



Universidad Nacional Autónoma de México

Programa de Maestría y Doctorado en Música

Facultad de Música

Instituto de Ciencias Aplicadas y Tecnología

Instituto de Investigaciones Antropológicas

La tridimensionalidad de la estereofonía.

TESIS

Que para optar por el grado de Maestría en Música (Tecnología Musical)

Presenta

Raúl Ignacio Moller Jensen

Tutor

Dr. Jorge Rodrigo Sigal Sefchovich

(Escuela Nacional de Estudios Superiores, Unidad Morelia)

Ciudad de México, agosto de 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Declaro conocer el Código de Ética de la Universidad Nacional Autónoma de México, plasmado en la Legislación Universitaria. Con base en las definiciones de integridad y honestidad ahí especificadas, aseguro mediante mi firma al calce que el presente trabajo es original y enteramente de mi autoría. Todas las citas de obras elaboradas por otros autores, o sus referencias, aparecen aquí debida y adecuadamente señaladas, así como acreditadas mediante las convenciones editoriales correspondientes.

Índice

Introducción	5
Capítulo 1: Antecedentes.....	7
1.1 La premisa de David Gibson	7
1.2 Espacio, escena e imagen.....	9
1.3 Estereofonía	10
1.4 Tridimensionalidad	12
1.5 Visualización del sonido.....	14
1.6 Recuperación de información musical.....	17
1.7 Análisis del contenido de audio y visualización del sonido en Max.....	19
1.7.1 Características físicas y sensaciones psicoacústicas.....	20
1.7.2 Algunas consideraciones	24
1.7.3 Utilización de descriptores en Max	25
1.8 Procedimientos para la investigación de la localización del sonido.....	26
1.9 Una mirada holística	28
Capítulo 2: Método.....	31
2.1 Diseño de la aplicación	31
2.2 Diseño del experimento.....	34
2.3 Estímulos	36
2.4 Cuestionarios.....	39
2.5 El experimento	41
Capítulo 3: Resultados	43
Capítulo 4: Conclusiones	53
Bibliografía	55
Índice de figuras	59
Índice de tablas	60
Anexo.....	61
Sección interna de la aplicación:.....	61

Dedicada a la memoria de mi abuela Elizabeth Josefa Widmer.

Introducción

El objetivo de esta investigación es examinar la solidez de la premisa del autor David Gibson (1997), quien en su obra titulada *The Art of Mixing: A visual guide to recording, engineering and production* presenta la posibilidad de pensar las mezclas estereofónicas en términos de tridimensionalidad sonora con su correspondiente visual.

Para cumplir con el objetivo fue necesario extraer esta premisa que surge en el contexto de la producción musical para luego abordarla desde el ámbito académico a través de una revisión de literatura y de un experimento, constituyendo así un estudio de tipo mixto. De este modo, mientras que la revisión de literatura me permitió ahondar en las teorías y conceptos que sustentan la investigación, el experimento me ayudó a constatar la solidez de la premisa en términos prácticos. Este último requirió del diseño y desarrollo de un *software* específico (a partir de ahora referido como aplicación), utilizado para la generación de estímulos audiovisuales mediante la obtención de visualizaciones tridimensionales a partir del análisis sonoro.

Existen dos motivaciones principales por las cuales hacer esta investigación, la primera de ellas radica en el hecho de que no es habitual hablar en términos de tridimensionalidad en el contexto de la estereofonía, y la segunda en que la obra donde surge la premisa es utilizada como material educativo en carreras afines a la música, la producción musical y la ingeniería en sonido, área en el cual me desempeño como docente y profesional.

Se obtuvieron como productos de la investigación una aplicación desarrollada en Max, una presentación de los avances en el simposio de alumnos del Posgrado en Música de la UNAM y una conferencia en el festival Visiones Sonoras exponiendo parte de las investigaciones realizadas.

En el primer capítulo trato los antecedentes del tema a través de la revisión de la literatura, en el segundo capítulo expongo el método utilizado para el experimento, en el tercer capítulo muestro los resultados del mismo, en el cuarto capítulo presento las conclusiones, y finalmente proveo un anexo donde se muestra el funcionamiento interno de la aplicación desarrollada para la presente investigación.

Capítulo 1: Antecedentes

En este capítulo expongo los conceptos principales que sustentan la investigación, comenzando por la premisa de Gibson que sirve como punto de partida para su contextualización teórica lo largo del capítulo, y finalizando con definiciones que sirven para respaldar el experimento.

1.1 La premisa de David Gibson

En este apartado explico la premisa de Gibson que es tomada como base para la presente investigación, se trata de una particular visión acerca de la estereofonía donde el autor propone una interpretación tridimensional de manera sonora y visual de ésta. Desde la perspectiva de la producción musical Gibson dice que, a través de la manipulación del volumen, paneo y tono de los sonidos es posible moverlos en el campo estereofónico, y que estos cambios se corresponden con los ejes visuales Z, X e Y respectivamente (pp. 9-13).

El autor asevera que nos relacionamos con el sonido de dos maneras, por un lado, sentimos y escuchamos las ondas sonoras físicas que salen de los altavoces y por otro lado imaginamos la ubicación aparente de ese sonido entre éstos. Además, afirma que esta ubicación de los sonidos es llamada imagen, ya que es producto de nuestra imaginación, dado que si no prestamos atención a la mezcla o si estamos a un lado de los altavoces la imagen no existe.

Gibson propone entonces una forma de representación visual a partir de estas nociones para representar la imagen asociada a la escucha. De modo que el paneo o posicionamiento entre izquierda y derecha de los sonidos entre los altavoces es mostrado visualmente como una ubicación entre izquierda y derecha correspondiendo al eje visual X, el volumen de un sonido puede ser mostrado como una ubicación desde el fondo hacia el

frente sobre el eje visual Z, las frecuencias agudas hacia arriba y las frecuencias graves hacia abajo en el eje visual Y (pp. 10-11). En cuanto a la disposición de la panorámica, el autor dice que cuando giras la perilla de paneo es como si pudiéramos ver el sonido moviéndose de izquierda a derecha entre los altavoces, mientras que, en cuanto al volumen de los sonidos explica que aquellos que están cerca de nosotros son más fuertes y los sonidos distantes son más suaves, dado que algunos sonidos en las mezclas están justo al frente mientras que otros instrumentos se ubican frecuentemente al fondo, por último, en cuanto al tono de los sonidos menciona que ocurre una interesante ilusión con las frecuencias agudas y graves en el mundo de la imagen puesto que pareciera ser que los agudos están más a lo alto y los graves más hacia abajo, e indica que esta ilusión existe en parte porque las frecuencias graves llegan desde el piso a nuestros pies y las agudas no.

Finalmente, el autor plantea la utilización de esferas para representar a los sonidos en el espacio tridimensional, considerando que una imagen redonda es la apropiada, dada la manera en que dos sonidos que están paneados se encuentran y se solapan hacia el centro. También argumenta que tiene sentido utilizar transparencias para los sonidos, dado que, si usamos esferas transparentes para representar la ubicación del sonido en el campo sonoro de la imagen, entonces dos sonidos pueden ser vistos y escuchados desde el mismo punto. Además, sugiere la utilización del color en las esferas aclarando que la función principal del color es diferenciar diferentes tipos de sonidos.

1.2 Espacio, escena e imagen

Es preciso señalar, luego de lo expuesto en el apartado anterior, que la premisa de Gibson se sostiene a través de una serie de estipulaciones idiosincráticas y arbitrarias. En este sentido Trevor Wishart (1996) menciona que la reproducción estereofónica del sonido mediante altavoces ha permitido la creación de un espacio acústico virtual en el cual es posible proyectar una imagen sonora (p. 136). Señala también que la existencia de este espacio acústico virtual formado por los altavoces, genera nuevas posibilidades creativas dado que el espacio acústico que estamos representando no tiene por qué ser real, lo que nos permite jugar con la percepción del oyente. El autor explica que la grabación del material musical y su representación han traído consigo una variedad de espacios acústicos que son convenciones de un modo de representación, y se refiere a ellos como espacios acústicos formalizados distinguiéndolos de los espacios acústicos reales (p. 142). Finalmente afirma que, en el estudio de grabación se dan por sentado varias manipulaciones formales del espacio acústico como convenciones de un paisaje musical formalizado.

Por su parte Jens Blauert (2013) dice que antes de la reproducción de la escena ésta es representada en términos de una escena grabada con ciertas técnicas o en función a determinadas manipulaciones posteriores que se basan en algún modelo, y cuando esa representación de escena se dispone en un método de reproducción dado el resultado es conocido como escena sonora virtual y se requiere de un sistema de reproducción específico para que la escena auditiva correspondiente se asemeje a la escena deseada por el creador (p. 256).

En este sentido Jonty Harrison (2023) describe la existencia un espacio musical intrínseco que se define por la interacción de otros espacios tales como: el espacio espectral donde se despliega el tono y el espectro de los sonidos, el espacio temporal gobernado por

las estructuras temporales, el espacio dinámico definido por las amplitudes y el espacio espacial o posición del sonido dentro del campo sonoro.

Finalmente, Denis Smalley (2007) asegura que el espacio de la imagen estéreo está centrado de manera implícita, y que la perspectiva de este espacio es tanto visual como física, siendo los altavoces quienes evocan la convención del marco de una ventana (pp. 49-50).

1.3 Estereofonía

El primero de los casos documentados de una técnica de microfonía estereofónica tuvo lugar en el Palacio de la Industria en París durante la primera Exposición Internacional de Electricidad de 1881, menciona Edwin Pfanzagl-Cardone (2020). El autor comenta que en este sitio de exhibiciones una de las habitaciones contaba con 10 audífonos que estaban conectados mediante vía telefónica a sendos micrófonos ubicados a lo largo del borde del escenario de la Opera de París a 3 km de distancia. Los presentes en aquella demostración pudieron notar que, al escuchar con ambos oídos utilizando dos audífonos el sonido tomaba un relieve y localización especiales que con un solo audífono no era posible producir (La Lumière Electrique and Scientific American, como se citó en Pfanzagl-Cardone, pp. 337-338).

En cuanto a la grabación estereofónica, el autor explica que la primera de ellas fue realizada mediante el *Multiplex Graphophone Grand* en 1889, un dispositivo que utilizaba tres embudos y tres cápsulas de captación independientes a través del principio mecánico de la grabación y reproducción del sonido, funcionando de manera similar al fonógrafo de Edison. Afirma además que otro caso bien documentado acerca de los inicios de la grabación estereofónica sucedió en Estados Unidos hacia 1928 en los Laboratorios Bell bajo la supervisión de Arthur C. Keller quien realizó las primeras grabaciones de voz hablada, piano

y orquesta a dos canales, como producto colateral de dichos experimentos se produjeron cerca de 6000 lacas de cera laminadas en oro con varias grabaciones, entre ellas 128 grabaciones de la Orquesta de Filadelfia bajo la conducción de Leopold Stokowski (pp. 338-339).

En las memorias biográficas escritas para la Academia Nacional de Ciencias, Stephen Fletcher (1992) cuenta que su padre Harvey Fletcher formó parte del equipo de trabajo de los Laboratorios Bell donde conoce a Stokowski, lo que posibilitó que en 1931 comenzaran a hacer pruebas de sonido estereofónico en la Academia de Música de Filadelfia, lugar donde la orquesta realizaba sus conciertos. El autor continúa explicando que su padre y Stokowski encontraron una gran sala libre que podía albergar a toda la orquesta donde había suficiente espacio para realizar los experimentos, de esta manera hicieron pruebas hasta que se percataron de que habían desarrollado un sistema estereofónico. Afirma además que originalmente la teoría de este sistema consistía en un número infinito de altavoces de un lado y el mismo número de micrófonos del otro lado, sin embargo, descubrieron que tres micrófonos, tres líneas de transmisión y tres altavoces eran suficientes (p. 183). El autor indica también que las grabaciones fueron hechas entre 1931 y 1932, con una respuesta en frecuencia de 30 Hz a 10 kHz utilizando una técnica de microfonía ABC amplia (p. 339), para la cual se emplean tres micrófonos, dos de ellos formando una técnica AB ubicados en paralelo mirando hacia al frente y un tercer micrófono al centro de este arreglo apuntando hacia la misma dirección.

Fletcher además comenta que su padre estaba listo para demostrar este efecto espacial conocido entonces con el nombre de perspectiva auditiva, cuya exhibición principal sucedió en la Feria Mundial de Chicago de 1932, donde se dispuso un semicírculo de mesas con varios pares de audífonos frente a un escenario, separando las mesas de éste mediante un

vidrio de manera que no se pudiese escuchar directamente lo que allí ocurría. El autor dice que su padre había colocado en el escenario una cabeza de maniquí con dos micrófonos (uno en cada oído de éste) mismos que estaban conectados a los audífonos disponibles para la audiencia, la experiencia consistía en una persona caminando alrededor de la cabeza de maniquí diciendo algunas palabras. Asegura también que la perspectiva auditiva lograda por este sistema permitía a los oyentes ubicar cada sonido en su posición correcta y seguir sus movimientos (pp. 182-184).

Es preciso señalar que para los fines de la presente investigación cuando hable de estereofonía me referiré a aquella que se forma por la escucha mediante altavoces, la cual como dice Gary Kendall (1995) crea la impresión de estar frente a un espacio sonoro (p. 23). Se trata de un arreglo estéreo tradicional donde el oyente se posiciona mirando hacia dos altavoces ubicados al frente de él, cada uno en un ángulo de 30°, mencionan Hugo Fastl y Eberhard Zwicker (2007). Asumiendo que el oyente está posicionado exactamente entre los altavoces, una señal estereofónica resulta en una imagen fantasma en la que los sonidos parecen originarse entre los dos altavoces, y dado el efecto de precedencia si el oyente está ligeramente cercano a uno de éstos parecerá que el sonido proviene de ese altavoz y por lo tanto la imagen fantasma será inestable, como asegura Barry Truax (2023).

1.4 Tridimensionalidad

Luego de una revisión de literatura cuya búsqueda estaba centrada en el término tridimensionalidad asociado a la estereofonía, pude encontrar varios autores hablando acerca de la existencia, o la inexistencia, de una tridimensionalidad sonora en el contexto estereofónico.

En primer lugar nos encontramos con Fletcher quien en su autobiografía de 1967 menciona que tras haber realizado sus primeros experimentos en 1931, podía asegurar que la perspectiva auditiva provocada por un sistema estereofónico no solo permitía disponer los sonidos en su posición correcta entre los altavoces o desde atrás hacia adelante, sino que también resultaba tridimensional, pudiendo reproducir sonidos incluso con su elevación aparente (Fletcher, 1967, como se citó en Fletcher, 1992, p. 184). Gibson por su parte asegura que el campo estéreo normal es en realidad tridimensional, y que un ingeniero de audio tiene el mismo rango de control que un escultor, dado que ambos trabajan tridimensionalmente (pp. 12-19). En este sentido, Fastl y Zwicker distinguen entre los conceptos de lateralización y localización, en primer lugar, la lateralización es estudiada mayormente a través de audífonos, donde la imagen sonora está ubicada entre la cabeza y la tarea del sujeto es describir el desplazamiento lateral percibido del sonido, y en segundo lugar mientras que la lateralización es descrita a través de una sola dimensión, la localización es un fenómeno reconocido como tridimensional, resultando de especial interés (pp. 308-310).

Por el contrario, Kendall menciona que si bien puede ser discutible que grabaciones estéreo convencionales produzcan efectos que pueden llamarse tridimensionales, en realidad cuando nos referimos a sonido tridimensional generalmente queremos decir que el oyente escucha en direcciones que no son experimentadas con la estereofonía convencional (p. 24). Además, asegura que en contraste con la escucha cotidiana la reproducción estéreo convencional provee algo de información espacial, pero no la suficiente para recrear la dimensionalidad experimentada al estar en un recinto con una banda tocando. También reconoce que ciertas grabaciones tienen asociaciones arquitectónicas que ayudan a establecer un contexto espacial, sin embargo, no logran dar la sensación de estar en un espacio tridimensional (p. 23). El autor señala que un sistema de sonido tridimensional debe exceder

el rango de dirección convencional estéreo y ser capaz de posicionar los sonidos por fuera de los límites de los altavoces. A pesar de ello y en concordancia con Gibson, comenta que hay evidencia de que la elevación de un sonido está influenciada por el contenido espectral de la fuente sonora que es recibida, así los sonidos agudos se ubican típicamente más arriba que los sonidos graves. Harvey White y Donald White (2014) mencionan que un sistema estéreo convencional solo es efectivo cuando la fuente sonora es distribuida a lo largo del plano horizontal frente al oyente, ya que consideran que percibe las fuentes sonoras distribuidas a través de una línea entre los altavoces, que combinado con la sensación de distancia que proviene del sistema estéreo produce una sensación de dos dimensiones del sonido, fallando completamente en reproducir la distribución del sonido en el plano vertical (pp. 355-356). Truax por su parte dice que en una reproducción electroacústica hay diversas formas en las que una imagen estéreo puede ser producida, y que cada método de reproducción variará el grado en el que pueda lograr una verdadera perspectiva estéreo de dos dimensiones formada por la lateralidad y la profundidad.

Dada esta revisión de autores se evidencia que no existe pleno acuerdo en cuanto a la asociación entre la percepción de la estereofonía y una determinada tridimensionalidad.

1.5 Visualización del sonido.

Como expuse anteriormente, Gibson propone desde la perspectiva de la producción musical que la estereofonía puede ser interpretada tridimensionalmente de manera sonora pero también visual. En este sentido, Richard Lyon (2017) asegura que existe una larga historia de la visualización del sonido dedicada particularmente a la asistencia de la voz hablada y de los problemas de la audición, por ejemplo, en el siglo XIX el fonógrafo y la llama manométrica fueron dispositivos utilizados por Alexander Graham Bell, en el intento de

realizar trazados o visualizaciones de formas de onda que una persona sorda podía tratar de igualar con su propia voz (p. 489).

