa riqueza que ofrece la captura binaural [6, 7, 8].

Otra cuestión importante es la distorsión del entorno y el contexto que puede generar el uso del HATS o micrófonos ambisónicos, y las diferencias incluidas en los PS debido a su presencia; dado que son elementos extraños a prácticamente la totalidad de ambientes sonoros reales. Además, la existencia de alternativas sencillas facilitaría la obtención de datos en múltiples escenarios para dinamizar la investigación, la difusión, la práctica en PS.

Ya en algunas universidades se utilizan los teléfonos móviles para familiarizar e introducir a los estudiantes en los PS, ya que la instrumentación asociada suele ser cara y difícil de usar [9]. Otros estudios comparan la inteligibilidad binaural en el habla mediante pruebas en remoto, donde los usuarios utilizan sus cascos personales, dándose una gran variedad tanto en la calidad como en el coste económico. El resultado de este estudio es una diferencia muy pequeña entre las pruebas remotas realizadas a través de Internet, y las pruebas perfectamente calibradas en el laboratorio. La conclusión del estudio lleva a valorar los métodos y herramientas que tenemos al alcance de la mano y explorar sus posibilidades de uso [10].

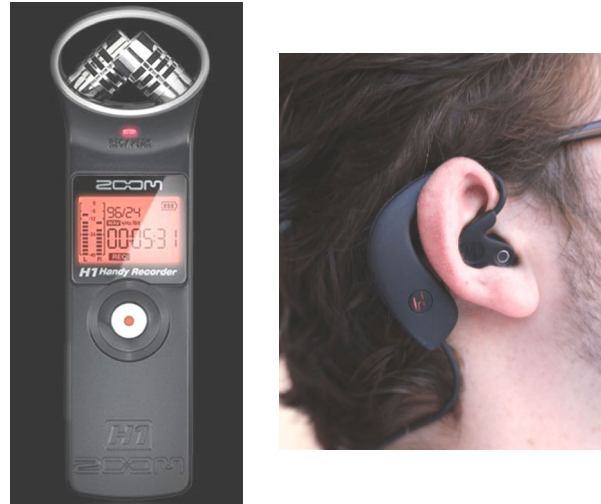
Esta comunicación presenta un experimento en el que se emplean y se comparan técnicas de grabación sencillas y asequibles para la evaluación de PS mediante encuestas a una muestra de población. Los resultados arrojan reflexiones interesantes que podrían ser útiles para mejorar y generalizar procedimientos y recomendaciones de investigación y normativos. Tras esta introducción, este documento se organiza como sigue: la Sección 2 expone la metodología empleada en el estudio, la Sección 3 presenta los resultados y los discute, y, por último, la Sección 4 recoge las conclusiones de este trabajo.

## 2. METODOLOGÍA

### 2.1. Equipamiento

Para realizar las grabaciones de los entornos sonoros, se emplearon tres sistemas de grabación, y un sistema de medición, común a las configuraciones de grabación.

Las grabaciones monoaurales se realizaron en primer lugar con el *smartphone Huawei Smart P 2019* [11]. En este caso, se debe contemplar el procesado de señal de la unidad, existente, pero desconocido: transductores, *firmware*, *software*, sistema operativo, filtros reductores de ruido,



**Figura 1.** Izquierda, grabadora estéreo; Derecha, auriculares binaurales.

*beamforming*, y control de automático de ganancia, por nombrar algunos.

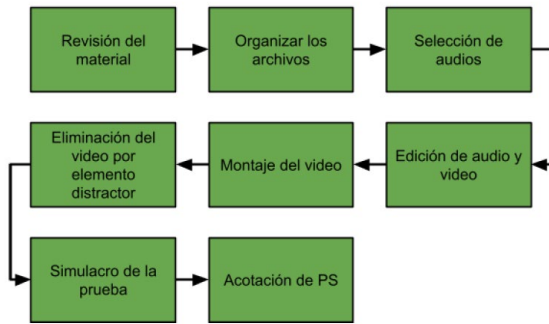
El siguiente sistema es una grabadora estéreo común y económica, *Zoom H1 Handy recorder*. Se basa en la técnica X/Y, con dos micrófonos omnidireccionales formando un ángulo de 90° [12], ver Figura 1.

