Acústica 2020
XI Congreso Ibérico de Acústica
51º Congreso Español de Acústica
TECNIACÚSTICA 2020

SOUND SCENE RECREATOR: HERRAMIENTA PARA INVESTIGAR LA PERCEPCIÓN DE PAISAJES SONOROS

Eguinoa R., San Martín R., Arana M

Universidad Pública de Navarra
Departamento de Ciencias
Campus de Arrosadia, 31006 Pamplona, Spain
{contact e-mail: ricardo.sanmartin@unavarra.es}

Resumen
Uno de los objetivos al analizar un paisaje sonoro es identificar sus componentes y su influencia en nuestra percepción del mismo. En esta comunicación se presenta una herramienta de software dedicada que facilita esta tarea. La herramienta permite controlar tanto el proceso de presentación de estímulos como la recopilación de datos y su exportación para el posterior análisis e interpretación.

La aplicación presenta una escena sonora previamente sintetizada. A partir de las fuentes individuales y mediante un controlador MIDI que las posiciona espacialmente, el sujeto recrea el paisaje original. La herramienta permite alternar ambas escenas, además de recoger datos de localización de las fuentes utilizadas y otros juicios de tipo semántico. Su aplicación más evidente es la realización de ensayos relativos a la percepción de paisajes sonoros, discriminación en localización de fuentes, evaluación de algoritmos de renderización de audio espacial, o análisis de los procesos de atención selectiva.

Palabras clave: paisajes sonoros, pruebas auditivas, audio espacial

Abstract
One of the main objectives when analyzing a soundscape is to identify its components and their influence in our perception. Aiming to help in this task, a dedicated software tool is presented in this communication. The tool allows controlling the stimulus presentation process and the collection and export of the data for its analysis and interpretation.

The application presents a previously synthesized sound scene. To recreate the original soundscape, the user must place the individual sources in the space with the help of a MIDI controller. The tool can alternate both scenes and, furthermore, it collects the localization data of the different sources and judgments on a semantic scale. The most evident application may be the accomplishment of different tests such as the perception of soundscapes, source localization discrimination, evaluation of spatial audio rendering algorithms or the analysis of selective attention processes.

Keywords: soundscapes, listening tests, spatial audio.

PACS nº. 43.66.Qp, 43.60.Bf

1 Introducción

La investigación en paisajes sonoros es interdisciplinar. Entre los factores que afectan a su evaluación se incluyen tanto sus características acústicas como elementos sociodemográficos. Existen diferentes estrategias para evaluar la respuesta subjetiva a un determinado paisaje sonoro. El enfoque más habitual es la realización de cuestionarios que engloban entrevistas cualitativas y escalas de valoración. De esa manera, los paisajes sonoros son evaluados 'in situ' [1], mediante grabaciones de campo presentadas en laboratorio [2][3] o, en menor medida, utilizando simuladores [4][5].

Esta última presentación controlada de paisajes sonoros en entornos de laboratorio permite una sistemática evaluación de sus diferentes componentes. Sin embargo, la confiabilidad de los campos sonoros estímulo presentados resulta un reto tecnológico para investigadores asociados a disciplinas como el urbanismo o las ciencias sociales.

La herramienta que se presenta pretende establecer el vínculo entre complejidad técnica y facilidad de uso. Se trata de una aplicación sencilla e intuitiva que permite reproducir paisajes sonoros tridimensionales en forma binaural y recoger datos respecto a las impresiones de los usuarios. Puede ayudar en la investigación de la percepción que tiene el ser humano sobre su entorno, para la recopilación de sonidos en espacios naturales y su evaluación, o incluso para vertientes más artísticas, como la composición musical, por lo que puede resultar atractiva para gran cantidad de usuarios.

2 Herramientas para la presentación inmersiva

La presentación o renderización de paisajes sonoros en un laboratorio pretende imitar nuestra percepción de escenas sonoras tridimensionales mediante el empleo de adecuadas técnicas de procesamiento y auriculares o, alternativamente, distribuciones de altavoces [6]. Una de estas ténicas se denomina Ambisonics [7]. Se basa en la descomposición del campo sonoro en funciones armónicas esféricas que constituyen una base ortogonal completa, lo que significa que el campo sonoro puede ser reconstruido partiendo del conocimiento de los coeficientes de cada función y viceversa.

Aunque desarrollado en los años 70, ambisonics ha ganado peso recientemente desde que YouTube, Oculus VR y Facebook lo adoptaron como un estándar para sus videos de 360 grados. En la actualidad, ambisonics se ha convertido en un estándar de audio espacial ampliamente utilizado en los sistemas de realidad virtual y audio inmersivo. La herramienta presentada implementa codificación y decodificación ambisonics de orden 3, incluyendo binauralización [8].

Varios grupos de investigación han desarrollado otras herramientas relacionadas con el audio espacial capaces de crear cualquier tipo de escenas a partir de grabaciones. Entre ellos destacan el Instituto de Música Electrónica y Acústica IEM (Austria) y el Laboratorio de Acústica de la Universidad de Aalto (Finlandia). Ambas entidades proporcionan paquetes con una gran variedad de plugins, todos ellos de código abierto tanto para uso académico como personal o profesional. IEM ofrece en su Plug-in Suite [9] una amplia gama de codificadores (MultiEncoder, StereoEncoder, RoomEncoder), decodificadores (AIIRADecoder, BinauralDecoder, ProbeDecoder) y visualizadores (EnergyVisualizer) de señales Ambisonics; así como compresores, delays, reverbs… Por su parte, el Laboratorio de Acústica de la Universidad de Aalto dispone de SPARTA (Spatial AudioReal-time Applications [10]), una colección de plugins VST para la producción, reproducción y visualización de audio espacial: sparta_ambiBIN, sparta_ambiDEC, sparta_ambiENC, compass_binaural o compass_decoder entre otros.

