

Edmar Olivares Soria

ESPACIO-TIMBRE

Hacia una teoría interdisciplinar
en la música electroacústica

EPUB


Casa abierta al tiempo



DOCUMENTOS





**Universidad Autónoma
Metropolitana**

Rector General

José Antonio de los Reyes Heredia

Secretaria General

Norma Rondero López

Coordinadora General de Difusión

Yisel Arce Padrón

Directora de Publicaciones y

Promoción Editorial

Freja Innina Cervantes Becerril

Subdirector de Distribución y

Promoción Editorial

Marco Antonio Moctezuma
Zamarrón

**Universidad Autónoma Metropolitana, Unidad
Lerma**

Rector

Gabriel Soto Cortés

Secretaria de la Unidad

Alma Patricia de León Calderón

*Director de la División de Ciencias Sociales y
Humanidades*

Raúl Hernández Mar

*Secretaria Académica de la División
de Ciencias Sociales y Humanidades*

Ana Carolina Robles Salvador

**Consejo Editorial de Ciencias Sociales y
Humanidades**

Coordinador General

Juan Carlos López García

Departamento de Procesos Sociales

Lidia Ivonne Blásquez Martínez

Departamento de Artes y Humanidades

Claudia Mosqueda Gómez

Departamento de Estudios Culturales

Gladys Ortiz Henderson

Espacio-Timbre

Hacia una teoría interdisciplinar en la música electroacústica

Edmar Olivares Soria



Universidad Autónoma Metropolitana
Unidad Lerma/División de Ciencias Sociales y Humanidades
Juan Pablos Editor
México, 2023

Olivares Soria, Edmar
Espacio-Timbre : hacia una teoría interdisciplinar en la música electroacústica /
Edmar Olivares Soria, autor. - - México : Universidad Autónoma Metropolitana : Juan Pablos
Editor, 2023
1a. edición
Ilustraciones

ISBN: 978-607-28-2787-5 UAM
ISBN: 978-607-711-680-6 Juan Pablos Editor

T. 1. Electroacústica

T. 2. Música y tecnología

ML1092 O45

Primera edición: enero de 2023

Espacio-Timbre. Hacia una teoría interdisciplinar en la música electroacústica
Edmar Olivares Soria

Diseño original: María Luisa Passarge

D.R. © 2023, Edmar Olivares Soria

D.R. © 2023, Universidad Autónoma Metropolitana
Prolongación Canal de Miramontes 3855
Ex Hacienda San Juan de Dios, Alcaldía Tlalpan, 14387, Ciudad de México

Unidad Lerma/División de Ciencias Sociales y Humanidades
Avenida de las Garzas 10, Col. El Panteón, 52005, Lerma, Estado de México
Consejo Editorial de la División de Ciencias Sociales y Humanidades, <cedcsh@correo.ler.uam.mx>

D.R. © 2023, Juan Pablos Editor, S.A. de C.V.
2a. Cerrada de Belisario Domínguez 19, Col. del Carmen
Alcaldía de Coyoacán, 04100, Ciudad de México, <juanpabloeditor@gmail.com>

ISBN: 978-607-28-2787-5 UAM
ISBN: 978-607-711-680-6 Juan Pablos Editor

Fotografía de portada: Giovanni Michelle Arroyo Torres, "Día 2. Pt. 4", fotografía digital, 2022.

Esta publicación no puede ser reproducida en todo ni en parte, ni registrada o transmitida por un sistema de recuperación de información en ninguna forma y por ningún medio, sea mecánico, fotoquímico, electrónico, magnético, electroóptico, por fotocopia o cualquier otro, sin el permiso previo y por escrito de los editores.

La presente publicación pasó por un proceso de dos dictámenes (doble ciego) de pares académicos avalados por el Consejo Editorial de la División de Ciencias Sociales y Humanidades de la UAM-Lerma, que garantizan su calidad y pertinencia académica y científica.

Distribución: TintaRoja <www.tintaroja.com.mx>

Juan Pablos Editor es miembro de la Alianza de Editoriales Mexicanas Independientes (AEMI)

Índice

15 Introducción

21 Capítulo I. Percepción como fundamento de la experiencia de la espacialidad

- 22 1. Fundamentos generales sobre percepción
- 26 1.1 Procesos *top-down* y *bottom-up*
- 28 1.2 Percepción y tiempos de reacción
- 30 1.3 Principios básicos de la teoría Gestalt en la percepción
- 33 2. Analogía relacional
- 35 3. Sumario de definiciones y conceptos

37 Capítulo II. Percepción sonora para la experiencia del Espacio-Timbre

- 38 1. Análisis de escena auditiva
- 41 1.1. Modos de agrupamiento: simultáneo y secuencial
- 47 1.2. Atributos de percepción en el ASA
- 50 2. La teoría Gestalt de la música de James Tenney
- 51 2.1. La complejidad sonora y los cúmulos tonales
- 53 2.2. La continuidad
- 54 2.3. Unidades temporales Gestalt

- 56** 3. Localización espacial del sonido
- 58** 3.1 Escucha espacial: resolución de distancia de las fuentes sonoras
- 59** 3.2. Escucha espacial: resolución de dirección de las fuentes sonoras
- 69** 3.3. Propiedades acústicas de la espacialidad en la experiencia aural
- 71** 3.4. Fuentes sonoras en movimiento y fuentes múltiples

75 Capítulo III. Espacio y espacialidad

- 76** 1. Construcción teórica del concepto de espacio
- 77** 1.1 Antecedentes sobre uso del espacio en la música
- 78** 1.2. Espacio de parámetros y espacio como experiencia
- 82** 1.3 Espacio interno-espacio externo y el problema de la disimilaridad serial/perceptual
- 83** 1.4. Espacio virtual, espacialidad e inmersividad
- 87** 2. Espacio virtual e inmersividad en la electroacústica multicanal
- 90** 2.1 Consideraciones generales sobre el espacio audible
- 94** 2.2. Diseño serial vs. percepción
- 96** 3. Componentes primarios de la espacialidad
- 100** 3.1. Objetos sonoros fijos
- 102** 3.2. Trayectorias continuas u objetos sonoros móviles
- 105** 3.3. Superficies sonoras y envolvimiento
- 106** 4. Observaciones finales sobre la descripción de espacialidad

111 Capítulo IV. Espacio-Timbre: un modelo teórico

- 113** 1. Perfil y espacio morfológico
- 123** 1.1. Distancia y medidas en el espacio espectromorfológico
- 125** 2. Espacio-Timbre: construcción abstracta
- 128** 3. Donde la espacialidad y el timbre se unen
- 131** 4. Ejemplos
- 132** 4.1. Ejemplo I
- 136** 4.2. Ejemplo II
- 137** 4.3. Ejemplo III

140	4.4. Ejemplo IV. Metodología computacional básica mediante aprendizaje máquina
178	5. Aplicaciones computacionales actuales sobre espacio y timbre
181	Conclusiones
185	Apéndice
185	Notación matemática
185	Fundamentos de sistemas dinámicos
190	Algoritmos computacionales
190	Dynamic Time Warping (DTW)
190	Alineación global de Kernel K-Medios
192	Bandas LB Keogh
193	Métricas de similitud para cadenas de caracteres
195	Entropía aproximada
197	Índice de figuras
201	Bibliografía

Agradecimientos:

A mis padres, María Teresa y Edmundo.

Elizabeth Anderson, Roberto Morales Manzanares, Rodrigo Sigal, Manuel Rocha
Iturbide, Esther Lamneck, Joao Pedro Oliveira, Panayiotis Kokoras.

Agradecimientos especiales:

Eduardo Morales Manzanares, Luis Alfonso Estrada Rodríguez, Ricardo Dal Farra,
Luz María Sánchez Cardona, Josué Martínez Alcántara, Montserrat GE.
Enrique Ruíz Velasco, Javier Álvarez, Juan Pampín, Daniel Teruggi, Diego Losa,
Beatriz Ferreira, Annete Vande Gorne.

Para vivir en el cruce de la imaginación de las formas y de la imaginación de las fuerzas, hay una obra particularmente eficaz; es la obra de un poeta y grabador, la obra de William Blake. [...] Podríamos llamar a ciertos poemas de Blake, poemas absolutos, es decir, poemas que no traducen ideas, pero que anudan a las palabras mismas la materia imaginaria y la forma de los fantasmas, el movimiento de la palabra y el movimiento del cuerpo “el pensamiento” y el “moviente”, o mejor aún, el parlente y el moviente. [...] Aquí son las imágenes verbales las que profetizan. No hay pensamiento profético subyacente.

Gaston Bachelard, *El aire y los sueños*

Introducción

La práctica e investigación de la música electroacústica ha sido desde sus inicios, dada su naturaleza, una formulación estética cuyo desarrollo histórico se encuentra íntimamente relacionado con la tecnología y con algunos campos del saber como la estética, la hermenéutica, la fenomenología, las ciencias cognitivas y por supuesto la acústica y la psicoacústica. Esta posición privilegiada la plantea como un punto de convergencia para la multi e interdisciplinariedad que le permite, además, nutrirse de manera directa de planteamientos, reflexiones y herramientas teórico-conceptuales propias de otros campos disciplinares, con el fin de extender su propio desarrollo, desde lo abstracto y lo práctico, desde lo técnico y lo estético.

A lo largo de la historia de la práctica e investigación de la música electroacústica resaltan dos parámetros compositivos primordiales en el lenguaje estético de esta expresión artística: el *espacio* y el *timbre*. Como es bien conocido, la exploración tímbrica se posicionó como un elemento predominante en la composición musical durante segunda mitad del siglo XX, y es posible afirmar que, en cierto modo, ha dirigido la composición del parámetro del espacio dentro de la electroacústica multicanal.

La investigación del espacio como elemento compositivo con una identidad individual se desarrolló históricamente de manera más bien fragmentada y no sería del todo incorrecto afirmar que no existe, al día de hoy, una consolidación homogénea respecto del mismo; ni desde un planteamiento teórico ni desde una formulación explicativa de éste, como fenómeno perceptual a través de un marco teórico o contextual específico. Al parecer el desarrollo más notable se concentra en el diseño de herramientas digitales de *especialización* y en las consecuencias directas de su uso, así como en la exploración estética de las diversas posibilidades desde la práctica y el imaginario experiencial de cada compositor/compositora. En este sentido, si bien existen numerosos compositores que desde la segunda mitad del siglo pasado

han planteado diversas conceptualizaciones estéticas y técnicas del espacio, en la actualidad no se distingue un desarrollo predominante de marcos teóricos que aborden el uso del espacio como elemento compositivo en la música electroacústica desde una mirada integral compuesta por elementos perceptuales, acústicos y teóricos.

Sería natural suponer que el espacio (o la espacialidad) como parámetro dentro de la música electroacústica puede ser considerado como algo intrínseco e inherente a casi cualquier obra que sea presentada en formato multicanal (y por supuesto, estéreo) debido a las cualidades, características y consecuencias acústicas y psicoacústicas derivadas de la propia naturaleza del formato en cuestión. Entonces, resulta fundamental plantear marcos teórico-contextuales que permitan desarrollar no sólo una explicación *a posteriori*, tanto de la obra como del proceso compositivo circunscritos en este contexto, sino que dichos planteamientos funcionen de igual modo, como una teoría interdisciplinar de la música electroacústica enfocada en el estudio y formalización del timbre y el espacio desde lo acústico, lo perceptual, lo estético y lo matemático.

Derivado de la problemática anterior, se propone que una comprensión alternativa al proceso compositivo electroacústico multicanal puede fundarse en el uso informado (y hasta cierto punto, sistematizado) de la acústica y la psicoacústica para la organización del material sonoro, tomando como referente su perfil espectral-morfológico en relación con la multiplicidad de elementos que involucran la percepción de la espacialidad del sonido. Es claro que esta aproximación es realizada de forma natural e innata por varios compositores; lo que aquí se plantea es el establecimiento de un marco teórico y contextual que aborde dicha práctica haciendo uso de las herramientas teóricas propias de la acústica, la percepción y la localización espacial del sonido. Todo lo anterior es además formalizado e inscrito en el lenguaje matemático de la teoría de conjuntos y los sistemas dinámicos, lo que posiciona a este libro como un aporte novedoso e inédito desde la perspectiva de la práctica e investigación multi-interdisciplinar o el denominado *cruce arte-ciencia-tecnología*, dentro del contexto ya mencionado.

Organización de la obra

El marco teórico propuesto en este libro incluye, entonces, la hipótesis fundamental de que el espacio y el timbre están inexorablemente ligados de modo tal que, dentro de un contexto es-

pecífico, pueden ser considerados como un sólo atributo compositivo: el *Espacio-Timbre*. A lo largo de la obra se desarrollarán los conceptos de espacio, de timbre (desde un punto de vista espectromorfológico) y, consecuentemente, del Espacio-Timbre; todo ello a partir de la integración de herramientas teórico-experimentales específicas provenientes de la psicoacústica y la percepción. La descripción del timbre estará dada por lo que se denominará *perfil espectromorfológico*, mientras que la espacialidad estará descrita por lo que se denotará como *perfil espacial*. En este sentido, la denominación *Espacio-Timbre* se utiliza para el parámetro único ya mencionado, mientras que *Espacio-Timbre* hará referencia a la propuesta teórica general presentada en esta obra.

Posteriormente se propone una caracterización general simple en el nivel teórico-matemático para cada uno de dichos componentes. Esta caracterización emplea el lenguaje de sistemas dinámicos desde la mirada de la topología, una herramienta matemática que resulta por demás útil para establecer una estructura teórica capaz de *organizar* y *formalizar* los conceptos de *timbre*, *espacio* y *Espacio-Timbre* de una manera general.

La obra presentada en este libro puede ser entendida como un modelo teórico original que describe el concepto de *Espacio-Timbre* desde una perspectiva construida a partir de tres aspectos primordiales:

- Aspecto perceptual.
- Aspecto acústico.
- Representación abstracta matemática.

Los dos primeros aspectos se entienden como la referencia acústico-perceptual de los diversos métodos, estrategias y procesos compositivos en torno al uso del timbre y el espacio, los cuales son bien conocidos dentro de la música electroacústica; es decir, representan una *explicación acústico-perceptual informada* de dichos elementos dentro del proceso creativo.

Estos elementos, aunque son bastante conocidos y manejados por los creadores relacionados con el formato multicanal, no cuentan con demasiadas formulaciones/organizaciones teóricas que expliquen el uso de estos elementos (los procesos y técnicas relacionadas al espacio y el timbre) mediante resultados perceptuales y acústicos, o bien, que los agrupen en un modelo teórico como el aquí propuesto.

El tercer elemento, la representación abstracto-matemática, es la estructuración de los dos elementos anteriores mediante una implementación simple a partir del lenguaje de la teoría de conjuntos y de los sistemas dinámicos. Estos objetos matemáticos son, a juicio propio, una de las herramientas más adecuadas para dicha representación, debido a su flexibilidad para modelar y representar el comportamiento de un sistema (entendido en su sentido más general) a través del tiempo, que es precisamente el objetivo de esta sección: representar al timbre (perfil espectromorfológico) y el espacio como entes abstractos dinámicos a través del tiempo. Toda vez que se han construido estos objetos matemáticos por separado, haciendo uso de la *propiedad universal del producto de conjuntos*, se obtuvo la representación matemática del *Espacio-Timbre* como el sistema dinámico derivado de la interacción conjunta de los dos anteriores. Esta formalización matemática es, por lo tanto, una manera única, inédita y original de presentar conceptos propios de la música electroacústica a través de un lenguaje simbólico que además abona para una formulación teórica del *Espacio-Timbre* que es a la vez estética y rigurosa.

Los dos primeros capítulos de este libro contienen los resultados acústico-perceptuales acerca de la audición, localización espacial y el fenómeno de percepción en general. En el primer capítulo se describen diversos enfoques relacionados con la percepción, abordando temas como el procesamiento humano de la información, la analogía relacional y la teoría Gestalt. Posteriormente se establecerán los fundamentos de la percepción auditiva y sus correspondientes procesos cognitivos a través de herramientas como el análisis de escena auditiva (ASA), la psicoacústica del sonido espacial y la teoría de análisis musical de James Tenney.

En los capítulos siguientes se plantea el modelo teórico central usando como elementos fundacionales todos los resultados, conceptos, definiciones y conjeturas expuestos en el capítulo anterior. Se construirá el concepto de *espacio virtual sonoro*, *espacio audible*, *espacio aural e inmersividad* y se planteará aquello que se ha denominado como los *componentes primarios de la inmersividad*. Se definirá de igual modo el *perfil espectromorfológico*, para concluir con la representación matemática de todos los elementos y del llamado *Espacio-Timbre*.

En la parte final se presentan diversos ejemplos que tienen el objetivo de mostrar cómo implementar esta propuesta teórica a casos prácticos, tanto los conceptos como las definiciones. Se concluye con un ejemplo particular que consiste en una descripción de una metodología computacional capaz de clasificar muestras de audio mediante una hibridación de procesos basada en el análisis de series de tiempo utilizando algoritmos como el Dynamic Time

Warping (DTW) y aprendizaje no supervisado (para series de tiempo) basado en Global Alignment Kernel (GAK). Estableciendo centroides derivados del cálculo de baricentros euclidianos y bandas LB Keogh, se define un espacio espectromorfológico de referencia (construido mediante el análisis de muestras de audio de instrumentos musicales convencionales). De manera posterior, cualquier muestra arbitraria de audio puede ser analizada en relación con este espacio de referencia y puede ser asociada, por grado de similaridad, a una espectromorfología de algún instrumento musical convencional de modo tal que dicha muestra puede ser tratada, en cierto grado, conforme a los principios de orquestación y acústica conocidos para tal instrumento; esto es lo que se nombrará *orquespectración virtual*. Este capítulo tiene la intención puntual de mostrar una implementación práctica y concreta del modelo teórico de modo tal que se puedan comprender de manera más clara tanto las propuestas teóricas como las formulaciones metodológicas planteadas en este trabajo.

Panorama general

En resumen, la composición electroacústica multicanal hace uso enfático de dos parámetros en particular: el timbre y el espacio. Si bien ambos parámetros han sido bastante explorados de manera heurística y práctica, no existen demasiados modelos estructurados que expliquen dichos elementos composicionales desde los fundamentos de la acústica, la psicoacústica y la percepción. El presente libro plantea un modelo teórico original que hace uso de resultados específicos provenientes de estas áreas de estudio, construyendo una definición de la espacialidad, como fenómeno, desde una perspectiva teórico-abstracta. Todo lo anterior es posteriormente estructurado dentro de un modelo matemático basado en el lenguaje de sistemas dinámicos el cual, debido a su generalidad, permitirá describir y explicar cualquier caso particular referente a la relación *Espacio-Timbre* o al *perfil espectromorfológico-espacial*. El desarrollo de los elementos del *Espacio-Timbre* están guiados por la búsqueda del desarrollo de la *inmersividad* en el proceso composicional como fin último de la experiencia aural de la obra. ■■■

Capítulo I. Percepción como fundamento de la experiencia de la espacialidad

La percepción sensorial (en particular de la percepción sonora) es un elemento clave que permite abordar el proceso compositivo relacionado a lo multicanal (y por lo tanto a lo espacial en la música) de una manera particular, ya que a través de ella el compositor o compositora puede conocer los diversos fundamentos acústicos y psicoacústicos que conforman el fenómeno tímbrico y de la espacialización sonora.

Si bien los compositores electroacústicos conocen perfectamente varios de estos fundamentos acústicos/psicoacústicos que explican y desarrollan la sensación de espacialidad en la experiencia aural de una obra, pareciera que no se ha prestado la suficiente atención a la construcción de un marco teórico o estructura conceptual que incorpore dichos elementos de manera cohesiva en el proceso creativo como parte fundamental del conocimiento compositivo.

El objetivo principal de este trabajo, tal y como se mencionó en la parte introductoria, es la construcción de un modelo teórico que explique el concepto de *Espacio-Timbre* basado en cuatro bloques fundamentales: la percepción del sonido (o experiencia aural), la escucha espacial, las propiedades acústicas inherentes de las fuentes (o espectromorfología) y la formalización de ambas a través de un lenguaje matemático basado en sistemas dinámicos.

En este capítulo se establecieron los fundamentos necesarios para desarrollar la parte asociada a la percepción y en particular a la percepción sonora general. Se detallarán, en primer término las características básicas del fenómeno de la percepción sensorial analizado desde enfoques particulares que luego permitirán construir conceptos y definiciones categóricas acerca de la percepción auditiva. Este primer análisis hará posible, en capítulos subsecuentes,

definir los conceptos de espacio, espacialidad e inmersividad y dotará al presente trabajo de un sustento formal respecto de la formulación de definiciones y conceptos originales propuestos.