El último sistema evaluado son unos auriculares inalámbricos binaurales asequibles: *Hooke Verse Mobile 3D Audio* [13], con micrófonos omnidireccionales *electret*, ver Figura 1.

Los sistemas de grabación comparados no son capaces de medir adecuadamente el nivel sonoro del entorno debido a su arquitectura, calibración, e incluso procesado de señal. El uso complementario del sonómetro nos ayudará a recrear los valores reales de presión sonora del ambiente acústico durante la grabación. Así, los participantes escucharán las tomas de PS al mismo nivel que *in situ*. Para esta tarea utilizamos el sonómetro SC310 de Clase 1 [14]. Para verificar la calibración del sonómetro, antes y después, de cada medición, utilizamos el calibrador acústico Rion, modelo NC74.

También se emplearon accesorios de relevancia como: auriculares, para verificar los registros de audio, pantallas anti-viento, al tratarse de grabaciones en el exterior; trípode ‘ad hoc’, para simultanear la medición con el sonómetro y los sistemas de grabación en un punto próximo para que sean equivalentes.

En la evaluación de los participantes, utilizamos el equipo *Creative Inspire T3300*, sistema de altavoces 2.1, compuesto por dos altavoces y un subwoofer con puerto de graves y salida inferior [17, 16]. Audacity se emplea para la edición de las tomas, modificación de ganancias, y normalización de duración a 30 segundos, principalmente.



**Figura 2.** Diagrama de bloques del procesado de datos y señales.

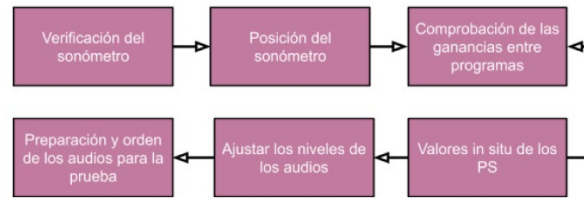
## 2.2. Grabación de Paisajes Sonoros

Antes de realizar cualquier medición, se verifica el sonómetro. El sonómetro se coloca en el trípode, apuntando en el mismo sentido que los sistemas de grabación, y en una posición típica del oyente del PS. El trípode permite una grabación con todos los dispositivos simultáneamente, dejando así las manos libres para manipularlos.

Las tomas duraron entre 1 y 3 minutos, dependiendo del lugar. Por ejemplo, en PS monótonos se reducía el tiempo de toma. La duración de los audios de las pruebas se normalizó a 30 s. La normativa [17], hace hincapié en que la duración de las grabaciones de PS tiene que durar un mínimo de 3 minutos, pero debemos tener en cuenta la capacidad de atención que puede tener un participante medio, cuya concentración se estima entre 10 y 20 minutos [19].

Tras la elección de los audios, su procesado y metraje, ver Figura 2 y [20] para más detalles, se realiza una prueba para evaluar la duración de una evaluación completa, es decir, escuchar los audios y contestar a las preguntas. Después del simulacro, decidimos acortar el número de PS, pasando de seis a tres, por lo que la duración pasó de una hora a treinta minutos aproximadamente, algo que favoreció al correcto desarrollo del experimento. Unos de los motivos por lo que se decidió acortar la prueba fue, de nuevo, por la capacidad de concentración que mostraron los participantes. Los PS seleccionados están situados en el centro la ciudad de Madrid (España): Puerta de Alcalá, Estación de Atocha-Renfe, Parque de El Retiro. Estas ubicaciones ofrecen diversidad de fuentes asociadas a su actividad y, por tanto, la respuesta afectiva variará hipotéticamente: animales, fuentes, personas, trenes y coches, entre otros.