Matthew Cooper et al. (2006) dicen que la idea de representar el sonido como un objeto visual en un espacio bidimensional o tridimensional con propiedades relacionadas al contenido de audio se origina en la psicoacústica, y mencionan que los rápidos avances en el rendimiento de las computadoras han permitido una variedad de esfuerzos creativos para conectar la imagen con el sonido, yendo desde simples espectrogramas a la elaboración de sistemas interactivos en tiempo real con gráficos tridimensionales (p. 42).

En cuanto a las visualizaciones de los datos musicales en la actualidad he encontrado una serie de artículos relacionados con la ciencia de la computación, en los cuales se realiza una descripción acerca de los tipos de representaciones visuales existentes, entre ellos: *Visual Display and Retrieval of Music Information* (Valle, 2018), *Music Information Retrieval: Recent Developments and Applications* (Schedl et al., 2014), *Music Information Retrieval and Visualization* (Lemström, K., y Tzanetakis, G., 2011), *Using 3D Visualizations to Explore and Discover Music* (Lamere y Eck, 2007) y *Visualization in Audio-Based Music Information Retrieval* (Cooper et al., 2006).

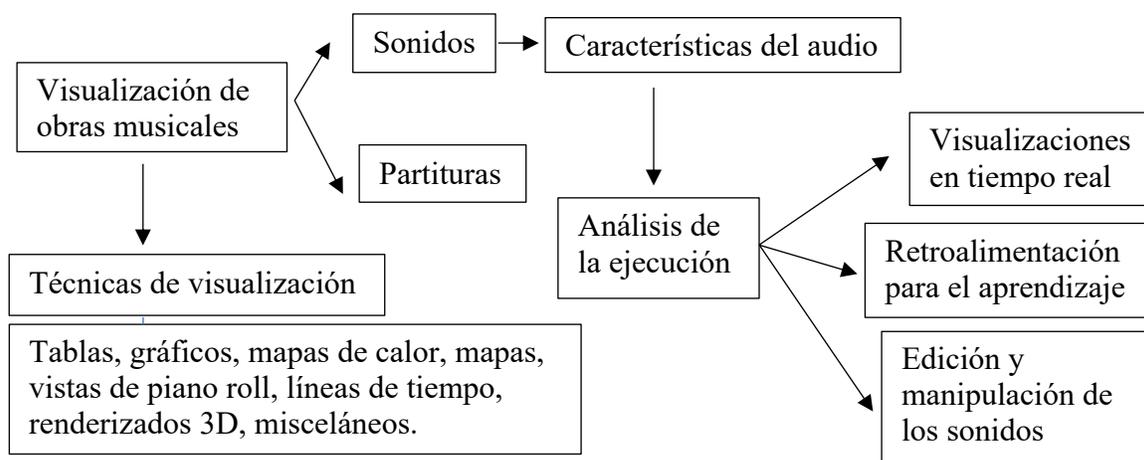
El artículo más destacado fue *A Survey of Visualizations for Musical Data* escrito por Richard Khulusi et al. (2020), se trata de una revisión de 129 visualizaciones las cuales fueron clasificadas de acuerdo a los tipos de datos visualizados, centrándose en la relación entre la musicología y la visualización de datos. Los autores indican que los tipos de datos a partir de los cuales se generan visualizaciones pueden tratarse de obras musicales (partituras y sonidos), colecciones, músicos e instrumentos, y pueden ser a su vez categorizados por su técnica de visualización en: tablas, gráficos, mapas de calor, mapas, vistas de piano roll, líneas de tiempo, renderizados tridimensionales y misceláneos. Durante su extensa revisión

encontraron una gran variedad de tipos de datos para los cuales se han diseñado visualizaciones para apoyar varias tareas de usuario en un contexto musical o musicológico, además hallaron que la mayoría de éstas han sido diseñadas para la visualización de obras musicales, distinguiendo entre aquellas que se desarrollan a partir de una forma escrita como las partituras y entre las que surgen de los sonidos, éstas últimas se tratan de características del audio, extraídas tanto en tiempo real como de audios pre grabados, tales como la altura, la sonoridad y el tempo. Además, los autores hallaron que las visualizaciones que surgen de los sonidos pueden agruparse en análisis de ejecución, análisis emocional del sonido y análisis de similitud sonora. Aseguran también que en cuanto al análisis de la ejecución las visualizaciones pueden tener diferentes propósitos, en primer lugar, generar visualizaciones en tiempo real de ejecuciones en vivo para clasificar o describir estilos de ejecución, en segundo lugar, desde una perspectiva de la enseñanza brindar retroalimentación a los estudiantes con la intención de mejorar sus ejecuciones, y en tercer lugar permitir la edición o manipulación el sonido de una ejecución.

Por otro lado, algo notable en la actualidad es la aparición de herramientas que permiten la visualización y manipulación del sonido en espacios tridimensionales elaboradas por instituciones y empresas desarrolladoras de *software*. Por ejemplo, Spat es un procesador de audio creado por el IRCAM (2023) para la espacialización del sonido en tiempo real, que permite a compositores, artistas sonoros e ingenieros de sonido controlar y visualizar la localización de las fuentes sonoras en espacios auditivos tridimensionales. También existe *Spaces 3D* creado por el GRM que permite al usuario mover y visualizar de manera tridimensional una fuente de sonido multicanal desde un canal hasta 24 canales, en un espacio de reproducción multicanal con la misma posibilidad de canales, el movimiento de las fuentes puede ser libre, aleatorio o puede seguir trayectorias configuradas por el usuario (INA, 2023).

Finalmente, *Fairlight* es un complemento desarrollado por Blackmagicdesign para el *software* de edición de video *Davinci Resolve 18*, que permite la post producción de audio de manera integrada a través de diversas herramientas, entre ellas incluye un espacializador tridimensional que facilita la ubicación visual de múltiples sonidos de manera simultánea (Blackmagicdesign, 2023).

Figura 1. Clasificación de las visualizaciones de las obras musicales según Khulusi et al.



Contextualizando la propuesta visual de Gibson según la revisión de Khulusi et al., puedo clasificarla como un tipo de visualización de obras musicales que a través de técnicas de renderizados tridimensionales analiza la ejecución de los sonidos y sus características para originar visualizaciones en tiempo real y retroalimentación para el aprendizaje.

1.6 Recuperación de información musical.

Para generar visualizaciones del sonido, como expuse en el apartado anterior, es necesario extraer información a través de su análisis, por ello a continuación realizo una revisión de las teorías relacionadas con la obtención de la información del sonido.

Según Alexander Lerch (2012) el análisis del contenido del audio o ACA (por sus siglas en inglés) refiere a la extracción de información de una señal de audio, tal como una grabación musical almacenada en un medio digital (p. 1). El autor explica que se han hecho esfuerzos para extraer de manera automática prácticamente cualquier cosa de una grabación, incluyendo *meta data* formal, perceptual, musical y técnica. Asimismo, indica que el análisis del contenido del audio se trata de un campo de investigación multidisciplinario, dado que requiere la utilización de conocimientos de diferentes ámbitos, tales como la musicología, la teoría musical, la psicología de la música, la psicoacústica, la ingeniería en audio, la biblioteconomía, y la informática. Incluso menciona que el término análisis del contenido del audio no es el único utilizado para referirse a sistemas de análisis de señales de audio, sino que también existe el concepto de recuperación de la información musical (p. 2).

Cooper et al. dicen que la recuperación de la información musical o MIR (por sus siglas en inglés) es un área de investigación que explora cómo la música almacenada digitalmente puede ser efectivamente organizada, buscada, recuperada y navegada (p. 42). Como explica Lerch la recuperación de información musical es entendida como un campo más general y amplio del cual el análisis del contenido de audio es parte, dado que también incluye el análisis de formatos simbólicos tales como partituras y archivos bajo el protocolo MIDI.

Rafael Valle (2018) indica que en el siglo XXI los métodos más comunes para el análisis musical son el estudio visual de la partitura, y la investigación aural y perceptual de grabaciones, y también menciona que la recuperación de información musical con medios computacionales ha jugado un gran rol ayudándonos en el aprendizaje y en el avance de nuestro entendimiento de la música, permitiendo la realización de análisis de diferentes tipos cuya ejecución manual sería incómoda (p. 1).

Tal como dice Lerch existen cinco clases generales que pueden ser identificadas para describir el contenido de una grabación musical: características técnicas o estáticas, de la cualidad sonora o timbre, relacionadas a la intensidad, tonales y temporales tales como el ritmo (p. 4). También menciona que la información básica agrupada en cada clase puede ser utilizada y combinada con otra para obtener un conocimiento más profundo de la música, tal como la estructura musical, el estilo, las características de la interpretación e incluso el estado de ánimo de la música o su efecto emocional. Además, el autor afirma que el análisis del contenido del audio puede ayudarnos con nuevas formas de visualización del sonido, interacción de usuario, y procesos musicales en *softwares* de música, en este último caso el objetivo del análisis del contenido del audio es permitir al usuario interactuar con una interfaz más musical y brindar un acercamiento más intuitivo a la visualización y edición de los datos del audio (pp. 2-3).

1.7 Análisis del contenido de audio y visualización del sonido en Max

Para desarrollar la aplicación que me permitió crear los estímulos audiovisuales utilizados en el experimento de la presente investigación, he optado por el *software* Max. Como menciona Francisco Colasanto (2010) “la versión actual, desarrollada por Miller Puckette y David Zicarelli, cuenta con la incorporación de MSP (tratamiento de audio digital) y Jitter (tratamiento de video) y es publicada por Cycling74” (p. 20). Además, menciona que Max está programado en lenguaje C pero que la interfaz de usuario utiliza un lenguaje gráfico que consta de pequeñas cajas llamadas objetos que realizan funciones determinadas que pueden ser interconectadas a través de cables virtuales, formando diagramas o *patches*.

De manera que para el desarrollo de la aplicación prevista para la presente investigación utilizaré en forma conjunta las posibilidades que proveen las herramientas de

MSP para el análisis sonoro y Jitter para la generación de visualizaciones mediante OpenGL, una herramienta para generar gráficas complejas y reactivas en tiempo real (Cycling '74, 2022), que es la interfaz de programación de aplicaciones multiplataforma más utilizada para la generación de gráficos bidimensionales y tridimensionales que permite a los desarrolladores de *software* generar gráficos de alto rendimiento por ejemplo para la creación de contenidos, entretenimiento, desarrollo de videojuegos, medicina y realidad virtual (Khronos Group, 2023). Por esto, OpenGL resulta ideal para la creación de los estímulos audiovisuales a ser empleados en el experimento de la presente investigación.

Como dice Colasanto, Max posee todas las funciones de procesamiento de otros lenguajes, pero brinda una interfaz gráfica amigable para compositores y creadores multimediales. Sin embargo, contemplé otras opciones para el desarrollo de la aplicación tal como el software TouchDesigner dada su capacidad para la generación de visualizaciones, así como también el software PureData como alternativa de código abierto, pero al carecer de descriptores no me permitían un análisis del sonido adecuado, en cambio Max posee la ventaja de utilizar librerías externas con funciones para el análisis y la extracción de información como las que se presentan en el apartado 1.7.3 Utilización de descriptores en Max

1.7.1 Características físicas y sensaciones psicoacústicas

En este apartado realizo un enlace entre las características físicas mencionadas por Gibson en su premisa y las técnicas de medición pertinentes asociadas a las sensaciones psicoacústicas producidas por cada una de dichas características. En este sentido Federico Miyara (2000) dice que “cada parámetro físico del sonido se corresponde de manera más o menos directa con un tipo de sensación psicoacústica específica” (p. 18), de manera que la

frecuencia está relacionada con la sensación de altura, la amplitud con la sonoridad, y el espectro con el timbre de un sonido. Por su parte, Juan Roederer (2000) menciona que la dirección espacial en la que llega un sonido al oyente es otra característica física relevante, que es codificada por el sistema nervioso para producir la sensación de direccionalidad sonora también conocida como lateralización (p. 13).

La asignación de altura, sonoridad y timbre a un sonido musical es el resultado de operaciones de procesamiento en el oído y en el cerebro. Esta asignación es subjetiva e inaccesible a la medición física directa ... Sin embargo, cada una de estas sensaciones primarias puede ser asociada en principio a una cantidad física bien definida del estímulo original, es decir, de la onda sonora, que puede ser medida y expresada numéricamente por métodos físicos. (Roederer, 2000, p. 12)

En preciso recordar que la premisa de Gibson, que se toma como base para esta investigación, contempla solamente tres aspectos de la estereofonía para representarla tridimensionalmente: el volumen, la panorámica y el tono de los sonidos.

En cuanto a la sonoridad, Fastl y Zwicker aseguran que ésta corresponde a la categoría de las sensaciones de intensidad y que no solo se trata de un valor de sensación, sino que pertenece a algún lugar entre la sensación y los valores físicos. Los autores continúan diciendo que la medición del nivel de sonoridad es la preferida para caracterizar la sensación de sonoridad, dado que las comparaciones de sonoridad pueden conducir a resultados más precisos que las estimaciones de magnitud (p. 203). Asimismo, afirman que la sonoridad fue presentada por Barkhausen hacia el 1920 y puede ser medida para cualquier sonido, pero la más conocida es aquella medida para sinusoides, llamada contornos de igual sonoridad.

Además, mencionan que los contornos de igual sonoridad para sinusoides o ruidos de banda estrecha indican la interesante dependencia de la sonoridad con respecto a la frecuencia, sin embargo, el volumen depende de muchas más variables, como el ancho de banda, el contenido espectral y la duración del sonido (p. 204). Según Fastl y Zwicker, derivado de los contornos de igual sonoridad se define que el nivel de sonoridad de un sonido corresponde al nivel de presión sonora de una senoide de 1 kHz que es perceptualmente tan intenso como él, cuya unidad es el *phon*, de manera que la función matemática de la sonoridad también puede ser trazada para otras frecuencias (pp. 203-207). En cuanto a esto último Alton Everest y Ken Pohlmann (2009) mencionan que, para emular la audición humana, los medidores de nivel usualmente ofrecen una selección de ponderaciones a modo de curvas de respuesta en frecuencia como las formadas por los contornos de igual sonoridad, designadas como ponderación de tipo A, B y C, que en términos generales reducen la medición del nivel de presión sonora para bajas y altas frecuencias (p. 30). Los autores también indican que estas curvas de respuesta fueron diseñadas para proveer a los medidores de nivel sonoro lecturas más acordes con la audición humana, y dado que la respuesta en frecuencia de nuestro sistema auditivo es más plana para mayores amplitudes, la ponderación más utilizada es la de tipo A que corresponde a la emulación de la escucha humana para bajas amplitudes donde el decaimiento en la sensibilidad de la audición para bajas y altas frecuencias es más pronunciado (pp. 30-31).

Acerca del tono de los sonidos, Lerch dice que la percepción humana del tono está directamente relacionada a la frecuencia de una señal, siendo que las altas frecuencias conducirán a la percepción de un tono más agudo y que si la señal es una combinación de componentes sinusoidales con frecuencias f , $2f$, $3f$, entonces la frecuencia fundamental f domina la percepción del tono (p. 79). Sin embargo, existen otras formas de interpretar el

tono de un sonido como por ejemplo a través del cálculo del centroide espectral, definida por el autor como la representación del centro de gravedad de la energía espectral de un sonido, siendo que para las secciones tonales se mueve junto con la frecuencia fundamental, además dice que esta concentración de energía también está correlacionada con la propiedad de brillo del timbre (p. 45). El centroide espectral se calcula como la media ponderada de las frecuencias presentes en la señal, determinadas mediante la transformada de Fourier, utilizando sus magnitudes como pesos, afirma Pfanzagl-Cardone (p. 317).

Con respecto a la panorámica, Bob McCarthy (2007) menciona que los sistemas estéreo son una hibridación de dos canales separados y dos canales emparejados, y que el grado de separación de las señales es una variable abierta a un control artístico (p. 321), asimismo dice que la imagen puede ser ubicada en cualquier punto entre los altavoces, cuya apertura se conoce como campo panorámico (p. 170). En este sentido Huber y Runstein (2017) aseguran que variando la proporción del sonido que se envía a cada altavoz el ingeniero de audio cambia las diferencias de intensidad interaurales relativas y de ese modo crea la ilusión de posición física entre éstos (p. 71). Como explican Fastl y Zwicker, si ambos altavoces irradian el mismo nivel de presión sonora al mismo tiempo el sonido es percibido como proveniente del centro y si el nivel de presión sonora del altavoz de la derecha es mayor al altavoz de la izquierda, el sonido será percibido como proveniente de la derecha (p. 310). Por último, George Tzanetakis et al. (2007) señalan que la ubicación de los sonidos utilizando el paneo estéreo es un elemento importante en el proceso de producción musical y debería ser incorporado en los sistemas de recuperación de información musical, sin embargo, los sistemas de clasificación de audio normalmente ignoran la información relacionada al paneo estéreo convirtiendo los audios a monofónicos antes de analizarlos (p. 441).

1.7.2 Algunas consideraciones

De acuerdo con la premisa de Gibson al momento de la escucha existe una relación directa entre el volumen de los sonidos y la distancia aparente de éstos, sin embargo, Trevor Wishart (1996) indica que la profundidad se representa a través de la disminución de la amplitud y de los componentes de alta frecuencia, así como también del aumento de la reverberación (p. 146). De acuerdo con esto, Jens Blauert (2013) menciona que la percepción auditiva de la distancia depende de una combinación de aspectos acústicos y no acústicos, siendo que puede verse influenciada por el contenido espectral del estímulo, el conocimiento previo que tengamos de la amplitud del estímulo, la lateralización de la fuente, las reflexiones sonoras del ambiente y la información visual que tengamos de la fuente sonora (p. 172).

Gibson establece también una relación entre el tono de los sonidos y la sensación de elevación, no obstante Robert Butler (1973) realizó un experimento donde pudo hallar que el timbre de los sonidos y no el tono influyen la percepción de elevación, sino que es gobernada por su composición espectral incrementando con aumentos en la frecuencia de dicha composición (p. 257).