3 Sound Scene Recreator

La herramienta Sound Scene Recreator se ha desarrollado utilizando el entorno JUCE [11], un marco de aplicación de código abierto multiplataforma escrito en lenguaje C++. Frente a otros entornos de programación similares, JUCE presenta funciones específicas para el procesado de audio, por lo que su uso se encuentra muy extendido para el desarrollo de aplicaciones móviles y de escritorio relacionadas con el audio. El objetivo de JUCE es crear código capaz de ser compilado y ejecutado en diferentes plataformas: Windows, Mac OS, Linux… Además de proporcinar su propio entorno de desarrollo, llamado Projucer, admite una gran variedad de IDEs y compiladores (Visual Studio, XCode, GCC…).

Para la creación de cualquier aplicación o plugin, JUCE genera un proyecto básico con diferentes scripts, funciones y módulos. Partiendo de esta base se deben añadir módulos, librerias y funciones dependiendo de las necesiades de cada trabajo, creando así proyectos propios. Gracias a todas las opciones y bibliotecas que ofrece, ha tenido mucho éxito en el campo de aplicaciones y plugins relacionados con el procesado del sonido y la música. Este framework se toma ya como un estándar para este tipo de desarrollos en código abierto, gracias a que se puede ejecutar y modificar en cualquier entorno de trabajo.

La herramienta desarrollada es escalable y puede ser particularizada según la aplicación deseada. Se ha simplificado al máximo el proceso de generación de escenas sonoras y se ha creado una sencilla interfaz que reune toda la información necesaria para manejarlas. La aplicación permite crear una escena desde cero o cargar una creada previamente, ya que pueden guardarse las características de las creaciones en un archivo .json (JavaScript Object Notation) para poder usarlas en cualquier momento.

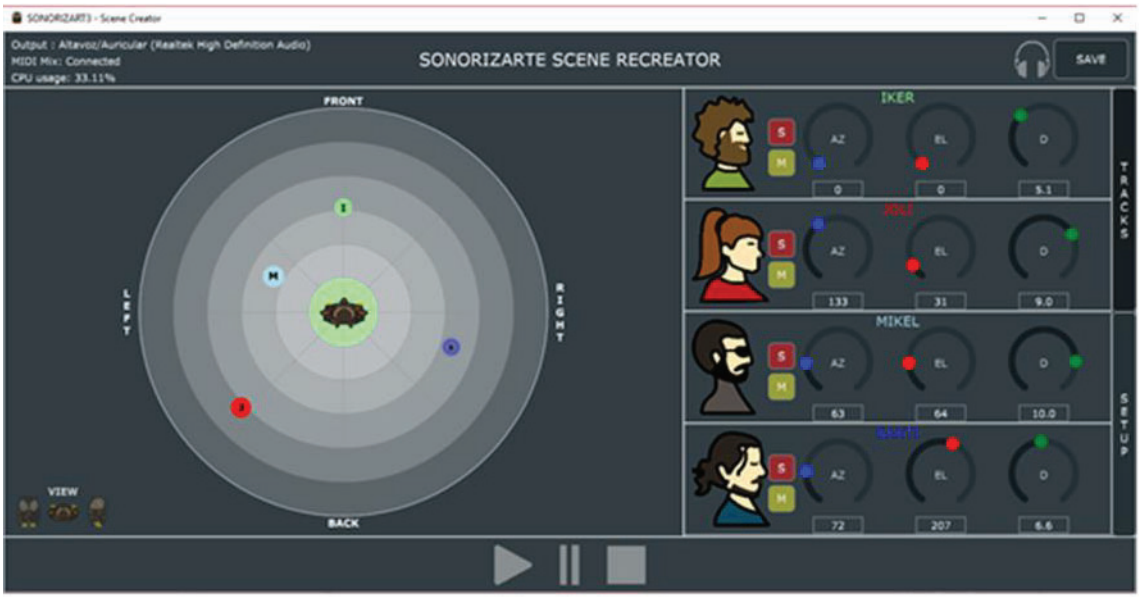

Figura 1 – Ejemplo de aplicación de la herramienta: composición lúdico-creativa.

Las escenas pueden contar con hasta ocho pistas de sonido diferentes, cada una asociada a una imagen, color y posición inicial en la escena. Una vez creada o cargada la escena, cuenta con una recreación del espacio que transmite una sensación esférica para poder visualizar la posición de cada fuente, que se muestran con un círculo de su color y la primera letra de su nombre (ver Figura 1). A su vez, cada pista cuenta con tres controles rotatorios para poder modificar en tiempo real el azimut, la elevación y la

distancia de la fuente sonora. Todos estos parámetros pueden controlarse con ratón o un dispositivo MIDI asociado.

La parte inferior de la ventana controla el flujo de reproducción mediante los botones clásicos de play, pause y stop. Las escenas pueden presentarse mediante auriculares o diferentes configuraciones de altavoces (ver Figura 2). El flujo ambisonics (codificación, manipulación de escena y decodificación), así como la binauralización se implementan en un segundo plano, no accesible para el usuario de la herramienta.