El lector o lectora podrá observar al final de este capítulo que, gracias al estudio de la percepción sensorial general (en vez de saltar directamente a temas específicos de percepción sonora), se obtiene un panorama y una comprensión particular respecto de las diversas formas en las que se establece una relación entre el compositor, la obra y el espectador en el siguiente sentido:



Comprender algunos principios básicos sobre la percepción sensorial general nos va a permitir, además, elaborar algunas ideas acerca de los procesos, elementos o incluso estrategias que podrían influir positiva o negativamente en la construcción (concepción/composición) y presentación (formato/display) de una obra; de este modo obtendremos algunas referencias a modo de guía sobre dichas aproximaciones.

Más aún, la concepción teórica de espacio que se presenta en este libro se construye a partir de la integración de todos estos elementos asociados a la percepción sensorial y a la percepción sonora.

1. Fundamentos generales sobre percepción

El fenómeno de la percepción sensorial implica una reconstrucción del mundo externo mediante un proceso cognitivo que revela una multiplicidad de aristas al momento de ser analizado. Existen numerosas posturas que intentan explicar dicho proceso en diversos campos: la filosofía de la mente, el idealismo objetivo-subjetivo, las neurociencias, la psicología cognitiva, entre otros. La percepción sensorial es la puerta que posibilita la interacción con el mundo externo, tal vez la primera, desde el punto de vista biológico-evolucionista.

Barlow (1961) describe la percepción como “el cálculo de una representación que nos permite realizar inferencias confiables y versátiles acerca de asociaciones que ocurren en el mundo a nuestro alrededor”. Por su parte, Fodor (1983) afirma que “lo que la percepción debe hacer es representar el mundo para hacerlo accesible al pensamiento”. Lo que estos autores aseveran

encierra de manera sintética una definición de este fenómeno de acuerdo con su funcionalidad, y también expresa de manera indirecta el aspecto de la percepción como herramienta primordial y básica en el conocimiento del mundo externo o, mejor dicho, de la frontera entre la consciencia y las sensaciones.

La *representación del mundo* a la cual se refiere Fodor, necesita implícitamente al menos dos elementos primordiales: el proceso de representación y aquello que va a ser representado. En otras palabras, para extraer información del mundo sensible, un proceso cognitivo sobre un objeto específico tiene lugar en un momento dado a través de un sistema multisensorial. En este sentido, y tal y como afirma Plumbley (2006), “el problema central de la percepción es el de construir una representación (del mundo) de una colección de señales que emanan de transductores sensoriales”. Sin embargo, es evidente que esta representación no es isomorfa al mundo externo debido a que está llena de imperfecciones y es plenamente vulnerable a errores provenientes de la *incertidumbre*¹ de los sistemas multisensoriales con los cuales aprehendemos los objetos; tanto o más como las representaciones mentales mismas que de ellos construimos. Estas imperfecciones tanto sensoriales como cognitivas, se traducen en ambigüedades respecto de las representaciones mentales finales. No obstante, dichas variaciones permanecen debajo de un nivel de tolerancia lo suficientemente funcional como para permitirnos adaptarnos y desenvolvemos de manera eficiente en nuestro entorno en la mayoría de los casos; es decir, nuestras respuestas a los estímulos externos son adecuadas.

Para Helmholtz (1962) la percepción ocurre “mediante un proceso de inducción inconsciente”, es decir, es “una actividad transparente y que no implica esfuerzo” (Plumbley, 2006). Lo anterior muestra a la percepción como un proceso inherente al sistema multisensorial, al menos en una primera etapa, caracterizado por estructuras que responden de manera más o menos reactiva o automática hacia los objetos físicos del mundo externo.

Por su parte, Gibson (1966) afirma que “los sentidos pueden obtener información acerca de los objetos en el mundo sin la intervención de un proceso intelectual”. Lo anterior refleja una idea de percepción *pasiva*, ya que los estímulos sensoriales se entienden como señales de infor-

¹ Incertidumbre entendida aquí desde su acepción proveniente de la física: “Toda medición física en investigación e industria está asociada con cierto grado de imprecisión. Incluso pequeñas fluctuaciones en las condiciones ambientales pueden influir en la medición y provocar desviaciones que hacen que el resultado de la medición no sea preciso. Para obtener resultados de medición significativos, resulta esencial conocer la incertidumbre asociada a los resultados” (consultado en <<https://www.kistler.com/es/glosario/termino/incertidumbre-de-medicion/>>).

mación que son absorbidas para obtener una representación del mundo externo sin la completa necesidad (al menos en una primera etapa) de un proceso cognitivo más complejo. En este sentido, Plumbley (2006) plantea una diferenciación entre *percepción pasiva* y lo que para él sería, dentro de este contexto, una *percepción activa*, la cual ocurre cuando “un sistema perceptual está involucrado no sólo en el análisis pasivo de cualquier estímulo que esté ocurriendo sobre los receptores de un organismo, sino en la exploración activa de un campo ambiental de estímulos”.

Para fines de claridad, es posible entender la percepción pasiva como la primer etapa de un proceso más complejo denominado percepción activa. La percepción pasiva involucra la absorción de la información producida por los objetos sensoriales del mundo externo en un primer nivel básico. En la siguiente etapa este flujo de información produce una respuesta fisiológica orientada a la obtención de mayor información acerca del objeto que produce la señal o sobre el ambiente, así como a la planeación de una reacción acorde al estímulo dependiendo de su naturaleza (reacción ecológica).² Dicha respuesta o reacción está direccionada por múltiples factores, entre los cuales destacan los condicionamientos evolutivos (estímulos asociados con la supervivencia como amenazas, ausencia o presencia de comida, identificación de posibles parejas reproductivas, etcétera) y la asociación directa o indirecta sobre estímulos percibidos en el pasado.

Al respecto de la percepción relacionada con la reacción ecológica, Shepard (1961) establece que “el cerebro ha sido moldeado por selección natural; sólo aquellos organismos que fueron capaces de interpretar correctamente lo que sucede en el mundo exterior y comportarse acordemente, han sobrevivido para reproducirse”.

En este sentido, la percepción activa no sólo es un proceso que contiene a la percepción pasiva, sino que es el fundamento del proceso de aprendizaje, entendido, a su vez, como procesamiento cognitivo consciente de las señales de la información obtenidas a través de los estímulos sensoriales que están primordialmente influidos por una necesidad de adaptación y supervivencia ecológica. De este modo, la percepción se puede dividir en etapas.

Dicho lo anterior, es necesario preguntarnos ¿qué significa *mundo externo*? En breve, es posible asumir que es el conjunto de *objetos y eventos* que tienen existencia fuera del cuerpo físico del organismo sensible y que generan *señales sensoriales* al momento de ser observadas

² Desde el punto de vista fisiológico, estas dos respuestas a los estímulos sensoriales están dadas por los *órganos efectores*.

por dicho organismo. Dado que la realidad ontológica o la formulación epistemológica de ellos no nos atañe en esta discusión, considerar su existencia *a priori* no representa ningún problema.

Sabiendo que el objetivo de esta sección no es el de realizar un análisis profundo y comparativo sobre distintas teorías o perspectivas de la percepción sensorial (ni mucho menos de la realidad del mundo externo), sino el de establecer una definición de dicho fenómeno adecuada para fundamentar las características de éste aplicadas a la comprensión de la percepción espacial del sonido, podemos resumir las siguientes definiciones.

DEFINICIÓN 1.1.1 La percepción sensorial es un proceso multifuncional cuyo objetivo principal es el de construir una representación cognitiva del mundo externo y que además cumple lo siguiente:

- Está moldeada en su estructura por un desarrollo evolutivo que responde a la adaptación y a la supervivencia del entorno; es decir, está delimitada por la funcionalidad ecológica.
- Puede ser fragmentado en dos etapas principales. La primera comprende la absorción primaria de la información emitida por los objetos y eventos que comprenden el mundo externo y ofrece una representación básica del entorno en un modo instintivo; está asociada al concepto de percepción pasiva. La segunda involucra un proceso cognitivo primariamente complejo e implica un procesamiento de la información más detallado y, a su vez, una respuesta fisiológica acorde a dicho procesamiento la cual es realizada por los órganos efectores. Estas dos fases en conjunto conforman la llamada *percepción activa*.
- El aprendizaje o primer etapa del conocimiento del mundo sensible es el proceso que hace uso de la representación perceptual primaria para desarrollar estructuras cognitivas más complejas que involucran además un análisis consciente de los estímulos externos y que, a modo de ciclo de retroalimentación, dan lugar a formulaciones socio-culturales que al mismo tiempo están delimitada por ellos. Esta primera etapa de conocimiento asocia los estímulos sensoriales con referentes aprendidos y vivencias experimentadas dirigiendo al proceso cognitivo a las primeras formulaciones de significación.

Por su parte, la *teoría de la atención de la integración de características* (Treisman, 1980) ha sido desarrollada y empleada principalmente para estímulos visuales; sin embargo, su aplicación a los demás exteroceptores no es extraña. Bajo esta perspectiva se considera que los procesos perceptuales que ocurren en un intervalo de entre 200 y 250 milisegundos (y que además son considerados como intrínsecos y fuera de control) corresponden a la *preatención*; mientras que aquellos procesos perceptuales que transforman los estímulos *preatención* en objetos estructurados son denominados *proceso de atención focalizada*. La relación de *preatención-atención focalizada* con la percepción activa-pasiva debería resultar clara en este punto.

Si la percepción sensorial del mundo externo puede ser entendida de acuerdo con la definición anterior, la siguiente pregunta podría formularse del modo siguiente: ¿cómo caracterizar o definir dichas representaciones de modo que se tenga una descripción adecuada del proceso de percepción? En otras palabras, “¿qué propiedades debería tener una representación para poder hacerla adecuada para su uso en subsecuentes mecanismos de aprendizaje?” (Barlow, 1990). Para plantear una resolución podemos aproximarnos a las estrategias *top-down/bottom-up* que representan una adaptación conceptual desarrollada en las áreas de las neurociencias y la psicología cognitiva a partir de modelos provenientes del área de la programación computacional.

1.1 Procesos top-down y bottom-up

Top-down y bottom-up son estrategias de procesamiento de información características de la informática, orientadas a diseños particulares de *software*. Los términos fueron introducidos a principios de los años setenta por los investigadores de IBM Harlan Mills y Niklaus Wirth. Mills desarrolló los conceptos de programación estructurada para usos prácticos y los probó en un proyecto de 1969 para automatizar ciertos indicadores del *New York Times*. Debido al éxito administrativo e ingenieril alcanzado en aquella ocasión, el enfoque *top-down* se esparció por IBM y por el resto de la industria de las computadoras. Por su parte, Niklaus Wirth, que entre sus logros se cuenta el desarrollo del lenguaje de programación Pascal, publicó en 1971 el artículo “Program Development by Stepwise Refinement”, el cual ejerció gran influencia en el desarrollo de dichas estrategias.

Poco tiempo después, tanto los términos como las perspectivas teóricas del *top-down* y el *bottom-up* se extendieron a otros campos como la administración, las finanzas y por supuesto la cognición y la percepción. Para este último (en particular la percepción visual), el uso del *top-down* y *bottom-up* puede rastrearse desde los trabajos de Gibson (1966) y de Gregory (1970). El primero propone una teoría de la percepción basada en el proceso *bottom-up* mientras que el segundo plantea una perspectiva constructivista basada en el proceso *top-down*.

El término *bottom-up* hace referencia a un proceso perceptual de tipo *raw data*, también conocido como *data-driven processing*, el cual plantea que la percepción comienza con el estímulo externo y se va transfiriendo linealmente hasta llegar a un cierto grado de cognición. Dicho proceso puede ser entendido en dos etapas: 1) recepción de la información sensorial en un nivel primario por los exteroceptores; 2) procesamiento de dicha información en el nivel más básico sin crear relaciones semánticas cognitivas.

El proceso *top-down*, por su parte, se entiende como el desarrollo de reconocimiento de patrones mediante el uso de información disponible dentro de un contexto específico. Siguiendo las afirmaciones de Von Helmholtz, Gregory (1997) suscribe que la percepción visual es un proceso derivado de *inferencias* provenientes de datos sensoriales y conocimiento derivado del pasado. En este sentido, las percepciones son más una suerte de hipótesis predictivas psicológicamente proyectadas en el espacio externo y aceptadas como nuestra realidad más inmediata. De hecho, el procesamiento *top-down* puede entenderse como percepción dirigida por la cognición.

Como bien lo describe McLeod (2008), la perspectiva *top-down* de Gregory implica que la información externa que recibimos del ambiente es con frecuencia ambigua, por lo que para poder interpretarla es necesaria información cognitiva a un nivel superior del estímulo primario, la cual puede provenir de experiencias pasadas o de conocimiento almacenado y permite hacer inferencias sobre dichos estímulos. Como se verá más adelante, ambos conceptos, *top-down* y *bottom-up*, son retomados directamente por Bregman para explicar ciertos aspectos del *análisis de escena auditiva* (ASA); más aún, dichos principios también pueden ser utilizados para explicar y categorizar diversas ideas propuestas por Tenney acerca de la percepción musical.

De manera incluso más directa en relación con el proceso creativo composicional, el conocer (y en un momento dado, utilizar) estas estrategias permitirá establecer y diferenciar los materiales o elementos sonoros que pueden ser asimilados a partir del *bottom-up* de los que

implicarían o detonarían un proceso tipo *top-down*. Lo anterior resulta pertinente para la práctica de la espacialización ya que permitiría al compositor o compositora diseñar y desarrollar escenas de espacio aural con el conocimiento adecuado sobre la capacidad perceptiva en el espectador y plantear, así, diversos modelos o alternativas que incluyan una suerte de *tradeoff* entre las decisiones estéticas o de concepto y las implicaciones perceptuales correspondientes.

1.2 Percepción y tiempos de reacción

Uno de los primeros modelos que estableció paralelismos entre las computadoras y el pensamiento humano dentro de la teoría psicológica es el *sistema de procesamiento de la información de Atkinson* (1968) y que podríamos asociarlo de manera directa con la teoría computacional de la mente (o computacionalismo) propuesto por Fodor a inicios de los años sesenta. En este modelo, el autor define tres componentes estructurales representativos para la descripción de la memoria:

Registro Sensorial (SR, sensory register). Registro inmediato de un estímulo dentro de las dimensiones sensoriales apropiadas; la información obtenida en esta etapa tiene una duración específica y desaparece después de dicho tiempo (decaimiento). Esta etapa puede ser comparada con los dispositivos de entrada de información como *mouse*, teclado, etcétera.

Almacenamiento a corto plazo (STS, short time store). Puede ser considerada como la *memoria de trabajo* del individuo. Algunos autores plantean como analogía a esta etapa, la memoria RAM de una computadora. El decaimiento³ de la información en esta etapa es de una duración mayor que en el SR y no depende necesariamente del tipo de entrada sensorial, como ocurre en el SR, por ello se utiliza la abreviación “a-v-l” (*auditory-verbal-linguistic store*). “Por ejemplo, una palabra presentada visualmente puede ser codificada del SR visual a un STS audible” (Atkinson, 1968).

Almacenamiento a largo plazo (LTS, long time store). En esta etapa, a diferencia de las anteriores, la información no se pierde y es almacenada de forma permanente (desde un punto de

³ El término decaimiento es utilizado para denotar el tiempo que transcurre entre la percepción del estímulo y la desaparición del mismo de la memoria del individuo.

vista relativo) aunque no de manera definitiva, ya que puede ser modificada o alterada por información subsecuente. La analogía natural con una computadora es la de un dispositivo de almacenamiento como un disco duro.

Derivado de lo anterior, se ha desarrollado el concepto de *tiempo de reacción*, el cual ha sido ampliamente utilizado en el campo de la psicología cognitiva y ha mostrado ser una herramienta útil en diversas pruebas relacionadas con la percepción, sobre todo dentro del intervalo de SR y STS, permitiendo con ello establecer una caracterización cada vez más definida de la etapa de *preatención*. Si bien el tiempo de reacción ha sido estudiado con mayor profundidad para el campo visual, su aplicación para otros sentidos, en particular para el campo audible, resulta válida tal y como se verá más adelante con el *análisis de escena audible*. A continuación se enuncia, a modo de definición, el concepto de tiempo de reacción de acuerdo con Shelton (2010).

DEFINICIÓN 1.1.2. El *tiempo de reacción* es el tiempo transcurrido entre el momento que ocurre un estímulo sensorial y la subsecuente respuesta de comportamiento.

- El *tiempo simple de reacción* es el tiempo que un observador tarda en detectar la presencia de un estímulo externo. Representa el nivel de coordinación neuromuscular en la cual el cuerpo, a través de diferentes procesos físicos, químicos y mecánicos, decodifica los estímulos visuales o auditivos, los cuales viajan a través de diversos caminos hasta alcanzar el cerebro como un estímulo sensorial.

Estudios llevados a cabo por Thompson (1992) han documentado resultados que indican que la media del tiempo de reacción para detectar un estímulo visual es del orden de entre 180 a 200 milisegundos, mientras que la misma media pero para estímulos auditivos se encuentra en el rango de 140 a 160 milisegundos. Por su parte, estudios realizados por Yagi (1999) y Verleger (1997) demuestran que, de manera contraria, los estímulos visuales son más rápidos que los auditivos. Estas diferencias en los resultados se deben tanto a las condiciones como a la metodología de los experimentos realizados; sin embargo, el punto principal de esta discusión es la concepción del tiempo de reacción como una medida cuantificable de la velocidad del procesamiento de la información presentada debido a estímulos sensoriales. Como se verá más adelante, los estímulos visuales son apoyados por los estímulos auditivos y viceversa; ello

representa un elemento adicional que puede ser incluido en la composición de la música electroacústica para reforzar la localización espacial de ciertos materiales sonoros.

Necesitamos ahora un contexto que nos permita organizar y estructurar los estímulos sensoriales para establecer estructuras básicas de agrupación y organización al momento en el que presentemos al espectador dichos estímulos (sonoros), toda vez que conocemos de manera general el modo en el que (dependiendo de su naturaleza y proceso cognitivo asociado) son aprehendidos por éste. Para tal fin la herramienta más directa y lógica es, por supuesto, la teoría Gestalt, de la cual se presentan algunas características generales a continuación.

1.3 Principios básicos de la teoría Gestalt en la percepción

La teoría Gestalt fue un movimiento teórico que surgió en Alemania a principios de la década de 1910 a partir de los trabajos de Max Wertheimer, Kurt Koffka y Wolfgang Köhler. Marcó un referente de suma importancia en la historia de la psicología. El objeto primordial de estudio de la Gestalt consiste en los mecanismos mediante los cuales tendemos a agrupar, perceptualmente, objetos externos (primordialmente visuales) en estructuras de mayor dimensión organizacional de acuerdo con las relaciones que mantienen dichos objetos con sus alrededores. La consideración de una tendencia natural de agrupamiento posiciona a la Gestalt como una teoría holística y, hasta cierto punto, sistémica en el sentido de Bertalanffy. Sin embargo, como bien señala Duero (2003):

La expresión que afirma que el todo es más que la suma de las partes, ha sido una especie de emblema empleado para caracterizar a la psicología de la Gestalt. Sin embargo, lo cierto es que ninguno de sus fundadores (ni Wertheimer, ni Koffka ni Köhler), la emplearon jamás. Lo que en cambio sostuvieron es que hay eventos (tanto psicológicos como físicos) que resultan fundamentalmente diferentes a una colección de sensaciones, piezas o sumatoria de elementos.

En el trabajo de 1920 de Köhler, *Physical Gestalten*, el autor introduce el término *Gestalten*, que puede ser definido como “procesos” o fenómenos que revelan propiedades que son más que la suma de sus partes. Lo perceptual se entiende, en este contexto, como un fenómeno emer-

gente de la interacción de los elementos individuales o partes, donde el tipo más importante de interacción es nuestra tendencia a agrupar dichos elementos de acuerdo con ciertos principios. Köhler (1920) resalta, por ejemplo, que “la función específica de una parte individual es percibida de acuerdo con su ubicación relativa”.

Del enorme corpus que conforma la teoría Gestalt, la descripción de la llamada *organización de las formas perceptuales* resalta como herramienta central para este trabajo. “Dado un número de estímulos, tendemos no a percibir objetos individuales, independientes o experiencias, sino *todos* más grandes separados y relacionados con otros al mismo tiempo” (Wertheimer, 1923).