Por último, Gibson menciona que la panorámica de los sonidos está correlacionada a la percepción de la lateralización de los sonidos. Con respecto a esto es posible destacar que en 1870 Lord Rayleigh publicó sus observaciones experimentales acerca de la localización de los sonidos, señalando que es notablemente sencillo decir si un sonido proviene de la izquierda o de la derecha, en relación con la cabeza. El autor calculó las diferencias de intensidad generadas por el efecto de sombra acústica producido por la cabeza para sonidos que provienen de ambos lados, y demostró que las diferencias de intensidad entre los oídos son suficientes para explicar el fenómeno de lateralización para la mayor parte del espectro, a excepción de las frecuencias graves por debajo de 200 Hz, y dado que esas frecuencias

tienen longitudes de onda mucho más largas comparadas con las dimensiones de la cabeza, no se provoca un efecto de sombra acústica significativo, sin embargo pueden ser fácilmente lateralizadas (Strutt 1877 como se citó en Lyon, 2017, pp. 379-380).

Finalmente, en relación a lo establecido en la premisa de Gibson pero estableciendo criterios más acertados según lo expuesto con anterioridad, Adrian Moore (2016) dice que durante el proceso de grabación pueden identificarse una cantidad de dimensiones espaciales con las cuales es posible trabajar: en primer lugar con la ubicación de los sonidos de izquierda a derecha que puede ser determinada al momento de la grabación o posteriormente en un *DAW* mediante el uso del paneo, en segundo lugar con la noción de lo cercano y lo lejano dada por la relación entre la amplitud, el espectro y la reverberación del sonido, y en tercer lugar la ubicación desde arriba hacia debajo de los sonidos debido a las inflexiones del tono y sus características espectrales (p. 26).

1.7.3 Utilización de descriptores en Max

En el caso del volumen y el tono utilicé un objeto externo a Max llamado *pipo~*, desarrollado por un equipo de investigadores del IRCAM, que como indica Norbert Schnell et al. (2017) se trata de una serie de módulos disponibles a través del paquete de objetos llamado MuBu. Este paquete fue creado por Schnell en el 2008, sus siglas significan *buffer* múltiple, y provee una caja de herramientas para el análisis multimodal del sonido y del movimiento, la síntesis de sonido interactiva y el aprendizaje de maquina e incluye procesamiento de datos en tiempo real y por lotes (descriptores de sonido, funciones de movimiento y filtrado), síntesis granular, aditiva y concatenativa, visualización de datos, reconocimiento estático y temporal, y algoritmos regresivos (IRCAM Forum, 2022). Como mencionan Schnell et al. los módulos de PiPo pueden correr en tiempo real en Max utilizando el objeto *pipo~* (p. 7). Los autores

también mencionan que la creación de PiPo fue motivada por la necesidad recurrente de utilizar múltiples módulos de procesamiento de señales para la extracción de descriptores de bajo nivel del audio en diferentes contextos de uso (p. 2). PiPo fue desarrollado para el diseño de aplicaciones de audio interactivas en Max, permitiendo el filtrado, transformación, extracción y segmentación de datos de audio, así como también la utilización de estos módulos para procesar datos de audio en tiempo real dedicados al control de procesos audiovisuales (IRCAM Centre Pompidou, 2022).

Por último, en cuanto a la panorámica de los sonidos he decidido crear un objeto propio en Max, en el cual se implementa una función para calcular qué lado (izquierdo o derecho) de la señal de audio estereofónica tiene una mayor amplitud que el otro y en qué proporción (ver Figura 35).

1.8 Procedimientos para la investigación de la localización del sonido

Los procedimientos más utilizados en la investigación de la localización del sonido pueden clasificarse de acuerdo con la técnica de presentación del estímulo y el procedimiento psicoacústico que se utilice, señalan Richard Fay y Arthur Popper (1993). Asimismo, explican que de acuerdo a la técnica pueden darse dos tipos de investigaciones, las que utilizan audífonos y las que utilizan altavoces para reproducir los estímulos sonoros, y en cuanto a los procedimientos psicoacústicos se pueden tratar de tres clases: la discriminación, la categorización y el juicio absoluto (p. 175).

Los autores mencionan además que una gran variedad de preguntas acerca de la localización y la posición aparente de los sonidos fueron abordadas a través de experimentos donde se utilizaban altavoces, mediante los procedimientos de categorización y juicio absoluto. También dicen que el término categorización es utilizado para procedimientos en

los cuales luego de la presentación de cada estímulo se les da a los oyentes unas pocas alternativas para responder, debiendo indicar cuales de las opciones disponibles coincide de manera más cercana con la posición aparente de la imagen sonora percibida (p. 176). Sin embargo, aseguran que existen algunos problemas con el procedimiento de la categorización, en primer lugar si la posición aparente de un estímulo no se encuentra disponible en las opciones de respuesta el oyente es forzado a tener que emitir un juicio que no estaba realmente presente en la posición aparente del sonido, y en segundo lugar si la cantidad de alternativas de respuestas disponibles es pequeña o si la diferencia entre los estímulos es sutil los oyentes podrían distribuir sus respuestas entre las diferentes opciones dependiendo más de una distribución del estímulo que de una distribución de las posiciones aparentes.

Por otro lado, Fay y Popper mencionan que, para medir la posición aparente de un sonido presentado a través de altavoces, el procedimiento del juicio absoluto ofrece algunas ventajas, dado que el oyente estima la posición aparente de la imagen auditiva directamente mediante escalas más o menos continuas y una cantidad de respuestas prácticamente ilimitadas. Asimismo, indican que en los procedimientos de juicio absoluto se utilizan dispositivos como punteros o pistolas para apuntar hacia la posición aparente del sonido, también se pueden utilizar dispositivos de seguimiento del movimiento de la cabeza o incluso se le puede pedir al oyente que señale con su nariz (pp. 176-177). Los autores a su vez advierten que también existen algunos problemas con el procedimiento de juicio absoluto, primero la distancia aparente del sonido no puede ser medida con esta técnica y segundo es difícil determinar cuánto de la variación en las respuestas es perceptual, resultado de imágenes difusas o posiciones dudosas, y cuánto puede ser atribuido a componentes del proceso de respuesta como la coordinación y la habilidad motora del oyente (p. 177).

En el caso de esta investigación y considerando lo anteriormente dicho, he decidido optar por el procedimiento de la categorización para la realización del experimento, dado que deseo indagar acerca de la percepción de la posición aparente de la distancia de los sonidos, y debido a que no quiero que los resultados dependan de la coordinación o habilidad motriz de las personas que participen en el experimento.

1.9 Una mirada holística

Como plantea Reinier Plomp (2001), el sistema auditivo no puede ser explicado solamente en términos de un proceso pasivo, sino que también involucra un proceso activo, siendo necesario un acercamiento más sofisticado a la escucha que considere ambos.

Desde su punto de vista la física del siglo XVII provocó un gran impacto en cómo la naturaleza debía ser estudiada, puesto que ésta sugería que la mejor forma de estudiar las relaciones de causa y efecto de un sistema complejo consistía en reducirlo en subsistemas más maleables, dotados de las siguientes características: a) aislar un pequeño subsistema del mundo físico, por ejemplo, al utilizar sinusoides como estímulo, b) eliminar todos los factores de confusión, por ejemplo las distorsiones, el ruido ambiente o del equipo, c) mantener todos los parámetros constantes menos a uno de ellos, por ejemplo la duración del sonido, y d) investigar el efecto de la variación de ese parámetro particular, por ejemplo la presión sonora (p. 3). Por consiguiente, considera que la preferencia por el estudio de sinusoides no es sólo un fenómeno accidental, sino que puede ser visto como la aplicación concreta de un ideal científico, el cual presume que sólo con la descripción de abstracciones es suficiente para explicar el sistema como un todo, un punto de vista que considera simplemente incorrecto. En este mismo sentido el autor menciona que en cuanto al estudio de la percepción sonora había una creencia común que estudiar los sonidos en un nivel elemental (a través de

sinusoides) nos daría una noción confiable de cómo se perciben los sonidos cotidianos, como consecuencia de ello la atención fue totalmente dirigida a los elementos individuales y los fenómenos globales no fueron suficientemente examinados (p. 6). Concluye entonces que el punto de vista macroscópico es esencial y por lo menos tan importante como el microscópico para explicar la percepción de los sonidos complejos de la vida cotidiana.

Asimismo, Plomp menciona que el proceso de percepción no sólo radica en la sensación de un estímulo, sino que también en una interpretación inconsciente en relación con experiencias previas. Sin este último aspecto nuestras habilidades perceptuales estarían limitadas puesto que la escucha implica tanto la audición como la cognición, de manera que el enfoque microscópico generalmente se centra en procesos auditivos mientras que el enfoque macroscópico también considera los factores cognitivos.

El autor explica que las estrategias de estudio de la audición y la cognición poseen distintas características, y que existe un riesgo de que estos dos aspectos de la escucha se estudien por separado, puesto que la percepción auditiva puede ser estudiada de manera psicofísica sin tener consideración alguna de los factores cognitivos. Sin embargo, advierte que la audición y la cognición no deben ser entendidos como diferentes estados de la escucha, de hecho, aclara que ambos procesos trabajan en paralelo, siendo que lo que escuchamos depende también de lo que esperamos escuchar (p. 7).

Posteriormente Plomp menciona que, si bien un investigador está cotidianamente expuesto a un mundo de los sonidos complejos, en el laboratorio sólo tolera los estímulos limpios, lo que provoca que se restrinja la vista del investigador, distorsiona las estimaciones acerca de los alcances del sistema auditivo y afecta la manera en la que los resultados pueden ser conceptualizados (pp. 7-8). Además, afirma que la mayor discrepancia entre las condiciones del laboratorio con respecto a la realidad acústica del mundo exterior se

evidencia en la exclusión del ruido, dado que el investigador trata de crear estímulos que sean lo más limpios posibles, es decir, libres de la interferencia del ruido. Dado esto, el autor asegura que la experiencia de escucha cotidiana es esencialmente eliminada del campo de estudio, y considera que aquellos estudios donde sonidos elementales sean utilizados deben ser complementados con estudios donde se investigue cómo un conjunto de sonidos es percibido (p. 9).

Capítulo 2: Método

En este capítulo expongo el método empleado para llevar a cabo el experimento que me permitió examinar la solidez de la premisa de Gibson en términos prácticos.

2.1 Diseño de la aplicación

Se presenta a continuación el esquema del diseño general de la aplicación desarrollada en Max para la generación de los estímulos audiovisuales utilizados en el experimento, las técnicas utilizadas para el análisis del contenido de audio, el esquema de generación de las visualizaciones, y la interfaz gráfica de usuario.

Figura 2. Esquema de diseño general de la aplicación.

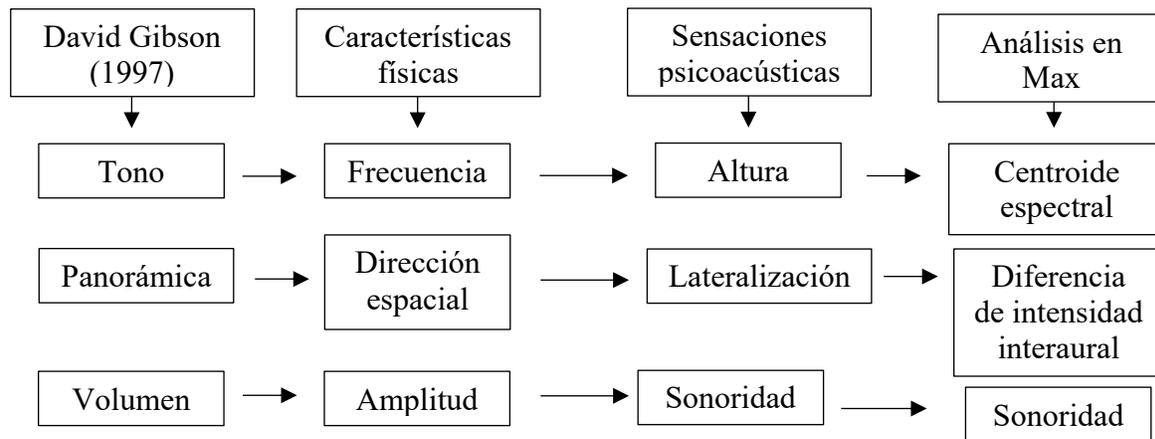


Figura 3. Propuesta para el análisis del contenido de audio en la aplicación.

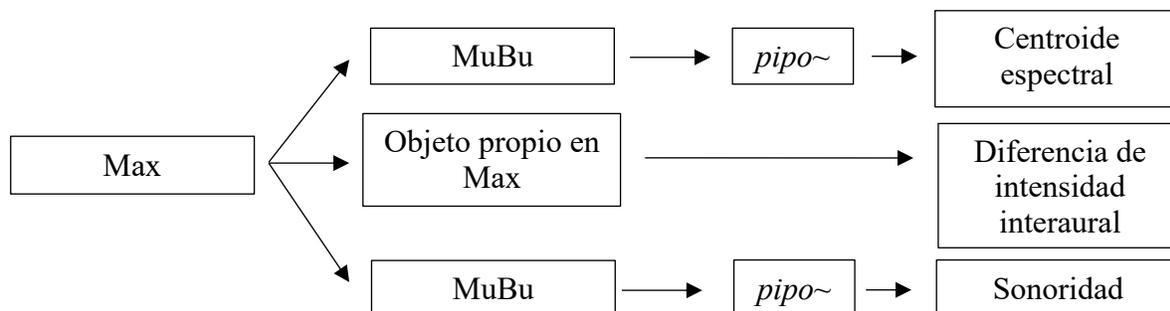
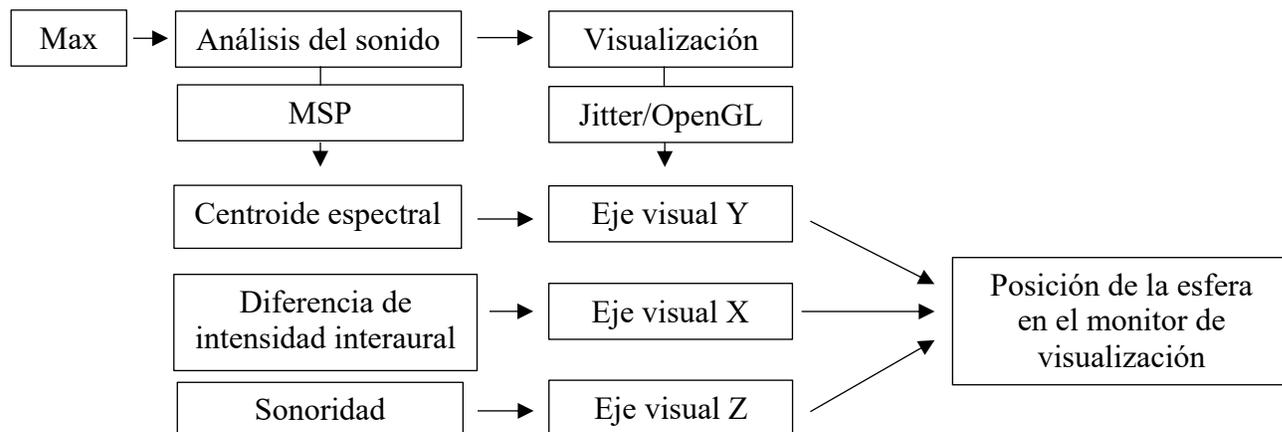


Figura 4. Esquema de generación de las visualizaciones de la aplicación.



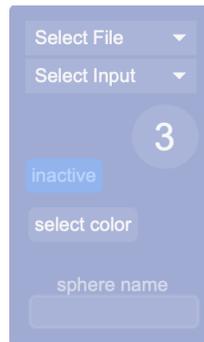
La sección externa fue diseñada para disponer de todos los controles de manera cómoda e intuitiva, cada audio a ser cargado en la aplicación se separa por *tracks*, los cuales son coloreados para identificarlos fácilmente.

Figura 5. Sección externa, vista general de la aplicación.



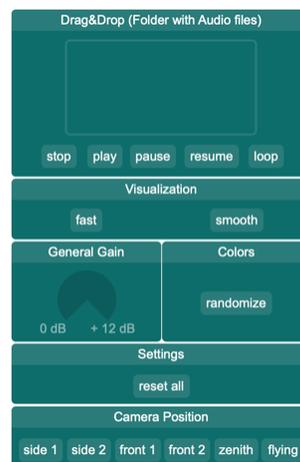
En la parte superior de cada *track* es posible elegir entre un archivo de audio cargado o una entrada de audio, además da la posibilidad de desactivar la monitorización de audio y la visualización de cada canal, así como también de elegir un color particular y poner una etiqueta de texto que acompañe la esfera.

Figura 6. Opciones de configuración para cada track de la aplicación.



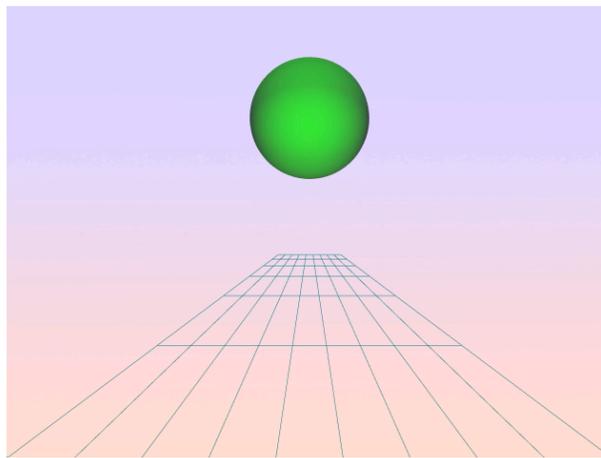
Posteriormente, en el sector derecho superior de la vista general de la aplicación se encuentra la sección de carga de la carpeta de archivos de audio mediante una función de *drag & drop* (arrastra y suelta), y las opciones de reproducción. Hacia abajo está la opción para elegir la velocidad de la visualización de las esferas, entre un modo rápido y uno suavizado. Después se encuentran las opciones de aumento de ganancia general para materiales que puedan estar demasiado bajos en amplitud, y también la posibilidad de elegir colores aleatorios para los distintos *tracks* a través de un botón. Continuando hacia abajo, se dispone de la opción para volver todas las configuraciones a su estado inicial y por último se permite variar las posiciones de la cámara.

Figura 7. Control y configuración general de la aplicación.