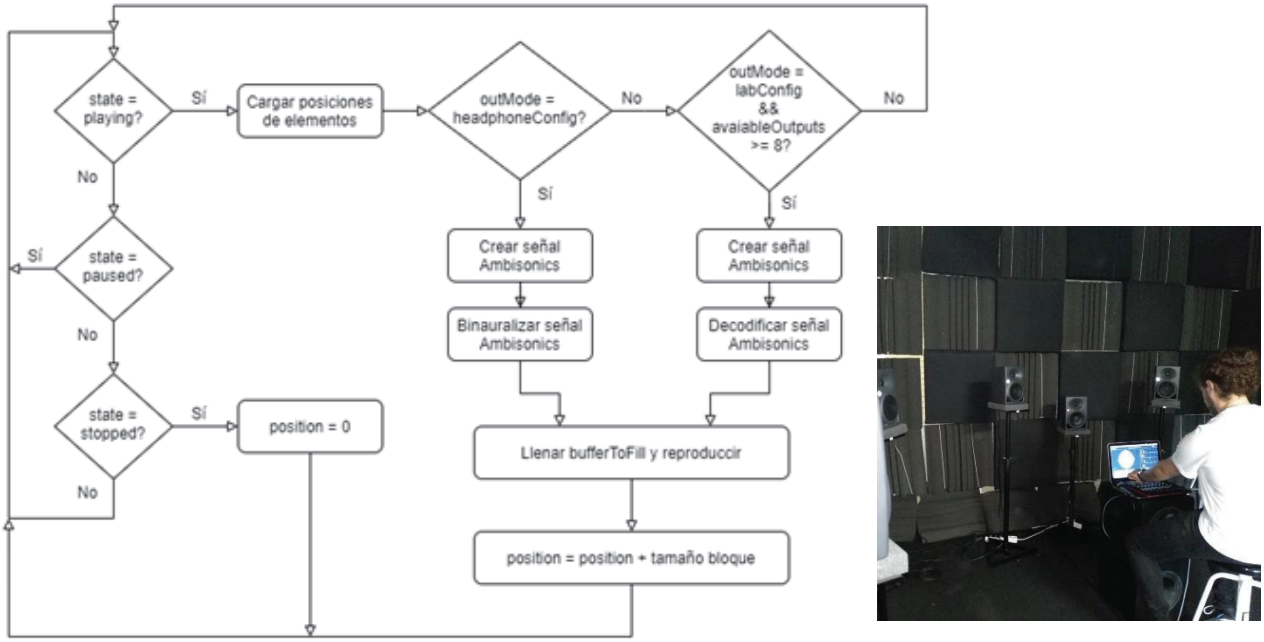

Figura 2 – Flujo de reproducción de la herramienta y ejemplo de presentación mediante configuración de 8 altavoces en la sala para la renderización de audio espacial del Laboratorio de Acústica de la Universidad Pública de Navarra.

4 Ejemplo de aplicación: diseño de listening tests

Un paisaje sonoro es definido como “el entorno acústico según es percibido, experimentado o entendido por una o varias personas, en contexto” [12]. Así pues, uno de los objetivos principales en los estudios relacionados con paisajes sonoros es identificar qué componentes influyen en su percepción [13]. Uno de ellos es su origen: geofónico, biofónico o antropofónico [14].

Las características espectrales de estos tres tipos de emisiones son habitualmente diferentes. Mientras los sonidos geofónicos son generalmente infrasónicos o de baja frecuencia, el rango espectral de los sonidos biofónicos o antropofónicos es mucho mayor y puede variar notablemente en cuanto a composición e intensidad temporal. Al ser posible encontrar sonidos de características similares pero de origen biofónico o antropofónico, cabe preguntarse si éstos son percibidos de manera diferente.

La probabilidad de aislar este tipo de sonidos en un paisaje real sonando simultáneamente para ser confrontados de forma objetiva es bastante remota. Por ello, la particularidad de la hipótesis planteada obliga a utilizar el mayor control sobre las variables estímulo que ofrece una presentación en laboratorio. La herramienta diseñada es de gran utilidad para realizar este tipo de listening tests. En este caso, se analizará la precisión en la localización de fuentes sonoras dependiendo de su origen biofónico o antropofónico.

4.1 Planteamiento

El conjunto de sonidos utilizado fue obtenido en Freesound, una base de datos colaborativa y bajo licencia Creative Commons [15]. Se agruparon según su origen, biofónico o antropofónico, y sus características acústicas, temporales y espectrales. Fueron editados hasta obtener ciclos de 20 segundos donde podían escucharse entre 20 y 25 repeticiones del sonido estímulo. Para facilitar la discriminación cuando se presentaran simultáneamente al usuario, las repeticiones fueron generadas de forma asíncrona (ver Figura 3)

