

## VISUALIZACIÓN DE DIFERENCIAS ENTRE SEÑALES ENVOLVENTES EN FORMATO AMBISÓNICO

**PACS:** 43.60.Ek, 43.60.Gk, 43.60.Jn

R. Eguinoa; A. Sagasti; R. San Martín; A.P. Pietrzak\*; M. Arana.

Universidad Pública de Navarra, Campus de Arrosadia, 31006 Pamplona, Spain  
ruben.eguinoa@unavarra.es, +34948168436 .

\* Warsaw University of Technology, Institute of Radioelectronics and Multimedia Technology,  
ul. Nowowiejska 15/19, 00-665 Warsaw

**Palabras Clave:** Ambisonics, Audio 3D, Visualización

### ABSTRACT

Ambisonics is a full-sphere sound format in which, in contrast to other surround formats, the channels do not distribute the signal feeding each loudspeaker. Therefore, to quantify differences between different encodings to evaluate the quality of different Ambisonics microphones, for example, direct comparison of the levels of each channel does not provide easily interpretable results. In this communication a Matlab application is presented, which facilitates the real-time comparison of signals in Ambisonics format up to seventh order. The differences in spatial distribution of sound energy are visualized in a modified azimuthal projection that preserves the proportions of the areas. For this purpose, each signal is decoded to a mesh of evenly spaced virtual loudspeakers. A subsequent calculation of the RMS values at each point makes it possible to represent the energy distribution of each signal and thus evaluate its spatial image. As examples of use, representations of different coding orders for the same sound scene and the effect of different parametric upmixing techniques in first order signals are shown.

### RESUMEN

Ambisonics es un formato de sonido de esfera completa en el que, en contraste con otros formatos envolventes, los canales no distribuyen la señal que alimenta a cada altavoz. Por ello, cuando se desea cuantificar diferencias entre diferentes codificaciones para evaluar la calidad de diferentes micrófonos ambisónicos, por ejemplo, la comparación directa de los niveles de cada canal no proporciona resultados fácilmente interpretables. En esta comunicación se presenta una aplicación en Matlab que facilita la comparación en tiempo real de señales en formato ambisónico hasta de séptimo orden. Las diferencias en cuanto a distribución espacial de energía sonora se visualizan en una proyección azimutal modificada que preserve las proporciones de las áreas. Para ello, se decodifica cada señal a una malla de altavoces virtuales espaciados uniformemente. Un cálculo posterior de los valores eficaces en cada punto permite representar la distribución de energía de cada señal y evaluar así su imagen espacial. Se muestran, a modo de ejemplo de uso, representaciones de diferentes órdenes de codificación para una misma escena sonora y del efecto sobre señales de primer orden de diferentes técnicas paramétricas de upmixing.

## 1. INTRODUCCIÓN

Ambisonics es una de las tecnologías utilizada más habitualmente para la grabación, codificación y reproducción de escenas sonoras tridimensionales. Al ser un formato de esfera completa, que cubre tanto el plano horizontal como el vertical, permite representar las denominadas señales envolventes, aquéllas que definen el campo sonoro alrededor de un punto central de referencia.

Ambisonics fue desarrollado por Michael Gerzon en la década de 1970 [1]. Se basa en una serie de expansión alrededor de un origen, en términos de armónicos esféricos. Cada canal en formato ambisónico corresponde a un coeficiente de dicha serie y, por tanto, a un armónico esférico. El número de armónicos esféricos utilizados define el orden de la señal ambisónica y precisa el número de canales necesarios para su codificación (el número de canales será igual a  $(N+1)^2$ , donde N es el orden). Cuanto mayor es el orden y, por tanto, el número de canales utilizados para codificar un campo sonoro, mayor será la resolución espacial de la representación de dicho campo [2].

Parece lógico pensar en una herramienta para la comparación entre diferentes codificaciones de forma objetiva. Sin embargo, los canales Ambisonics no son la señal de alimentación de cada altavoz y requieren un proceso de decodificación para poder reproducirse correctamente en sistemas de sonido. Por ello la comparación directa de los niveles de cada canal en formato ambisónico no proporciona resultados fácilmente interpretables y demanda un modelo o software para poder apreciar las diferencias.