Para comenzar a utilizar la aplicación es necesario arrastrar una carpeta con archivos de audio en el recuadro blanco, esto carga automáticamente cada archivo presente en la carpeta en un *track* distinto, luego para comenzar a generar las visualizaciones sólo es necesario presionar el botón *play*. Por otro lado, en el monitor de visualización se observa una malla formada por varias líneas en los ejes visuales X y Z que indican la superficie sobre la cual se desplazarán las esferas, que delinea un punto de fuga visual hacia el centro. El área donde se mueven las esferas está determinada por la malla y los límites visuales del monitor.

Figura 8. Monitor de visualización de la aplicación.



2.2 Diseño del experimento

El experimento estuvo destinado a una población conformada por jóvenes estudiantes afines a la música y la tecnología dado que poseen relación con la música y el sonido en cuanto a conocimientos y experiencias, además de la predisposición a la percepción de estímulos sonoros y audiovisuales, aspectos cognitivos relevantes para la presente investigación. (ver 1.9 Una mirada holística). Otro punto destacable es que la obra de Gibson suele utilizarse en

la enseñanza de carreteras relacionadas a la música y la tecnología, de modo que la población forma parte del público al que va dirigida su obra donde surge la premisa que se examina.

Considerando lo anterior se obtuvo una muestra de 40 estudiantes de la Licenciatura en Música y Tecnología Artística de la Escuela Nacional de Estudios Superiores, UNAM, unidad Morelia pertenecientes al segundo, cuarto, sexto y octavo semestre. Para participar del experimento le solicité a cada persona su consentimiento a través de un formulario, donde se explica de qué trata el experimento, a qué están expuestos y cuáles son las condiciones generales. Luego cada participante tuvo que completar un cuestionario para estimar su grado de experiencia y conocimiento acerca del tema (básico de 0 a 4 puntos, intermedio de 5 a 10 puntos y avanzado de 11 a 15 puntos) sin develarles el resultado, posteriormente se anotaron sus direcciones de e-mail en una planilla obteniéndolas previamente del propio cuestionario, designándolos según el orden de llegada y según la clasificación obtenida, a los grupos A (de control) y B (experimental), equilibrando en cada grupo el puntaje de conocimiento y experiencia previa de los participantes que formaron parte. Al grupo de control se le presentaron los estímulos sonoros, mientras que al grupo experimental los estímulos audiovisuales creados a partir de los anteriores.

Como he dicho en el apartado 1.8 Procedimientos para la investigación de la localización del sonido, elegí el procedimiento psicoacústico de la categorización mediante el cual se presentan una serie de estímulos a través de altavoces y luego se realizan preguntas respecto a la ubicación aparente de los sonidos, dando una cantidad de posibles respuestas a elegir. Dichas preguntas las realicé a través de un cuestionario diseñado en la plataforma Google Forms, donde primero se presentan los sonidos y/o visualizaciones, para luego dar una serie de opciones a responder. Los audios y los videos que sirvieron como estímulo los subí a mi canal de YouTube en modo oculto y se insertaron en el cuestionario mediante sus

enlaces correspondientes para facilitar su reproducción. Para acceder a la carta de consentimiento informado, al cuestionario de conocimientos y los cuestionarios de control y experimental, diseñé una página web (<https://sites.google.com/view/3dstereo/>), desarrollada en la plataforma Google Sites.

2.3 Estímulos

Los estímulos sonoros que se diseñaron para el experimento consisten en un total de seis sonidos estereofónicos de los cuales tres de ellos son sinusoides y los tres restantes se tratan de ondas complejas sintetizadas, todos ellos con una duración aproximada de 10 segundos. Cada uno de los primeros tres estímulos poseen variaciones particulares en tono, volumen y panorámica, mientras que los siguientes tres sonidos tienen claramente marcada una variación de alguna de estas características, pero incluyen leves variaciones de las restantes. Se detalla a continuación la construcción de cada estímulo sonoro:

- Estímulo 01: senoide de 1 kHz, desde - 96 dBFS hasta - 6 dBFS, paneado al centro.
- Estímulo 02: senoide de 1 kHz, a - 6 dBFS de amplitud, desde el canal izquierdo hasta el canal derecho.
- Estímulo 03: senoide con barrido de frecuencia desde 63 Hz hasta 4,000 Hz, a - 6 dBFS de amplitud, paneado al centro.
- Estímulo 04: sonido complejo conformado por: tres osciladores (diente de sierra ascendente y escalera ascendente sumadas a una onda cuadrada a una octava por debajo, mezclados con la misma intensidad), con modulaciones de afinación fina en el segundo oscilador y en la frecuencia de corte del filtro pasa altos mezclados a la misma intensidad, envolvente con ataque y *release* rápidos (manteniéndose en el *sustain* a - 6dBFS), desde el canal derecho hasta el canal izquierdo.

- Estímulo 05: sonido complejo formado por: tres osciladores (onda triangular y diente de sierra ascendente en modulación cruzada sumadas a una senoide a una octava por debajo, con mayor presencia de la senoide) y un generador de ruido, con modulaciones en la frecuencia de corte del filtro y en la amplitud del generador, envolvente aplicada a la amplitud del primer oscilador y a la modulación cruzada, envolvente del amplificador con ataque rápido y *release* lento (provocando un barrido en amplitud de - 6 dBFS a - 53 dBFS), paneado al centro.
- Estímulo 06: sonido complejo formado por dos osciladores (onda triangular y diente de sierra ascendente con modulación cruzada, con mayor presencia del segundo oscilador) y un generador de ruido, con modulaciones en la amplitud del primer oscilador y en la amplitud del generador, envolvente aplicada a la amplitud del primer oscilador, a la transposición de los dos osciladores (provocando un barrido de 3 kHz a 260 Hz) y a la modulación cruzada, envolvente del amplificador con ataque rápido y *release* lento, manteniéndose a -6 dBFS y paneado al centro.

Para la generación de los estímulos sonoros se utilizaron dos procedimientos diferentes, en primer lugar, para la creación de ondas sinusoidales se utilizó el *software* SuperCollider (de código abierto) a través de la programación de instrumentos e instrucciones simples, una vez exportados los sonidos desde SuperCollider en un único archivo procedí a cortarlos en Pro Tools 12.7. En segundo lugar, para el diseño de sonidos complejos utilicé el sintetizador virtual Helm (de código abierto) y la estación de trabajo de audio digital Pro Tools 12.7.

A partir de los estímulos sonoros generé las visualizaciones destinadas al grupo experimental, utilizando la aplicación desarrollada en Max para la presente investigación, obteniendo los estímulos audiovisuales mediante la grabación de video del monitor visual de Max empleando la herramienta de captura de pantalla nativa del sistema operativo *macOs*

Mojave de Apple. Al mismo tiempo y con la misma herramienta registré el sonido configurando la salida de audio de Max por un *driver* virtual llamado SoundFlower (de código abierto), de manera de no perder la sincronización de ambas partes. A su vez, los estímulos sonoros empleados para el grupo de control debieron convertirse a video y ser alojados en Youtube para obtener una ventana de reproducción en el cuestionario de Google Forms (sin necesidad de abrir una nueva pestaña en el navegador), esta adaptación a video de los estímulos sonoros la realicé mediante el software de edición de video Shotcut (de código abierto) generando un video con fondo negro para cada uno de éstos.

Por último, mediante la utilización de otro cuestionario en la plataforma Google Forms (<https://forms.gle/L6Ze1nmLgENkR9dj7>), solicité a dos colegas con conocimientos teóricos y prácticos en cuestiones técnicas, musicales, sonoras y audiovisuales, que respondieran una serie de preguntas para validar la consistencia y claridad de los estímulos, y de esa manera afianzar la rigurosidad del experimento. En dicho cuestionario se presentan los estímulos audiovisuales y para cada uno de ellos se les pregunta acerca de la claridad y consistencia tanto sonora como visual, y finalmente se les solicita un comentario general. El resultado del cuestionario de validación arrojó que todos los estímulos con excepción del primero resultan claros y adecuados para el experimento, este estímulo se trata de una senoide que comienza en una amplitud baja y llega hasta una amplitud máxima, que debería dar la percepción de que el sonido realiza una trayectoria de atrás hacia el frente. Uno de los consejos recibidos para solucionar este problema fue el utilizar un efecto de reverberación para aumentar la sensación de distancia, aunque no consideré que fuese adecuado, ya que escapa a los supuestos de la premisa de Gibson y puede resultar tendencioso para el experimento.

2.4 Cuestionarios

A continuación presento los cuestionarios que se utilizaron para el experimento, en primer lugar, el cuestionario que sirvió para determinar el grado de experiencias y conocimientos de los participantes acerca del tema, y en segundo lugar los cuestionarios empleados para el grupo de control y el grupo experimental.

En el cuestionario de conocimientos se le solicita a la persona que indique su nombre y su edad, y posteriormente se realizan cinco preguntas:

- 1- ¿Utilizaste alguna vez una estación de trabajo de audio digital?
- 2- ¿Realizaste alguna vez una mezcla estereofónica?
- 3- ¿Conoces acerca del concepto de localización, en el marco de la psicoacústica?
- 4- ¿Generaste alguna vez visualizaciones a partir del análisis sonoro?
- 5- ¿Encuentras familiar la idea de que en la estereofonía existe una tridimensionalidad sonora?

Con respecto a los cuestionarios para los grupos de control y experimental, en las primeras seis preguntas hay una respuesta correcta para cada caso, en consecuencia, esto provoca una puntuación por cada participante donde el máximo a conseguir son seis puntos. Para cada estímulo se dan las siguientes indicaciones: luego de reproducir cada sonido, basándote en tu percepción de la localización sonora, elige la respuesta que más se adecúe a la trayectoria y ubicación aparente de los sonidos. Y se otorgan las siguientes opciones de respuesta:

- a) El sonido realiza un trayecto de atrás hacia adelante.
- b) El sonido realiza un trayecto de adelante hacia atrás.
- c) El sonido realiza un trayecto de izquierda a derecha.
- d) El sonido realiza un trayecto de derecha a izquierda.
- e) El sonido realiza un trayecto de abajo hacia arriba.

- f) El sonido realiza un trayecto de arriba hacia abajo.
- g) Ninguna de las anteriores.
- h) No lo sé.

En cuanto al cuestionario realizado al grupo experimental los estímulos fueron exactamente los mismos, sólo que se presentaban en su versión audiovisual generada por la aplicación desarrollada para la presente investigación, las opciones de respuesta también son las mismas, y los enunciados cambian levemente.

Para poner a prueba los cuestionarios solicité la participación a compañeras y compañeros de la asignatura de Seminario de Investigación II, y a un grupo pequeño de estudiantes de la carrera de Música y Tecnología Artística de la ENES, unidad Morelia, UNAM, distribuyéndose azarosamente al grupo de control y al experimental, con un total de 12 personas participantes. Cabe destacar que para esta prueba hubo algunos detalles operativos del experimento real que no fueron aplicados, por un lado, no se realizó un equilibrio de los grupos (control y experimental) por niveles de conocimiento o experiencia de los participantes, y por otro lado tampoco se controló la instalación desde la cuál recibieron los estímulos sonoros y visuales, sino que cada quien utilizó el sistema de reproducción que tenía a su disposición.

Los resultados de la puesta a prueba de los cuestionarios son los siguientes (recordando que el grupo de control solo recibió estímulos sonoros y el grupo experimental estímulos audiovisuales):

- El promedio de aciertos o coincidencias con las respuestas marcadas como correctas para cada pregunta, el grupo de control (denominado grupo A) recibió un puntaje de 4,67 de un total de 6 puntos, mientras que el grupo experimental (denominado grupo B) obtuvo una puntuación de 4,33 sobre un total de 6 puntos.

- En ambos grupos el valor medio de las puntuaciones es de 4 puntos.
- En ambos grupos solo dos participantes obtuvieron la puntuación ideal.
- El quinto estímulo (un sonido complejo que disminuye su volumen hasta una amplitud mínima), en el grupo B, fue el que obtuvo mayores desaciertos.

2.5 El experimento

Para llevar a cabo el experimento se utilizó un cubículo de la Escuela Nacional de Estudios Superiores, UNAM, Unidad Morelia, dado que se trataba de un punto clave para asegurar una buena cantidad de participantes, es decir, una muestra que resulte suficiente. Dicho cubículo disponía de una mesa sobre la cual se encontraba una pantalla ViewSonic modelo VX2270Smh-LED de 21.5”, dos altavoces a los lados de la pantalla de la marca Genelec 8330A de 5” para el *woofer* y un tamaño de 3/4” para el *tweeter*, una interfaz de audio Focusrite Scarlett 18i8, y una computadora Apple Macbook Pro 11.3 que se utilizó para el llenado de los cuestionarios y contaba con conexión a internet.

Tomé consideraciones acerca de la posición de los altavoces y de la silla donde se ubicaban los participantes adoptando cierta la angulación y distancia hacia el oyente, según lo mencionado acerca de las características de la estereofonía en el apartado 1.3 Estereofonía, y también ajusté el nivel de los altavoces de manera de no generar presiones sonoras que puedan ser molestas para los participantes. Además, realicé marcas con cinta adhesiva de papel para asegurar la misma posición de los equipos en cada jornada del experimento.

Al momento de iniciar con la experiencia le daba las indicaciones generales a cada participante, y de ser necesario también los asistía para el correcto llenado de los cuestionarios. Para cada paso del experimento los participantes tomaban el control de la selección de cada respuesta y de la reproducción de los estímulos.

El primer día del experimento hubo un total de once participantes, el segundo día solo un participante, el tercer día un total de doce participantes, el cuarto día un total de cinco participantes, el quinto día un total de siete participantes y el sexto y último día cuatro participantes, para dar un total de 40 participantes.

Con la finalidad de promover la participación en el experimento ofrecí como retribución una *masterclass* llamada *Puntos técnicos clave en el mastering digital*, dictada en la Escuela Nacional de Estudios Superiores, Unidad Morelia, UNAM, con una duración total de dos horas y una afluencia de la mayoría de las personas que formaron parte del experimento.

Capítulo 3: Resultados

Todos los datos obtenidos de los cuestionarios quedaron alojados en la plataforma de Google Forms, posteriormente para extraer la información de ésta utilicé la opción disponible para exportar los datos en un archivo de Microsoft Excel, lo que me permitió tener mayor flexibilidad en el manejo de los datos y la creación de representaciones gráficas de éstos.

En primer lugar, es posible observar que, en las puntuaciones obtenidas en cuanto al éxito en las respuestas correctas para cada estímulo, hay una tendencia a mejorar en las personas que recibieron los estímulos audiovisuales, es decir para el grupo B (experimental).

Figura 9. Puntuaciones obtenidas para el grupo A (sólo estímulos sonoros).

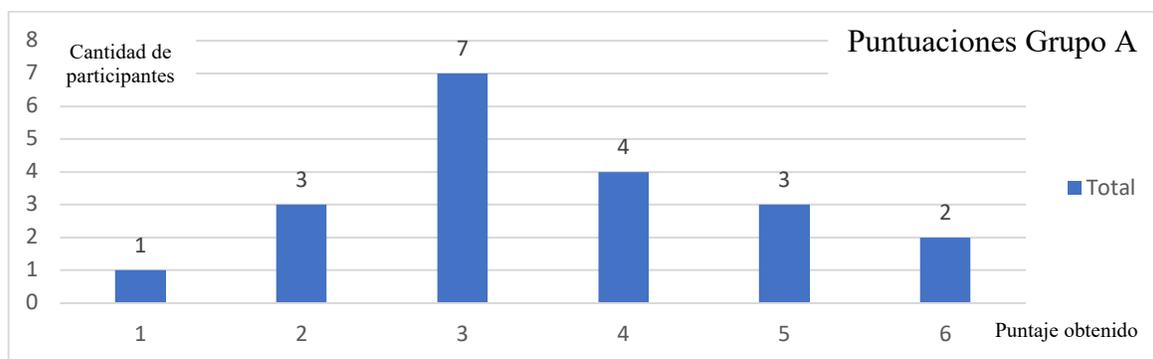
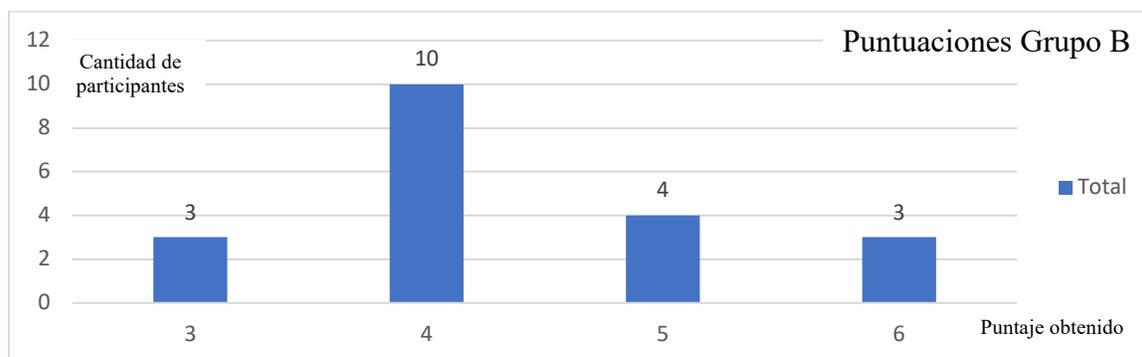


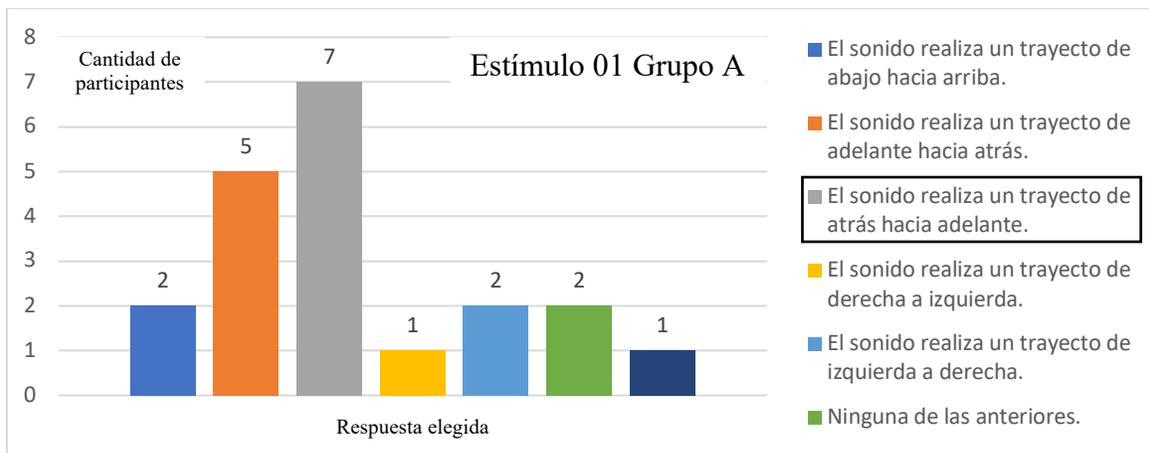
Figura 10. Puntuaciones obtenidas para el grupo B (estímulos audiovisuales).