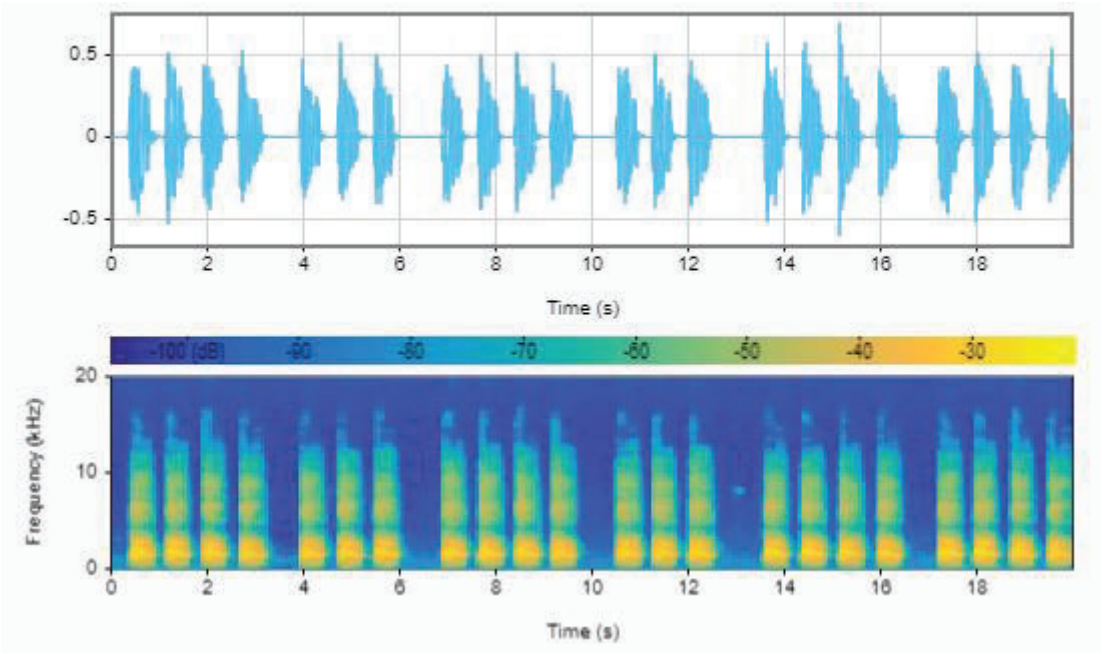

Figura 3 – Sonido biofónico “raven”, señal temporal (arriba) y espectrograma (abajo)

Los niveles fueron ajustados para que presentaran una misma sonoridad de 35.6 sones evaluada según el método Zwicker [16]. Además fueron agrupados por parejas según su sonoridad específica. En la Figura 4 se muestran las parejas finalmente seleccionadas y que fueron confrontadas posteriormente en el test: teléfono-rana (phone-frog), martillo-perro (hammer-dog), despertador-grillo (alarm-cricket), bocina-cuervo (klaxon-raven).