Uno de los aspectos primarios dentro de la teoría Gestalt es la idea del *fondo vs. figura*, la cual establece que ciertos estímulos sensoriales son percibidos en primer plano por destacar de un fondo explícito. Al variar la atención voluntariamente sobre dichos estímulos se puede intercambiar la posición (y por lo tanto, la importancia perceptual) de dichos elementos. Incluso, es posible crear una ambigüedad sobre la diferenciación *fondo-figura* a través del llamado *factor de inclusividad* el cual, en general, provoca un grado de homogeneidad entre el fondo y la figura, tal y como sucede en el conocido ejemplo de los vasos y los rostros.

Wertheimer detalla y estructura categóricamente la descripción anterior a manera de leyes o principios los cuales se enuncian de la forma en que son presentados en Brownie (2006):

- *Factor de proximidad*. Este principio establece que los objetos que están localizados cercanos unos con otros serán percibidos con una asociación entre ellos, *i.e.* pertenecientes a un mismo grupo o a partes de un todo más grande.
- *Factor de similaridad*. Este principio afirma que, para los objetos que comparten propiedades similares, se asume que tienen una asociación entre ellos.
- *Factor de destino uniforme* o dirección común. Sucede cuando perceptualmente se agrupan, como figura única, un conjunto de objetos que presentan factores descriptivos altamente similares:

Implica que los elementos que parecen construir un patrón en la misma dirección, son percibidos como una figura. Por ejemplo, cuando visualizamos una bandada de pájaros que vuelan como un todo como si se tratara de un ente único que va formando figuras en movimiento en el aire (Torreblanca, 2017).

- *Ley de la buena configuración* o ley de Prägnanz. Establece que el arreglo de elementos más simple es el más plausible de ser percibido en primer lugar. “Cuando se nos presenta una serie de elementos complejos, tendemos a percibirlos como partes de un todo más simple” (Wagner *et al.*, 2006).
- *Factor de dirección* o ley de la continuidad. Afirma que cuando un conjunto de objetos se encuentra acomodado de modo tal que hace referencia a una forma lineal, es percibido con ciertos tipos de *ramificaciones*. La rama que sigue la dirección de la línea original de manera más sucinta es percibida como la continuación del camino original, mientras que los demás elementos son percibidos como agregados o anexos.
- *Factor de cerradura*. Los objetos que están cercanos entre sí son percibidos como partes de un todo, en la medida en que las brechas o divisiones entre ellos puedan ser imaginadas como *cerradas*; “es la tendencia de percibir material no terminado como objetos más terminados de lo que realmente son” (Skjærven, 2008).
- *Factor de la buena curva*. Si la continuidad de una línea a otra es percibida, esto puede anular otras posibles formas o agrupamientos; “la influencia de la tendencia hacia una buena Gestalt” (Wertheimer, 1923).
- *Factor del contraste*. Un objeto es capaz de distinguirse de otro mediante la especificidad de sus propiedades o atributos perceptuales; esto es, a través de su singularidad.

Wertheimer (1923) explica también que es posible encontrar diversas situaciones donde dos o más principios se pueden aplicar simultáneamente, ya sea de una manera constructiva (es decir reforzándose entre ellos) o sustractiva (compitiendo y anulándose entre ellos):

Por ejemplo, si existe un grupo de objetos de apariencia variada y que están localizados en proximidad cercana y otro grupo conteniendo objetos que son similares en apariencia a aquellos en el primer grupo pero que se encuentran alejados, entonces surge un conflicto entre el factor de proximidad y el factor de similaridad (Brownie, 2006).

Sin embargo, también es muy importante señalar que estos principios no son inamovibles y que su jerarquía y aplicación es totalmente variable dependiendo tanto del individuo que percibe como del contexto.

Los principios Gestalt se han utilizado (o mejor dicho referenciado) en ciertas ocasiones en el campo de la música electroacústica debido a su natural conexión con la idea de asociación

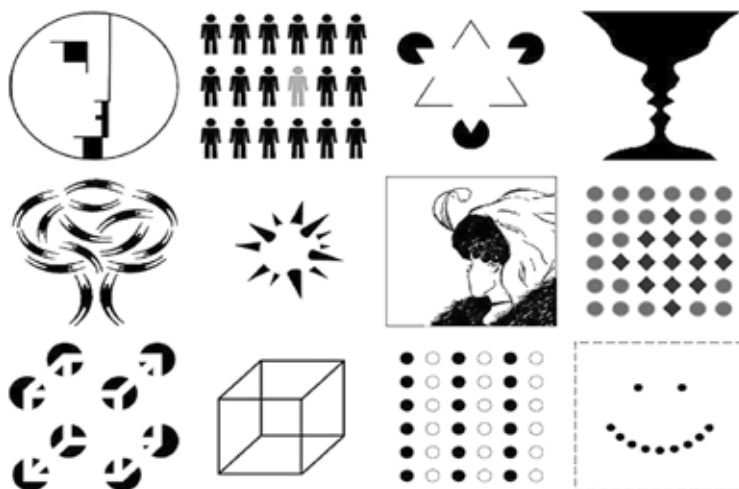


Figura 1 | Principios Gestalt (Soegaard, 2020. Autor: Impronta. Copyright: CC BY-SA 3.0).

perceptual de materiales sonoros para la construcción de ensambles aurales o texturas. Aquí se propone que estos principios pueden ser utilizados para el diseño y creación de espacialidad sonora cuando se trabaja con electroacústica multicanal y que, más allá de ser una formalización teórica, resulta (al menos desde mi práctica personal) una herramienta adecuada para establecer componentes espaciales que puedan tener una intención dinámica clara en el proceso compositivo, de modo tal que el diseño espacial no sea sólo el resultado directo y explícito del uso de los *plugins* y herramientas computacionales de espacialización.

2. Analogía relacional

Para finalizar esta sección se mostrará un panorama general básico de la llamada *analogía relacional*, la cual permite establecer un marco referencial para comprender procesos cognitivos más complejos (en comparación con las reacciones inmediatas a los estímulos sensoriales) asociados al razonamiento. En este sentido, el lector o lectora podrá observar, en esta sección, el tránsito de una comprensión básica del estímulo sensorial (al inicio del capítulo) al establecimiento de una perspectiva específica para entender dichos estímulos y, a partir de ello, establecer relaciones contextuales con nuestro conocimiento y experiencias previas. Nuevamente esto resultará útil para el momento en el que busquemos definir el espacio de manera

conceptual, y también para plantear los procesos compositivos asociados a la espacialidad apelando justamente a la experiencia previa del espectador. Por supuesto, el lector o lectora podrá establecer de manera inmediata una asociación de la analogía relacional con la escucha reducida de Schaeffer y, por lo tanto, la concepción de la escucha acústica y la escucha ecológica de Gibson (1966, posteriormente planteada por Erick F. Clarke).

La *analogía relacional* puede considerarse como un subcampo de estudio de las ciencias cognitivas, en particular del área que se enfoca en el estudio de la *similaridad* y la *analogía* como componentes estructurales del razonamiento, tal como dan cuenta los trabajos de Gentner (1980, 1983, 1989), Gick, Holyoak (1980, 1983) o Hofstadter (1995), por mencionar algunos. Por ejemplo, la visión ofrecida por Gentner sobre el concepto de analogía es la siguiente:

[...] puede ser vista como un tipo de razonamiento donde el conocimiento es transferido de una situación inicial (llamada *fuentes*) a otra final (llamada *objetivo*) sobre la base de algún tipo de similaridad entre ambas situaciones (Kokinov, 1996).

Para Hofstadter (1995), “la analogía puede ser vista como un tipo de percepción de alto nivel donde una situación es percibida como (en términos de) otra”. Por su parte (Hall, 1989) enumera cuatro procesos abstractos que son ampliamente considerados para el razonamiento por analogía:

- 1) Reconocimiento de la fuente, dada una descripción del objetivo.
- 2) Elaboración y evaluación del mapeo entre fuente y objetivo.
- 3) Transferencia de la información de la fuente al objetivo.
- 4) Consolidación del resultado; *i.e.* aprendizaje.

De acuerdo con Holyoak y Thagard (1995), existen al menos tres criterios específicos bajo los cuales es posible identificar dos situaciones análogas:

- *Correspondencia estructural*. La necesidad de encontrar un mapeo inyectivo (uno a uno) entre los elementos de ambas situaciones y de preservar las relaciones entre ellos.
- *Similaridad semántica*. Relación establecida entre los elementos de ambas situaciones.
- *Correspondencia* establecida entre los aspectos más importantes de ambas situaciones con respecto al objetivo del razonador.

Cabe señalar que gran parte de los estudios sobre analogía se han realizado como implementaciones computacionales para el reconocimiento automático de elementos visuales, así como para el modelado computacional de la percepción y aprendizaje humano (sobre todo en el campo visual). Sin embargo, la esencia primordial de la analogía relacional para este trabajo radica en la formulación concreta del modo en que absorbemos nuevos conceptos de acuerdo con la relación que hacemos de conceptos familiares aprendidos previamente; dicha absorción involucra un proceso perceptual-cognitivo en el sentido de *ciclo acción-percepción* de Foster. En el ámbito audible, que es el que atañe a esta investigación, la analogía relacional funciona como una herramienta de destacada utilidad en el proceso perceptual sonoro-musical; esto es, la manera en cómo procesamos nuevos sonidos o configuraciones de sonidos, lo que en general ocurre a través de este tipo de asociaciones preestablecidas. Este problema es central en el concepto de *la escucha reducida* de Schaeffer, *la emancipación de la disonancia* de Schoenberg y en lo que James Tenney denomina *función musical extendida de los sonidos*. Reforzando lo anterior, French (2002) afirma que:

La analogía relacional dirige nuestra habilidad de explicar nuevos conceptos en términos de otros previamente familiarizados, para enfatizar aspectos particulares de situaciones, para generalizar, para caracterizar, para explicar o describir nuevos fenómenos, para servir como base de cómo actuar en entornos no familiares y para comprender muchos tipos de humor.

3. Sumario de definiciones y conceptos

En esta sección se establece, a modo de conjeturas y definiciones, un sumario de los conceptos mostrados en las secciones anteriores, así como una estructuración de los mismos que servirá para los fines propios de este trabajo. Estas conjeturas representan interpretaciones propias sobre los resultados mencionados y son presentados de modo tal para los fines ya descritos. Se hace la aclaración aquí que el término *conjetura* se utiliza en el sentido literal de la palabra: una afirmación derivada de observaciones o procesos lógicos que no ha sido completamente probada.

DEFINICIÓN 1.2.1. Una *escena perceptual* es una ubicación espacio-temporal de duración delimitada en la que un individuo recibe un *torrente multisensorial* a través de sus extero-

ceptores. Aquí, torrente multisensorial se entiende como el conjunto de estímulos externos percibidos a través de cualquiera de dichos exteroceptores. La *focalización de una escena perceptual* es la delimitación de dicho torrente multisensorial a un único exteroceptor con fines de análisis por un observador externo o por el mismo individuo. Por lo tanto, cualquier escena analizada de ese modo se denominará *escena perceptualmente focalizada*. En este sentido se puede hablar, por ejemplo, de focalización auditiva de la escena perceptual o escena auditivamente focalizada.

DEFINICIÓN 1.2.2. Dada una escena perceptualmente focalizada, se denominará *proceso fundacional cognitivo* al compuesto por las etapas de percepción pasiva, percepción activa y primera etapa del conocimiento.

CONJETURA 1.2.1. Dentro de una escena perceptualmente focalizada, la *preatención* es equivalente a la *percepción pasiva*, mientras que la *atención focalizada* es equivalente a la *percepción activa*. De este modo, el *proceso fundacional cognitivo* puede ser descrito también como aquel que está compuesto por la preatención, la atención focalizada y la primer etapa del conocimiento.

CONJETURA 1.2.2. Dada una escena perceptualmente focalizada, puede afirmarse que el proceso *bottom-up* está compuesto de la preatención en conjunto con la atención focalizada. Por su parte, el proceso *top-down* puede explicarse como componente primario de la primera etapa del conocimiento.

CONJETURA 1.2.3. Dada una escena perceptualmente focalizada, la preatención está compuesta por el *registro sensorial* (SR) seguido del almacenamiento a corto plazo (STS). De este modo, la preatención ocurre dentro del intervalo del *tiempo simple de reacción*. El almacenamiento a largo plazo (LTS) es el componente de transición entre la preatención y la atención focalizada; es elemento que permite que la preatención se convierta en atención focalizada.

CONJETURA 1.2.4. Dada una escena perceptualmente focalizada, la preatención en conjunto con la atención focalizada ocurren en el intervalo del *tiempo de reacción*.

CONJETURA 1.2.5. El proceso *top-down* es equivalente al conjunto de los procesos de *analogía relacional* de Hall (1994). ■ ■ ■

Capítulo II. Percepción sonora para la experiencia del Espacio-Timbre

En las secciones anteriores se introdujeron los fundamentos de las diferentes aproximaciones teóricas al fenómeno de la percepción a partir de las cuales se formularon conceptos y definiciones específicas. Dichas definiciones funcionarán como referentes básicos para abordar el fenómeno de la percepción sonora y, en particular, para establecer categóricamente las características acústico-perceptivas que definen y distinguen a la composición electroacústica, sobre todo aquellas relacionadas con el timbre y el espacio.

Como se mencionó, el objetivo de estas secciones no fue el de realizar un análisis profundo y comparativo sobre las distintas teorías de percepción, sino únicamente documentar algunos referentes y antecedentes formales conocidos que justifiquen las definiciones utilizadas en el presente trabajo. Posteriormente se desarrollarán los principios relativos a la percepción sensorial auditiva que fundamentarán el concepto de espacialidad e inmersividad espacial desde un punto de vista acústico y psicoacústico, con el fin de establecer claramente los elementos primarios que soporten definiciones posteriores de Espacio-Timbre.

Esta sección comienza con una revisión de los conceptos de la *teoría del análisis de la escena audible* o ASA (Auditory Scene Analysis), la cual permite estructurar los diversos aspectos, mecanismos y escenarios que constituyen los numerosos procesos de percepción del sonido en su nivel primario. Es esta característica (la de tratar al sonido en su nivel básico, anterior a la organización compleja que es la música en el sentido tradicional) es la que hace al ASA una herra-

mienta por demás adecuada para abordar un análisis teórico-práctico profundo de la música electroacústica, no sólo desde el punto de vista del espectador sino también desde la perspectiva del creador. Si bien dicha teoría fue formulada desde hace bastante tiempo, sus implicaciones siguen siendo válidas y referenciadas al día de hoy; por ello, a pesar de la distancia temporal, las afirmaciones y resultados poseen objetividad actual. Más adelante se desarrollarán los conceptos e ideas clave de la teoría de análisis musical de James Tenney (la cual deriva de los principios Gestalt y se relaciona de manera indirecta con los resultados del ASA).

1. Análisis de escena auditiva

El sonido proveniente del entorno que es captado por nuestros oídos constituye una mezcla compleja de una cantidad enorme de ondas sonoras. A través de esta ingente mixtura de componentes espectrales, nuestro sistema auditivo es capaz de agrupar subconjuntos de ondas de modo tal que puede identificar fuentes individuales dentro de una escena que, por lo general, está compuesta por una multiplicidad de ellas. Dicho de otro modo, “el sistema auditivo enfrenta el problema de descomponer de algún modo la onda de presión que llega a nuestros oídos con el fin de construir un número de representaciones separadas” (Bregman y Woszczyk, 2004).

El ASA (Auditory Scene Analysis) es un modelo psicofísico que constituye la base teórica de la investigación de la percepción auditiva; fue propuesto por Albert Stanley Bregman en 1990 en su famoso trabajo *Auditory Scene Analysis: the Perceptual Organization of Sound*. Esta investigación sirvió de referente para posteriores desarrollos en los campos de la percepción del habla, sistemas auditivos en humanos y animales, prótesis auditivas y, sobre todo, en el campo de la modelación computacional, ya que generó toda una línea de investigación denominada “análisis de escena auditiva computacional” (Computational Auditory Scene Analysis, CASA).

De acuerdo con Bregman y Woszczyk (2004), “ASA es el proceso mediante el cual el sistema auditivo agrupa de manera conjunta, a través de la frecuencia y tiempo, todos los componentes espectrales que pertenecen a cada fuente individual sonora para que pueda ser reconocida independientemente de las propiedades de sonidos concurrentes”. El proceso del ASA no es para nada trivial aunque pueda parecerlo en un principio; sobre todo porque, como bien puede corroborarse en un espectrograma, los componentes espectrales mezclados de to-

das las fuentes sonoras que conforman una escena audible cualquiera no ocupan por lo general un espacio separado en el espectro.

El correcto funcionamiento del ASA como sistema perceptual implica, además, una funcionalidad ecológica importante, ya que la discriminación de las fuentes sonoras individuales dentro de un entorno de complejidad espectral es vital para la supervivencia del organismo en dicho entorno. Esta funcionalidad ecológica desemboca entre otras habilidades; por ejemplo, en la de construir representaciones de los sonidos individuales y, mediante ellas, formar abstracciones relacionadas con los eventos que produjeron tales sonidos. Además conlleva, entre otras cosas:

- *Distinguir un mismo sonido o timbre en diferentes entornos o ambientes.* Entornos distintos poseen diferentes conjuntos de componentes frecuenciales; por tanto, existe una alta posibilidad de que el conjunto de componentes frecuenciales se traslape con los componentes espectrales del sonido en cuestión.
- *Asociar correctamente cadenas de palabras con sus respectivas fuentes.* De este modo se previene que los escuchas perciban sílabas habladas por distintas personas como si fueran parte de una misma palabra; de manera inversa, también se previene que una palabra única se perciba como si fuera emitida por distintas personas.

En términos generales, puede pensarse que el ASA, tras haber realizado el proceso de análisis antes descrito, concluye en alguna de dos acciones en cada instante: *segregar* o *agrupar*. El ASA entonces, segrega los subconjuntos de componentes espectrales (de la mezcla enorme proveniente de la onda compleja de sonido que incluye todos los eventos acústicos de una escena audible en particular) que corresponden a cada una de las fuentes individuales cuando las *evidencias físico-perceptuales* apuntan hacia ello, o bien, agrupa ciertos subconjuntos espectrales como un sólo evento acústico si la diferencia cuantitativa de dichas evidencias es muy pequeña.

El ASA opera mediante una ponderación dinámica de estas evidencias físico-perceptuales que incluyen diversos aspectos acústicos: contenido espectral, localización espacial, sincronía de *onsets*, morfología de la envolvente de amplitud, etcétera. De hecho, existe una fuerte evidencia de que la tendencia primaria del sistema auditivo, al percibir una masa compleja de componentes espectrales, es la de la integración o fusión:

Sólo cuando los detalles de la señal dan al sistema audible alguna razón para segregar partes de él, esto sucederá, por lo que cualquier transformación de la señal que difumine las propiedades, favorecerá la integración de la señal (Bregman y Woszczyk, 2004).

Bregman (1993) define ciertos procesos básicos utilizados por el escucha para descomponer mezclas sonoras complejas provenientes de cualquier escena audible, los cuales se encuentran diferenciados en función de una característica primaria: la presencia o ausencia de *esquemas previamente aprendidos* (la analogía relacional). En el caso en que la escucha ocurra a partir de esquemas aprendidos previamente se puede hablar, por un lado, de *reconocimiento automático* y, por otro, de *reconocimiento voluntario*. Cuando el proceso de escucha implica una ausencia de esquemas previos, el individuo realiza un análisis básico de la escena audible de acuerdo con características acústicas primarias y genéricas; este proceso se denomina *análisis de escena audible primitivo*.

El *reconocimiento automático* implica la activación de algún esquema previamente aprendido de manera no voluntaria y hasta cierto punto reactiva. Esto es, ante un estímulo auditivo externo, el escucha realiza una asociación perceptual de dicho estímulo con algún concepto, idea o definición (esquema) que ha aprendido previamente de manera automática o instintiva; de este modo segrega dicho esquema de la mezcla sonora. Según Bregman (1993), este proceso resulta aproximado, en el sentido de que no es necesario que el estímulo externo sea el equivalente exacto al esquema aprendido para que este último se active en el proceso perceptual del individuo.

El *reconocimiento voluntario* o *atención selectiva*, como proceso, es idéntico al *reconocimiento automático* con la única diferencia de que en este caso el escucha está concentrado en aislar de manera intencional un esquema de una mezcla compleja sonora. Es decir, dados un esquema previamente aprendido y una escena audible, el escucha busca segregar dicho esquema de la mezcla compleja audible.

El *análisis de escena audible primitivo* presupone que el escucha no tiene referencia previa alguna acerca del estímulo externo auditivo; por tanto, para poder segregar eventos sonoros específicos de una escena audible (o integrarlos con otros), el ASA emplea únicamente las características acústicas primarias, las cuales serán descritas en las secciones siguientes. Los esquemas de reconocimiento, tanto automáticos como selectivos, pueden considerarse como un caso especial del marco contextual de percepción de la analogía relacional descrita en el capítulo previo.