Nota: sólo se muestran los puntajes obtenidos.

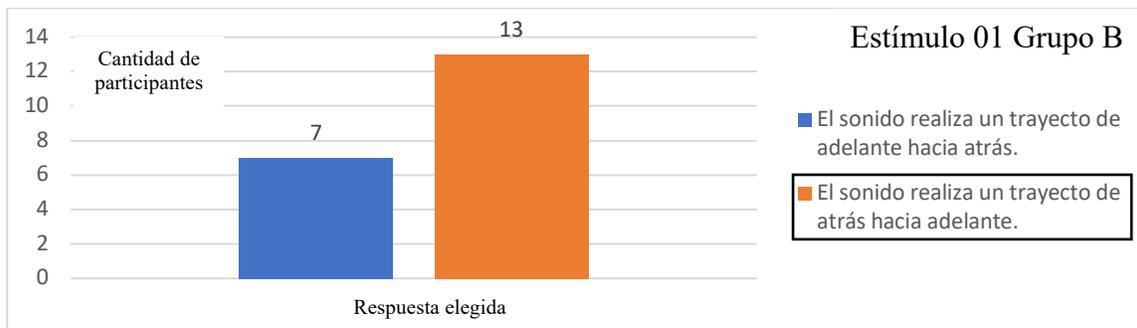
En el primer estímulo la opción correcta es: el sonido realiza un trayecto de atrás hacia adelante. Los resultados del grupo A se encuentra más dispersos, aunque la mayoría elige la respuesta correcta, el grupo B en cambio provoca menos variación en las elecciones, aunque parece no quedar del todo claro dado que se lo confunde con la trayectoria opuesta, es decir de adelante hacia atrás.

Figura 11. Respuestas del primer estímulo, para el grupo A (sólo sonido).



Nota: se marca con un recuadro negro la respuesta correcta, y de todas las opciones sólo se muestran las respuestas elegidas.

Figura 12. Respuestas del primer estímulo, para el grupo B (audiovisual).



Para el segundo estímulo la opción correcta es: el sonido realiza un trayecto de izquierda a derecha. En este caso ambos grupos lograron muy buenos resultados, también es posible apreciar que la cantidad de respuestas erróneas se reduce con la presencia de las visualizaciones.

Figura 13. Respuestas del segundo estímulo, para el grupo A (sólo sonido).

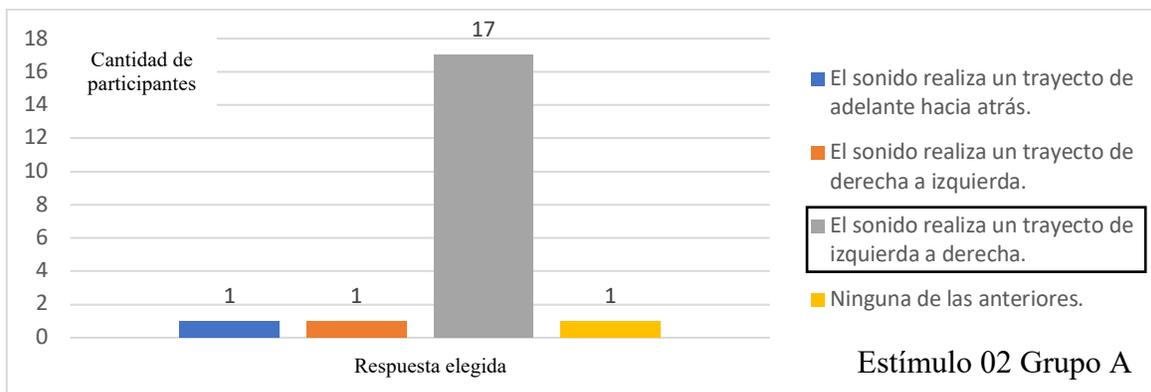
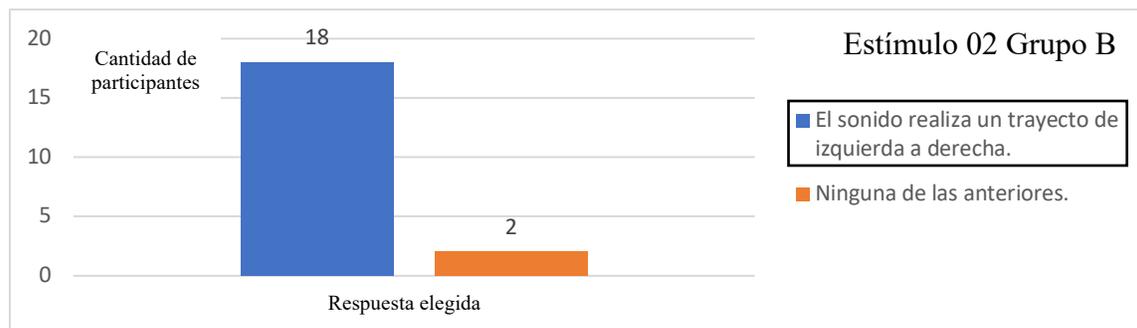


Figura 14. Respuestas del segundo estímulo, para el grupo B (audiovisual).



En cuanto al tercer estímulo, donde la opción correcta es: el sonido realiza un trayecto de abajo hacia arriba. Aquí sucede lo mismo que en el estímulo anterior, puesto que ambos grupos logran un buen desempeño y existe un mayor acierto en presencia de las visualizaciones donde se reduce el margen de error.

Figura 15. Respuestas del tercer estímulo, para el grupo A (sólo sonido).

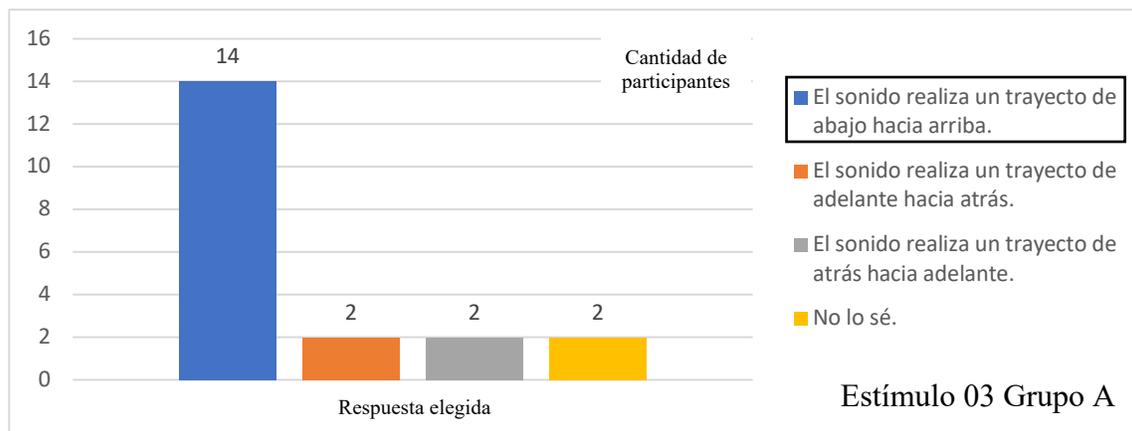
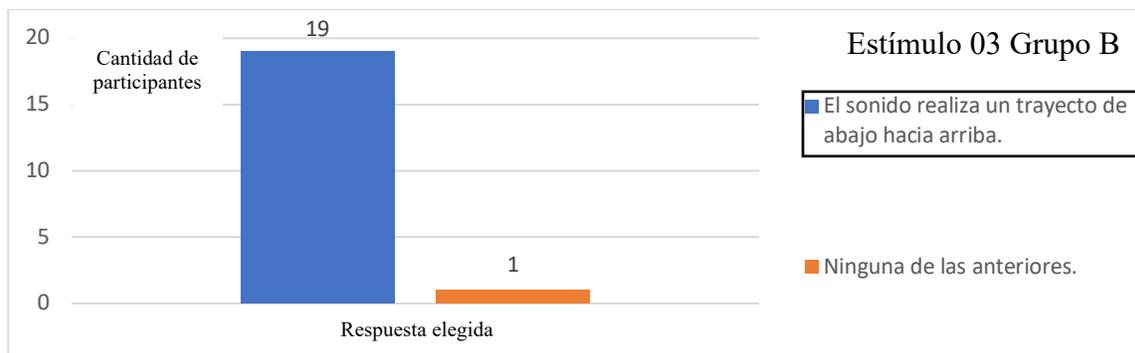


Figura 16. Respuestas del tercer estímulo, para el grupo B (audiovisual).



El cuarto estímulo donde la opción correcta es: el sonido realiza un trayecto de derecha a izquierda. En este caso se evidencia un porcentaje de acierto casi perfecto para ambos grupos.

Figura 17. Respuestas del cuarto estímulo, para el grupo A (sólo sonido).

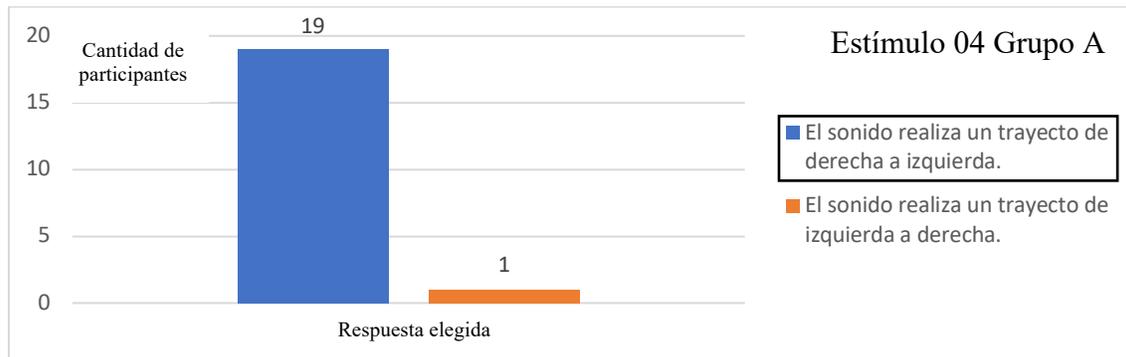
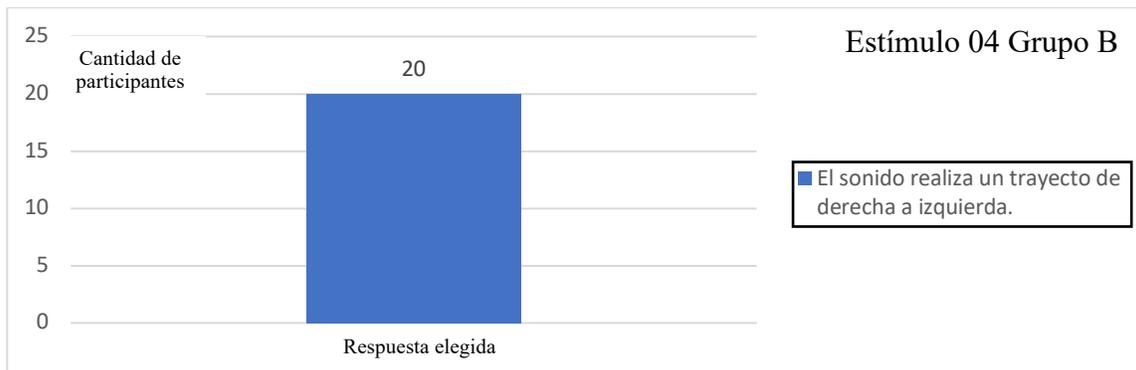


Figura 18. Respuestas del cuarto estímulo, para el grupo B (audiovisual).



Para el quinto estímulo la opción correcta es: el sonido realiza un trayecto de adelante hacia atrás. En este caso, la cantidad de desaciertos y el margen de error de las respuestas incorrectas ha aumentado considerablemente para ambos grupos, sucediendo algo similar al primer estímulo que realiza un trayecto en el mismo eje, donde el sonido realiza un trayecto de atrás hacia adelante.

Figura 19. Respuestas del quinto estímulo, para el grupo A (sólo sonido).

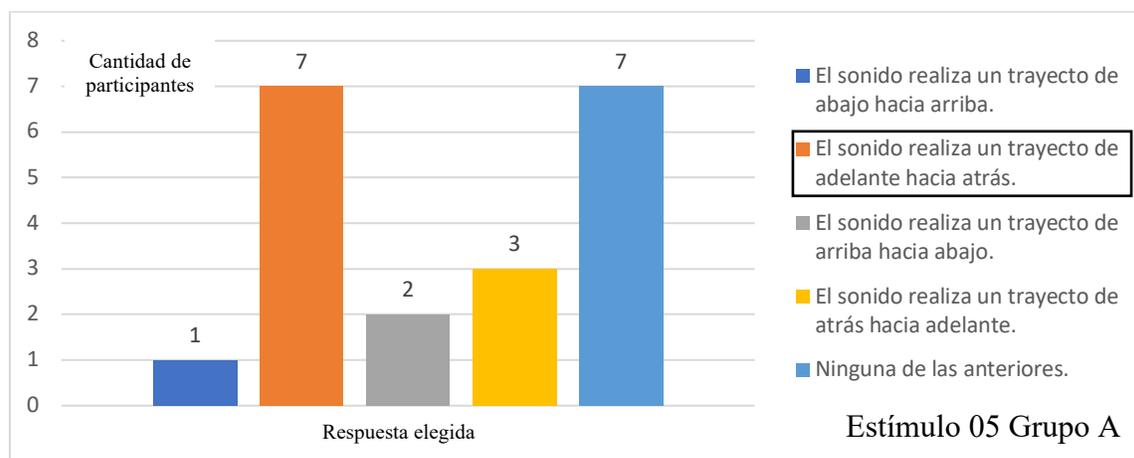
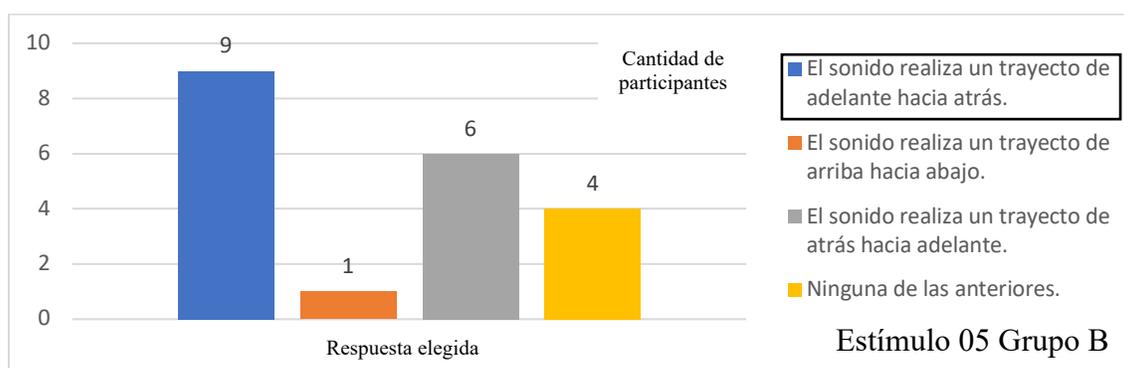


Figura 20. Respuestas del quinto estímulo, para el grupo B (audiovisual).



El sexto estímulo tiene como opción correcta: el sonido realiza un trayecto de arriba hacia abajo. Este es otro caso peculiar, donde también se dan una gran cantidad de desaciertos en ambos grupos, como siempre con mayor dispersión de las respuestas en el grupo A. Esto en contraste con el tercer estímulo que se desarrolla en el mismo eje, donde el sonido realiza un trayecto de abajo hacia arriba, para el cual no hubo mayor confusión.

Figura 21. Respuestas del sexto estímulo, para el grupo A (sólo sonido).