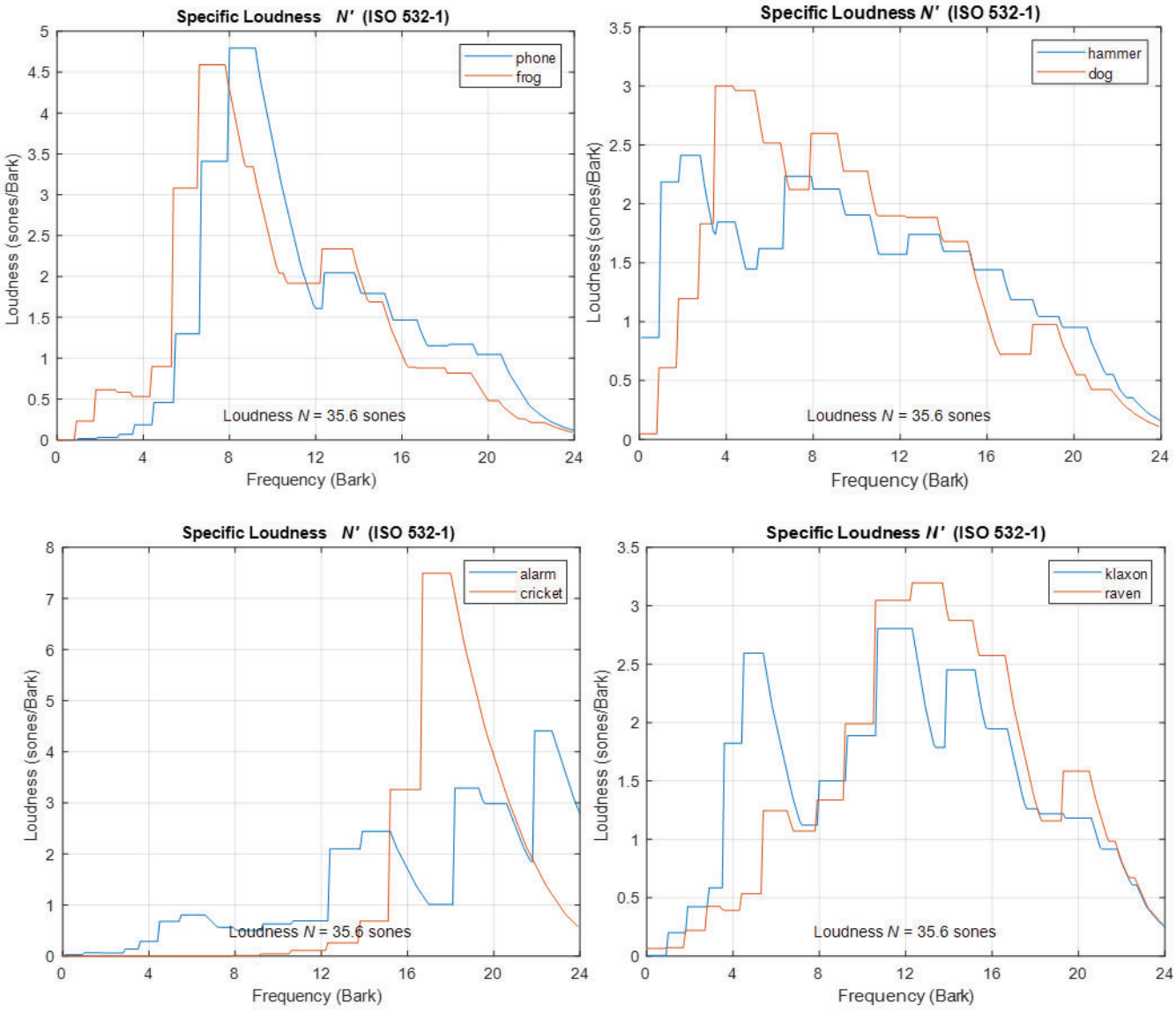

Figura 4 – Sonoridad específica de las parejas de sonidos seleccionados, antropofónicos (azul) y biofónicos (rojo). De izda a derecha y arriba a abajo las parejas son teléfono-rana, martillo-perro, despertador-grillo, bocina-cuervo.

4.2 Desarrollo del test

El test se ejecuta en tres fases diferentes. Todas ellas consisten en posicionar las fuentes sonoras presentes en una escena sonora (S) con respecto a otra escena incógnita (X) de referencia con las mismas fuentes pero en posiciones desconocidas y elegidas de forma aleatoria. Los ocho sonidos seleccionados se reproducen o no dependiendo de la fase del test en la que el usuario se encuentre. Cada uno cuenta con un canal en el dispositivo MIDI que se utiliza como controlador, el cual ha sido recreado en la mitad derecha de la ventana (ver Figura 5). En cada canal solo están habilitados dos potenciómetros para variar el ácimut y la elevación, y un fader vertical para la distancia. Además, su color es más brillante si se está escuchando dicho sonido y más apagado en el caso que no se esté reproduciendo.

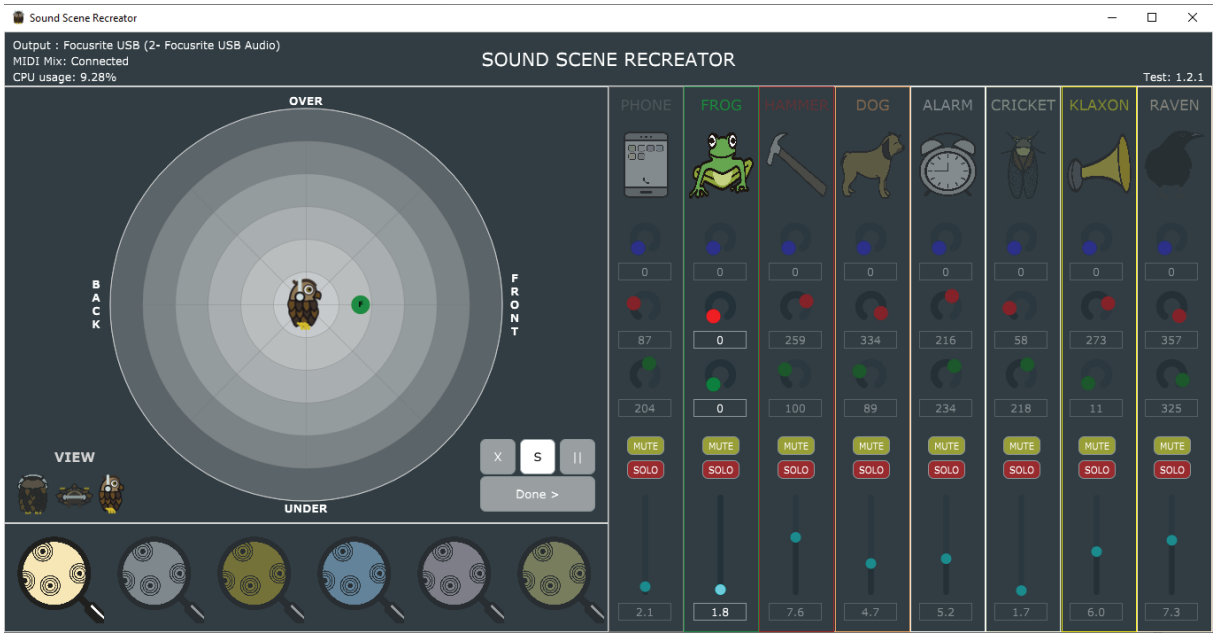

Figura 5 – Interfaz de la aplicación durante el desarrollo del test

La primera fase sirve para familiarizarse con los sonidos, la herramienta y su control mediante el dispositivo MIDI. Cada fuente se escuchará individualmente y seis veces, en diferentes posiciones. Así, para evaluar ocho sonidos se necesitan ocho subfases, una por sonido.

En la segunda fase los sonidos se muestran por duplas de sonoridad especifica similar enfrentando procedencia antropofónica a biofónica (móvil-rana, martillo-perro, alarma-cigarra, bocina-cuervo). Ambas fuentes suenan al mismo tiempo. Al igual que en la fase anterior, cada una de las fuentes debe ser posicionada seis veces, por lo que en este tramo del test, las sub-fases cuentan con sus propias seis fases.