Todo lo anterior no implica que, en el caso de que exista reconocimiento mediante esquemas previos, el ASA no utilice o elimine por completo el análisis de las características acústicas mencionadas; más bien el proceso de segregación y/o integración se produce de una manera condicionada a dichos esquemas.

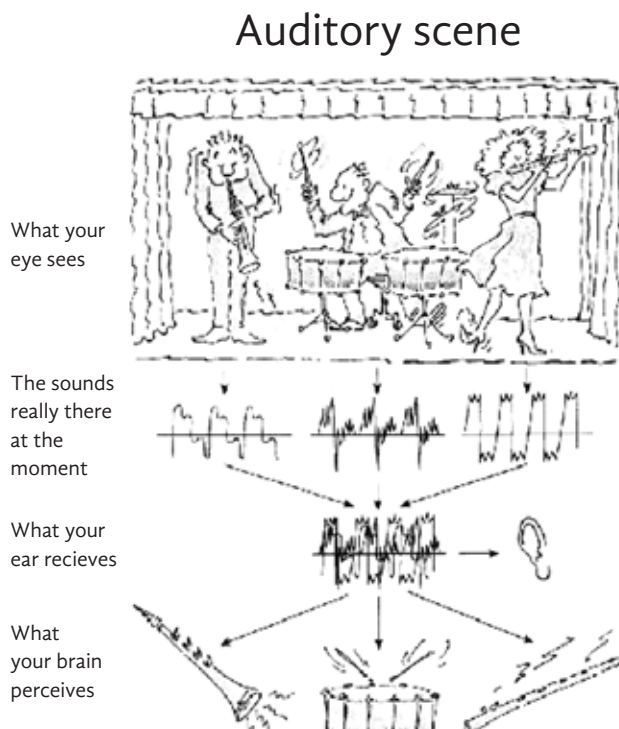


Figura 2 | Representación del funcionamiento del ASA (Fuente: Diederich, 2005).

1.1. Modos de agrupamiento: simultáneo y secuencial

Las escenas audibles que ocurren en cualquier entorno presentan una multiplicidad de fenómenos acústicos que deben ser analizados por el ASA. En cada caso, de acuerdo con la ponderación de las evidencias físico-perceptuales, se realiza alguna de las dos acciones antes mencionadas: segregar o agrupar. Este proceso ocurre para eventos que suceden *simultáneamente* o *sucesivamente* en la escena audible. Para el primer caso, cuando existe *agrupamiento simultáneo*, es posible poner como ejemplo la situación en la que tres personas hablan al mismo

tiempo; en este escenario, el ASA agrupa los componentes acústicos relacionándolos en tres representaciones perceptuales asociadas a cada una de las voces, las cuales poseen características propias como altura y timbre.

Para el segundo caso, el *agrupamiento secuencial*, el ejemplo básico consiste en la capacidad del ASA de asociar o agrupar en un mismo evento acústico partes de una misma voz a través del tiempo.

Vale la pena hacer notar y recalcar que el desarrollo teórico del ASA está basado en la premisa de que el sistema auditivo ha evolucionado, se ha adaptado y se ha desarrollado a través de la observación y asimilación de las regularidades y detección de patrones en el entorno; ello le ha permitido asociar la información sensorial detectada a la *causa más probable* de acuerdo con el patrón observado. Lo anterior se denomina *principio de validez ecológica* (Bregman y Woszczyk, 2004) y coincide con el concepto de *reacción ecológica* de Shepard (1961).

Agrupamiento simultáneo

Como ya se mencionó, el agrupamiento simultáneo implica que, dada una escena audible, ocurre la acción de fusionar distintos subconjuntos de componentes frecuenciales que suceden en un mismo instante de tiempo, así como su asociación con fuentes individuales para crear las representaciones perceptuales correspondientes. Esta acción se lleva a cabo en función de distintos factores propuestos por la teoría ASA, los cuales, de acuerdo con ciertas circunstancias y restricciones específicas (relacionadas con las evidencias físico-perceptuales), funcionan en conjunto (a modo de una ecuación de combinación no lineal) para realizar dicha tarea:

DEFINICIÓN 2.1.1. Los factores de agrupamiento simultáneo del ASA son los siguientes:

- Principio de armonicidad.
- Asincronía de *onsets* y *offsets*.
- Independencia de envolventes.
- Separación espacial.
- Separación espectral.

Principio de armonicidad

Este principio se sustenta en la cualidad física de los *sonidos armónicos* (también llamados *periódicos*) cuyo contenido espectral está compuesto por una serie de múltiplos enteros de una frecuencia fundamental. En la vida cotidiana, una gran multiplicidad de sonidos presentan esta propiedad; ejemplos de ellos son la voz humana, sonidos de animales, instrumentos musicales, etcétera.

Usando esta propiedad, el ASA agrupa todos los componentes frecuenciales que son múltiplos de una misma frecuencia base y genera una representación perceptual separada asociada a dicho subconjunto segregándolo del resto de la masa sonora proveniente de la escena audible. Ésta es una de las observaciones concluyentes más sobresalientes de la teoría del ASA y debe notarse que sólo es válida para los sonidos que poseen las características antes mencionadas.

Uno de los experimentos más conocidos que ejemplifica el principio de armonicidad es la demostración 18 de Bregman y Ahad (1996), en la que el tercer armónico de un tono complejo es percibido como separado del evento audible completo tras ir disminuyendo conjuntamente y de manera gradual los demás componentes frecuenciales. Esta observación deriva, de nuevo, de una funcionalidad ecológica ya que en cualquier entorno es altamente improbable que un conjunto de componentes frecuenciales múltiplos de una fundamental provengan de fuentes distintas.

Asincronía de *onsets* y *offsets*

Esta característica es de suma importancia para el autor del ASA (Bregman, 1993) y la presenta como la primera de las cuatro *regularidades ambientales* que conforman la explicación fundamental de su teoría. Estas regularidades son utilizadas por el sistema auditivo para resolver escenas audibles de manera genérica. Esta característica podría definirse como sigue:

OBSERVACIÓN 2.1.1. Dada una escena audible, los componentes acústicos derivados de eventos ambientales independientes tienden a no empezar y terminar sincrónicamente. Por esta razón, asincronías de más de unos cuantos milisegundos en los *onsets* y *offsets* de los componentes acústicos son tomadas como evidencia por el ASA de que provienen de diferentes sonidos.

Una consecuencia directa de la característica anterior es la habilidad que tiene el sistema auditivo para separar el sonido directo de las reflexiones dentro de un recinto reverberante; de hecho, el sistema auditivo parece ser capaz incluso de diferenciar las primeras reflexiones de las tardías a través de esta característica. Debe notarse que este principio pierde resolución en ambientes altamente reverberantes debido a que el detalle granular de los *onsets-offsets* de diversos sonidos queda difuminado por el efecto de la reverberación.

Independencia de envolventes

OBSERVACIÓN 2.1.2. Dada una escena audible, la asincronía en la morfología de las fluctuaciones de intensidad (envolventes de amplitud) de los componentes acústicos es tomada por el ASA como evidencia de que dichos componentes forman parte de sonidos que provienen de fuentes distintas. De manera inversa, si las envolventes de amplitud están sincronizadas a través del tiempo, el ASA resuelve el estímulo sensorial como proveniente de un mismo sonido.

El mismo Bregman anota respecto de la evidente similitud entre las dos características hasta ahora descritas: “es posible que las características de asincronía de *offsets* y *onsets* junto con la independencia de envolventes, puedan ser reducidas a un mismo principio puesto que los *onsets* y *offsets* pueden ser vistos simplemente como un cambio en la envolvente de amplitud”. Como se mostrará al final de esta sección, es posible plantear una generalización de los principios presentados por Bregman utilizando la espectromorfología de Dennis Smalley.

Separación espacial

OBSERVACIÓN 2.1.3. La percepción de los componentes acústicos con localización espacial diferente sirve como evidencia para el ASA para segregar los sonidos como producidos por fuentes distintas.

Es notable que en la mayoría de los escritos de la teoría del ASA se comente que la localización espacial, si bien puede ayudar al proceso de segregación de fuentes distintas dentro de una

escena audible, no es un factor determinante ni con la suficiente fuerza como para ser considerado por separado; *i.e. la localización espacial tiene funcionalidad para el proceso perceptual del ASA únicamente como soporte de otros factores*. Este hecho es descrito con mayor detalle para el caso del *agrupamiento secuencial*.

Como se verá enseguida, sonidos con cierta espectromorfología tienen por lo regular zonas restringidas de localización espacial en el azimut y el plano de elevación. Un ejemplo comúnmente referido es el que trata sobre sonidos que ocurren detrás del escucha; puesto que la fisiología de las orejas tienen una disposición frontal, la resolución espacial en la ubicación posterior resulta disminuida en comparación al azimut frontal. De igual modo, en ambientes cerrados o semiabiertos, para ciertos casos, las reflexiones del sonido pueden distorsionar la percepción de la ubicación real de la fuente. Debido a estas y otras razones, el ASA no puede valerse únicamente de la localización espacial para realizar la segregación de fuentes simultáneas.

Separación espectral

Este principio se refiere a sonidos cuyo contenido espectral se encuentra más o menos bien delimitado en bandas definidas (por lo general asociados a la inarmonicidad o al ruido); entonces, el agrupamiento simultáneo estará en función de la distancia en el espectro entre dos sonidos de este tipo. Sonidos con bandas espectrales concurrentes serán agrupados como un solo objeto perceptual, mientras que sonidos separados en el espectro serán percibidos como individuales.

Agrupamiento secuencial

El agrupamiento secuencial que realiza el ASA se vincula con la asociación de características sonoro-perceptuales *a través del tiempo* dada una mezcla compleja de contenido espectral. El ASA agrupa los eventos sonoros en sucesiones específicas separándolas de otras que presenten características acústicas distintas, de tal modo que el organismo sea capaz de inferir las diversas fuentes individuales presentes en la escena audible. Las sucesiones percibidas cuyas partes forman una entidad perceptual única se llaman *torrentes auditivos*.

DEFINICIÓN 2.1.2. Un *torrente auditivo* (*auditory stream*) es la unidad perceptual que representa un evento individual. Puede incorporar más de un sonido simultáneo y contiene la representación mental de la información perceptual recibida debido al evento físico. Esta representación tiene además como objetivo, agrupar secuencialmente cualidades relacionadas con el evento. La causa física es denominada *evento acústico* y las unidades auditivas físicas son denominadas sonido.

La definición anterior aclara la diferencia entre *sonido* y *torrente auditivo*; el primero está ligado a la parte física y representa la unidad en ese ámbito, mientras que el segundo está ligado a la parte perceptual-cognitiva representando la unidad en ese otro ámbito. “Nuestras representaciones mentales de los eventos acústicos pueden ser tan múltiples en un modo que la simple palabra sonido no lo representa. Al acuñar un nuevo término, *torrente*, somos libres de cargarlo con cualesquiera propiedades teóricas que nos parezcan apropiadas” (Bregman, 1990).

Uno de los experimentos más conocidos respecto del agrupamiento secuencial es el llamado *efecto streaming* (Bregman, 1990), el cual permite la introducción de un concepto de suma importancia: la *distancia acústica*. Si bien Bregman no clarifica una definición para dicho concepto, establece un marco lo suficientemente adecuado para comprender esta idea: “la distancia acústica d es la diferencia entre dos tonos simultáneos, A y B , como la combinación (no necesariamente lineal) de sus separaciones en un número de dimensiones físicas” (Bregman y Woszczyk, 2004). De acuerdo con la métrica planteada por los autores, valores bajos de d favorecen la integración de los eventos A y B en una sola secuencia perceptual, mientras que a la inversa, valores altos de dicha distancia representan una tendencia a la segregación.

Al igual que en el caso del agrupamiento simultáneo, para el agrupamiento secuencial existen factores explícitos que contribuyen a la formación de secuencias perceptuales:

- *Separación frecuencial*. A mayor separación del contenido espectral entre dos sonidos, mayor tendencia de segregación perceptual.
- *Separación temporal* entre los *onsets* de los eventos sonoros.
- *Separación de la frecuencia fundamental*. Para el caso de sonidos periódicos, la separación de la fundamental entre dos sonidos favorece la segregación perceptual.
- *Diferencias en la forma del espectro*. La morfología de la amplitud del contenido espectral. Esta característica se asocia además al concepto del timbre y, por lo tanto, implica

que el escucha será capaz de agrupar secuencialmente un estímulo externo si su grado de variación no es muy alto a través del tiempo.

- *Diferencia en la localización espacial.*
- *Diferencias en los centros de gravedad espectrales.* Para el caso de sonidos de ruido.
- *Diferencias en la intensidad.*
- *Diferencias en la duración del ataque de la envolvente de amplitud.*
- *Diferencia en el grado de ruido de los sonidos.*
- *Efecto acumulativo.* El ASA agrupa secuencias perceptuales las cuales confirman, después de cierto lapso de tiempo, que dichos estímulos siguen siendo escuchados de manera más o menos continua. Se ha calculado que la percepción separada de dos tonos en distintas regiones frecuenciales se construye máximo en cuatro segundos, y que puede tardar otros cuatro segundos para que esta predisposición desaparezca.

Al igual que en el agrupamiento simultáneo, no es posible establecer una jerarquía lineal sobre los factores anteriores y muchas veces sucede una *competencia* entre ellos. De acuerdo con los autores, los últimos tres principios son factores más bien débiles que pocas veces resultan relevantes por sí mismos; no obstante, en conjunto con otros factores pueden ayudar al ASA a segregar o agrupar correctamente los estímulos perceptuales externos.

Existe una propiedad acústico-perceptual de las fuentes sonoras que es de suma importancia para el agrupamiento secuencial del ASA: *la suavidad de cambio*.⁴ Dicha característica tiene una base ecológica y de adaptación evolutiva en el sentido de que en la mayoría de los entornos, “cuando los objetos físicos interactúan para producir sonidos, las propiedades de estos objetos físicos, tenderán en promedio, a cambiar relativamente lento y continuo a través del tiempo” (Bregman y Woszczyk, 2004).

1.2. Atributos de percepción en el ASA

Un funcionamiento eficiente de la teoría del ASA, se relaciona con la creación de representaciones mentales de torrentes auditivos bien definidos; en este sentido, tendemos a adjudicar ca-

⁴ La cual generaliza el principio anterior de la *forma del espectro*.

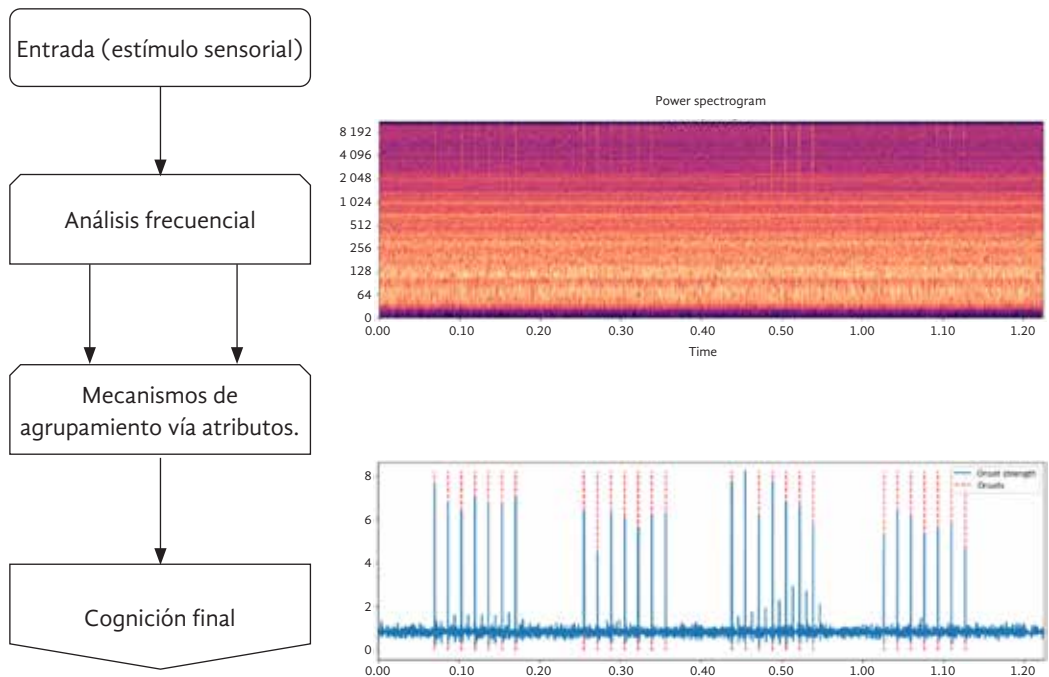


Figura 3 | Flujo general de procesamiento del ASA.

racterísticas particulares a dichos eventos de acuerdo con preconcepciones que nos son familiares. En otras palabras, utilizamos el *proceso de analogía* (asociado a la analogía relacional) para poder establecer relaciones con propiedades específicas de los torrentes auditivos de modo tal que sean interpretados por nuestro cerebro como entidades separadas. Cada torrente o *stream* lleva implícito un cúmulo o *cluster* de propiedades asignadas por el escucha, por lo que este *cluster* variará según el individuo y los diferentes contextos. En este sentido, una de las formas más básicas de representación mental de los objetos externos (y de descripción verbal de los mismos) consiste en la descripción de sus propiedades específicas o atributos tal y como los aprehendemos en un momento dado.

La analogía en el campo visual de la idea anterior puede resumirse que “en nuestras descripciones verbales de lo que vemos, decimos que un objeto es rojo, o que se está moviendo rápido, que está cerca o que es peligroso” (Bregman, 1990). De igual modo, la representación mental de los eventos acústicos externos está dada por la atribución de propiedades descriptivas en un momento dado; un sonido puede ser, por ejemplo, *lejano, cercano, fuerte, bajo, móvil, fijo, consonante, disonante o ruidoso*. El *stream* es entonces una suerte de pagamento que cohe-

siona estas propiedades; al mantenerlas unidas, permite la abstracción de eventos individuales y, por lo tanto, identificar las fuentes separadas que les corresponden.

Las propiedades o atributos de los objetos son abstracciones mentales que se crean en el proceso cognitivo; son etiquetas que colocamos en los objetos del mundo externo para poder describirlos y realizar representaciones mentales de ellos. Esto significa que “las propiedades necesitan pertenecer a algo” (Bregman, 1990). En este sentido, como ejemplifica el autor, el *stream* o torrente nos permite realizar agrupaciones de estas propiedades de modo tal que podemos ser capaces de distinguir las fuentes origen de donde provienen.

Lo anterior puede ejemplificarse si, como describe Bregman (1990), suponemos que existen dos fuentes acústicas de sonido, una “alta-y-cercana” y la otra “baja-y-lejana”. Debido a que los atributos de *cercanía* y de *altura* son agrupados como propiedades de un *stream*, y los de bajo y lejano como propiedades de otro, podemos experimentar la unicidad de dos sonidos individuales sin que parezca una masa amorfa de cuatro atributos. Lo anterior puede ser enunciado del modo siguiente:

DEFINICIÓN 2.2. El *Principio de pertenencia de atributos* es uno de los mecanismos ASA realizados para ser capaces de identificar eventos acústicos individuales mediante la descripción mental *conjunta de atributos descriptivos* a dicho evento en un instante dado.

Para finalizar esta sección se hablará de la heurística *old-plus-new*, la cual ha probado ser una de las herramientas fundamentales dentro del ASA para poder resolver escenas auditivas de manera adecuada. En general, puede ser entendida como el proceso a través del cual el ASA agrupa secuencias de eventos a través del tiempo (dado el análisis de las características acústicas de las secuencias previas) que se encuentran escondidas en la mezcla compleja:

Si es de algún modo posible interpretar cualquier sección de un grupo de componentes acústicos como la continuación de un sonido que ha ocurrido, hacerlo y removerlo de la mezcla sonora. Tomar entonces la diferencia entre el sonido actual y el sonido previo como el nuevo grupo que deberá ser analizado (Bregman, 1990).

Lo anterior puede además ejemplificarse con un caso muy simple. Cuando un espectro se vuelve más complejo a lo largo de un intervalo de tiempo, la heurística *old-plus-new* establece la siguiente solución: se trata de un sonido que continúa pero acompañado de los componen-

tes adicionales que se añadieron a la mezcla. Esto es, la percepción constará de dos elementos: el sonido primario más los componentes añadidos posteriormente. Puede observarse que esta heurística se relaciona de cierto modo con el factor de dirección de la Gestalt.