Los audios escuchados por los participantes mantienen el nivel de presión sonora de la ubicación a través de procedimiento de verificación del nivel de audio descrito en la Figura 3. Con Audacity se genera un tono puro de 1 KHz y ajustamos la ganancia de la señal y de los altavoces hasta medir con el sonómetro 94 dB. Tras exportar el audio en



**Figura 3.** Diagrama de bloques de la verificación del nivel de audio de los PS.

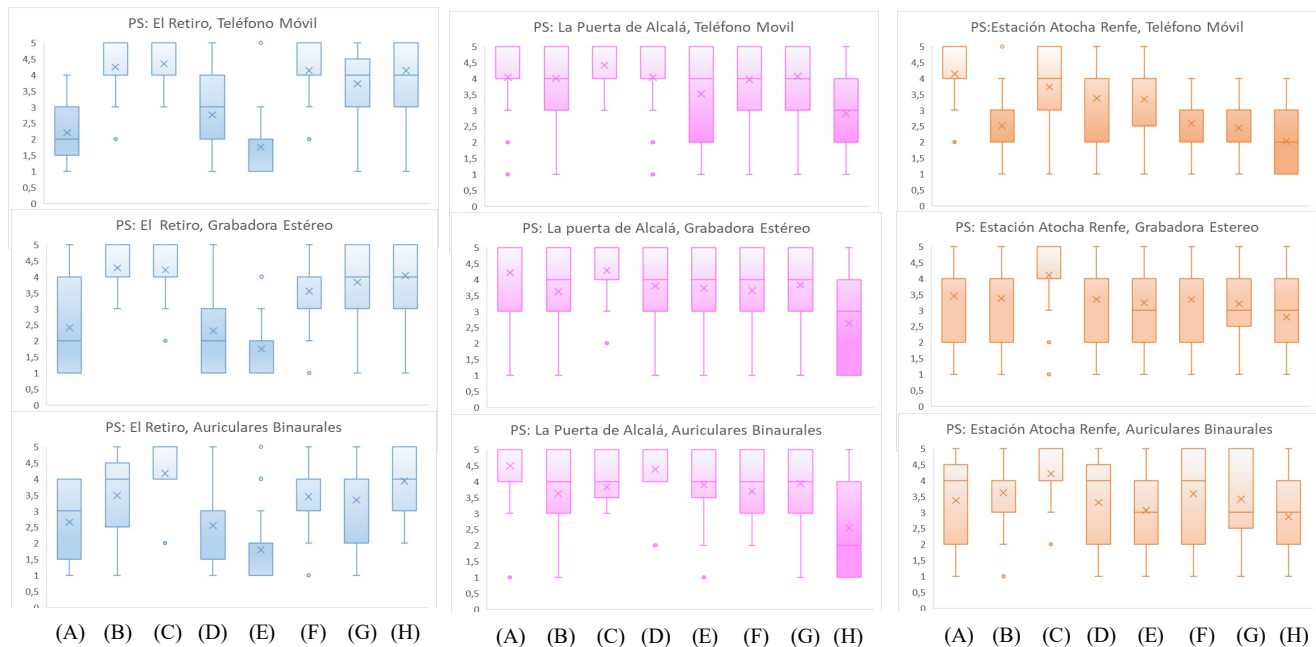
formato WAV de Audacity, se reproduce con VLC y comprobamos el nivel. VLC aumenta la ganancia de los archivos alrededor de 5 dB, sabiendo está diferencia de ganancias ya podemos calibrar nuestros audios de forma correcta con los datos de la campaña de grabación.

En cada prueba se siguió la misma secuencia: reproducción del primer PS calibrado con los tres dispositivos; tras esto se procedía con el siguiente PS. Se contó con un entorno controlado donde el ruido de fondo en la sala satisfacía un  $L_{Aeq} < 30$  dBA.