La fase final de las pruebas es más complicada y se ha incorporado para potenciar el caracter lúdico de la herramienta y evaluar la percepción de dificultad encontrada para el posicionamiento de fuentes sonoras en entornos muy ruidosos. En ella se presentan a la vez los ocho sonidos, lo que complica mucho la discriminación de cada posición. En este caso, solo se posicionan todas las fuentes en una ocasión.

La parte izquierda de la ventana reproduce una esfera de 12 metros de radio, reproduciendo el espacio alrededor de la persona, representada mediante la figura de un búho. En ella se dibujan, mediante un círculo, la procedencia de los sonidos. Además, en la parte inferior, seis lupas marcan las seis diferentes posiciones que habrá en cada escena (ver Figura 5).

En este tipo de aplicaciones es importante que la interfaz de usuario, típicamente en 2D, facilite un posicionamiento tridimensional. Para ello, se llevan a cabo diferentes técnicas con el círculo que representa la posición de cada fuente sonora (ver Figura 6.izda). Su tamaño (el área) varía conforme lo hace su posición. Si se encuentra cerca de la visión del usuario, tienen mayor diámetro, y disminuyen conforme se alejan hacia el fondo. Todos los elementos parten de un diámetro fijo, el que tienen cuando se encuentran en el corte central de la esfera. Con el uso de razones trigonométricas se amplía o estrecha el perímetro. Además, cuando la posición se encuentra en la parte trasera de la esfera, se transparenta su color, quedando más marcado su perímetro, y si la ubicación está en la parte trasera del búho, quedará

ocultada. Además, gracias a los botones de la parte inferior izquierda, los usuarios pueden variar la vista de la esfera, seleccionando una visión frontal, lateral o superior.

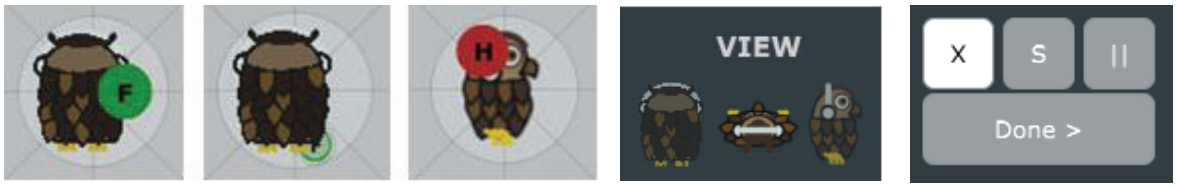

Figura 6 – Selección de vistas y posicionado tridimensional de fuentes (izda) y panel de interacción entre fases y alternancia entre escena sonora y escena incógnita (dcha)

En la zona inferior derecha de la esfera, se dibuja un conjunto de cuatro botones: X, S, pause (II) y DONE >, los cuales sirven para pasar de fase en los tests y alternar la escucha. Gracias a los tres botones de la parte superior, se puede cambiar el estado de la aplicación (ver Figura 6.dcha). Dichos estados son tres: incógnita (X), escena (S) y pause (II).

- Incógnita (X): Se escuchan los sonidos a localizar. En la esfera no se dibuja ninguna fuente, aunque sí se pueden modificar las posiciones que se están buscando.
- Escena (S): Se escuchan los sonidos en las posiciones elegidas, las cuales se representan en la esfera y se mueven en directo. Estas posiciones serán las que se registren.
- Pause (II): El programa pausa la reproducción de todos los sonidos y espera la vuelta a uno de los dos estados anteriores.

Por último, una vez encontradas las posiciones, se pulsa el botón inferior (Done). Este desencadena varias acciones: se guardan las posiciones marcadas y las de la incógnita, se crean nuevas posiciones aleatorias tanto para los sonidos como para los parámetros a buscar y, finalmente, pasa el programa a la siguiente fase.

Un total de seis personas de edades comprendidas entre 22 y 61 años respondieron al test. Ninguna de ellas manifestó sufrir algún tipo de lesión o pérdida auditiva. La duración total de cada test fue de unos 60 minutos, pudiendo descansar entre fase y fase si lo solicitaban.

4.3 Resultados (REC)

Al finalizar cada fase, la dificultad encontrada para la localización de cada estímulo se valoró en una escala de 0 (muy fácil) a 10 (muy difícil). La fase 3, aquella en la que todos los estímulos suenan al mismo tiempo, es tomada como referencia de máxima dificultad para cada sujeto y se normalizaron los valores con respecto al obtenido en esa fase.

Los valores promedio de todos los estímulos para las evaluaciones realizadas por las 6 personas que realizaron el test se muestran en la Figura 7.izda. Además de observarse la esperada tendencia creciente en cuanto a dificultad entre fases, los elevados valores relativos a la fase 1 (entre 0.4 y 0.7) indican que la tarea no resultó sencilla ni siquiera cuando los estímulos se presentaban por separado.