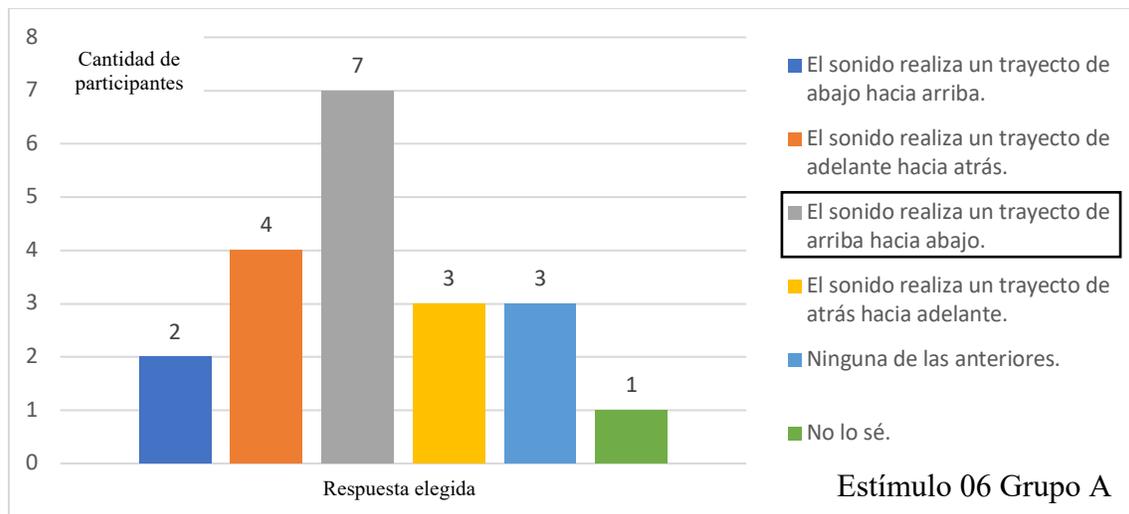
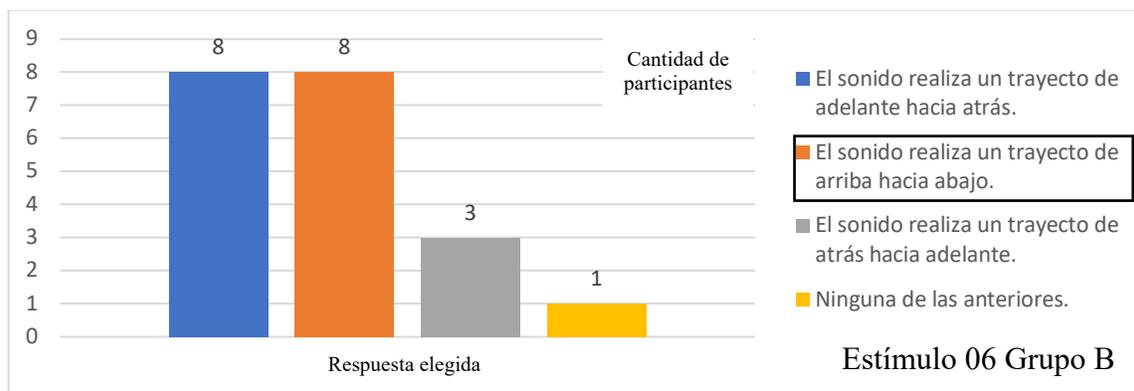


Figura 22. Respuestas del sexto estímulo, para el grupo B (audiovisual).



Finalmente, para la última pregunta del cuestionario donde se requerían observaciones generales acerca del experimento las respuestas obtenidas fueron muy variadas.

En el grupo A en general opinaron que la experiencia les resultaba interesante, como menciona por ejemplo el quinto participante (en orden cronológico): “Me resulta muy interesante el análisis de diferentes parámetros sonoros para experimentar y conocer más de la estereofonía y la tridimensionalidad. En algunos audios, me fue más difícil identificar la dirección pero en general experimenté estas sensaciones de movimiento”. Otros participantes dicen que el experimento permite percibir y comprender mejor el aspecto tridimensional de la estereofónica, que a primera escucha es difícil percibirlo, y que puede haber sesgos en la percepción dada la formación y conocimientos previos, como dice el treceavo participante: “creo que algunas de mis respuestas están sesgadas por mi formación como músico. específicamente en la idea de relacionar altura con subidas y bajadas”. Por otro lado, en lo particular opinaron que es más clara la trayectoria cuando el timbre es más complejo, que se podrían incluir más opciones de trayectoria de los sonidos como opciones de respuesta, y que es fácil predisponerse e imaginar otras trayectorias, como opina el sexto participante: “Me parece que es fácil imaginar distintas proyecciones de sonido y predisponerse. Por ejemplo, si imaginas una trayectoria de sonido de abajo hacia arriba, puedes escuchar al sonido moviéndose en esa dirección. La única dirección que no se puede alterar voluntariamente es la de izquierda/derecha”.

En el grupo B en general también opinaron que la experiencia les resultaba interesante, que gracias a los estímulos visuales fue más fácil percibir la tridimensionalidad planteada y resulta más clara, por ejemplo, como observa el séptimo participante: “Me parece muy interesante como da la perspectiva de tridimensionalidad aun con solo dos bocinas, y siento que la visualización también juega un papel”. Además, opinan que es llamativo como

las variaciones de altura también influyen en la percepción de la localización, como observa el catorceavo participante: “Es muy interesante la manera en la que el software es capaz de interpretar la espacialidad del sonido, y cómo las variaciones de altura y timbre igualmente influyen en la percepción de la localización y como esto es interpretado visualmente de manera automática”. En términos particulares opinaron que la presencia de diferentes timbres fue acertada para el experimento, y que a pesar que algunos sonidos parecían tener más de una trayectoria siempre había un desplazamiento principal, como dice el décimo participante: “Precisamente la variedad de timbres me pareció una decisión acertada para los fines del experimento. Hubo algunos sonidos en donde se podían percibir más de un movimiento de este mismo, pero no considero que haya representado un problema ya que había siempre alguno que predominaba”. Así como también que, si bien la percepción de la lateralización resulta evidente, la percepción de la sonoridad fue mucho más subjetiva, y que normalmente uno no está consciente de la percepción de la estereofonía en cuestiones de tridimensionalidad.

Tabla 1. Resumen de los aciertos y desaciertos de los cuestionarios.

Estímulo	Grupo A		Grupo B		Total	
	Aciertos	Desaciertos	Aciertos	Desaciertos	Aciertos	Desaciertos
1	7	13	13	7	20	20
2	17	3	18	2	35	5
3	14	6	19	1	33	7
4	19	1	20	0	39	1
5	7	13	9	11	16	24
6	7	13	8	12	15	25

Nota: se marca con color verde los estímulos que tuvieron mayor cantidad de aciertos y con color rojo los estímulos que tuvieron una menor cantidad de aciertos.

Figura 23. Cantidad de aciertos totales de los cuestionarios.

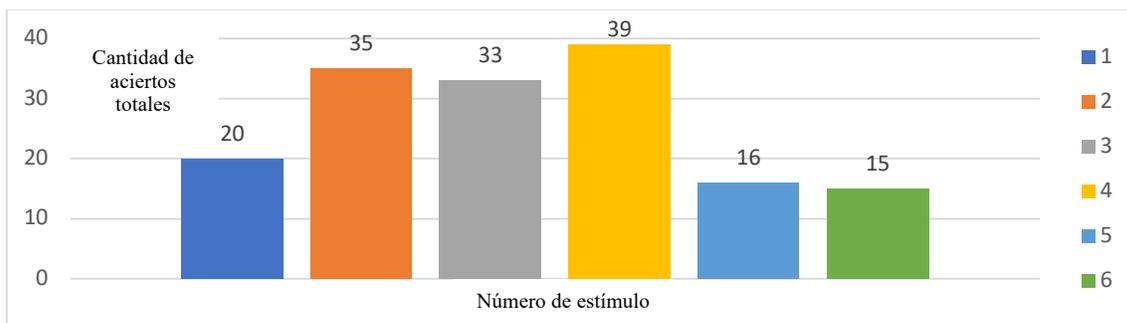


Figura 24. Cantidad de desaciertos totales de los cuestionarios.

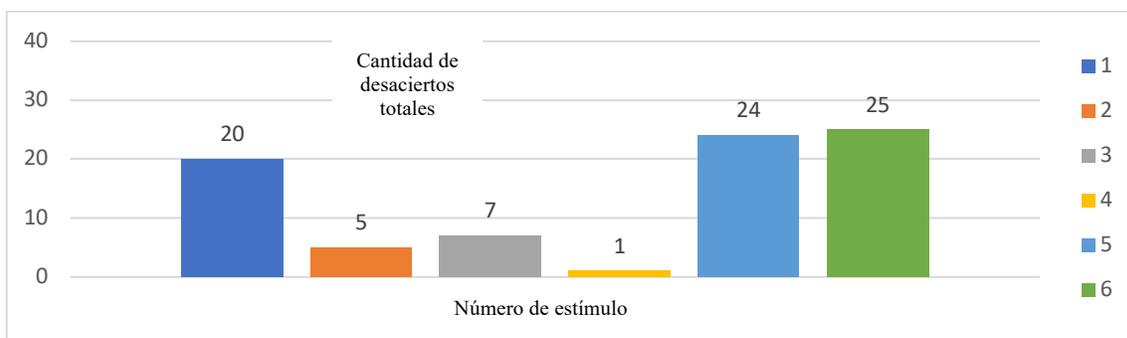
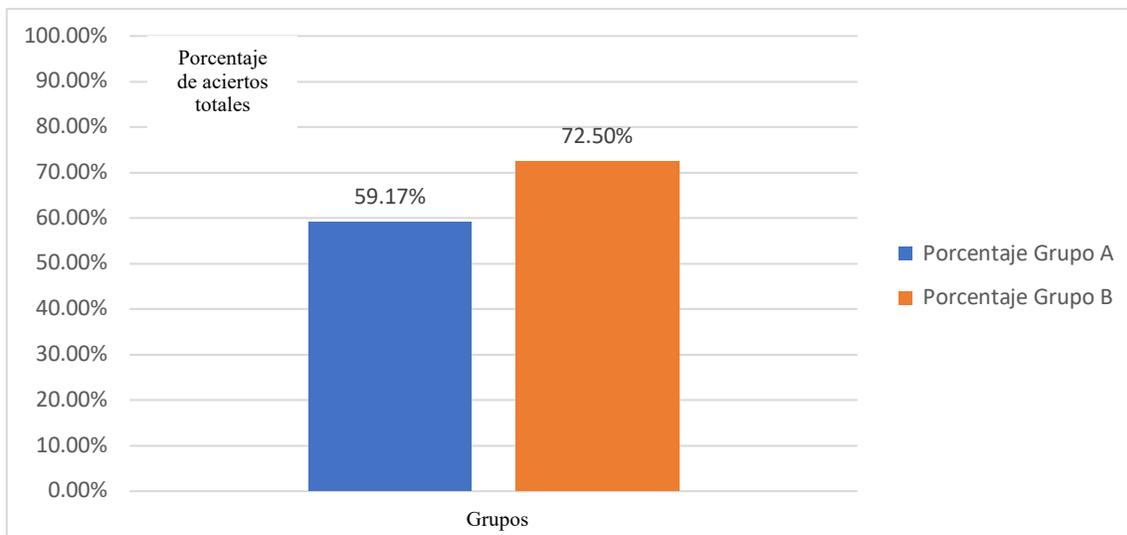


Figura 25. Porcentaje del promedio de aciertos para cada grupo.



Capítulo 4: Conclusiones

Si bien desde los orígenes de la estereofonía se hace referencia a cuestiones de relieve, localización espacial y efecto espacial, y se habla en términos de tridimensionalidad, no todos los autores han estado de acuerdo con esto dado que algunos de ellos argumentan que la estereofonía es un fenómeno bidimensional y otros aseguran que se necesitan de otras tecnologías para lograr una cierta tridimensionalidad sonora. La premisa de Gibson resulta en una convención, una visión arbitraria de la percepción auditiva de la estereofonía, aunque no ha sido el único que ha tratado de definir un espacio, escena o imagen sonora como formalización de un espacio acústico. En cuanto a la asociación entre el volumen de los sonidos y la sensación de distancia Gibson no es lo suficientemente riguroso dado que deja de lado otras características claves del sonido que forman parte de dicha sensación, tales como el contenido espectral del sonido y la reverberación. Luego, la relación que establece Gibson del tono de los sonidos con la sensación de elevación resulta ineficiente, aunque es más acertada que en el caso del volumen, ya que dicha sensación está dada por variaciones en el contenido espectral pero no existe una relación determinadamente estricta con el tono de los sonidos. Por último, la correspondencia entre panorámica y lateralización propuesta por Gibson es un poco más apropiada que las establecidas para el volumen y el tono, puesto que la mayor parte de la sensación de dirección espacial proviene de las diferencias de intensidad interaurales.

En el caso de la muestra obtenida para el experimento, la premisa propuesta por Gibson no parece ser tan evidente dado que se percibe con mayor dificultad para algunas de sus partes componentes, y con mayor facilidad para otras (variaciones de volumen, tono y panorámica respectivamente). Además, es notorio que los resultados mejoran con la

visualización de los estímulos. La percepción de la sonoridad y su asociación con una trayectoria de atrás hacia adelante y viceversa, en el contexto estereofónico, parece no estar lo suficientemente clara ni para el caso del grupo que experimentó los estímulos sonoros ni para el grupo que tuvo la experiencia visualizada, tanto para sonidos simples como para sonidos complejos. La percepción de la altura resulta en un caso intermedio, siendo que para el tercer estímulo donde se presenta una senoide ascendente la mayoría de los participantes de ambos grupos estuvieron de acuerdo en su percepción, sin embargo para el sexto estímulo que se trata de un sonido complejo que realiza un barrido descendente, las respuestas se dispersaron en ambos grupos, y el margen de error fue alto, de manera que, la percepción de la altura y su asociación una trayectoria en el eje vertical resulta en un caso intermedio. Finalmente, la percepción de la lateralización de los sonidos, por el contrario, fue el aspecto de la premisa de Gibson donde menos dudas hubo, siendo que la asociación de la lateralización con una trayectoria de izquierda a derecha y viceversa, en ambos grupos parece estar bastante clara y ser lo suficientemente evidente.

En términos generales puedo afirmar que la premisa de Gibson posee debilidades tanto teóricas como prácticas, siendo que su solidez está mayormente comprometida por la asociación del volumen de los sonidos con la sensación de distancia. De manera que, si se quisiera continuar con el desarrollo de la premisa se podría generar una nueva propuesta, tomando las consideraciones presentadas en esta investigación, y así acercarse a un modelo más sólido. Un siguiente paso podría ser la creación de un *plugin* contemplando dichas consideraciones para ser utilizado por la comunidad estudiantil en el DAW de su preferencia a la hora de trabajar en sus proyectos estereofónicos.

Bibliografía

- Babbie, E. (2000). *Fundamentos de la investigación social*. International Thomson Editores.
- Bernal, C. (2012). *Metodología de la investigación: administración, economía, humanidades y ciencias sociales*. Pearson.
- Blackmagicdesign. (6 de junio de 2023). Fairlight. Davinci Resolve 18. <https://www.blackmagicdesign.com/mx/products/davinciresolve/fairlight>
- Blauert, J. (2013). *The Technology of Binaural Listening. Modern Acoustic and Signal Processing*. Springer
- Butler, R. A. (1973). *The relative influence of pitch and timbre on the apparent location of sound in the median sagittal plane*. Perception & Psychophysics. Vol. 14 (2), 255-258.
- Cha, S. et al. (2019). *Handbook of Software Engineering*. Springer Link.
- Cohen, L. et al. (2018). *Research Methods in Education*. Routledge.
- Colasanto, F. (2010). *Max/MSP: guía de programación para artistas*. CMMAS.
- Cooper, M. et al. (2006). *Visualization in Audio-Based Music Information Retrieval*. Computer Music Journal. Volumen 30 (2). pp. 42- 62. <https://www.jstor.org/stable/3682003>
- Cycling '74. (7 de agosto de 2022). *Working with OpenGL. Max v8.3.1 Documentation*. https://docs.cycling74.com/max8/vignettes/working_with_opengl_topic
- Everest, F. A., Pohlmann, K. C. (2009). *Master Handbook of Acoustics*. McGraw Hill
- Fastl, H., Zwicker, E. (2007). *Psychoacoustics Facts and Models*. Springer.
- Fay, R., Popper A. (1993). *Springer Handbook of Auditory Research*. Springer.
- Fletcher, S. H. (1992). *Harvey Fletcher 1884 - 1981: A Biographical Memoir by Stephen H. Fletcher*. National Academy of Sciences.
- Gibson, D. (1997). *The Art Of Mixing: A Visual Guide to Recording, Engineering, and Production*. MixBooks.

Harrison, J. (3 de junio de 2023). *Imaginary space: Spaces in the imagination*. Australasian Computer Music Conference Keynote Address.

https://econtact.ca/3_2/ACMConference.htm

Huber, D., Runstein R. (2017). *Modern Recording Techniques*. Routledge.

INA. (6 de junio de 2023). *Spaces*. GRM. <https://inagrm.com/en/store/product/15/spaces>

IRCAM Centre Pompidou. (20 de noviembre de 2022). *PIPO: Plug-In-Plug-Out/Programming Interface for (Afferent Stream) Processing Objets*. <http://ismm.ircam.fr/pipo/>

IRCAM Forum. (20 de noviembre de 2022). *MuBu: Toolbox for multimodal analysis of sound and motion, interactive sound synthesis and machine learning*. <https://forum.ircam.fr/projects/detail/mubu/>

IRCAM. (6 de agosto de 2022). *MuBu for Max: A toolbox for Multimodal Analysis of Sound and Motion, Interactive Sound Synthesis and Machine Learning*. Sound Music Movement Interaction. <https://ismm.ircam.fr/mubu/>

IRCAM. (6 de junio de 2023) *Spat*. IRCAM Forum. <https://forum.ircam.fr/projects/detail/spat/>

Kendall, G. S. (1995). *A 3-D Sound Primer: Directional Hearing and Stereo Reproduction*. The MIT Press.

Khronos Group (17 de junio de 2023). OpenGL: The Industry's Foundation for High Performance Graphics. <https://www.khronos.org/opengl/>

Khulusi, R. et al. (2020). *A Survey on Visualizations for Musical Data*. COMPUTER GRAPHICS forum. Volume 39 (6). 82–110. DOI: 10.1111/cgf.13905

- Lamere, P., y Eck, D. (2007). Using 3D visualizations to explore and discover music. *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007*, pp. 173–174.
- Lemström, K., y Tzanetakis, G. (2011). Music information retrieval. *Understanding Information Retrieval Systems: Management, Types, and Standards*, pp. 409–418. <https://doi.org/10.4018/978-1-60566-010-3.ch216>
- Lerch, A. (2012). *An Introduction to Audio Content Analysis*. IEEE Press
- Lyon, R. F. (2017). *Human and Machine Hearing: Extracting Meaning from Sound*. Cambridge University Press
- McCarthy, B. (2007). *Sound Systems: Design and Optimization*. Focal Press.
- Miyara, F. (2000). *Acústica y Sistemas de Sonido*. UNR Editora.
- Moore, A. (2016). *Sonic Art: An Introduction to Electroacoustic Music Composition*. Routledge
- Pack, B. (4 de agosto de 2020). *Using Your Eyes While Mixing and Mastering*. *Sonarworks Blog*. <https://www.sonarworks.com/blog/uncategorised/using-your-eyes-while-mixing-and-mastering/>
- Pfanzagl-Cardone, E. (2020). *The Art and Science of Surround and Stereo Recording: Including 3D Audio Techniques*. Springer
- Plomp, R. (2001). *The Intelligent Ear: On the Nature of Sound Perception*. Psychology Press.
- Roederer, J. G. (2000) *Acústica y Psicoacústica de la Música*. Ricordi.
- Schedl, M. et al. (2014). Music information retrieval: Recent developments and applications. In *Foundations and Trends in Information Retrieval* (Vol. 8, No. 2–3). <https://doi.org/10.1561/1500000042>
- Schnell, N. et al. (2017). *PiPo, A Plugin Interface for Afferent Data Stream Processing*

Modules. International Symposium on Music Information Retrieval (ISMIR).