2. La teoría Gestalt de la música de James Tenney

A continuación se explica la perspectiva Gestalt de James Tenney al respecto de la composición musical, en la cual se pueden vislumbrar diversos elementos de la percepción sonora, el ASA y posibles extrapolaciones para el desarrollo de la conciencia audible espacial. Sin embargo, el motivo principal para presentar este tema consiste en establecer un referente composicional que, desde lo conceptual y lo composicional, pueda ser trasladado a la noción de timbre en el contexto del *Espacio-Timbre*, entendiendo que el ASA funciona como un primer acercamiento perceptual a la relación formal entre el contenido espectral del sonido y la capacidad humana de percepción (lo que podríamos entender como una noción prototípica de timbre).

James Tenney (1934 - 2006) fue uno de los compositores y teóricos de la música contemporánea más importantes del siglo XX. Alumno de Cage, Ruggles y Varesse, su obra transita entre diversos estilos que abarcan la música algorítmica, el minimalismo, la música por computadora, el espectralismo y el microtonalismo por mencionar algunos. Como investigador, el compositor norteamericano trabajó profundamente en el campo de la psicoacústica y la composición algorítmica y sus contribuciones más conocidas son las concernientes a su teoría musical basada en la Gestalt, expuesta por primera vez en su tesis de maestría *Meta-Hodos* de 1961.

Catorce años después, Tenney publicó *META Meta+Hodos*, documento que representa “un intento de organizar ciertas ideas presentadas por primera vez en *Meta+Hodos*, incorporando reflexiones y revisiones que han emergido desde entonces” (Tenney, 1975). Desde este momento profundiza en el tema con la colaboración de Larry Polansky, desarrollando material teórico de gran importancia entre el cual destacan, para propósitos del presente trabajo, los conceptos de *métricas morfológicas* y *percepción jerárquica gestalt musical*.

La teoría de Tenney está mayormente enfocada a la música instrumental contemporánea; su impacto ha sido enorme en la exploración y desarrollo de ciertas corrientes dentro de dicha área.

2.1. La complejidad sonora y los cúmulos tonales

Una de las primeras afirmaciones del *Meta+Hodos* es que el desarrollo de la complejidad de la música en el siglo XX se debe no sólo a una nueva forma de organizar los elementos o materiales musicales tradicionales, sino a una evolución de los mismos; esto, a su vez, ha afectado la percepción musical del espectador, quien demanda nuevos niveles de organización y asimilación sonora.

Tenney propone un tipo de composición analítica de la música en dos direcciones: vertical y horizontal. La primera tiene que ver directamente con la complejidad armónica-sonora y la segunda con el cambio de la complejidad de los parámetros musicales a través del tiempo. Por supuesto, es posible relacionar la dirección vertical con el agrupamiento simultáneo y la horizontal con el agrupamiento secuencial del ASA.

Para el caso de la complejidad sonora, el autor puntualiza sobre el espectro sonoro continuo, a partir del cual es posible clasificar al menos tres puntos de referencia. Al respecto, considérese la densidad sonora como la propiedad de acumulación tímbrica de un elemento musical. Bajo la perspectiva de Tenney, los elementos menos densos (y por lo tanto menos complejos) son aquellos conformados por acordes consonantes o por la percepción de consonancia en sí. El extremo contrario, esto es, los elementos más densos, son aquellas configuraciones sonoras cuya complejidad tímbrica impide percibir una altura definida y, por lo tanto, la asimilación por parte del escucha queda directamente relacionada con el ruido. Esto es lo que podría llamarse *escala de complejidad tímbrica*. En el punto medio de esta escala se encuentran los elementos musicales que son parcialmente percibidos como armónicos pero que no pueden ser completamente resueltos en tal sentido por el escucha.

Tenney introdujo en su momento, el uso de lo que él llama *tone clusters* o *cúmulos tonales*, dentro de la música contemporánea del siglo XX, y que son definidos como elementos musicales que pueden ser localizados en la segunda mitad de la escala antes descrita (es decir entre los elementos parcialmente armónicos y el límite de la complejidad sonora cercana al ruido) y cuya función armónica es relegada a un segundo plano para ser muchas veces sustituida por una función más bien percusiva y por lo tanto rítmica.

En este sentido, la concepción de los elementos o materiales musicales deja de estar definida de manera negativa con respecto de los parámetros tradicionales (atonal, matemático, etcétera) y establece un nuevo paradigma en el cual cualquier elemento a lo largo de la escala de

complejidad tímbrica tiene el mismo potencial de ser utilizado como componente musical. Lo anterior se vincula con la afirmación de Schoenberg (1950) en *Estilo y la Idea*:

Lo que distingue a la disonancia de la consonancia no es un mayor o menor grado de belleza, sino un mayor o menor grado de comprensibilidad [...] El término emancipación de la disonancia hace referencia a su comprensibilidad, la cual se considera equivalente a la comprensibilidad de la consonancia. Un estilo basado en esta premisa trata a la disonancia como consonancia y renuncia a un centro tonal.

Bajo este planteamiento, Tenney extiende la idea propuesta por Schoenberg y establece uno de los primeros elementos constitutivos de su teoría: el *principio de equivalencia*. Este principio afirma que todos los diferentes tipos de sonidos (específicamente los descritos conforme la escala de complejidad tímbrica) pueden ser usados como materiales elementales de construcción musical con igual potencialidad (Tenney, 1961). La consecuencia inmediata de este principio, según Tenney, es que ahora los elementos mínimos de construcción musical en la dirección vertical (altura-armonía) ya no son los tonos (o notas) sino una generalización conceptual de los mismos: *configuraciones de sonidos*.

Sobre la complejidad en la dirección horizontal, Tenney destaca dos factores principales que por sí mismos son capaces de caracterizar dicha configuración respecto de los parámetros musicales (altura, timbre, dinámica, densidad temporal):

- La extensión o rango de los parámetros composicionales a través del tiempo. Por ejemplo, en el caso de la dinámica: [ppp,pp,p,mp,mf,f,ff,fff].
- La tasa de cambio de los parámetros composicionales a través del tiempo. Esto es, qué tanto se introducen variaciones del rango del parámetro dentro de intervalos de tiempo específicos.

Bajo esta mirada, el timbre es el parámetro cuyo rango se ha extendido mucho más en la exploración composicional, sobre todo a lo largo del siglo XX, tomando en cuenta las diversas y numerosas manifestaciones de técnicas extendidas que han proliferado para la mayoría de los instrumentos.

La tasa de cambio de los diferentes parámetros musicales también constituye un referente distintivo de la música del siglo XX. Las exploraciones composicionales respecto de los cam-

bios de densidad, dinámica y timbre en intervalos relativamente cortos de tiempo son resalta- bles y su uso tiende a consolidarse como material compositivo y método de tratamiento de los materiales sonoros que integran una obra.

Las exploraciones sonoras que actúan de acuerdo con el rango y a la tasa de cambio de los diversos parámetros musicales a través del tiempo generan elementos nuevos de complejidad inusual cuya integración a la percepción del escucha no es tan directa la mayoría de las veces. Quizá, el mayor obstáculo que se presenta en tal situación es lo que Schoenberg llamó *efecto de interrupción de sentidos* con respecto a la asimilación perceptual del tratamiento de las disonancias como material musical al mismo nivel de las consonancias. Mientras exista mayor asimilación perceptual de los elementos musicales novedosos por parte del escucha, este efecto de interrupción de los sentidos disminuirá y dará paso a una nueva configuración de escucha y percepción musical que permitirá incorporar, cada vez mejor y con mayor énfasis y frecuencia, diversas manifestaciones de complejidad musical en cualquier parámetro: la emancipación de la disonancia (Schoenberg, 1941). De nuevo, puede verse aquí un reflejo tanto de la analogía reclacional como de los procesos *top-down* y *bottom-up*.

En este punto es pertinente notar que todos los conceptos introducidos en el capítulo anterior pueden ser invocados para comprender y extender las ideas propuestas por Tenney y autores afines. Esto se verá reflejado en las conjeturas listadas al final del capítulo.

2.2. La continuidad

Se acepta comúnmente que la continuidad es el atributo que le da unidad a una obra desde el punto de vista de la percepción. Es aquella característica inherente a una pieza que no sólo mantiene, sino que dirige la atención del espectador a través de una narrativa o discurso estético. En este sentido, lo opuesto, es decir la discontinuidad, sería la incapacidad de encontrar relaciones concretas entre los elementos constitutivos de una obra. Dado que la percepción de la continuidad está directamente condicionada por los distintos contextos socioculturales, una de las soluciones para generar las nuevas configuraciones de escucha que se han mencionado, surge de la aceptación, comprensión y asimilación por parte del público de la extensión del rango y la tasa de cambio de los parámetros musicales, tal y como se expuso en la sección anterior. Es necesario por lo tanto, como audiencia (sobre todo audiencia no familiariza-

da del todo con estas expresiones), no sólo saber reconocer la existencia de estos recursos como parte inmanente de la música contemporánea, sino permitirles entrar dentro de nuestro marco referencial de acercamiento conceptual a la música. Dicho de otro modo, es necesario crear nuevos puentes cognitivos y no limitarse por los procesos básicos de la analogía relacional. Tenney (1965) hace explícita esta idea al afirmar que el escucha que sólo puede aceptar la altura como factor primario de articulación de las ideas musicales, se limitará a escuchar espacios vacíos en gran parte de la música del siglo XX y podría eventualmente rechazar de manera conjunta algunas de las expresiones más avanzadas del arte musical.

Tenney señala además dos factores primordiales al momento de establecer la continuidad en la percepción: *la escala y la atención*. El factor de la atención lo subdivide en dos categorías que parecen directas desde el punto de vista lógico: *atención textural* y *atención paramétrica*.

Para el caso de la atención textural es necesario, dentro del razonamiento de Tenney, hacer una distinción entre los elementos importantes y aquellos menos importantes que componen una estructura, una sección o una obra completa en sí misma para un determinado lapso de tiempo de percepción. Es posible asociar esta formulación (a modo de clarificación al menos) con los conceptos de figura y fondo de un paisaje sonoro de Moles (1991) e incluso desde el punto de vista de la Gestalt. De este modo, la atención textural está asociada a la percepción de una estructura de cierto grado de complejidad dentro de una obra (o la obra misma), lo que deja en segundo plano los elementos individuales o aislados que son constitutivos de ésta. Mientras más atención se dirija a uno o varios de estos elementos de menor importancia, más se perderá la comprensión de las estructuras fundamentales de mayor escala de la obra en sí.

Por último tenemos la atención paramétrica, que es la que está dirigida hacia la articulación de los diversos parámetros musicales a través del tiempo dentro de pasajes o secciones específicas de una obra.

2.3. Unidades temporales Gestalt

Para Tenney la música puede ser considerada como una red ordenada y jerarquizada de sonidos, motivos, pasajes, frases, secciones, etcétera (Tenney, 1980). Dentro del contexto teórico de Tenney, la percepción de tales elementos sonoros como entidades individuales musicales en el tiempo es de vital importancia. Así, estos elementos son configuraciones de sonidos percep-

tualmente identificables como entidades individuales compuestas por sus elementos constitutivos, cualesquiera que sean. Esta percepción temporalmente delimitada de dichos sonidos o configuraciones de sonidos implica una aprehensión conceptual jerárquica: las configuraciones de sonidos individuales pueden ser consideradas como elementos constitutivos de subsecuentes configuraciones más grandes; es decir, con una delimitación temporal de mayor duración y así sucesivamente. Tenney denomina a dichas configuraciones *Temporal Gestalt-Units* (TGU).

Las TGU son temporalmente delimitadas; esto significa que es posible percibir esta individualidad en relación con otras TGU en la escala horizontal, es decir, con respecto al tiempo. Dicha identificación surge de dos características primarias: la percepción de la *cohesión* y la percepción de la *segregación*. En este sentido, para que una configuración de sonidos pueda ser entendida como una TGU, esta debe ser internamente cohesiva y externamente segregada. La figura 2.2.1 ilustra esta idea.

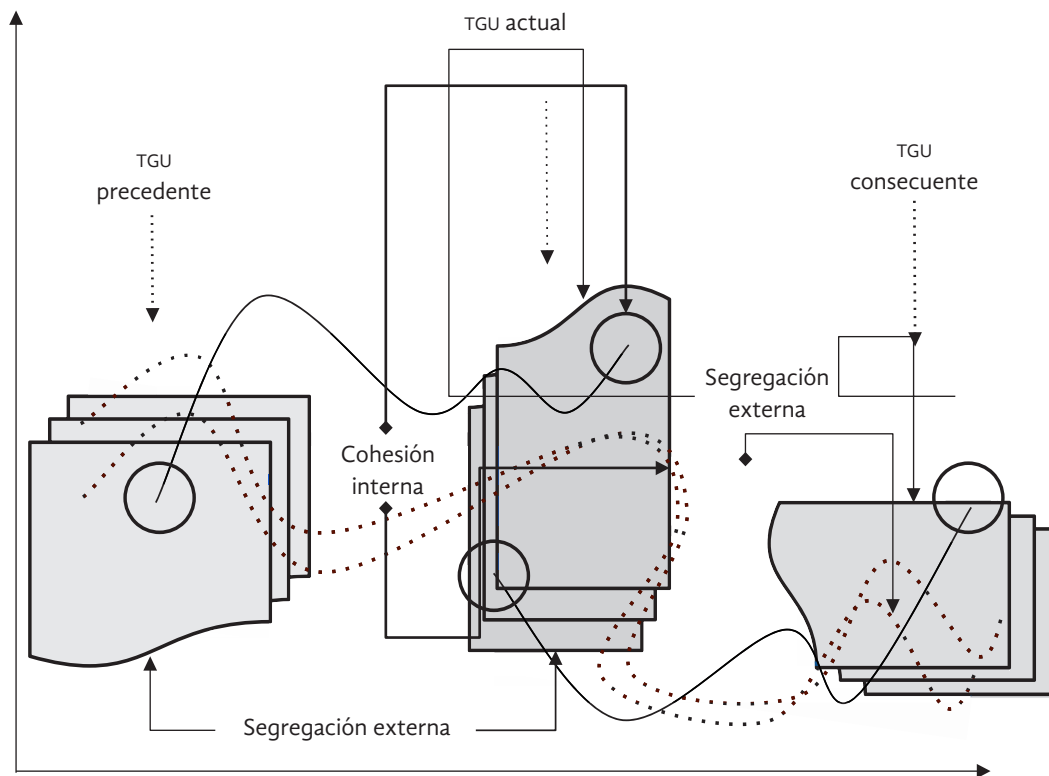


Figura 4 | Representación diagramática de las TGU.

Tenney define tres niveles principales de jerarquía.⁵ Dichos niveles se relacionan con la organización de las TGU de acuerdo con la escala de tiempo considerada al momento de ser asimiladas. El nivel primario y fundamental de dicha jerarquía corresponde a lo que él denomina *clang*. Desde la perspectiva de este primer nivel, el *clang* es la unidad musical mínima o lo que Tenney denomina *gestalt aural singular* y que representa el gestalt con mayor fuerza de cohesión. En el segundo nivel la percepción cohesiva disminuye y el gestalt asimilado es menos fuerte. Este segundo nivel es denominado por Tenney como *serie* y corresponde a una agrupación o sucesión de *clangs*; dicha sucesión, como TGU en sí misma, posee un grado reconocible de unidad y coherencia aunque, como se ya mencionó, de menor magnitud que el *clang*. En términos de Köhler, el *clang* representaría un *starke Gestalt* y la serie un *schwache Gestalt* (Köhler, 1958).

Organizaciones temporales de series dan lugar a segmentos e incluso a secciones completas dentro de una obra; sin embargo, Tenney considera como tercer y último nivel de percepción el correspondiente a la obra misma, donde la percepción temporal está claramente delimitada por el inicio y el final de la música de dicha obra.

3. Localización espacial del sonido

El problema de la localización espacial del sonido por el oído humano es un área de estudio bastante amplia que ha logrado resultados importantes a lo largo de su desarrollo como campo del conocimiento. Sin embargo, hasta ahora no se tienen resultados totalmente homogéneos capaces de resolver y contestar la interrogante de los mecanismos completos a través de los cuales el oído humano es capaz de realizar una localización aural tanto en espacios abiertos como en recintos cerrados. No obstante, esto no ha sido un impedimento para que los expertos en el área continúen desarrollando experimentos, teorías y propuestas formales para establecer una explicación completa de dicho fenómeno.

Al igual que en las secciones anteriores, el objetivo de estos apartados no es hacer una revisión extensa y detallada de la temática de la localización espacial audible, ya que existe bastante literatura que aborda esto de manera amplia y refinada. Lo que se pretende es la exposición

⁵ Tenney reemplazaría años después el término *jerarquía* por la expresión en inglés *holarchy*, haciendo referencia a una categorización u organización de elementos completos en sí mismos con respecto a la percepción de las TGU.

de los conceptos y resultados más sobresalientes y pertinentes para establecer los elementos constitutivos del modelo teórico propuesto.

De tal forma, el primer elemento cualitativo de la localización espacial es lo que se conoce como “conciencia audible espacial, la cual puede entenderse como un fenómeno perceptual que deriva en la conciencia multisensorial de estar inmerso en un ambiente específico, real o virtual, el cual puede ser dinámico o estático” (Letowski, 2012). Esta conciencia pudiera entenderse como aquella que engloba todas las propiedades de localización espacial de las fuentes acústicas del ambiente donde ocurre la escucha. En general, las anteriores se consideran reducibles a tres: *a*) evaluación de la distancia de la fuente sonora al espectador, *b*) dirección de la que proviene la fuente sonora y *c*) características del espacio físico en el cual se propaga dicha fuente.

Además, dicha conciencia audible espacial, tal y como sucede en el caso de la percepción (como se presentó en los capítulos anteriores), está también directamente relacionada con los factores individuales del escucha que Letowski (2001) enuncia y que se plantean aquí como definición debido a que se hará uso de ellos más adelante.

DEFINICIÓN 2.3. Se denominarán, *condiciones perceptual-cognitivas intrapersonales* del escucha al conjunto de características fisiológicas, perceptuales y cognitivas propias e individuales que afectan de manera directa el acto de comprensión y aprehensión de una experiencia sensorial aural. Entre ellas pueden incluirse ejemplos como predisposición positiva/negativa a la escucha, grado de atención, estatus fisiológico del sentido de escucha del espectador, conjunto de experiencias o memorias sonoras, conocimiento de estrategias de escucha, familiaridad con el ambiente circundante y grado de involucramiento con la actividad de la escucha.

El problema de la localización espacial del sonido se asocia a dos áreas específicas: *a*) la percepción de la distancia (de la fuente al espectador) y *b*) la capacidad de ubicar la dirección de la fuente (la cual también es denominada *localización espacial audible* y que es un problema abordado en dos secciones: ubicación horizontal y elevación).

Tanto la percepción de la distancia como la dirección de una fuente sonora son problemas que resueltos por el sistema aural humano a partir de una relación compleja y dinámica de una serie de descriptores (o *cues*) específicos para cada uno. El problema de resolver la distancia es

mucho más complejo y posee más ambigüedad en el sistema aural humano que el de la dirección y, como se verá a continuación, esto posee diversas implicaciones para la resolución espacial de las fuentes sonoras en general. Se abordarán de manera separada estos dos aspectos iniciando con el que presenta mayor problemática: la distancia.

3.1 Escucha espacial: resolución de distancia de las fuentes sonoras

Para poder localizar sonidos, el sistema auditivo humano debe resolver dos parámetros esenciales, la distancia y la dirección (horizontal y vertical). Los estudios hasta ahora desarrollados han permitido una comprensión bastante extensa al respecto de la resolución de la dirección en el azimuth, de la localización espacial del sonido; no así para el caso de la distancia. Las señales más básicas que el sistema auditivo utiliza para resolver en cierta medida la distancia de una fuente sonora corresponden al filtrado del contenido frecuencial alto que es indicativo de este parámetro (la distancia), debido a la dispersión del aire en relación con las ondas cortas. Del mismo modo, para entornos cerrados (o semiabiertos) la energía de reverberación será más pronunciada para sonidos más distantes que para los más cercanos, debido a los efectos en torno al directo *vs.* reflexiones.