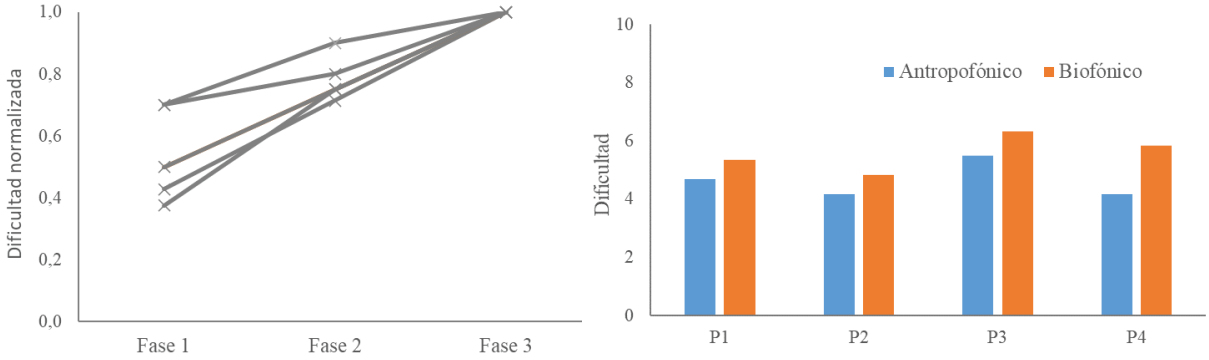

Figura 7. Dificultad declarada por las personas que realizaron en test según fases y tipo de estímulo presentado.

Agrupando las evaluaciones recibidas por cada estímulo según las parejas de similares características en cuanto a sonoridad presentadas (Figura 7.dcha), los sonidos de procedencia biofónica presentan una dificultad de localización en torno a un 15% superior a los de procedencia antropofónica, alcanzando la diferencia un valor máximo del 40% para el caso de la dupla bocina-cuervo.

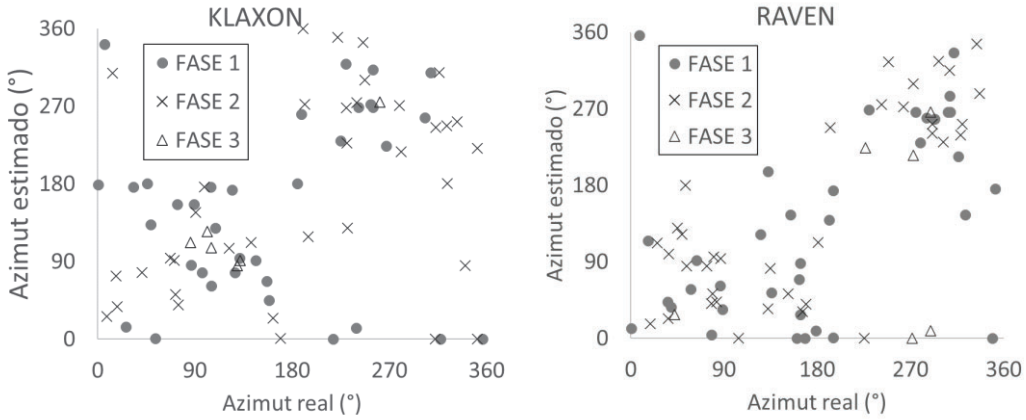

Figura 8. Precisión de localización (azimut) para los estímulos bocina (izda) y cuervo (dcha) en las tres fases de ejecución del test.

Sin embargo, esta dificultad no se ve reflejada en los resultados obtenidos en cuanto a localización. Así puede observarse en la Figura 8 para el caso de esta última pareja, donde no existen diferencias significativas en precisión de estimación del azimut ni entre estímulos ni entre las distintas fases, lo que nos lleva a suponer que, si bien el esfuerzo para realizar la discriminación espacial es mayor en presencia de otros estímulos, no repercute en la precisión obtenida. Los resultados obtenidos para la distancia confirmaría esta hipótesis. Las correlaciones entre distancia estimada y distancia real son muy similares en las tres fases analizadas (ver Figura 9).

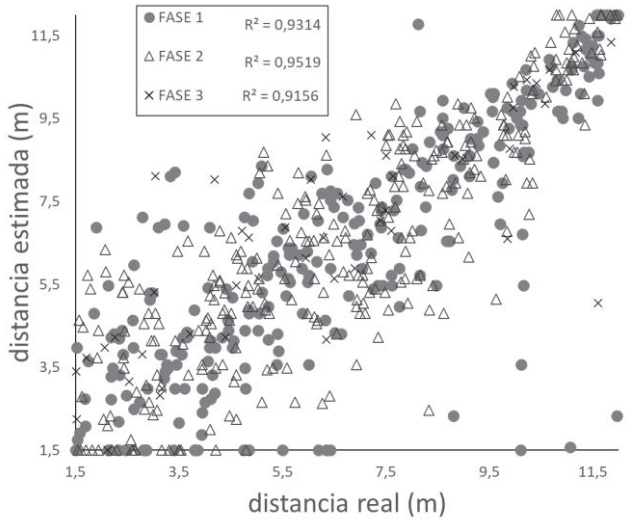

Figura 9. Precisión de localización (distancia) para todos los estímulos y sujetos en las tres fases de ejecución del test

Por útlimo, en el análisis individualizado de los resultados de localización se encontraron diferencias significativas entre los diferentes sujetos. Así, por ejemplo, los resultados obtenidos por el sujeto 1 y el sujeto 5 para la elevación son muy diferentes (ver Figura 10). De hecho, el sujeto 1 apenas percibía esta variable.