Smalley, D. (2007). *Space-form and the acousmatic image*. *Organised Sound*, 12 (1), 35–58.

<https://doi.org/10.1017/S1355771807001665>

Truax, B. (3 de junio de 2023). *Handbook for Acoustic Ecology*. <http://www.sfu.ca/sonic-studio-webdav/handbook/index.html>

Tzanetakis, G. et al. (2007). *Stereo Panning Features for Classifying Recording Production Style*. Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR), pp. 441-444.

Valle, R. (2018). *Visual Display and Retrieval of Music Information*. CNMAT, UC Berkeley.

<https://doi.org/10.48550/arXiv.1807.10204>

White, H. E., y White, D. H. (2014). *Phycis And Music: The Science of Musical Sound*.

Dover Publications

Wishart, T. (1996). *On Sonic Art*. Routledge

Índice de figuras

Figura 1. Clasificación de las visualizaciones de las obras musicales según Khulusi et al..	17
Figura 2. Esquema de diseño general de la aplicación.	31
Figura 3. Propuesta para el análisis del contenido de audio en la aplicación.	31
Figura 4. Esquema de generación de las visualizaciones de la aplicación.	32
Figura 5. Sección externa, vista general de la aplicación.	32
Figura 6. Opciones de configuración para cada track de la aplicación.	33
Figura 7. Control y configuración general de la aplicación.	33
Figura 8. Monitor de visualización de la aplicación.	34
Figura 9. Puntuaciones obtenidas para el grupo A (sólo estímulos sonoros).	43
Figura 10. Puntuaciones obtenidas para el grupo B (estímulos audiovisuales).	43
Figura 11. Respuestas del primer estímulo, para el grupo A (sólo sonido).	44
Figura 12. Respuestas del primer estímulo, para el grupo B (audiovisual).	44
Figura 13. Respuestas del segundo estímulo, para el grupo A (sólo sonido).	45
Figura 14. Respuestas del segundo estímulo, para el grupo B (audiovisual).	45
Figura 15. Respuestas del tercer estímulo, para el grupo A (sólo sonido).	46
Figura 16. Respuestas del tercer estímulo, para el grupo B (audiovisual).	46
Figura 17. Respuestas del cuarto estímulo, para el grupo A (sólo sonido).	47
Figura 18. Respuestas del cuarto estímulo, para el grupo B (audiovisual).	47
Figura 19. Respuestas del quinto estímulo, para el grupo A (sólo sonido).	48
Figura 20. Respuestas del quinto estímulo, para el grupo B (audiovisual).	48
Figura 21. Respuestas del sexto estímulo, para el grupo A (sólo sonido).	49
Figura 22. Respuestas del sexto estímulo, para el grupo B (audiovisual).	49
Figura 23. Cantidad de aciertos totales de los cuestionarios.	52
Figura 24. Cantidad de desaciertos totales de los cuestionarios.	52
Figura 25. Porcentaje del promedio de aciertos para cada grupo.	52
Figura 26. Sección interna general de la aplicación.	61
Figura 27. Sección interna de la aplicación, carga de archivos.	62
Figura 28. Sección interna de la aplicación, velocidad de visualización.	62
Figura 29. Sección interna de la aplicación, ganancia general.	63
Figura 30. Sección interna de la aplicación, control de posición de la cámara.	63
Figura 31. Sección interna de la aplicación, configuraciones de inicio.	64
Figura 32. Sección interna de la aplicación, estructura de los canales de audio.	64
Figura 33. Gestión de la carga de archivos e ingreso de la señal de audio.	65
Figura 34. Análisis de sonoridad y centroide espectral.	66
Figura 35. Cálculo de las diferencias de intensidad interaurales relativas.	67
Figura 36. Sección interna de la aplicación, objeto de generación de las visualizaciones.	68
Figura 37. Generación de las visualizaciones, mundo visual.	68
Figura 38. Generación de las visualizaciones, cámara y luces.	69
Figura 39. Generación de las visualizaciones, datos de la aplicación.	70
Figura 40. Generación de las visualizaciones, cubemap.	70
Figura 41. Generación de las visualizaciones, malla.	70
Figura 42. Generación de las visualizaciones, esferas.	71
Figura 43. Creación y control de la posición, color y transparencia de las esferas.	72

Índice de tablas

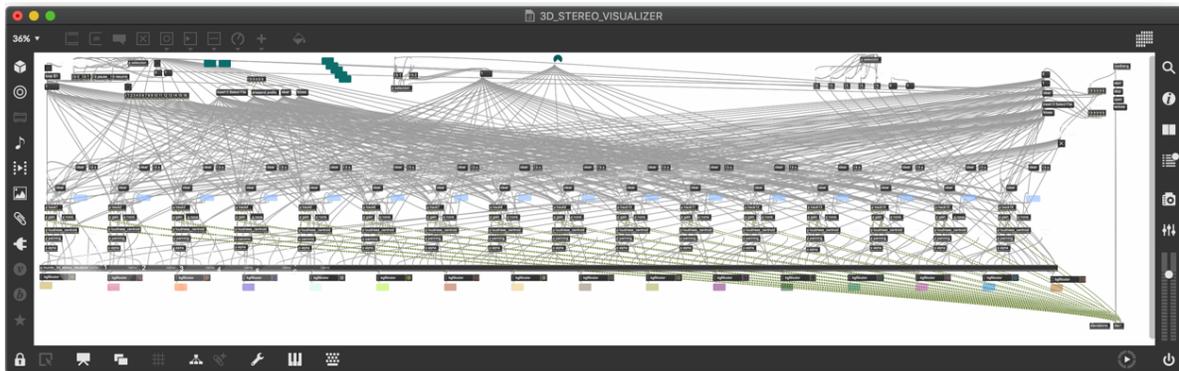
Tabla 1. Resumen de los aciertos y desaciertos de los cuestionarios.	51
---	----

Anexo

Sección interna de la aplicación:

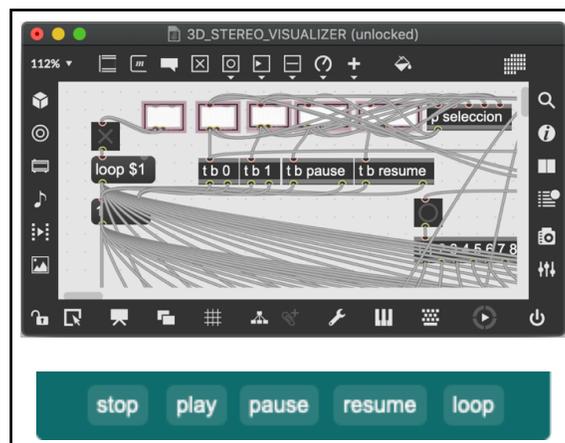
A continuación, muestro la sección interna de la aplicación comenzando por la vista general y siguiendo con cada parte relevante de ésta.

Figura 26. Sección interna general de la aplicación.



En la parte superior izquierda de la sección interna de la aplicación se encuentra el control de la reproducción de los archivos de sonido, tales como: comenzar a reproducir, detener, pausar, continuar y el modo *loop*.

Figura 27. Sección interna de la aplicación, control de la reproducción.



Siguiendo por la parte superior hacia la derecha está la sección de carga de los archivos de audio (Figura 27), luego los botones que permiten la selección de la velocidad de la visualización (Figura 28), después la adaptación de la ganancia general (Figura 29), posteriormente la selección de la posición de la cámara (Figura 30), y por último la sección que permite la inicialización de los parámetros al abrir el programa (Figura 31).

Figura 27. Sección interna de la aplicación, carga de archivos.

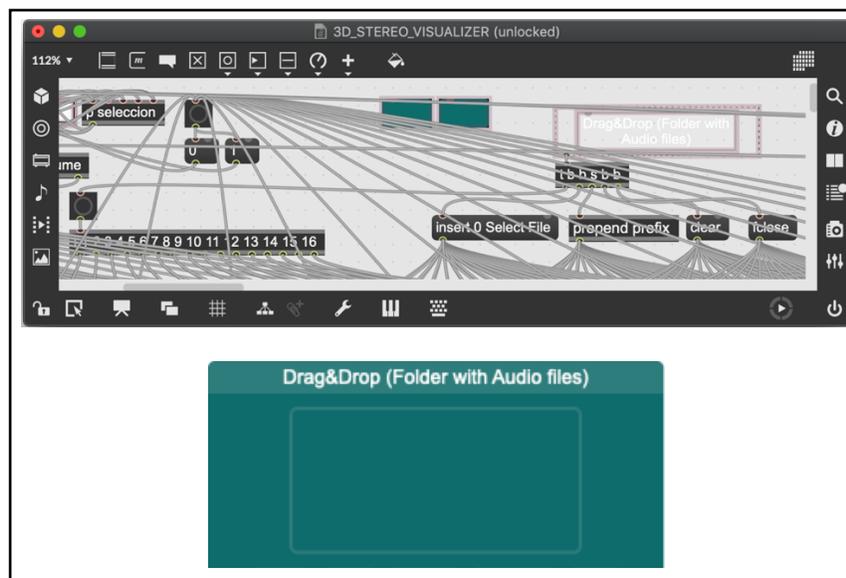


Figura 28. Sección interna de la aplicación, velocidad de visualización.

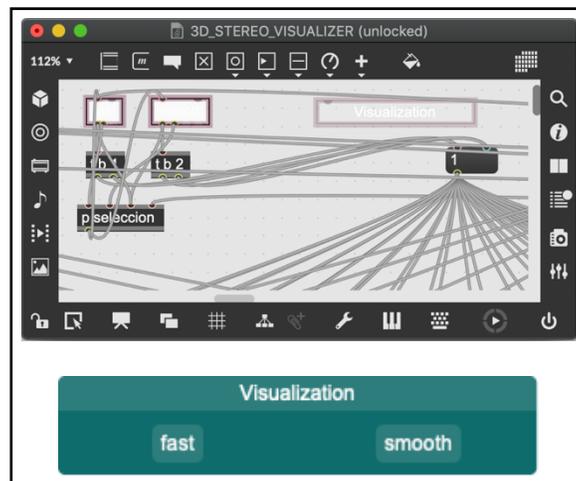


Figura 29. Sección interna de la aplicación, ganancia general.

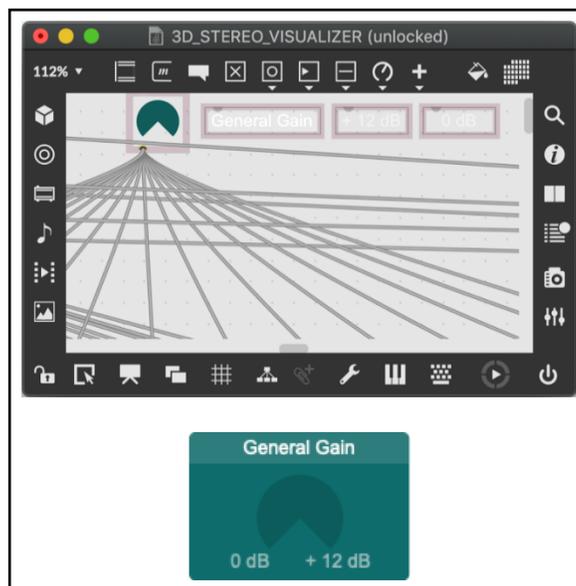


Figura 30. Sección interna de la aplicación, control de posición de la cámara.

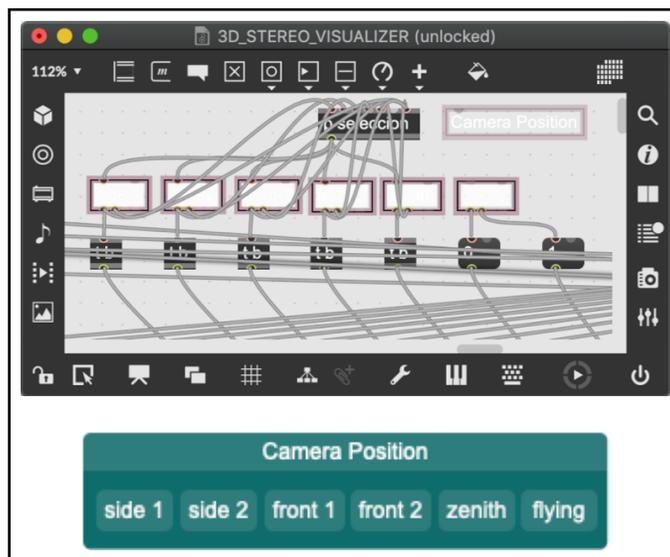
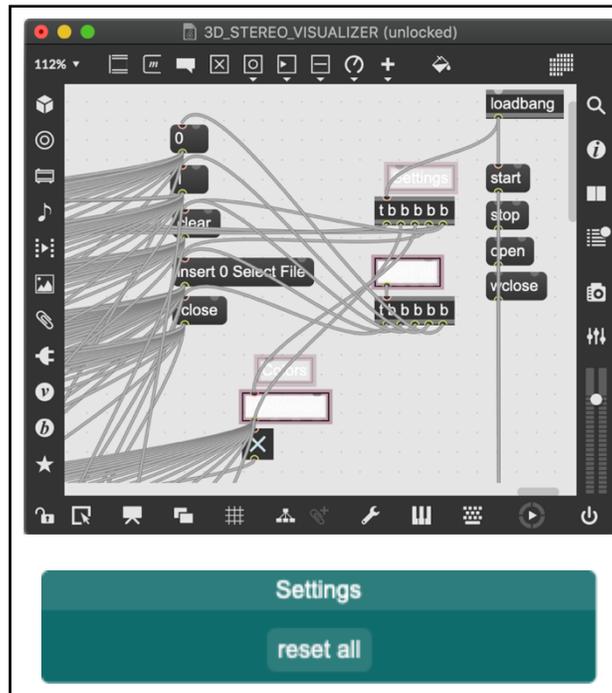
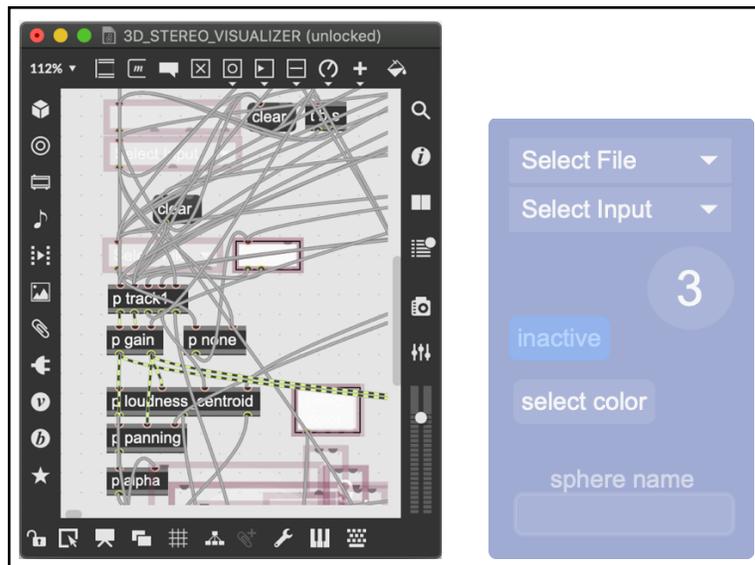


Figura 31. Sección interna de la aplicación, configuraciones de inicio.



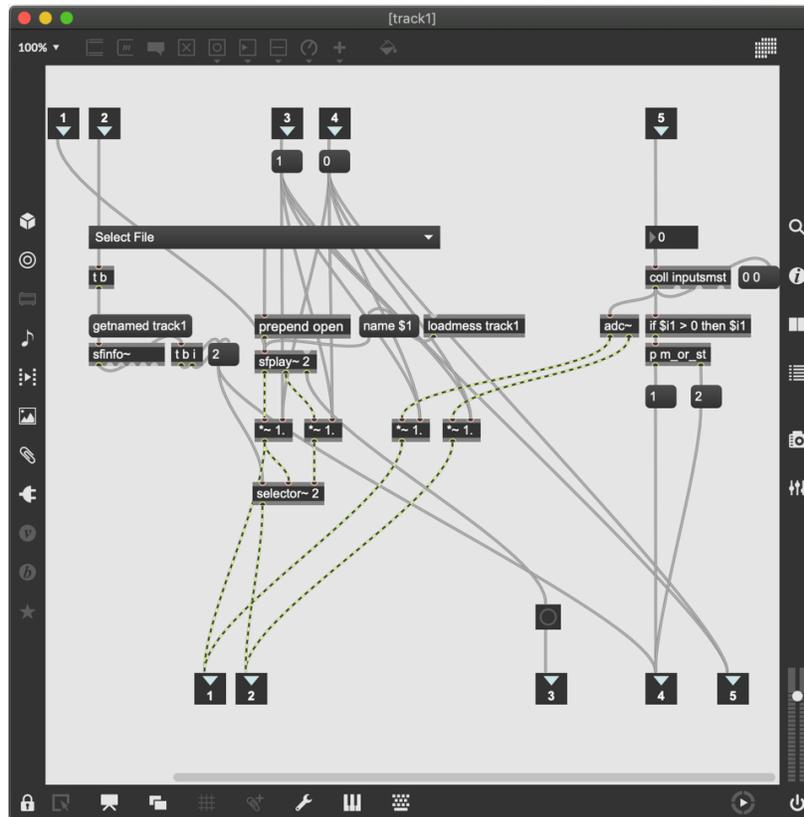
Luego, hacia abajo se encuentra una estructura repetida 16 veces, se trata de toda la programación del ingreso, reproducción y análisis de la señal de audio para cada una de las 16 posibles esferas a ser ubicadas tridimensionalmente.