## 2.3. Cuestionarios y Evaluación de PS

En la recopilación de datos se toma como referencia la norma ISO 12913-2 [17]. La evaluación se realiza con una escala de cinco puntos y el cuestionario se presenta convenientemente a los participantes, asegurando su comprensión.

En cuanto al reclutamiento de participantes, fue voluntario, por lo que pudimos excluir la participación de personas con problemáticas en su salud otológica. Dado el carácter exploratorio e inicial de la investigación, así como el alcance en términos materiales, se empleó un modelo de muestreo no probabilístico por bola de nieve [21]. Éste consiste en que los individuos seleccionados para ser estudiados reclutan a nuevos participantes entre sus conocidos. El proceso de creación de una muestra mediante bola de nieve se fundamenta en obtener contactos iniciales, pedirles su participación y solicitar, una vez finalizada la prueba, el acceso a otros contactos. Estos contactos iniciales procederán de personas del entorno de los investigadores de este estudio y los criterios de selección son: salud otológica, interés por el estudio y buena actitud para la participación en el mismo, y capacidad de realización del cuestionario asociado sin asistencia. Este trabajo fue aprobado por el Comité de ética de la URJC pues recabó datos personales de los participantes.



**Figura 4.** *Boxplots* de los cuestionarios en todos los PS estudiados. Las características presentadas son: (A) Ruidoso, (B) De calidad, (C) Realista, (D) Alto, (E) Desagradable, (F) Definido, (G) Espacial, (H) Agradable.

### 3. RESULTADOS Y DISCUSIÓN

Se obtuvieron 29 encuestas válidas. En primer lugar, se realiza un análisis estadístico descriptivo de las respuestas de las encuestas. Además, se aplica la prueba de hipótesis no-paramétrica de Kruskal-Wallis para comparar los dispositivos, de acuerdo con investigaciones en el análisis de PS en [8]. La Figura 4 presenta una síntesis de los resultados obtenidos mediante diagramas de cajas y bigotes (*boxplot*) de los cuestionarios de los 3 PS evaluados.

El realismo que han aportado los tres dispositivos podría calificarse como notable, destacando el teléfono móvil en PS abiertos y los auriculares binaurales en PS en recintos (PA Atocha Renfe). Se trata de una de las características fundamentales para el objetivo de este experimento, puesto que el realismo se relaciona directamente con la ‘naturalidad’ y ‘fidelidad’ del audio según [22]. Todos los dispositivos ofrecen una calidad notable de sonido para los participantes. Respecto a la calidad del sonido, Gabrielsson y Sjören [23] detectaron que para los factores que la condicionan son: claridad, sensación de espacio y cercanía en la reproducción. La Estación de Atocha Renfe es un escenario reverberante, y la calidad del PS se manifiesta más baja que en espacios abiertos. Observando las puntuaciones de todos los dispositivos, la grabadora estéreo y los auriculares binaurales tienen una clara diferencia con respecto al teléfono móvil, dado que dotan de más espacialidad al audio en su toma, siendo los auriculares binaurales el método mejor valorado.

Por el contrario, en los PS abiertos como el Parque El Retiro y la Puerta de Alcalá, el dispositivo que acumula más puntuaciones en estos parámetros, es el teléfono móvil; presentando puntuaciones más altas en calidad, espacialidad o realismo.

Los resultados acerca de cómo de ruidoso es el PS en la Estación Atocha Renfe y La puerta de Alcalá son mayores con respecto al escenario del Parque El Retiro. Esto lleva a deducir que, si el ruido fuera generado por los equipos utilizados, la toma del Parque debería ser la toma con más ruido, ya que, al ser el PS con menor volumen, sería la toma donde mejor se escucharía el ruido introducido por los dispositivos. Al no ser así, se llega a la conclusión de que los participantes relacionaron el parámetro ruido con elementos sonoros molestos. Los PS más ruidosos son los que contienen más elementos sonoros no deseados por los participantes, probablemente.