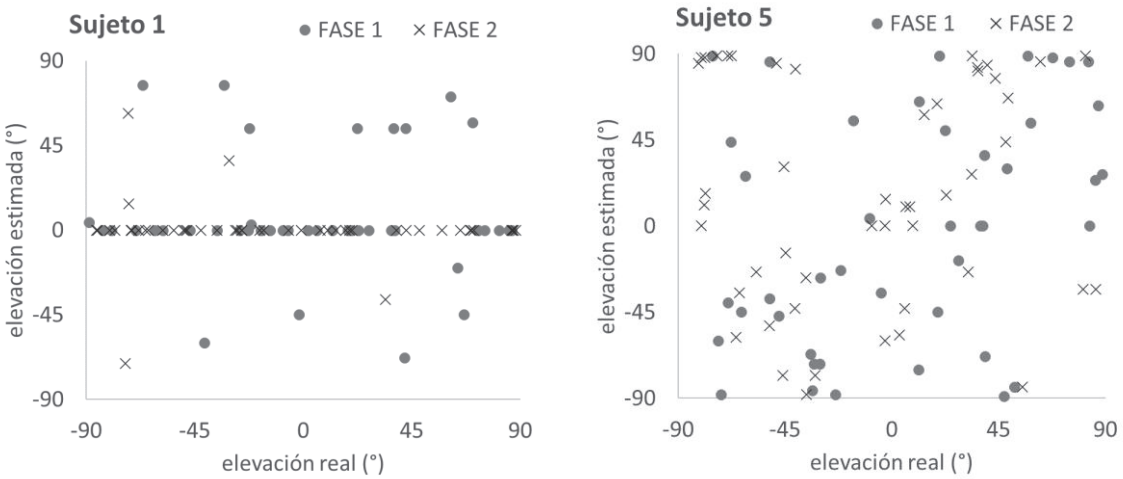

Figura 10. Precisión de localización (azimut) para el sujeto 1 (izda) y el sujeto 5 (dcha) en las tres fases de ejecución del test.

Estos resultados confirman la dificultad para reproducir artificialmente sonidos en un espacio tridimensional mediante auriculares, sobre todo en elevación. Queda por investigar si la binauralización de ambisonics de tercer orden utilizada [8] es lo suficientemente fiel en comparación con el método de las HRTFs de fuentes virtuales. Sobre todo porque este segundo método permite la implementación de funciones de transferencia individualizadas para cada sujeto; lo cual, a priori, debería ofrecer mejores resultados en los posicionamientos de las fuentes sonoras.

5 Conclusiones

Con el objetivo de disminuir la brecha tecnológica entre los investigadores de diferentes disciplinas, se ha desarrollado una herramienta software que permite realizar estudios relativos a la percepción de paisajes sonoros de forma sencilla y confiable. Utiliza en segundo plano ambisonics de tercer orden, aunque los algoritmos de presentación pueden modificarse, y permite recoger juicios de tipo semántico. Resultados preliminares obtenidos para un caso práctico de localización de fuente sonora en función de su origen biofónico o antropofónico confirman su potencial.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio de Economía y Competitividad a través del proyecto de Investigación I+D+I con referencia BIA2016-76957-C3-2-R.

Referencias

[1] Schulte-Fortkamp, B.; Fiebig, A. Soundscape Analysis in a Residential Area: An evaluation of noise and people's mind. *Acta Acustica United With Acustica*, Vol. 92, 2006, 875-880.

[2] Guastavino, C.; Dubois, D. Ecological Validity of Soundscape Reproduction, *Acta Acustica United with Acustica*, Vol. 91, 2005, 333-341.

[3] Lavandier, C.; Defreville, B. The Contribution of Sound Source Characteristics in the Assessment of Urban Soundscapes, *Acta Acustica united with Acustica*, Vol. 92, 2006, 912-921.

[4] Bruce, N.S.; Davies, W.J. The effects of expectation on the perception of soundscapes, *Applied Acoustics*, Vol. 85, 2014, 1-11.

[5] Lafay, G.; Rossignol, M.; Misdariis, N.; Lagrange. M; Petiot, J.F. Investigating the perception of soundscapes through acoustic scene simulation. *Behavior Research Methods,* Vol 51, 2019, 532-555.

[6] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, F. Zotter, (2013) Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State. IEEE Proceedings 101. 1920-38.

[7] F. Zotter, M. Frank. Ambisonics, Springer 2019.

[8] C. Schörkhuber, M. Zaunschirm, R. Höldrich, Binaural rendering of ambisonics singals via magnitude least squares, in *Fortschiritte der Akustik* – DAGA (Munich, 2018)

[9] IEM Spatial Audio Plug-in Suite. Institute of Electronic Music and Acoustics, Austria.

[10] SPARTA Spatial Audio Real-time Applications. Acoustics Lab at Aalto University, Finland.

[11] JUCE open-source cross-platform C++ application framework: github.com/juce-framework/JUCE

[12] ISO 12913-1: 2014 Acoustics Soundscape. Part 1: Definitions and Conceptual Frameworks.

[13] Aletta, F.; Kang, J.; Axelsson, O. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape and Urban Planning*, Vol.149, 2016, 65-74.

[14] B.C. Pijanowski, A. Farina, S.H. Gage, S.L. Dumyahn, B.L. Krause. What is soundscape ecology?. An introduction and overview of an emerging new science. *Landsc. Ecol.*, Vol. 26, 2011, 1213-1232

[15] Fonseca, E., Pons, J., Favory, X., Font, F., Bogdanov, D., Ferraro, A., Oramas, S., Porter, A., and Serra, X. (2017). "Freesound datasets: A platform for the creation of open audio datasets, in Proceedings of the 18th ISMIR Conference, Suzhou, China, International Society for Music Information Retrieval, Canada, 486–493.

[16] ISO 532-1:2017: Acoustics - Methods for calculating loudness - Part 1: Zwicker method.