Figura 32. Sección interna de la aplicación, estructura de los canales de audio.



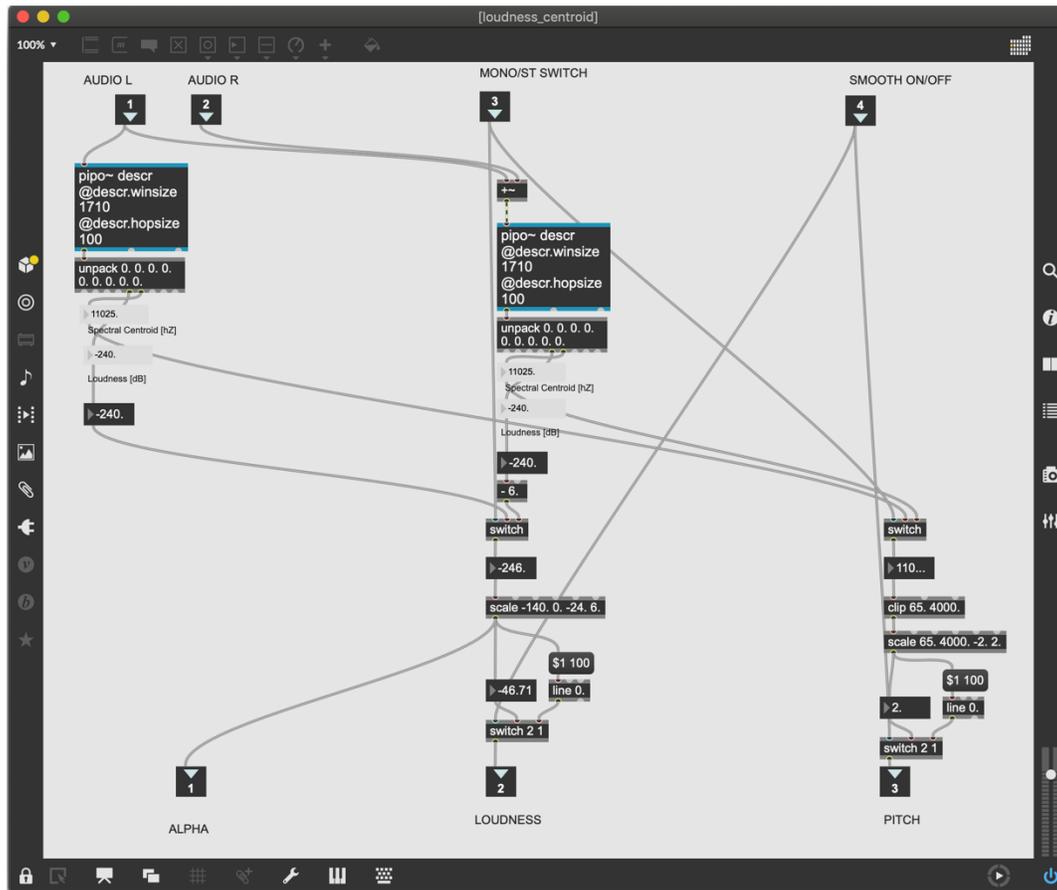
Para cada una de las 16 estructuras que se repiten, creé el siguiente objeto llamado *track1*, *track2*, *track3* respectivamente, en donde se reciben instrucciones de carga de archivos y de selección de entrada de la señal.

Figura 33. Gestión de la carga de archivos e ingreso de la señal de audio.



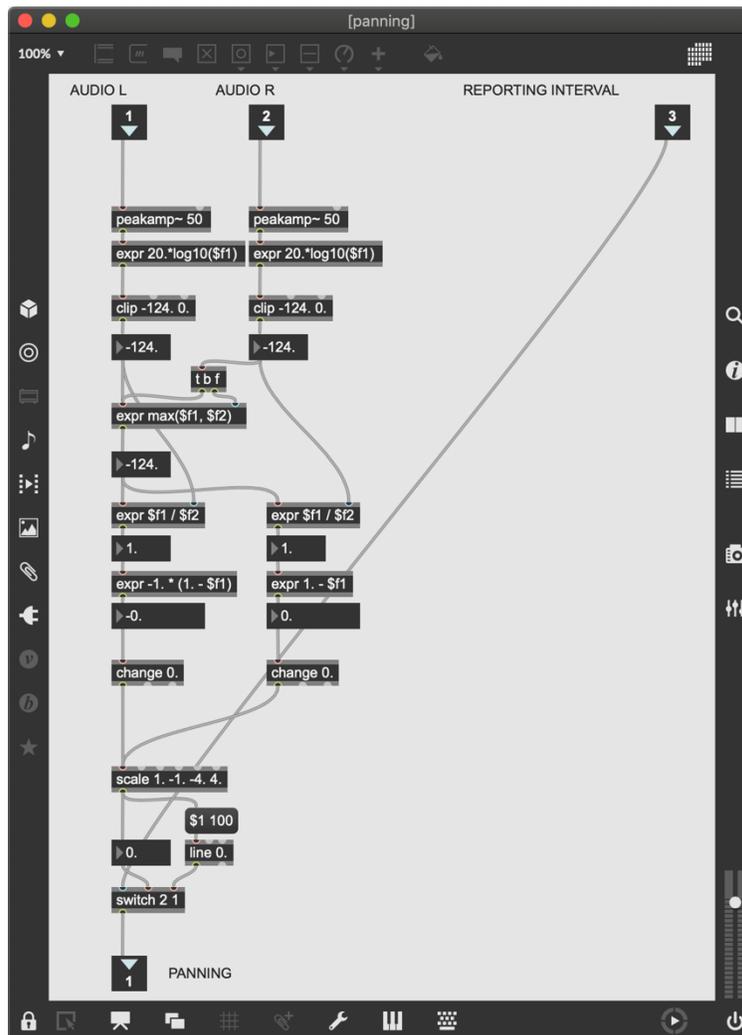
El siguiente objeto llamado *loudness_centroid* utiliza los procesos del objeto *pipo~* para la obtención del valor de frecuencia para el centroide espectral y el valor en decibeles para la sonoridad. También realiza el escalado de ambos valores a números flotantes y limita sus rangos, por ejemplo, el centroide llega hasta 4,000 Hz debido a que es una frecuencia cercana a la última fundamental posible en la notación musical convencional.

Figura 34. Análisis de sonoridad y centroide espectral.



El objeto *panning* recibe las señales de ambos canales (izquierda y derecha) y mide sus amplitudes pico, luego las convierte a decibeles a través de una operación logarítmica y limita los valores ingresados a través del objeto *clip*. Después identifica el valor máximo entre ambas mediciones y calcula la proporción entre el máximo valor obtenido y cada uno de los valores de los canales, posteriormente escala esa proporción a números flotantes entre -1 y 0 para el lado izquierdo, y entre 0 y 1 para el lado derecho. Finalmente se realiza el escalado final para ser interpretado por el objeto que se encargará de ubicar la esfera tridimensionalmente.

Figura 35. Cálculo de las diferencias de intensidad interaurales relativas.



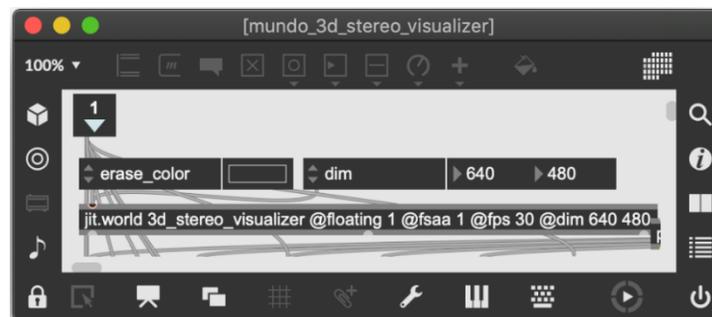
Volviendo a la vista general de la aplicación el objeto *mundo_3d_stereo_visualizer* (Figura 36) recibe los datos analizados y escalados para ubicar cada esfera en el plano tridimensional, también se utiliza este objeto para la inicialización del mundo visual del OpenGL que genera la visualización. En el interior de dicho objeto hacia el sector izquierdo superior nos encontramos con el objeto *jit.wold* (Figura 37) que realiza el *render* de los objetos de

OpenGL y los muestra en la ventana de monitorización para su visualización, para este objeto se indican por ejemplo la cantidad de fotogramas por segundo y tamaño de la ventana.

Figura 36. Sección interna de la aplicación, objeto de generación de las visualizaciones.

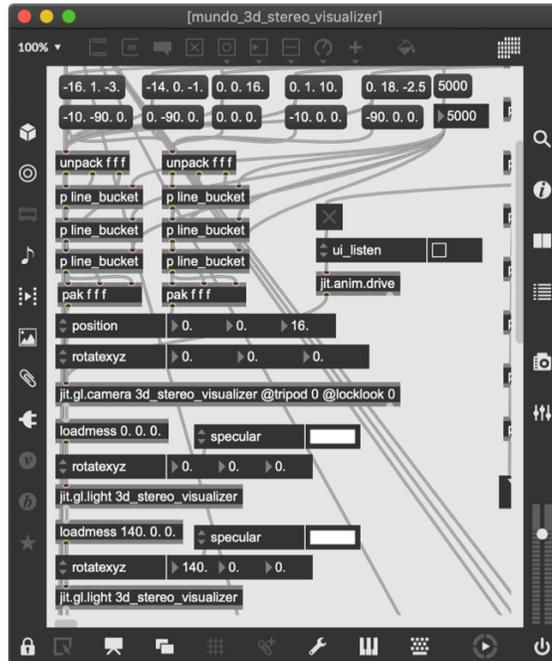


Figura 37. Generación de las visualizaciones, mundo visual.



Continuando del lado izquierdo hacia abajo nos encontramos con instrucciones que están relacionadas a las diferentes opciones de posicionamiento de la cámara y propiamente con el objeto que se encarga de esto llamado *jit.gl.camera*. Luego se encuentran los objetos que inician las luces y permiten ver las esferas con una determinada profundidad y perspectiva tridimensional llamados *jit.gl.light*, siendo que de modo contrario solo se verían círculos planos.

Figura 38. Generación de las visualizaciones, cámara y luces.



Si siguiendo hacia abajo, nos encontramos con tres estructuras importantes: en primer lugar con el objeto llamado *jit.gl.videoplane* que permite la inclusión (en el monitor de visualización) de una imagen con transparencia que contiene el nombre de la aplicación, su versión y el nombre del desarrollador (Figura 39). En segundo lugar podemos ver los objetos *jit.gl.cubemap* y *jit.gl.skybox* que generan el espacio visual a través de una imagen de tipo *cubemap* (Figura 40), esto genera un cubo de imágenes que sirven como fondo en 360°. En tercer lugar están los objetos *jit.gl.gridshape* y *jit.gl.mesh* que se encargan de generar la malla de líneas que define la superficie sobre la que se disponen las esferas (Figura 41).

Figura 39. Generación de las visualizaciones, datos de la aplicación.

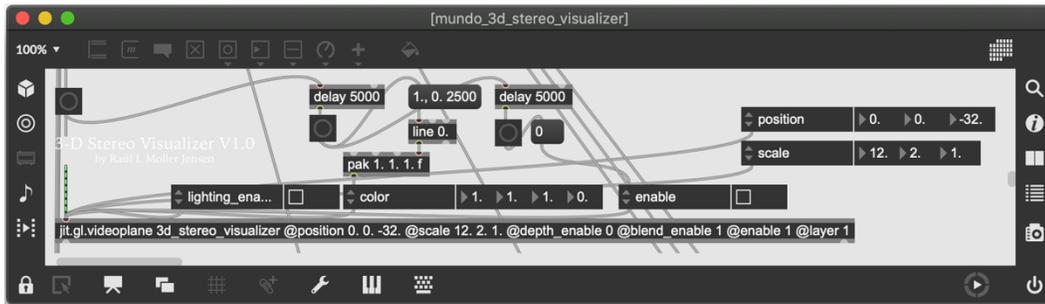


Figura 40. Generación de las visualizaciones, cubemap.

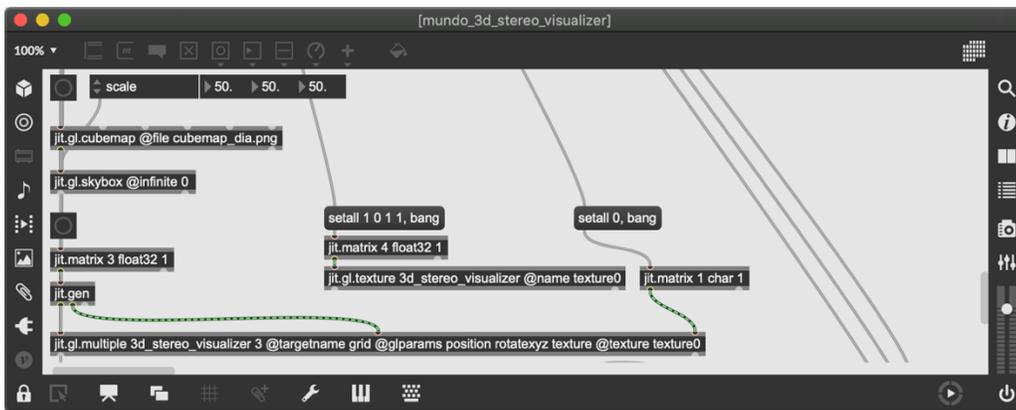
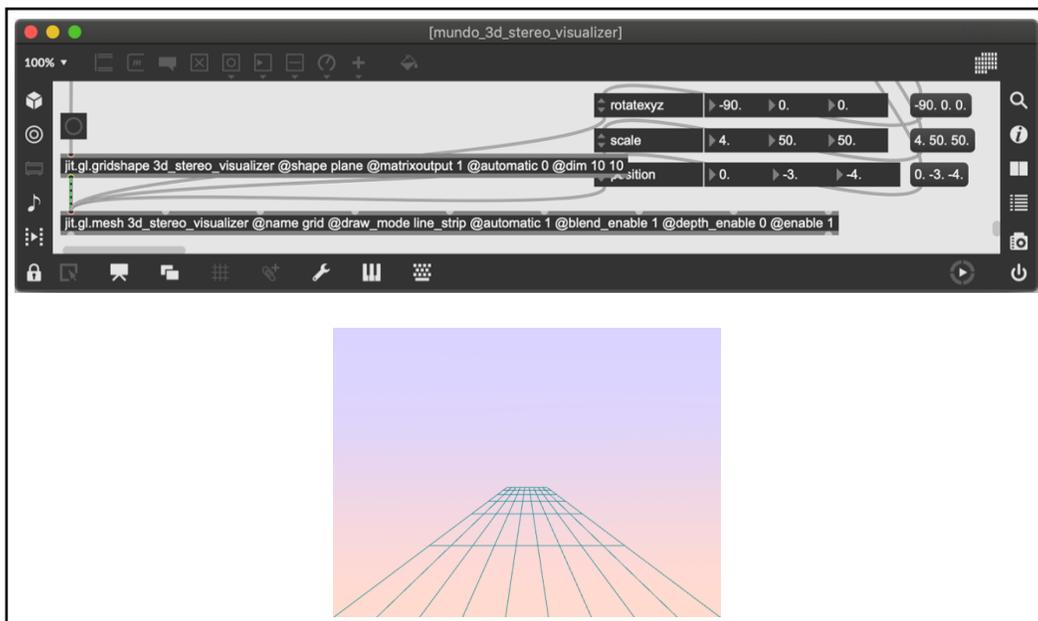
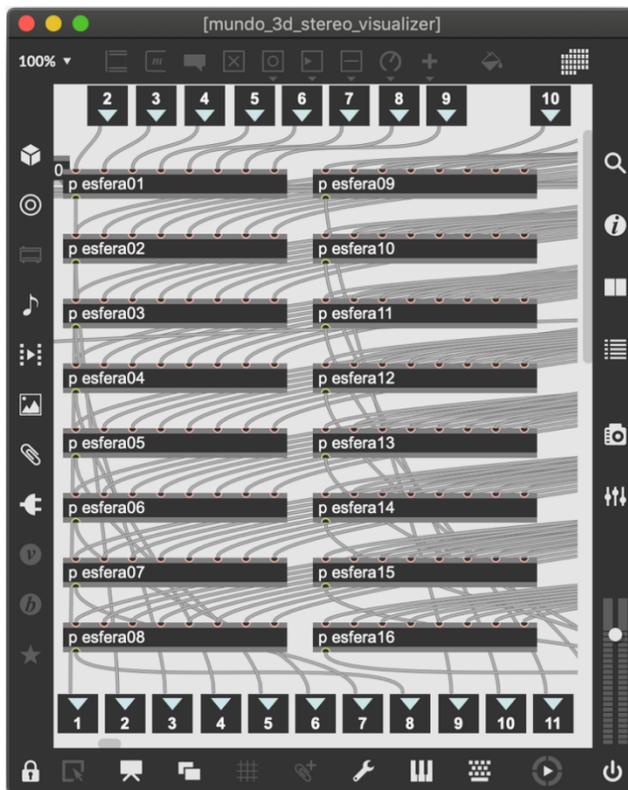


Figura 41. Generación de las visualizaciones, malla.



Volviendo a la parte izquierda superior del objeto *mundo_3d_stereo_visualizer*, yendo un poco hacia la derecha nos encontramos con los objetos donde se generan y ubican las esferas en el plano tridimensional nombrados como *esfera01*, *esfera02*, *esfera03*... respectivamente.

Figura 42. Generación de las visualizaciones, esferas.



Cada objeto *jit.gl.gridshape* genera la esfera que representará a cada audio en particular controlando su posición (eje X para la panorámica, eje Y para el tono, eje Z para el volumen), la transparencia y el color. Además, se añade mediante el objeto *jit.gl.text* la posibilidad de etiquetar cada esfera con un texto para mejorar su localización, en el caso de estar viendo varias esferas a la vez.

Figura 43. Creación y control de la posición, color y transparencia de las esferas.