Xie y Yu (2019) explican, a partir de diversos experimentos, que el sistema auditivo humano tiende a subestimar significativamente las distancias de las fuentes sonoras cuando éstas se encuentran a una distancia física mayor que 1.6 metros (en promedio) y que cae en errores de sobreestimación de fuentes cercanas menores a dicho valor. De acuerdo con lo anterior se ha llegado a plantear una relación matemática entre la distancia física real de una fuente (r_f) y su distancia percibida (r_s) mediante la siguiente ecuación (Zahorik, 2005):

$$r_f = k r_s^\delta$$

donde k es una constante que en promedio tiene un valor ligeramente mayor a 1 y δ es un exponente cuyo valor se relaciona con las condiciones experimentales y otras subjetivas. De acuerdo con el mismo Zahorik (2005), la percepción de distancia de las fuentes sonoras en el contexto de la escucha espacial depende de una relación compleja (no del todo descifrada hasta ahora) de diversas señales (*cues*), de las cuales el autor enlista las siguientes:

- *Amplitud o volumen.* Para espacios abiertos se conoce la relación existente entre la distancia de la fuente sonora y su presión sonora (suponiendo una fuente sonora con presión constante). Esta relación es llamada la *ley 1/rS*: el nivel de presión sonora (SPL) se reduce en 6dB por cada duplicamiento de la distancia de la fuente sonora respecto del escucha.
- *Señal espectral.* Asociado al efecto de filtrado de las frecuencias altas de la fuente sonora, debido a la absorción del aire, en espacios abiertos y para fuentes con una distancia mínima de 15 metros.
- *Curva del frente de onda y señales HRTF de campo cercano.* Estas señales están asociadas a los efectos de dispersión y difracción de los frentes de onda de las fuentes sonoras debido a los diversos elementos relacionados con la estructura anatómica humana (la cabeza, el pabellón auditivo, el tronco, los hombros, etcétera) y, por otro lado, a los efectos de la teoría duplex de Rayleigh al respecto de las ILD y las ITD.
- *Relación sonido directo a energía reverberante (DRR).* Puede ser calculada de acuerdo a la fórmula:

$$DDR = \frac{D_s \Sigma_s}{16 \pi r_s^2} \frac{\alpha_{abs}}{1 - \alpha_{abs}}$$

donde r_s es la distancia entre la fuente sonora y la posición del escucha, D_s es el factor direccional del sonido, Σ_s es el área total de absorción y α_{abs} es el coeficiente de absorción promedio. Al respecto de lo ya mencionado, Xie y Yu (2019) nos dicen que:

En resumen, la percepción audible de la distancia está sesgada. Es una consecuencia integral de múltiples señales que dependen de múltiples factores de las fuentes sonoras y los entornos acústicos. Las señales de la percepción de la distancia acústica en un recinto, dependen especialmente de las reflexiones y por lo tanto, son distintas a aquellas relacionadas a los entornos abiertos. La dependencia multifactorial de estas señales complican la situación. Algunos mecanismos de la percepción de la distancia permanecen inciertos.

3.2. Escucha espacial: resolución de dirección de las fuentes sonoras

En esta sección se presentarán diversos conceptos y resultados al respecto de la resolución de la direccionalidad de las fuentes sonoras en la escucha espacial humana. Como ya se mencio-

nó, el parámetro de la distancia es el que ha sido abordado con mayor éxito y se tienen numerosos recursos de diversos investigadores al respecto, quienes han contribuido de manera sustancial a la comprensión del fenómeno de la escucha espacial en general. Con el fin de evitar confusiones en el uso de los términos anteriores con otros que serán presentados más adelante, se ofrece a continuación una forma explícita de nombrarlos.

DEFINICIÓN 2.4. Se denominará *localización direccional* a la localización espacial audible. De este modo, la localización direccional en el plano horizontal será nombrada como *localización direccional horizontal* y la referente a la ubicación en la elevación será denominada *localización direccional vertical*. La resolución de la distancia se denominará *localización métrica*.

Planteado lo anterior, es posible pensar en la *localización espacial* como la capacidad perceptual de ubicar una fuente en un espacio en términos generales, tanto en dirección (horizontal/vertical) como en distancia. Al respecto, se consideran usualmente dos tipos de evaluación de la localización espacial: *relativa* o *discriminatoria* y *absoluta* o *categorica*. El primero de estos tipos se relaciona con la capacidad del escucha de comparar la ubicación espacial de un sonido respecto a otro, ya sea de manera simultánea o secuencial, mientras que el segundo implica la habilidad del escucha de ubicar la posición, de la forma más precisa posible, de una única fuente sonora. Esta última, la localización absoluta, requiere un proceso mucho más complejo de atención por parte del escucha.

Una de las premisas en el modelo que se propone en este trabajo (véase la definición 2.5) es que la sensación mínima de distancia espacial desde el punto de vista composicional está dada por la comparación de dos sonidos; esto es, mediante la *localización relativa*.

Usualmente, el sistema de coordenadas esféricas se toma como convención para la localización de las fuentes sonoras por parte del escucha. En este sentido, se tienen tres componentes que el oído debe resolver: dirección en el plano horizontal (azimut), dirección en el plano medio (elevación) y la distancia. Cada uno de estos elementos es resuelto por el sistema auditivo mediante mecanismos distintos. Otro sistema de coordenadas bastante utilizado es el llamado “sistema de coordenadas interaural polar” (Morimoto, 2001), el cual requiere de los siguientes parámetros (tal y como se muestra en la figura 5): *ángulo lateral* α , *ángulo de creciente* β , el azimut θ , el *ángulo de elevación* φ y la distancia a la fuente sonora d .

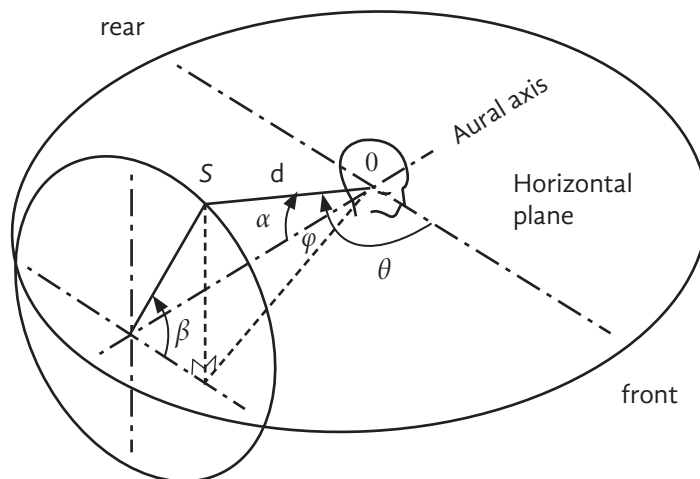


Figura 5 | Sistema de coordenadas interaural polar (Letowski, 2001).

Este sistema de coordenadas interaural polar, como bien describe Letowski (2001), presenta la ventaja principal de que la longitud de arco entre dos ángulos laterales es independiente de la elevación; además, la descripción de la ubicación espacial para el *espacio interno* formado dentro del sistema multicanal es más adecuada con este sistema. Se opta por la convención de $\pm 180^\circ$ para la definición de los ángulos tanto en el plano medio como en el horizontal y en el frontal (coronal).

Al inicio de esta sección se planteó el concepto de conciencia audible espacial y se mencionó que a partir de éste se desprendían, si no formalmente al menos sí de manera indirecta, los demás elementos que conforman la estructura de la localización de la escucha espacial. Esta conciencia está sujeta a diferencias entre las ubicaciones percibidas y las localizaciones reales de una o más fuentes sonoras; es decir, a errores de localización. Se mencionan a continuación los tipos de errores de localización más comunes dentro de la literatura de la percepción psicoacústica espacial según se muestran en Letowski (2012):

- El *error de localización* LE es la diferencia entre la dirección estimada y dirección real de una fuente sonora en el espacio.
- El *error local* o *genuino* es aquel que se encuentra en un intervalo de tolerancia de $\pm 45^\circ$ con respecto a la media y que en la práctica real implica una desviación de entre $\pm 20^\circ, \pm 25^\circ$.

- El *error de confusión* es la discrepancia entre la localización percibida y la localización real de una fuente sonora para valores mayores a $\pm 90^\circ$ y por lo general cercanos a $\pm 180^\circ$ a lo largo del plano medio; esto es, errores del tipo frente-atrás (FBE, *front-back errors*) y viceversa (BFE, *back-front errors*). Dichos errores son más comunes para fuentes sonoras espectralmente limitadas a menos de 8 KHz, así como para fuentes con ancho de banda limitada, todas ellas localizadas a lo largo del plano medio.
- El *error constante* CE es la desviación que surge en la medición estadística debido a la falta de simetría en el proceso o condiciones de audición del escucha.
- El *error aleatorio* es aquel que se considera debido a la incertidumbre, imprecisión y cambios aleatorios en las condiciones de escucha, así como en las fluctuaciones de la atención del escucha.

Considerar estos errores para plantear un marco explicativo *a posteriori* de la espacialización en la música electroacústica resulta bastante favorable desde una mirada tanto compositiva como musicológica. Nótese, además, que es posible considerar ponderaciones de importancia en cada uno de los errores de acuerdo con el contexto particular del escucha. Así, por ejemplo, para formatos de concierto en los cuales, por lo general, la arquitectura del espacio no es siempre la idónea para el formato multicanal, los errores aleatorio y constante tendrán una mayor participación en la resolución espacial final del escucha. Para el caso de estudio (o de recintos adecuadamente adaptados para formatos multicanal), la resolución final estará más ligada al error de confusión y el error aleatorio en segundo plano.

Difusión espacial y localización audible del sonido

Rumsey (2001) hace una distinción clara respecto a las características espaciales de las fuentes sonoras en ambientes naturales; por lo general, éstas se dividen en dos categorías: fuente y ambiente. Las primeras son entidades relativamente discretas y localizables, mientras que las segundas poseen un carácter difuso y de difícil localización. La descripción anterior, aunque muy general, establece una estructura primaria para dividir conceptualmente el fenómeno de la localización espacial y la experiencia de inmersividad.

En esta sección se enlistan los aspectos y conceptos más importantes de la escucha espacial en relación con los objetivos del presente trabajo. Existe abundante literatura que aborda la es-

cucha espacial y la acústica, tanto de campos abiertos como de recintos cerrados o reverberantes, por lo que se remite al lector a referencias tales como Susuki *et. al.* (2011), Rumsey (2001), Theile (1980) y Toole (2008), para una consulta más detallada.

Existen diferencias sustanciales en la localización espacial del sonido entre escenarios de campo abierto y recintos cerrados las cuales están descritas de manera detallada en las referencias antes mencionadas. La convención usual es que, en campos abiertos, la localización espacial resulta mucho más sencilla para el oído humano debido a que no está presente el efecto de reverberación, el cual podría generar confusión al respecto para dicha tarea; sin embargo, y por esta misma razón, la resolución de distancia y profundidad no es tan exacta y resulta más complicada de realizar que en la contraparte de recintos cerrados. Lo anterior queda también influenciado por el hecho de que la mayoría de las fuentes sonoras no tienen un patrón de radiación perfectamente esférico sino que presentan una forma direccional que varía en función de la zona del contenido frecuencial de la misma.

Respecto de la categoría de fuentes y la localización espacial en el plano horizontal, se tienen primordialmente dos escenarios: campos sonoros libres y campos sonoros reverberantes. Uno de los mecanismos analíticos básicos para estudiar la escucha espacial de esta índole se relaciona con la teoría duplex de Rayleigh en la que los indicios binaurales ITD (*interaural time difference*) e IID (*interaural intensity difference*) juegan un papel primordial; dichos elementos describen bastante bien el fenómeno en cuestión para situaciones ideales compuestas por una sola fuente sonora.

Las ITD son más dominantes en las frecuencias bajas, mientras que las IID los son para las frecuencias altas, donde “la frecuencia de corte es determinada por la distancia entre los dos oídos (típicamente entre 22 a 23 cm) y se considera usualmente con un valor alrededor de los 1500 Hz” (He, 2017). Por su parte, Rumsey (2001) establece que “el oído es sensitivo a diferencias interaurales de fase sólo para frecuencias bajas y dicha sensibilidad comienza a deteriorarse a los 1000 Hz”. Lo anterior tiene antecedentes más antiguos como las pruebas realizadas por Mills (1958), en las que reporta que para “estímulos dicotómicos, la resolución de la dirección de una fuente es determinada mediante ITD o IPD⁶ para frecuencias por debajo de los 1400 Hz y para frecuencias más altas mediante IID”.

⁶ IPD o *interaural phase difference* es el término utilizado en cierta literatura como sustitutivo de las IID, para el caso de tonos puros.

Debe notarse que, al igual que como ocurre con la teoría del ASA, las conclusiones anteriores y la mayoría de los resultados sobre localización espacial han sido plenamente demostrados para pruebas en ambientes controlados y con fuentes sonoras relativamente simples (tonos puros y ruidos con un contenido de banda específico); sin embargo, de estos mismos resultados han derivado conjeturas bastante formales para establecer la validez de las mismas en el caso de fuentes sonoras más complejas.

De acuerdo con diversos investigadores (Blauert, 1997; Begault, 2000; Algazi y Duda, 2011, y Xie, 2003), se acepta que los siguientes mecanismos son los “más significativos para la localización direccional del sonido” (He, 2017):

- ITD-IID.
- Indicios espectrales (monoaurales y relevantes a la antropometría del escucha).
- Indicios de movimiento de la cabeza.
- Indicios de intensidad y volumen.
- Familiaridad con la fuente sonora.
- DRR (*direct-to-reverberation ratio*).
- Señales e indicios visuales y de propiocepción.

Mientras que los indicios binaurales; esto es, las IID y las ITD, resuelven la dirección de la fuente en el plano horizontal, los indicios espectrales son necesarios para resolver la elevación en el plano medio (resultado del llamado *cono de confusión*) y la discriminación frente-atrás. Estos indicios están asociados a la antropometría del escucha y se relacionan con el coloramiento espectral del sonido (filtrado) causado por la cabeza, el torso y el pabellón auditivo antes de entrar al canal del oído; son formulados mediante las HRTF (*head related transfer function*). “La mayoría de los indicios espectrales debidos al pabellón auditivo ocurren para frecuencias arriba de los 3Khz, mientras que aquellos relacionados a la cabeza y el torso ocurren por debajo de ese valor” (He, 2017).

Previamente, para el caso particular del pabellón auditivo, Batteau (1967) proponía que “las reflexiones producidas por el oído externo eran capaces de proporcionar indicios adicionales para la localización y que los retardos temporales entre el sonido directo y esas reflexiones, variaban en función del ángulo de incidencia”.

Para escenarios en los que ocurren diversas fuentes sonoras simultáneas o *campos sonoros superimpuestos*, estos mecanismos no son directamente aplicables y la mayoría de los estudios

apuntan a que otros procesos complementan o incluso sustituyen dichas funciones tales como el *efecto de precedencia* (también conocido como *efecto Haas*). De hecho, la discriminación de localización espacial para fuentes simultáneas parece estar ligada a los mecanismos del ASA.

Como es bien sabido, la capacidad humana de localización espacial del sonido tiene una mayor resolución en el horizonte frontal debido, en gran medida, a que es apoyada por el sentido de la vista. Al respecto, Blauert (1997) apunta que el MAA (ángulo audible mínimo), “puede llegar a tomar un valor tan pequeño entre 1° y 3° para sonidos de banda amplia (como el ruido) en el plano frontal horizontal ($\pm 90^\circ$ azimut) y decrece en resolución para otras direcciones y sonidos de banda reducida”. Derivado de esto y aunado, por ejemplo, al efecto de las reflexiones en recintos reverberantes, una gran variedad de fuentes cuya ubicación física se encuentra en la parte posterior de la cabeza del escucha pueden conducir a localizaciones no sólo deficientes sino incluso erróneas por parte del oído humano (percibir que la fuente proviene de otro lugar distinto, como del frente horizontal).

Como bien retoma Cetta (2003), la localización espacial de fuentes sonoras de banda estrecha presenta además una anomalía particular, y es que para ciertos casos, dicha fuente es percibida en una posición simétrica respecto del eje interaural. Lo anterior resulta de que, como dichos sonidos tienen un ancho de banda limitado, el sistema auditivo no es capaz de utilizar eficientemente los indicios espectrales para resolver la posición espacial.

Del mismo modo, “típicamente, las fuentes ubicadas en la parte posterior, dan lugar a una respuesta reducida para frecuencias altas, debido a la forma del pabellón de la oreja que está orientada hacia el frente” (Rumsey, 2001), o como comenta Cetta (2003), “hacia los lados, este valor (el MMA) se incrementa de tres a diez veces, y en la parte posterior decrece nuevamente hasta aproximadamente el doble del valor en el frente”.

Sin embargo, algunas conjeturas sugieren que debido a esta limitación fisiológica (la forma del pabellón de la oreja y la concentración de la percepción de la información audiovisual hacia el horizonte frontal), el proceso auditivo humano ha desarrollado adaptaciones ecológico-cognitivas con el fin de aumentar la resolución de la localización espacial de las fuentes sonoras ubicadas en la parte posterior (Rumsey, 2001).

Bandas direccionales y localización vertical

El mecanismo de bandas direccionales para resolver la ubicación espacial vertical de ciertas fuentes sonoras se encuentra en función, exclusivamente, del contenido espectral de las mis-

mas. Las bandas propuestas por Blauert (1997) encuentran su antecedente en las investigaciones de Pratt (1930), quien afirma que “antes de cualquier adición asociativa, existe en cada tono, un carácter intrínseco espacial el cual dirige directamente al reconocimiento de diferencias en la altura y profundidad junto con el continuo de alturas”. Como consecuencia particular de esa conclusión, el mismo Pratt (1930) afirma que los tonos tienen una relación directa entre su contenido espectral y su ubicación vertical en el espacio. Años después, Mudd (1963) encontró evidencia de que la percepción espacial de la frecuencia no sólo estaba en función de la ubicación vertical sino que también existía un componente horizontal; a este resultado le denominó *estereotipo espacial asociativo*.

Estudios más recientes han confirmado lo anterior en el sentido de que el sistema cognitivo humano asocia el tono aural con una representación mental del espacio. Rusconi *et al.* (2005) mostraron que existe “una naturaleza integral del procesamiento de los estímulos audibles tanto espectrales como espaciales”. A continuación se citan sus resultados:

La integración de los atributos espaciales y espectrales pueden ocurrir en una etapa de procesamiento (cognitivo) [...] Aún cuando la ubicación física de una fuente sonora se mantiene constante, el tono (altura) todavía interactúa con el espacio, afectando la velocidad y precisión de las respuestas manuales mediante un pareo preferencial con ubicaciones de respuesta. Llamamos a este fenómeno, el efecto SMARC en analogía con el efecto SNARC, el cual ha sido descrito primero utilizando números.

Blauert estableció el concepto de *bandas direccionales*, las cuales “son regiones del espectro frecuencial que parecen estar atenuadas o aumentadas para posiciones particulares en el plano medio” (Rumsey, 2001). En su trabajo, Blauert (1969) demostró que la dirección de una imagen sonora para una banda de ruido de octava, es una función únicamente del centro de la frecuencia y no depende del ángulo de elevación de la fuente; de este modo, la *banda direccional* es la banda frecuencial mediante la cual es determinada la dirección de la imagen sonora. Junto con las bandas direccionales de Blauert, Hebrank y Wright (1974) identificaron picos y valles relacionados con cada banda para localizaciones específicas *estableciendo, de este modo, relaciones entre las características espectrales y la localización en el plano medio*.

A continuación se enlistan las bandas direccionales de Blauert junto con los picos-valles antes mencionados tal y como son descritos en Rumsey (2001):

1. La región frecuencial centrada en los 8 KHz parece corresponder de manera muy acertada a la percepción espacial por encima de la cabeza con un pico de un cuarto de octava entre 7 KHz y 9 KHz.
2. Las regiones entre los 300 y 600 Hz, así como aquellas entre los 3 KHz y los 6 KHz corresponden de manera óptima a la localización frontal con un valle entre 4 y 8 KHz, y un incremento en la energía superior a los 13 KHz.
3. Las regiones centradas en los 1.2 KHz y los 12 KHz están relacionadas con la localización posterior con un pico entre los 10 KHz y los 12 KHz junto con un decremento de energía arriba y abajo de dicho intervalo.
4. La percepción de la elevación varía de acuerdo a la frecuencia baja de corte de un valle de una octava localizado ente los 5 KHz y los 11 KHz.

A continuación se enlistan diversos resultados obtenidos de acuerdo con distintos autores, acerca de la localización espacial tanto en el plano horizontal como en el plano medio o vertical. Estos resultados han sido recopilados de manera detallada en Letowsky (2012).

Plano horizontal-vertical

1. La región correspondiente al intervalo [800, 1600] Hz es considerada en términos generales como aquella en la cual el desempeño de la localización espacial es más deficiente.
2. De acuerdo con Kuhn (1977), la ITD se puede calcular mediante la fórmula:

$$ITD^{ar} = \sin(\theta)$$

Para $a = 3$, frecuencias por debajo de los 500 Hz. Para $a = 2$ frecuencias arriba de los 2000 Hz y $c = 331 + 0.6T$ donde T es la temperatura ambiente.