Los resultados de este estudio están en línea con otras investigaciones previas. Por ejemplo, podemos observar cómo la evaluación del PS reverberante es mucho mejor cuando se realiza con grabaciones que dotan de espacialidad las tomas captadas, como las tomas binaurales o las estereofónicas, al igual que en [2]. El teléfono móvil ha sido el mejor método en PS abiertos pero el peor método en PS reverberantes; donde lideran las puntuaciones los auriculares binaurales.

Los resultados de la prueba de Kruskal-Wallis indican que las diferencias entre los métodos de grabación y los parámetros evaluados no son significativas en un

porcentaje elevado de casos y de características, por lo que es independiente el método de grabación con el resultado. Sin embargo, es importante destacar que, en la Estación Atocha Renfe, los parámetros de calidad, definición y espacialidad sí presentan diferencias significativas entre registros monoaurales y binaurales, y por lo tanto el dispositivo puede generar diferencias en la evaluación de PS; de nuevo corroborando las conclusiones de [2].

Durante la realización de las pruebas surgieron varias dudas entre los participantes, muchas de ellas comunes. Se observaron varios patrones que a nuestro juicio deberían descartarse en futuras líneas de investigación. Una de las dudas más recurrentes fue la definición de “ruido”, a lo que respondíamos con una breve descripción: “sonido desagradable, indeseado y dañino.” [1]. Contextualizamos el ruido y lo ejemplificamos para facilitar su entendimiento, pero no siempre quedaba completamente claro. Otro término que resultaba extraño o novedoso a los oyentes era la espacialidad. Acostumbrados a términos como estéreo, pensaban que un audio sólo se podía interpretar desde la izquierda o derecha, desconociendo otros métodos. Para futuras investigaciones concluimos en que una solución a esta situación podría ser la de añadir una introducción inicial ejemplificada y con una espacialidad notable entre audios para que el oyente pueda familiarizarse con el término e interiorizarlo mejor. Muchos de los participantes preguntaron si el cuestionario debía contestarse pensando en si les gustaba el sonido que escuchaban o si este estaba bien grabado. Ante esta duda explicamos que la prueba consistía en comparar los tres sistemas de audio, por lo que preferiblemente deberían evitar sus respuestas más afectivas.

#### 4. CONCLUSIONES

Este trabajo sugiere que los tres métodos empleados en la captación de PS son válidos, pero hay que tener en cuenta si el PS es reverberante, porque el método seleccionado influiría en la toma. La mayoría de los parámetros utilizados como calidad, realismo o la valoración personal de cada toma han obtenido unos resultados positivos, llegando a obtener una calificación de notable. Estos resultados abren un camino a utilizar herramientas más económicas para la grabación de PS en vez de usar costosos sistemas HATS. Estos métodos o herramientas son de bajo coste, algunos ya integrados en el día a día, como es el caso de un teléfono móvil, teniendo los mejores resultados en PS abiertos. Se trata de medios más accesibles lo que puede favorecer la recolección y el número de datos; de especial relevancia en investigación de PS.

Es relevante poner de manifiesto la falta de familiarización de las personas que participaron en el experimento con la terminología y comprensión del objetivo de la prueba. Los participantes confundían términos como ruido con volumen o con elementos sonoros que no eran de agrado, desconocían conceptos como espacialidad, o no lograban diferenciar el término de calidad con el de su

valoración subjetiva sobre los audios. La evaluación perceptual de los sonidos ambientales es un factor que cobra mucho peso en este ámbito de investigaciones, es importante de cara a plantear nuevos estudios contemplar estos factores desde un primer momento y asegurarse de la preparación de los participantes, y no confiarlo todo a una referencia estandarizada.

#### AGRADECIMIENTOS

Los autores agradecen el soporte y la financiación parcial de esta investigación a los siguientes organismos: Agencia Estatal de Investigación - AEI (proyecto POLIGRAPH, ref. num. PID2022-136887NB-I00), AEI (proyecto SPGRAPH, ref. num. PID2019-105032GB-I00), y Comunidad de Madrid y Universidad Rey Juan Carlos (Proyecto I+D Jóvenes Doctores, AUTO-BA-GRAPH, ref. num. F861).