La aplicación presentada facilita la comparación en tiempo real de señales en formato ambisónico hasta de séptimo orden. Las diferencias se trasladan a una distribución espacial de energía sonora que se visualizan en una proyección azimutal. Se pretende así facilitar comparaciones entre, por ejemplo, codificaciones creadas con distintos softwares o codificaciones de diferente orden.

## 2. HERRAMIENTAS PARA LA COMPARACION DE SEÑALES AMBISÓNICAS

Existen implementaciones en tiempo real para la codificación, espacialización, edición y producción de escenas sonoras tridimensionales utilizando Ambisonics. Varias de ellas han sido desarrolladas originalmente por grupos de investigación, como los del Institute of Electronic Music and Acoustics austriaco (IEM Plug-in Suite [3]) o del Acoustics Lab de la Universidad de Aalto, en Finlandia (SPARTA - Spatial Audio Real-Time Applications [4]). Las más sofisticadas incluso van más allá de los algoritmos lineales convencionales y emplean procesamiento paramétrico dependiente de la señal.

Habitualmente, estas herramientas se implementan en formato plug-ins que pueden insertarse en el flujo de trabajo de cualquier DAW (estación de trabajo de audio digital). En concreto, para la generación de mapas visuales que ayudan a la medición de la imagen espacial existen diferentes enfoques.

Uno de ellos es trabajar dibujando la distribución de energía de la señal. SPARTA ofrece 'PowerMap', un plug-in que representa la energía sonora relativa, o la probabilidad estadística de que una fuente llegue a la posición de escucha desde una dirección determinada [5]. 'EnergyVisualizer', de IEM, permite visualizar la distribución de energía en la esfera de una señal Ambisonics de entrada.

Por otro lado, la suite de SPARTA presenta también 'SLDoA' [6, 7], un estimador de la dirección de llegada espacialmente localizado y que representa una opción mucho más eficiente desde el punto de vista computacional, en comparación con los algoritmos integrados en 'PowerMap'. Utiliza patrones de haz VBAP para direcciones que se distribuyen uniformemente en la superficie de una

esfera, para obtener señales de orden cero y primer orden que luego se usan para la estimación del vector intensidad. Permite de esta manera la estimación de la dirección de llegada en varios sectores restringidos espacialmente para cada subbanda.

Otro método para conseguir una alta resolución en las visualizaciones en bajos órdenes ambisónicos se presenta por los mismos autores en el plug-in 'DirASS' [7]. Se basa en un principio de reasignación donde la energía direccional de cada punto en una malla se asigna a una nueva dirección, correspondiente a la estimación de dirección de llegada dentro de una región espacialmente localizada centrada alrededor de dicho punto de la malla. Se consigue así una concentración de energías alrededor de las verdaderas fuentes, resultando en mapas más nítidos incluso para órdenes bajos.

Finalmente, respecto a este tipo de plug-ins, McCormack et al. presentan en [8] una completa evaluación de tres alternativas (entre ellas las nombradas en párrafos anteriores) para la visualización de campos sonoros. Todas ellas emplean vectores de intensidad activa espacialmente localizados (SLAI), de gran interés ya que permiten extraer estimaciones de la dirección de llegada en múltiples sectores localizados espacialmente.