3. Derivado del punto anterior, existe un valor crítico de ITD, el cual es el referente que se toma como el valor máximo que puede tener dicho parámetro: $ITD_{max} = 794\mu s$.

4. Para frecuencias por debajo de los 1.5 KHz, las IID son demasiado pequeñas como para contribuir a la localización espacial.

5. Dependiendo de la posición lateral y el contenido espectral de la fuente, las IID pueden alcanzar valores de hasta 10 dB para sonidos centrados en los 3 KHz y de hasta 35 dB para sonidos centrados en los 10 KHz.

6. Debido al comportamiento no-monótono de las IID, existe una incertidumbre considerable de localización para el rango frecuencial [1, 1.6] KHz para sonidos con desviaciones de hasta $\pm 30^\circ$, 40° con respecto de la línea frontal media.

7. Para el campo cercano (esto es a una distancia de aproximadamente 5 veces el radio de la cabeza del escucha), las IID dependen del contenido espectral y del azimuth de la fuente sonora y esta dependencia afecta positivamente de manera directa la localización de fuentes sonoras en dicho campo cercano. Por su parte, las ITD no presentan una relación con la distancia en el campo cercano y su contribución a la localización espacial es casi negable para tal situación.

8. Numerosos estudios han demostrado que las ITD son particularmente activas para la localización izquierda-derecha para sonidos únicamente por debajo de los 800 Hz. Por su parte, las IID lo son para sonidos por arriba de los 1.6 KHz.

9. El MMA o *ángulo audible mínimo* (diferencia detectable en el azimuth de dos fuentes sonoras idénticas no simultáneas) para estímulos de banda ancha y tonos puros de frecuencias bajas, se encuentra en el orden de 1° o 2° para la posición frontal y del orden de 6° o 7° para ubicaciones por detrás.

10. El MAA vertical para la posición frontal en el plano medio está entre 3° y 9° grados.

Indicios monoaurales espectrales

1. Las reflexiones debido al torso afectan el espectro frecuencial en el rango [2, 3] KHz. Los efectos debido al pabellón auricular son más pronunciados en el rango de [3, 4] KHz, pero también tiene influencia para espectros localizados arriba de los 5 KHz. De los resultados anteriores se deben tomar en cuenta también las ISD o *diferencias interaurales espectrales*.

2. Los indicios espectrales más importantes para la diferenciación de la localización frente-atrás están centrados en el rango de [4, 16] KHz. Para el caso de la diferenciación arriba-abajo, dicho rango corresponde a [6, 12] KHz.

3. Los indicios monoaurales de frecuencias altas son predominantes para la discriminación de ubicación vertical y frente-atrás. Sin embargo, algunos estudios han mostrado que indicios monoaurales de frecuencias menores a 2 KHz pueden contribuir de manera importante a dicha localización; sobre todo para elevaciones mayores a 45° grados.

Indicios dinámicos

1. Los movimientos de la cabeza causan cambios momentáneos en el patrón de picos del espectro sonoro de cada oído y este tipo de indicios tienen mayor impacto para frecuencias por debajo de los 2 KHz, siendo además considerado por varios autores como uno de los mecanismos primordiales para la discriminación de la localización frente-atrás.

2. Sonidos con una duración mínima de entre 600-800ms pueden ser localizados con precisión aceptable mediante los indicios dinámicos. Sin embargo, para sonidos cortos y movimientos de cabeza demasiado rápidos, tales indicios pueden conducir a errores de localización.

3. Si la cabeza se mantiene fija, los sonidos necesitan tener una mayor duración para poder resolver su localización. Un aumento en la duración de 3ms a 1s, disminuye el error promedio de localización de 10° a 20 grados.

4. El sistema de escucha humano reacciona fundamentalmente a la parte del ataque de la señal u *onset* para resolver la localización espacial.

5. Tonos en el intervalo [400, 600] Hz con duraciones menores a 1ms pueden ser localizadas con menos de 3° de error a lo largo de todo el rango frecuencial.

6. En general, se ha observado que impulsos cortos del orden entre 5-200ms con tiempos de ataque de 5ms son los más sencillos para localizar en espacios cerrados.

3.3. Propiedades acústicas de la espacialidad en la experiencia aural

Según diversos estudios, y de acuerdo con Rumsey (2001), para experiencias aurales en sistemas multicanal existe una tendencia de preferencia en escuchas por parámetros como la *profundidad de la imagen sonora*, el *ancho de la imagen sonora* y la *sensación de involucramiento*. Lo anterior plantea de entrada una discusión sobre una posible perspectiva de las características puntuales al momento de realizar una creación electroacústica multicanal y, sobre todo, cuando se desea poner especial énfasis en el desarrollo de una sensación de inmersividad.

Algunos autores definen el término *espaciosidad* (*spaciousness*) como la sensación del espacio en el que se encuentra localizado el escucha, la cual está directamente relacionada con el tamaño del recinto y el sentido de externalización del sonido por parte del escucha (Rumsey, 2001). Por otro lado, el término *involucramiento* está ligado a la sensación de inmersividad y

al efecto opuesto de localización espacial; esto es, el hecho de percibir el sonido proveniente de todas las direcciones alrededor del escucha.

Medidas tales como el “envolvimiento del escucha (LEV), la fracción lateral (LF) y la ganancia lateral L_{g80} , relacionan la proporción de la energía lateral en recintos comparada con una medida omnidireccional de presión sonora en la posición de escucha” (Rumsey, 2001). Se ha demostrado, de igual modo, que el ancho de la imagen sonora o ASW (*apparent source width*) está directamente relacionado con el coeficiente interaural de correlación cruzada (IAC-CE), la fracción lateral (LFE) y la ganancia lateral.

El elemento final que se considera de importancia para el presente trabajo es el descrito en Rumsey (2001) como *naturalidad*, el cual está relacionado con el grado cualitativo subjetivo de realismo de una experiencia aural espacial. La naturalidad, por lo tanto, si bien es una descripción perceptual subjetiva, debido a las características propias de la función ecológica de la escucha humana, puede considerarse más o menos homogénea y de este modo puede confiarse como una descripción cualitativa que afecta directamente al desarrollo de la inmersividad. La naturalidad, de nuevo, puede vincularse con la analogía relacional y con el proceso *top-down*.

Este factor (la naturalidad) es de una importancia bastante específica ya que en el proceso de la creación electroacústica se pueden presentar diversos elementos, escenas u objetos sonoros que generen en el espectador una sensación de *antinaturalidad* en la experiencia sonora debido a múltiples razones, entre las cuales pueden resaltar un uso equivocado/ambiguo de la reverberación o del movimiento espacial y construcciones tímbricas altamente artificiales con bajo grado de resolución sonora. Tales situaciones se presentan de manera inmediata cuando dichos elementos (objetos o escenas sonoras) son percibidos con un comportamiento inusual o incluso contrario a la experiencia aural ecológica usual; este factor llega a ser (el de la naturalidad) uno de los más importantes en ciertas pruebas, al momento de evaluar la preferencia general en la calidad de la reproducción del audio (Mason y Rumsey, 2000).

Puede surgir una pregunta en este punto: ¿por qué es importante considerar todos estos factores acústicos, perceptuales y psicoacústicos dentro de un modelo de creación electroacústica multicanal? La respuesta radica en uno de los objetivos de la investigación propuesta: *la búsqueda de la inmersividad como experiencia aural en sistemas multicanal*. Por lo tanto, una creación sonora, si no basada del todo en este *corpus* de conceptos y datos acústico-perceptuales, pero sí asistida y dirigida por ellos, estará fundamentada en una construcción sólida y dinámicamente informada que le permitirá al creador-compositor emplear dichas herramientas

para lograr resultados concretos al momento del diseño espacial en su obra. A continuación se agrupan, a modo de definición, todos los conceptos antes descritos.

DEFINICIÓN 2.5. La *conciencia audible espacial* puede dividirse en dos subgrupos:

- Conciencia audible espacial de ubicuidad.
- Conciencia audible espacial de dispersión o de ambiente.

La primera engloba todas aquellas características mediante las cuales el escucha es capaz de describir las propiedades de ubicuidad (fija o dinámica) y extensión limitada de una fuente específica dentro de un espacio o recinto y entre las cuales pueden listarse las siguientes: localización direccional, percepción de la distancia, ancho de la imagen sonora y profundidad de la imagen sonora. La segunda trata sobre las propiedades de espacialidad que son percibidas difusas y distribuidas a través de un espacio o recinto: espaciosidad, envolvimiento y reverberación.

OBSERVACIÓN 2.3.1. Derivado de lo anterior, el concepto de *naturalidad* puede entenderse tanto en el nivel grupal (conciencia audible espacial) como en el local de cada uno de los componentes; por ejemplo, naturalidad de la distancia o naturalidad de la espaciosidad.

3.4. Fuentes sonoras en movimiento y fuentes múltiples

El estudio de la percepción de las fuentes sonoras en movimiento no ha tenido tanto desarrollo si se le compara con el que se ha hecho para las fuentes fijas; sin embargo, se han obtenido diversos resultados que han permitido establecer marcos explicativos y de hipótesis con un nivel considerable de efectividad. La importancia de la comprensión de dicho caso (fuentes sonoras en movimiento) tiene aplicaciones muy variadas que van desde las militares hasta los ingenieriles y científicas.

Como se verá más adelante, uno de los componentes principales para el desarrollo de la inmersividad dentro de un espacio aural en el marco de la composición electroacústica es, precisamente, el tratamiento adecuado de los objetos sonoros móviles; es por ello que en esta sección se presentan algunos resultados psicoacústicos importantes al respecto.

Existen algunos parámetros o indicios primordiales que el escucha toma como referencia para resolver el problema de localización de las fuentes sonoras móviles: *velocidad angular*, *velocidad radial*,⁷ el efecto Doppler, la intensidad del sonido y las diferencias interaurales.

Así como el MAA es la métrica utilizada para la diferenciación en la localización, el *ángulo de movimiento mínimo audible* (MAMA) es la correspondiente para las fuentes en movimiento y se define como *la distancia angular más pequeña que la fuente sonora debe recorrer para que la dirección de su movimiento sea identificada*.

Como describe Letowski (2012), existen dos teorías principales que explican la percepción de las fuentes sonoras en movimiento: *la teoría de la instantánea* y *la teoría del movimiento continuo*. La primera supone que la percepción del movimiento de la fuente sonora no necesita ser real sino potencial, en el sentido de que el escucha es capaz de establecer esa sensación de movimiento comparando únicamente las posiciones iniciales y finales de dicha fuente; a este efecto se le conoce como *movimiento audible aparente* o AAM. La segunda teoría establece que el escucha realiza un proceso de atención-percepción sobre toda la trayectoria de la fuente sonora.

De las investigaciones más conocidas al respecto se encuentran las de Perrott y Musicant (1977), de las cuales se obtuvieron las primeras conclusiones sobre las fuentes móviles: las posiciones de los *onsets* y los *offsets* son desplazadas primordialmente en la dirección del movimiento. Esto fue recopilado por Getzmann y Lewald (2007):

La magnitud del desplazamiento del onset se incrementa con la velocidad de la fuente sonora (entre 90°/seg y 600°/seg) [...] En general, las posiciones de los offsets fueron menormente desplazadas que las de los *onsets* para sonidos de corta duración (50 y 100 ms), pero más desplazadas para sonidos de mayor duración (150 y 300 ms).

Lo anterior parece relacionarse directamente con el *efecto Fröhlich* y el llamado *momentum representacional*; propios de la percepción en el campo visual. Una de las principales posibles causas del primero tiene que ver con el nivel de atención requerido por un individuo para percibir una fuente visual en movimiento al inicio de la trayectoria, ya que se asume que esta acción recibe, por lo general, menos atención que el subsecuente trayecto: “Dado que la atención es un prerequisite necesario para el procesamiento consciente de la información espacial, la pri-

⁷ La velocidad angular es aquella a la cual el sonido rota alrededor del escucha. La velocidad radial se refiere a la velocidad con la que la fuente sonora se aleja o se acerca respecto del escucha (Letowski, 2012).

mera parte de la trayectoria se pierde y la posición percibida del *onset* parece haber cambiado hacia la dirección del movimiento” (Müsseler y Aschersleben, 1998).

Por su parte, *el momentum representacional* se define como el efecto de percibir una fuente sonora más lejos de su trayectoria real cuando dicha trayectoria ha concluido. En cualquiera de los dos casos anteriores, el problema a estudiar es la distorsión perceptual de la trayectoria de la fuente sonora en cualquiera de sus etapas: inicio, durante o final.

Por otro lado, los experimentos de Aschoff (1963) y de Plath (1970) mostraron que la percepción del movimiento de una fuente sonora en una configuración multicanal es inversamente proporcional a la velocidad de dicho movimiento.

Sumario de resultados experimentales

1. El MAMA es por lo general dos veces más grande que el MAA cuando se trata de la misma fuente sonora y la misma dirección inicial de referencia, siendo además independiente de la dirección del movimiento en el plano horizontal y de la intensidad de la señal.

2. El MAMA presenta valores mínimos al frente del escucha; de aproximadamente $2^\circ - 8^\circ$ para fuentes con una velocidad menor a $10^\circ/\text{seg}$, y va incrementándose de manera directa tanto en función de la velocidad como de la lejanía lateral con respecto al escucha; hasta 1.5 a 3.0 veces más grande para $\pm 60^\circ$ a lo largo del azimut.

3. La resolución óptima del MAMA en el plano horizontal es de $9^\circ - 16^\circ/\text{seg}$ y de $7^\circ - 10^\circ/\text{seg}$ en el plano vertical.

4. Para fuentes sonoras cuya velocidad excede los $10^\circ/\text{seg}$, el MAMA horizontal está linealmente relacionado con la velocidad de la fuente sonora.

5. El *umbral de velocidad*, el cual se considera como la velocidad mínima necesaria para que una fuente sonora sea detectada en periodo constante de observación, depende de dicho periodo T y del perfil espectral de la fuente. Para un periodo $T=500\text{ms}$ y tonos de 250 500 y 1000 Hz, se ha observado que dicha velocidad oscila entre los $10^\circ - 15^\circ/\text{seg}$ y de $40^\circ/\text{seg}$ para tonos de 2000 Hz.

6. La separación espacial entre dos fuentes sonoras no afecta la fuerza de la sensación del AAM y sólo tiene efecto en la velocidad percibida del movimiento. ■ ■ ■

Capítulo III. Espacio y espacialidad

El *Espacio-Timbre*, tal y como se propone en el presente trabajo, será construido a través de dos etapas. La primera será la formulación abstracta y generalizada de las características de *espacialidad* y de *timbre* de un objeto sonoro cualquiera; en ambos casos considerando los aspectos acústicos y perceptuales. Una vez construidas estas definiciones, se establecerá el concepto de *espacio aural* y *espacio spectromorfológico* mediante sistemas dinámicos. Este capítulo corresponde a la construcción de la primera parte; el de la formalización de los conceptos de *espacio* y *espacialidad*.

Aunque puede resultar algo confuso, en esta etapa la palabra *espacio* hace referencia a un objeto teórico con representación matemática y no al *espacio* desde el punto de vista sonoro; de este modo, *espacio aural* es el objeto abstracto que representa la espacialidad acústico-perceptual de un objeto sonoro, mientras que *espacio spectromorfológico* es el objeto abstracto que representa las características spectromorfológicas acústico-perceptuales de un objeto sonoro.

Toda vez que se ha especificado lo anterior, la siguiente etapa consiste en formular de manera abstracta y general el *Espacio-Timbre* haciendo uso de la propiedad universal del producto de conjuntos y tomando como elementos de dicho producto al *espacio aural* y al *espacio spectromorfológico*. El capítulo concluye con una definición formal de *Espacio-Timbre* que abarca una concepción teórica y generalizada por medio de la cual cualquier aplicación práctica y concreta podrá ser explicada y desarrollada, a través de dicho modelo.

La idea de establecer esta relación indisoluble entre las características espaciales y tímbricas puede parecer evidente si se toma en cuenta que en la composición de la música electroacústi-

ca y la música acusmática se tiene conciencia constante de dicha relación. Sin embargo, no existe una convención o unificación de términos y conceptos respecto al espacio y al timbre dentro de la música electroacústica y existen pocos trabajos académicos que aborden esta temática de manera explícita.

Se encuentran antecedentes para el estudio de la relación Espacio-Timbre en el *espacio circumspectral* de Smalley (2007): “la distribución o división del espacio espectral de lo que es percibido como espectromorfología coherente o unificada”. O en la espacialización tímbrica mencionada por Normandeu (2008): “no es una concepción de espacio que es agregada al final del proceso compositivo (una aproximación muy frecuente en estos días, sobre todo debido al uso del software multitrack), sino una espacialización verdaderamente compuesta”. Al respecto de esa misma técnica (la espacialización tímbrica), James y Hope (2013) afirman que “se centra ampliamente en la localización del espectro sonoro con respecto al azimut en el plano medio”.

Se comenzará este capítulo con una breve reseña histórica sobre el uso del espacio como parámetro compositivo en la música. A continuación se establecerán los conceptos de *perfil morfológico*, *perfil espectral* y *perfil espectromorfológico*. Seguido de lo anterior se plantean los fundamentos teóricos de la concepción de espacio utilizada dentro del contexto de música electroacústica para continuar estableciendo los elementos primarios que conforman, a juicio propio, la caracterización primordial de la espacialidad en la música electroacústica. Finalmente, se establecerán los componentes matemáticos tanto del *espacio aural* como del *espacio espectromorfológico* para concluir con la derivación generalizada de *Espacio-Timbre*.

1. Construcción teórica del concepto de espacio

A lo largo de esta sección se construirá una definición original del concepto de *espacialidad* siguiendo en tres aspectos específicos:

1. *Componente heurístico*. Derivado de ciertos antecedentes explícitos acerca del uso y formulación de la idea de espacio en la música electroacústica; en particular aquellas relacionadas con el trabajo de Dennis Smalley, Francis Dohmont, Francois Bayle y Michel Chion.

2. *Componentes acústicos-psicoacústicos*. Derivados de los resultados mostrados en el capítulo II.
3. *Componente epistemológico*. Derivado de una reflexión fenomenológica de la idea de espacio y los elementos que construyen la inmersividad.

De este modo, al final de la sección se habrá desarrollado una construcción teórica sobre el concepto de espacio, la cual estará soportada por las tres bases angulares ya mencionadas con el fin de sustentar dicha formulación de manera formal.

1.1 Antecedentes sobre uso del espacio en la música

A principios de los años cincuenta del siglo pasado, se estableció en Nueva York un grupo llamado The Project for Magnetic Tape, del cual surgieron tres obras representativas: *William's Mix* (John Cage, 1952), *Octet* (Earl Brown, 1952) e *Intersection* (Morton Feldman, 1953). Surgen como creaciones que buscan explorar y explotar la utilización del espacio para la música en el formato de soporte fijo (cinta magnética) como un parámetro compositivo.

Por el lado acústico, Charles Ives y Henry Brant habían establecido las bases de lo que sería denominado posteriormente *música acústica espacial* (cada quien con distintos enfoques) en piezas como *Fourth Symphony* (Ives, 1916) y *Antiphony I* (Brant, 1953), cuyos antecedentes pueden ser rastreados hasta la escuela veneciana del siglo XVI con las *Antifonías* de Adrian Willaert y Andrea y Giovanni Gabrielli, pasando por ejemplos como la *Sinfonía No. 2* (Mahler, 1892), *El Pájaro de Fuego* (Stravinsky, 1910) o la *Grande Messe des Morts* (Berlioz, 1837).

Por su parte, también en la década de 1950 en Europa, Schaeffer introducía la sistematización de la *musique concrète* y los precedentes de la difusión en tiempo real con el *potentiomètre du space*, además de establecer, junto con Pullin y Pierre Henry, los conceptos de *relief spatial* y *projection sonore*. Xenakis por su parte, exploraba de manera sistematizada las posibilidades intrínsecas de la relación del sonido con el espacio arquitectónico (al igual que Varese) y desarrollaba los conceptos de contrapunto espacial y geometría espacial. Stockhausen sentaba además las bases de la llamada *serialización del espacio* en *Gesang der Jünglinge* (1956) y desarrollaba en formas alternativas las ideas de la música espacial de Henry Brant, con *Gruppen* (1957). Hacia 1966 François Bayle introducía el *Acousmonium* en el GRM de París y junto

con él una formulación conceptual de lo que se conocería popularmente como orquesta de bocinas.