#### REFERENCIAS

- [1] Aletta, F., & Xiao, J. (2018). What are the current priorities and challenges for (urban) soundscape research? *Challenges*, 9(1), 16.
- [2] Xu, C., & Kang, J. (2019). Soundscape evaluation: Binaural or monaural? *The Journal of the Acoustical Society of America*, 145(5), 3208-3217.
- [3] Rumsey, F. (2012). *Spatial audio*. Taylor & Francis. Boston, USA.
- [4] Begault, D. R. (1990, September). Challenges to the successful implementation of 3-D sound. In *Audio Engineering Society Convention 89*. Audio Engineering Society.
- [5] Møller, H., Sørensen, M. F., Hammershøi, D., & Jensen, C. B. (1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society*, 43(5), 300-321.
- [6] Coensel, B. D., Vanwetswinkel, S., & Botteldooren, D. (2011). Effects of natural sounds on the perception of road traffic noise. *The Journal of the Acoustical Society of America*, 129(4), EL148-EL153.
- [7] Genuit, K., & Fiebig, A. (2006). Psychoacoustics and its benefit for the soundscape approach. *Acta acustica united with acustica*, 92(6), 952-958.
- [8] Hermida, L., & Pavón, I. (2019). Spatial aspects in urban soundscapes: Binaural parameters application in the study of soundscapes from Bogotá-Colombia and Brasília-Brazil. *Applied Acoustics*, 145, 420-430.
- [9] Satoh, F., Sakagami, K., & Omoto, A. (2016). Application of a smartphone for introductory teaching of sound environment: Validation of the precision of the devices and examples of students' work. *Acoustical Science and Technology*, 37(4), 165-172.
- [10] Padilla-Ortiz, A. L., & Orduña-Bustamante, F. (2021). Binaural speech intelligibility tests conducted remotely over

the internet compared with tests under controlled laboratory conditions. *Applied Acoustics*, 172, 107574.

[11] Huawei (2019). HUAWEI P smart 2019. <https://consumer.huawei.com/es/phones/p-smart-2019/>

[12] Zoom (2019). Handy Recorder. <https://www.zoom.co.jp/products/handy-recorder/h1-handy-recorder>

[13] Kathleen Chapman. (2019). Hooke Audio Hooke Verse review. <https://the-gadgeteer.com/2019/10/07/hooke-audio-hooke-verse-review/>

[14] Cesva (2022). Cesva SC310. <https://www.cesva.com/es/productos/sonometros/sc310/>.

[15] Creative. (2013). Creative Inspire T3300. <https://es.creative.com/p/speakers/inspire-t3300>

[16] Audacity (2022) Audacity, editor de audio libre. <https://audacity.es/>

[17] ISO/TS 12913-2:2018. Acoustics – Soundscape - Part 2: Data collection and reporting requirements. Geneva, Switzerland.

[18] Axelsson, Ö., Nilsson, M. E., & Berglund, B. (2010). A principal components model of soundscape perception. *The Journal of the Acoustical Society of America*, 128(5), 2836-2846.

[19] Bligh, D. A. (1998). *What's the Use of Lectures?*. Intellect books. Exeter, England.

[20] López-Peñalver, I. (2022). *Análisis de Técnicas de Grabación de Paisajes Sonoros*. TFG, Universidad Rey Juan Carlos. Móstoles, España.

[21] López-Roldán, P., & Fachelli, S. (2017). *El diseño de la muestra. Metodología de la investigación social cuantitativa*. Universitat Autònoma de Barcelona. Cerdanyola del Vallès, España.

[22] Berg, J., & Rumsey, F. (2000, February). In search of the spatial dimensions of reproduced sound: Verbal protocol analysis and cluster analysis of scaled verbal descriptors. In *Audio Engineering Society Convention 108*. Audio Engineering Society.

[23] Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, 65(4), 1019-1033.