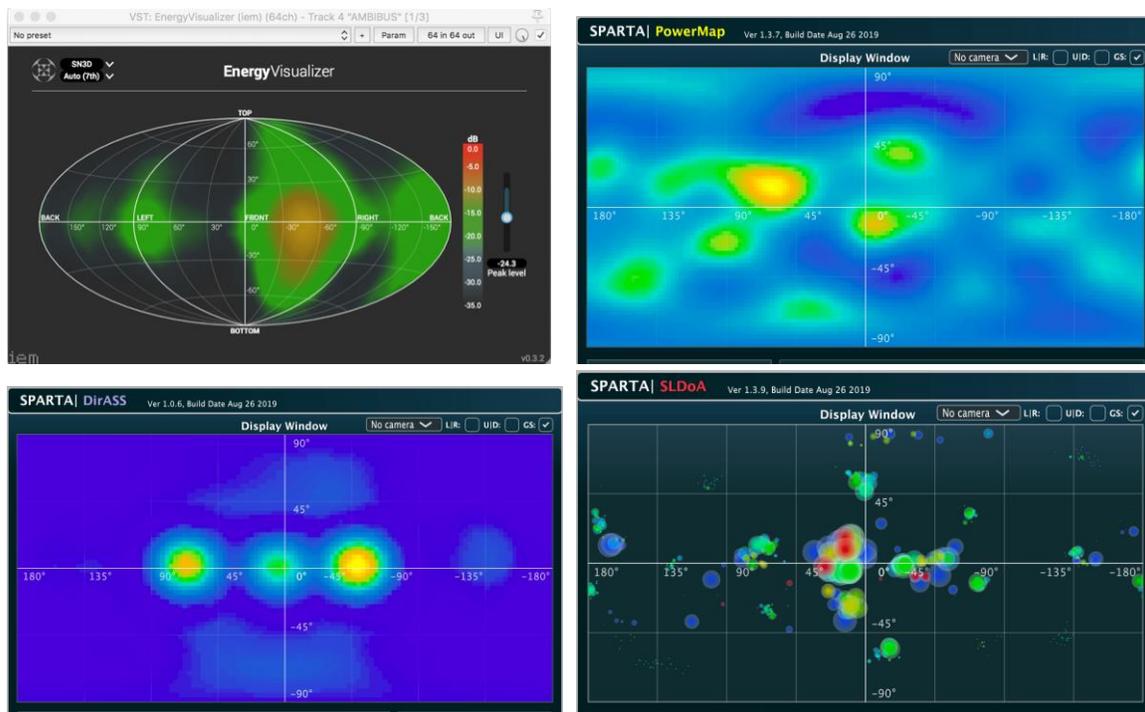


Figura 1 – Visualización de campos sonoros en los plug-ins analizados. De izquierda a derecha y de arriba abajo: EnergyVisualizer, PowerMap, DirASS y SLDoA. Fuentes: [3, 4]

La comparación entre técnicas o programas de codificación o upmixing, así como la de codificaciones en diferentes órdenes, es necesaria para poder cuantificar la calidad y las diferencias objetivamente. Habitualmente estos estudios se hacen de forma subjetiva, diseñando listening tests para un grupo de sujetos [9, 10], los cuales son tediosos y costosos.

Sin embargo, no existen herramientas para una comparación objetiva en tiempo real que prescindiera de un DAW. En este artículo se presenta una aplicación independiente, que permite comparar las distribuciones de energía de dos señales Ambisonics cualesquiera de forma visual.

### 3. IMPLEMENTACIÓN DEL SOFTWARE

La aplicación se ha diseñado para comparar dos codificaciones Ambisonics de una misma escena sonora. Cada una de las señales Ambisonics cargada en la aplicación se representa en una proyección azimutal modificada que preserva las proporciones de las áreas (proyección Hammer). Dicha representación se obtiene mediante la decodificación a una matriz virtual de 426 puntos y un cálculo de los valores RMS en cada uno de ellos. Se consigue así la visualización de la distribución de energía. Tanto la representación como los cálculos que se realizan con las señales han sido adaptados del plug-in 'Energy Visualizer' de IEM [3].

'Ambisonics Energies Comparer' ha sido completamente desarrollada en Matlab, tanto la interfaz de usuario como los algoritmos necesarios para el manejo y representación de las señales. Es de código abierto y se encuentra en: [github.com/jaulab/AmbisonicsEnergiesComparer](https://github.com/jaulab/AmbisonicsEnergiesComparer)

#### 3.1. Proyección Hammer

La proyección Hammer está inspirada en la proyección Aitoff, pero utiliza la forma ecuatorial de la proyección acimutal equidistante de Lambert en lugar del uso de la proyección acimutal equidistante de Aitoff. En la aplicación se calculan 426 puntos uniformemente espaciados de dicha proyección para posicionar los altavoces virtuales (ver Figura 2).

#### 3.2. Cálculo de la distribución de energía

La distribución de energía se calcula como el valor eficaz, o RMS, de un número finito de muestras aisladas de las codificaciones. Para ello es necesario seleccionar un frame rate (por defecto de 0.5 segundos), el cual se puede modificar en cualquier momento. La tasa de imagen marca también el tamaño de muestras de cada buffer de audio en el que se dividen las señales, que será igual al producto entre el periodo y la frecuencia de muestreo del audio. Cuanto menor sea este valor, mayor será la definición temporal; pero también la carga computacional.

El primer paso es decodificar la señal de entrada, obteniendo así las 426 salidas de los altavoces virtuales. El método de decodificación utilizado es de  $\max-r_E$ , el cual maximiza el vector de energía optimizando la reproducción en altas frecuencias [11]. En este paso también se tiene en cuenta la normalización utilizada en la codificación (N3D o SN3D). El cálculo de los valores RMS se repite para cada frame, resultando así tantas matrices de 426 puntos como secciones (frames) en las que se ha dividido el flujo de audio.

Finalmente, una vez obtenidos los valores eficaces, se procede a dibujar la distribución de energía, la cual se muestra en decibelios. Para ello se hace una interpolación de estos puntos (Figura 2 izquierda) a una malla de 380x310 puntos (se pasa así de tener 426 puntos a 117800, creando así una superficie continua, ver Figura 2). El tamaño de esta malla final se ha propuesto para optimizar tanto la resolución del dibujo como la carga computacional.

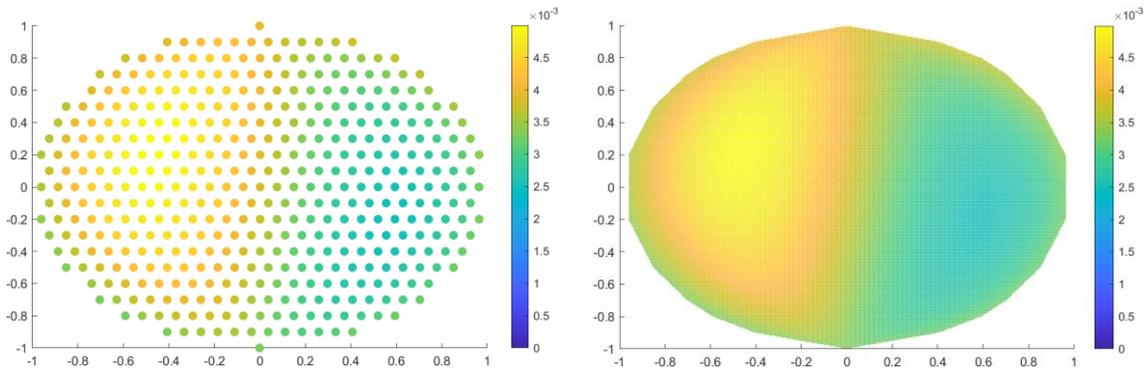


Figura 2 – Ejemplo del paso de los 426 puntos iniciales a la superficie de 117800.

### 3.3. Interfaz

La interfaz se divide en cuatro zonas: dos para la carga de los archivos a comparar (A y B), la que muestra la comparación (C) y la de reproducción (D) (ver Figura 3).

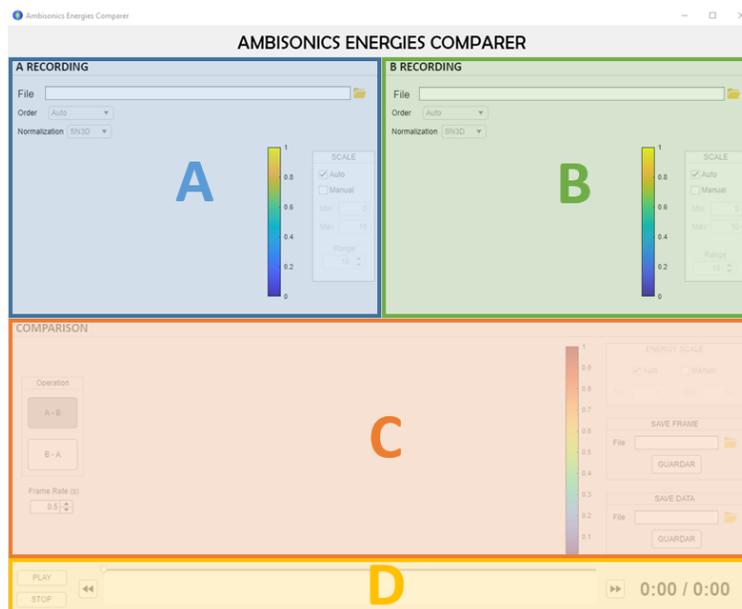


Figura 3 – Interfaz de la aplicación con las cuatro zonas marcadas.

#### 3.3.1. Carga de archivos

Para cargar los dos archivos se lanza una búsqueda en el ordenador. Solo se pueden seleccionar archivos wav, y ambos deben tener la misma longitud. Una vez seleccionado el archivo de audio, la aplicación calcula automáticamente el orden basándose en el número de canales del archivo. Si se desea, se puede seleccionar el orden manualmente, pero nunca órdenes superiores al calculado por

la aplicación (no aparecerán en la lista). También es necesario seleccionar qué tipo de normalización se ha utilizado a la hora de la codificación (N3D o SN3D).

La visualización se puede hacer de dos formas: automática y manual. Con el modo automático el programa calculará el rango óptimo para cada visualización, por lo que será independiente en cada figura. Con el modo manual se puede seleccionar el rango y el valor mínimo a dibujar. Dicho rango será igual para las dos figuras de las señales de entrada.

### 3.3.2. Comparación

La comparación se plantea como la resta de los valores de energía entre A y B o B y A. Una vez cargados los dos archivos a comparar, y habiéndose comprobado que ambos tienen la misma longitud, el programa procede a calcular la diferencia de energías (por defecto A-B). Para mostrarla de una forma más comprensible, el rango se centra en 0.

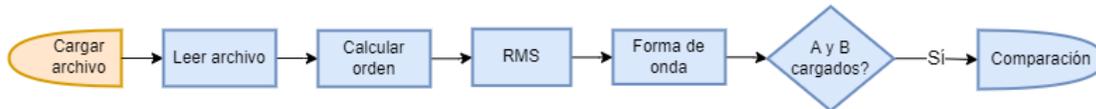


Figura 4 – Diagrama de carga de archivos. En naranja las acciones ejecutables y en azul las internas.

### 3.3.3. Reproducción

La aplicación permite reproducir la comparación como si se tratase de un archivo simple de audio (nótese que la reproducción estará marcada por el frame rate seleccionado). Se marca en todo momento en qué tiempo de la reproducción se está, así como la duración total de las señales. Además, se puede seleccionar manualmente el frame a visualizar y moverse a los contiguos para ver los cambios que se van produciendo.

Para ayudar en el posicionamiento se dibuja la forma de onda del primer canal Ambisonics de los dos archivos cargados; ya que, al ser el canal omnidireccional, recoge todo el audio codificado.

### 3.4. Funcionamiento

En este apartado se detallan las actuaciones que se desencadenan una vez inicializada la aplicación:

1. Cálculo de las 426 posiciones de los altavoces virtuales (ver Figura 2). Dicho cálculo se hace siguiendo las directrices de la proyección Hammer.
2. Cálculo de la matriz de decodificación correspondiente para la disposición virtual de 426 altavoces.
3. Espera a la carga de ambos archivos. Si no son de la misma longitud no continúa.
4. Cálculo de la diferencia entre las matrices de energía de A y B.
5. Visualización de las disposiciones de energía.

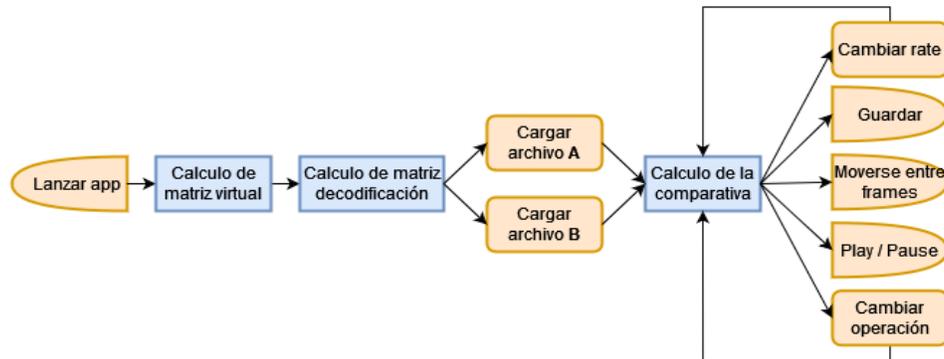


Figura 5 – Diagrama del funcionamiento de la aplicación. En naranja las acciones ejecutables y en azul las internas.

#### 4. ALGUNOS EJEMPLOS DE USO

En la Figura 6 se compara una misma escena sonora codificada en primer y tercer orden con el codificador 'StereoEncoder' de IEM [3]. La escena del ejemplo se trata de una cantante a capela, codificada en círculos en el plano horizontal y posteriormente en el vertical. Se puede apreciar como la codificación en tercer orden es menos ancha, más directiva comparada con la de primer orden. La escala que aparece en cada figura nos permite apreciar también la diferencia entre niveles.

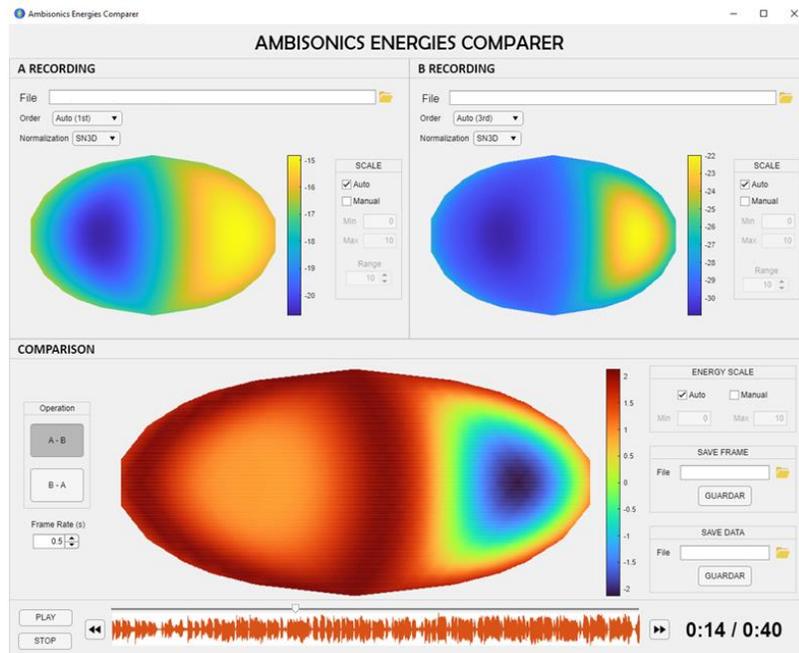


Figura 6 – Ejemplo de aplicación: comparación entre una codificación de un archivo de audio en primer orden (A) y en tercer orden (B). La visualización se hace de forma automática. La coloración roja en la comparación significa que habrá un nivel mayor en la codificación de primer orden, mientras que la azul lo marca en la de tercero; con un máximo de 2 dBs.

Dicha codificación en tercer orden también se puede obtener mediante el uso de upmixers. Utilizando técnicas paramétricas se pueden conseguir señales en altos órdenes partiendo de grabaciones o codificaciones de ordenes inferiores. Entre los existentes, DirAC [12], Harpex [13] o Compass [14] son populares.

En la Figura 7 se comparan dos técnicas de upmixing: DirAC y Compass. Ambas han partido del mismo audio en primer orden para obtener una decodificación en tercer orden. En este caso, la escena se trata de una grabación con un micrófono Ambisonics de cuatro personas teniendo una conversación en diferentes posiciones alrededor del punto de grabación. En comparación con el ejemplo anterior, se ve como el campo generado es más difuso, ya que no hay una única fuente de sonido. Aun así, la diferencia de niveles no es tan acusada como en el caso de la codificación directa a tercer y primer orden.

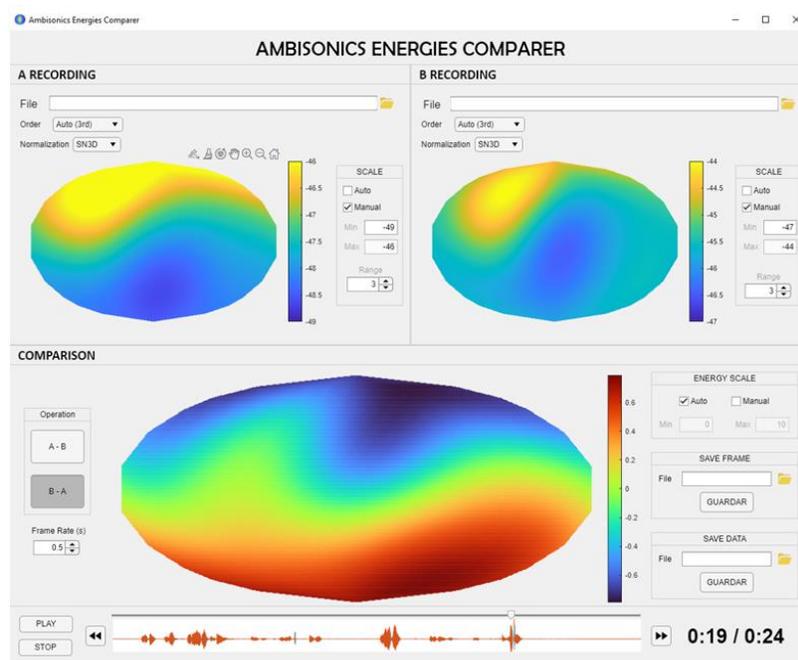


Figura 7 – Ejemplo de la aplicación: comparación entre dos codificaciones creadas con las técnicas de upmixing DirAC (A) y COMPASS (B). En este caso, la visualización se hace de forma manual con un rango de 3 dBs y la comparación se ha hecho de B hacia A. El rojo vuelve a marcar diferencia positiva y el azul la negativa, aunque en este ejemplo solo llega a 1 dB de diferencia máxima.

## 5. CONCLUSIONES

Se ha desarrollado una aplicación con el objetivo de poder comparar en tiempo real dos señales Ambisonics cualesquiera de forma visual. Para ello se calcula la distribución de energía de ambas señales sobre una malla de altavoces virtuales de alta resolución, creando así un mapa de colores en una proyección azimutal que preserva las proporciones de las áreas. La aplicación es útil para comparar diferentes codificaciones, órdenes o algoritmos de upmixing.

## AGRADECIMIENTOS

Este trabajo ha sido financiado por el Ministerio de Ciencia e Innovación y la Agencia Estatal de Investigación, Proyecto PID2020-118829RB-I00 / MCIN/ AEI / 10.13039/501100011033

## REFERENCIAS

- [1] M. A. Gerzon, «The design of precisely coincident microphone arrays for stereo and surround sound,» *Journal of the audio engineering society*, Marzo 1975.
- [2] F. Zotter y M. Frank, *Ambisonics*, Springer, 2019.
- [3] Institute of Electronic Music and Acoustics - IEM, «IEM Plug-in Suite,» [En línea]. Available: [plugins.iem.at](http://plugins.iem.at). [Último acceso: 27 09 2022].
- [4] Aalto University, «SPARTA,» [En línea]. Available: [leomccormack.github.io/sparta-site](https://leomccormack.github.io/sparta-site). [Último acceso: 27 09 2022].
- [5] L. McCormack, S. Delikaris-Manias y V. Pulkki, «Parametric Acoustic Camera for Real-time Sound Capture, Analysis and Tracking,» 2017.
- [6] L. McCormack, S. Delikaris-Manias, A. Farina, D. Pinaridi y V. Pulkki, «Real-Time Conversion of Sensor Array Signals into Spherical Harmonic Signals with Applications to Spatially Localised Sub-Band Sound-Field Analysis,» 2018.
- [7] L. McCormack, A. Politis y V. Pulkki, «Sharpening of Angular Spectra Based on a Directional Re-assignment Approach for Ambisonic Sound-field Visualisation,» de *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.
- [8] L. McCormack, S. Delikaris-Manias, A. Politis, D. Pavlidi, D. Pinaridi y V. Pulkki, «Applications of Spatially Localized Active-Intensity Vectors for Sound-Field Visualization,» *Journal of the Audio Engineering Society. Audio Engineering Society*, vol. 67, p. 840, 2019.
- [9] R. Eguinoa, R. San Martin, D. Arteaga y D. Scaini, «Subjective Evaluation of the Localization Performance of the Spherical Wavelet Format Compared to Ambisonics,» de *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 2021, pp. 1-8.
- [10] F. Zotter y M. Frank, «Localization Experiments Using Different 2D Ambisonics Decoders,» *Tonmeistertagung*, Leipzig, 2008.

- [11] J. Daniel, «Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions,» de *105th Convention of Audio Engineering Society*, San Francisco, Septiembre 1998.
- [12] V. Pulkki, «Directional Audio Coding in Spatial Sound Reproduction and Stereo Upmixing,» de *Audio Engineering Society Conference: 28th International Conference: The Future of Audio Technology—Surround and Beyond*, Piteå, Suecia, Enero 2006.
- [13] S. Berge y N. Barrett, «HIGH ANGULAR RESOLUTION PLANEWAVE EXPANSION,» de *2nd International Symposium on Ambisonics and Spherical Acoustics*, París, Mayo 2010.
- [14] P. Archontis, T. Sakari y V. Pulkki, «COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes,» de *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Alberta, Canada, Abril 2018.