Hacia 1988, Michel Chion estableció los conceptos teóricos de *espacio interno* y *espacio externo*, resaltando la necesidad de formular una interrelación adecuada entre ambos desde el punto de vista compositivo-performático. Fue el primero en usar el término *spatialization*; junto con Risset, resaltaron las relaciones de la manipulación del espacio externo durante la ejecución de la música en vivo.

Por último, en su teoría de la espectromorfología, Dennis Smalley definió los conceptos de *subrogación* (de primer y segundo orden) y el del *gesto* como movimiento energético. Estableció también la idea de que el movimiento espacial es, desde el punto de vista compositivo, una forma de apoyar y enfatizar la espectromorfología inherente de los objetos sonoros dentro de la composición.

1.2. Espacio de parámetros y espacio como experiencia

Desde finales del siglo XIX y durante la mayor parte del XX se han desarrollado diversas concepciones de *espacio* en la música. Macedo (2015), por ejemplo, mediante un recuento histórico, plantea una taxonomía de cinco categorías: *metáfora*, *espacio acústico*, *espacialización sonora*, *referencia y localización*. Henriksen (2002), por su parte, recoge y estructura un análisis con seis tipos de espacios, divididos en dos grandes categorías: el espacio musical en su base perceptual y el espacio musical en su perspectiva compositiva-performática. La primera categoría consta de *espacio intrínseco*, *espacio extrínseco* y *espacio espectral*, mientras que la segunda se conforma por el *espacio compuesto*, *espacio de escucha* y el *espacio percibido*.

Cabe resaltar que la mayoría de las taxonomías actuales sobre espacio y espacialidad pueden considerarse combinaciones de las concepciones de autores y compositores como Schaeffer, Henry, Dhomont, Bayle, Smalley, Chion, Chowning, Varese, Stockhausen y demás grandes personajes de la música contemporánea. De aquí que no exista una homogeneidad teórica y conceptual sobre lo que significa *espacio* y *espacialidad* dentro de la música electroacústica, ya que las aportaciones al respecto corresponden en gran medida a las miradas individuales de los compositores y su particular forma de comprender el *espacio* conforme a su estética y proceso compositivo subjetivo.

En este sentido, para fines del presente trabajo, es necesario diferenciar dos tipos generales de concepción de espacio (muy parecido a la división propuesta por Henriksen). La primera se relaciona con el *espacio de parámetros musicales* y la segunda con la *espacialidad del sonido*.

El *espacio de parámetros musicales* busca establecer una cuantificación (abstracta o práctica) de un conjunto de cualidades del sonido o de la música (según el contexto y el autor), mientras que la *espacialidad* está asociada a todas aquellas características que conforman tanto el proceso compositivo como el de experiencia aural de la *sensación espacial* de una obra y sus elementos (objetos sonoros). Entonces, el término *espacio* tiene una doble connotación dependiendo del contexto: *espacio* como representación paramétrica de cualidades sonoras/musicales y *espacio* como conjunto de características metafóricas, perceptuales, acústicas y experienciales respecto de la escucha aural.

En este trabajo, el modelo teórico de *Espacio-Timbre* conjuga estas dos perspectivas: mediante el desarrollo de una representación paramétrica de ciertos aspectos sonoros (espectromorfología y localización espacial) se construye una formulación experiencial-perceptual de la *espacialidad*.

La caracterización o definición de un *espacio musical* o *espacio de parámetros sonoros* ha sido abordada de manera exhaustiva por numerosos autores y compositores a lo largo de la historia y ha tenido un particular auge en el transcurso del siglo XX. Helmholtz, Riezler, Conrad, Stofft, Kagel, McDermott, Rochberg o Morgan han contribuido al desarrollo de propuestas de formulaciones de espacios de parámetros sonoros. Para una revisión detallada al respecto, consúltese Harley (1994).

El propio Wishart (1998) plantea un modelo tridimensional de *espacio sonoro* cuyos elementos son: el continuo de alturas o pitch continuum, la dimensión del ruido-coloración y el timbre. “La discusión de las características del continuo de alturas es parte de la descripción de espacio sonoro de Wishart y está relacionada con lo que él denomina rejillas sonoras” (Macedo, 2015).

Por el lado de concepción de espacio como *experiencia de espacialidad*, tal y como lo describe Zelli (2010), para Francis Dhomont existen tres categorías de espacio en la música acusmática: *le figuratif*, *le symbolique* y *l'artificiel*. El *espacio figurativo* puede ser entendido como el primer nivel de relación semántica en el cual el material sonoro mantiene sus propiedades anecdóticas pero explora nuevas formas de generación musical instrumental, siendo la música concreta el mejor ejemplo de ello, a juicio del propio autor, debido a que existe una asociación

del material sonoro con la fuente (aunque no del todo con la forma de producción del sonido). El *espacio simbólico* podría entenderse como la experiencia perceptual que conduce a la analogía relacional y, por lo tanto, hacia una construcción semántica más compleja del material sonoro con metáforas y experiencias no necesariamente asociadas a las propiedades anecdóticas del mismo. Ambos, el espacio figurativo y el simbólico, “se usan como una metamorfosis para construir un mundo virtual de ideas que está más allá de la realidad tangible y que establece su propio criterio” (Zelli, 2010). Finalmente, Dhomont asocia el concepto del espacio artificial a la experiencia aural derivada de los sonidos sintéticos o puramente electrónicos. “Espacio artificial significa espacio simulado. La realización de tales espacios se ha hecho posible solamente en la música hecha por computadora” (Zelli, 2010).

Para Clifton (1983), el espacio se entiende como una experiencia de “estar dentro de un lugar” con características específicas. “El espacio musical tiene significado porque una persona se encuentra a sí mismo en ese lugar, como un lugar para ocupar un hábitat temporal”.

Françoise Bayle propone tres *especies de espacios*, en función de aspectos perceptuales y elementos cognitivos de la escucha: *présentification*, *identification* e *interprétation*. La *presentificación* tiene que ver con el proceso preatención y percepción pasiva descrito en el capítulo I, así como con la idea de conciencia audible espacial: “está caracterizada por un estado de escucha pasiva en el cual la percepción sensomotora del sonido irradiado, deviene en un escenario intuitivo de escucha mediante la diferenciación de la percepción de la dirección y distancia” (Zelli, 2010). En la etapa de *identificación* tiene lugar un proceso cognitivo más complejo en el que surgen asociaciones y discriminaciones semánticas con mayor detalle respecto a la conciencia de la espacialidad de los objetos sonoros y de la escena aural general. En la etapa final, la de *interpretación*, el escucha realiza asociaciones sobre la espacialidad de acuerdo con experiencias enteramente subjetivas, “los escuchas interactúan con el espacio a través de una percepción activa, componiendo el mismo de acuerdo a sus propios conceptos y afectando su configuración a través de la subjetividad” (Zelli, 2010).

Para Vande Gorne (1988) es posible identificar cuatro tipos de espacios: *espacio envolvente*, *espacio/sonido fuente*, *espacio geométrico* y *espacio ilusorio*. El *espacio envolvente* es aquel que genera una sensación de inmersión en el espectador y que sería opuesto al espacio fuente, el cual tendría la función de localización sonora puntual. El *espacio geométrico* se relaciona con un prediseño de la espacialización de acuerdo con volúmenes o estructuras específicas. Finalmente, el *espacio ilusión* tiene que ver con el trabajo compositivo realizado en el formato estéreo de tal modo que se pueda percibir una ilusión de profundidad de campo.

Existen numerosas propuestas y definiciones sobre espacio (desde el punto de vista paramétrico y desde la espacialidad), por lo que se recomienda al lector el extenso trabajo realizado al respecto por Harley (1994) para una detallada revisión histórica de dichas formulaciones. Para efectos del presente trabajo basta con reafirmar dos aspectos de la concepción de espacio: *espacio como representación cuantificable/metafórica de parámetros sonoros y espacio como representación experiencial de localización/ubicación /inmersión espacial*.

Es claro que los antecedentes previamente descritos han sido reutilizados, mezclados y reformulados por los compositores de música electroacústica interesados en el componente espacial; para el caso de una clasificación sobre espacialidad desde la perspectiva meramente composicional/performativa, es posible establecer una pauta general:

- *Música espacial acústica*. Distribución predeterminada de instrumentistas en la música acústica dentro de un recinto para generar inteligibilidad en texturas sonoras complejas y provocar efectos sonoros relacionados con el timbre (Ives, Brant, Stockhausen, etcétera) y con la ubicación espacial del sonido.
- *Serialización y concepción geométrica del espacio*. Diseño predefinido de trayectorias espaciales del sonido dentro del proceso composicional a partir de ciertos elementos, heurísticas, lógicas o fórmulas matemáticas (Xenakis, Stockhausen).
- *Difusión espacial*. Distribución en vivo de la obra en sistemas multicanal.

Es posible afirmar que en la actualidad, gran parte de los compositores interesados en la espacialidad del sonido tienen un acercamiento al uso del espacio basado en la teoría de Smalley y en los conceptos de Chion de manera más o menos consciente. En otras palabras, la composición del parámetro espacial está ligada a las características espectromorfológicas de cada objeto sonoro; ello genera un espacio interno que después adecúa o modifica de acuerdo con el espacio externo en el performance de la obra. En este sentido, el espacio está “sometido” a la funcionalidad tímbrica. De hecho, desde una perspectiva reduccionista, salvo la música espacial acústica, todos los acercamientos composicionales descritos crean espacios artificiales en el espacio interno de la obra (usando la terminología de Chion) a través de la espacialización y/o la localización mediante *plug ins*, o bien, crean la sensación espacial haciendo uso del espacio externo al implementar distintos métodos de difusión en tiempo real como es el caso de difusión con *mixer* (los cuales pueden estar directamente relacionados con las características del recinto como en el caso de Varese o Xenakis).

Una revisión histórica detallada del uso del espacio en la composición musical queda fuera de los objetivos de este trabajo, pero el lector puede encontrar un excelente referente en Bates (2009).

1.3 Espacio interno-espacio externo y el problema de la disimilaridad serial/perceptual

Uno de los principales problemas que enfrenta la creación electroacústica es, sin duda, la traslación de la obra desde el estudio al recinto de concierto; se trata de un proceso complejo ya que cada recinto es por lo general completamente distinto en sus características arquitectónicas y acústicas y, por supuesto, difiere del lugar en el cual se compuso la obra, el cual seguramente posee condiciones más o menos ideales o al menos fijas y controladas.

Esta problemática ha sido abordada por diversos compositores de renombre como Michel Chion, Jean Claude Risset, John C. Chowning y Dennis Smalley, por mencionar algunos. Sus descripciones e investigaciones son, hasta la actualidad, fundamentos cruciales para conservatorios y universidades donde tal conocimiento es presentado como referencia principal para la escuela acusmática-concreta inglesa y francesa.

Michel Chion describe dos tipos de espacio en la música acusmática: el *espacio interno* y el *espacio externo*. El primero se refiere a todos los elementos espaciales desarrollados composicionalmente y que quedan fijos en la obra misma. Es decir, “son los aspectos espaciales de la música que están composicionalmente predeterminados en la grabación como la reverberación, localización, planos y distancias” (Chion 1988). El *espacio externo*, por otro lado, es la adaptación del espacio interno en el recinto de concierto mediante la interpretación del compositor de la obra misma. Esta adaptación es muy común para difusión de obras estéreo (espacio interno) a configuraciones multicanal (espacio externo), aunque también se presentan con gran frecuencia adaptaciones de obras multicanal a otros sistemas multicanal. La traslación descrita es, en cierto modo, una reinterpretación del espacio composicional trabajado en el estudio, adecuada a las circunstancias específicas del performance en vivo de la obra en la sala de concierto; ello es realizado casi siempre mediante difusión por *faders* en mezcladoras multicanal.

Estableciendo analogías un tanto arbitrarias con el espacio interno y externo, es posible relacionar dichos conceptos con lo que Risset denomina *espacio ilusorio* y *espacio real* y con la in-

teracción que ambos establecen para desarrollar una experiencia final sobre el escucha. Para él, los sonidos pueden ser espacializados; esto es, “distribuidos en el espacio en una variedad de patrones y movidos en el mismo espacio a lo largo de diferentes trayectorias” (Harley, 1994), “para jugar en la localización del sonido, en su desplazamiento y en su cinética” (Chion, 1988).

Por su parte, para Dennis Smalley existen muchas categorías distintas o *tipos de espacios* en el nivel perceptual del compositor, en primera instancia, y del espectador, en segunda. Tal y como lo menciona Harley (1994), para Smalley el *espacio compuesto* es el imaginario espacial considerado por el compositor, mientras que el *espacio audible*, que puede ser personal o público, se encuentra fuera del control del o la compositora. El espacio audible puede pensarse como la experiencia de la percepción del espacio compuesto desde el punto de vista del espectador y de hecho, “lo que un escucha percibe durante un concierto es un espacio superimpuesto, un anidado de espacios compuestos dentro de un espacio audible” (Smalley, 1991).

Retomando el binomio espacio interno-espacio externo, la traslación del primero al segundo implica un número de circunstancias técnicas que afectan considerablemente la percepción de la obra al momento del performance; quizá, una de las más importantes es la distorsión de la imagen espacial, la cual afecta las distancias virtuales (y por lo tanto los tamaños de las escenas espaciales virtuales), la localización fija, el involucramiento sonoro y, en particular, la *percepción de la continuidad* (estos conceptos se explicarán en las secciones siguientes). Como se vio en el primer capítulo, desde el punto de vista perceptual la continuidad sensorial es muy importante para establecer representaciones mentales coherentes y convincentes de los objetos sonoros. Disrupciones en la continuidad provocan ambigüedades perceptuales que son interpretadas como errores y, más aún, podrían provocar consecuencias como alejar al espectador de la experiencia inmersiva debido a posibles indicios de antinaturalidad ecológica.

1.4. Espacio virtual, espacialidad e inmersividad

Para plantear un modelo teórico que aborde la problemática de lo que se denominó *Espacio-Timbre* es necesario, en primer término definir claramente el concepto de *espacio* en sí mismo. Como se mencionó con anterioridad, esta definición estará orientada hacia una mirada del fenómeno experiencial por lo que no intenta desarrollar una ontología del *espacio* sino un

acercamiento epistemológico del mismo; todo ello con el objetivo de estructurar este concepto dentro del campo de la electroacústica.

En Soria (2020) se presenta una propuesta teórica original para definir *espacio, espacialidad e inmersividad*, a partir de un marco teórico fundamentado en la ciencias de la cognición y la percepción, y mediante la definición de *experiencia estética* (entendida desde dichas áreas).

Haciendo uso de los conceptos *bottom-up* y *top-down*, la teoría del procesamiento mental y de lo que se definió como “condiciones perceptual-cognitivas intrapersonales”, se planteó una definición inicial de *inmersividad*:

La inmersividad puede ser entendida como el fenómeno que deriva en un nivel elevado de abstracción de la ubicuidad espacio temporal que experimenta un individuo al realizar una tarea específica; implica el uso y la focalización de su atención plena, provocando posibles efectos de distorsión en la percepción del tiempo y de la localización física del individuo.

En función de lo anterior se distinguieron categorías específicas para entender este concepto como fenómeno en sí:

- *Inmersividad virtual*. Abstracción fenoménica de la ubicuidad espacio-temporal que surge debido a la contemplación de un objeto específico. Implica la atención focalizada en un nivel elevado; es decir, abstracción de la realidad espacio-temporal física guiada por la experiencia sensorial.
- *Inmersividad de localización*. Abstracción fenoménica de la percepción de la temporalidad debido a la contemplación o atención focalizada en un nivel elevado de la ubicuidad del espacio físico.
- *Inmersividad de temporalidad*. Abstracción fenoménica de la percepción de la ubicación del espacio físico debido a la contemplación o atención focalizada, en un nivel elevado, del sentido de la temporalidad.
- *Inmersividad propioceptiva inversa*. Abstracción fenoménica de la percepción de la corporalidad física del individuo debido a la contemplación o atención focalizada, en un nivel elevado, del sentido de la espacio-temporalidad.

A partir de estas bases se desarrolló una discusión conceptual que derivó en una definición de espacio inscrita en el marco teórico ya mencionado. A continuación se transcribe esta parte del trabajo de Soria (2020) con el fin de que el lector (a reserva de que pueda consultar dicho capítulo posteriormente) comprenda el planteamiento y la lógica deductiva utilizada para definir dichos conceptos:

Una posible perspectiva de la idea de espacio (quizá la más simple desde la cual es posible partir), es aquella en la cual éste puede definirse como una región delimitada o como la acción misma de delimitar una región. En este sentido, el espacio como fenómeno, surge de la acción de observar o experimentar sensorialmente un ente con una realidad ontológica *a priori*. Desde una perspectiva materialista, este ente que ya existe por sí mismo, independientemente del observador, forma entonces lo que podría denominarse espacio físico.

Como punto de partida se tienen entonces dos tipos de espacio: el físico y el fenoménico, siendo este último, el resultado de experimentar el primero.

Se propone que el espacio físico sea entendido como una región con una realidad independiente del observador, que además posee como característica primordial, la extensión. Esta propiedad se plantea simultáneamente tanto desde la fisicalidad como desde la temporalidad del espacio mismo. A esta extensión se le puede denominar dimensionalidad.

Para el otro tipo de espacio, el que es abordado como fenómeno, cabe enfatizar de nuevo que su caracterización no estará completa sino hasta el momento mismo de la experiencia estética. En este sentido, se puede proponer como primera aproximación, que el espacio fenoménico es la acción de delimitar el espacio físico a través del fenómeno de la experiencia sensorial subjetiva.

Dicho lo anterior, resulta necesario establecer: ¿qué se entiende por delimitación? para posteriormente enfatizar ¿qué elementos se consideran para establecer esa delimitación dentro de un marco específico para definir la espacialidad cuando el objeto de contemplación es precisamente uno de tipo artístico?

Delimitar, en el sentido anteriormente expuesto, no significa otra cosa que establecer los límites del fenómeno de la percepción de la experiencia estética del espacio físico en un intervalo de tiempo dado. Dicho de otro modo, delimitar implica establecer todo aquello que el observador es capaz de percibir y aprehender del espacio físico en un momento dado por medio de un proceso sensorial subjetivo condicionado por un marco perceptual-cognitivo. Por lo tanto, cuando el observador experimenta sensorialmente el espacio físico, como consecuencia, está al mismo tiempo delimitando la dimensiona-

lidad y esta delimitación fenoménica de la extensión del espacio físico (entendida tal y como fue descrita) puede denominarse espacialidad física.

En este sentido, si se toma como premisa la realidad *a priori* de la dimensionalidad, entonces la espacialidad física puede ser entendida como una abstracción perceptual-cognitiva de la dimensionalidad de la región que experimenta el observador.

Si la espacialidad se considera como una abstracción perceptual subjetiva de la dimensionalidad, entonces como se dijo anteriormente, el elemento principal que la construye es la experiencia estética del observador en un instante de tiempo dado, a través de un proceso perceptual-cognitivo. Dicho de otro modo, la espacialidad de una región cualquiera está asociada a la extensión física del mismo, en medida en como es percibida por el observador en un instante dado y, por esta razón, la construcción perceptual-cognitiva de esa espacialidad depende directamente de lo que se estableció como condiciones perceptual-cognitivas intrapersonales.

Para un recinto cerrado, por ejemplo, la formulación subjetiva de la espacialidad de ese recinto, descrita por distintos individuos, coincidirá en gran medida debido a que la percepción de dicho espacio es asistida por la delimitación física natural del recinto mismo. Para un espacio físico abierto es más complicado establecer delimitaciones claras; sin embargo, debido precisamente a esta falta de capacidad de establecer fronteras delimitantes específicas, en tal caso es también probable que la espacialidad coincida en gran medida para distintos individuos.

Una vez que se planteó la idea de espacio, se propuso introducir el concepto de *espacio ficticio*, entendido como una serie de construcciones, reconstrucciones o incluso deconstrucciones virtuales de cualquier tipo de recinto o región, las cuales pueden ser consecuencia de procesos naturales o desarrollados a través de métodos artificiales específicos intencionales. Dicho de otro modo, si el individuo experimenta un fenómeno sensorial que difiere de la espacialidad del recinto físico donde se encuentra, se puede decir que dicho individuo está dentro de un espacio ficticio o virtual.

Lo anterior permitió, por lo tanto, plantear una noción de *virtualidad* basada en todas las consideraciones anteriores, bajo el razonamiento de que todo espacio virtual se construye y existe dentro de un espacio físico específico y, además, que la virtualidad es la ubicuidad alterna y separada de espacio-tiempo que se experimenta debido a la abstracción de la inmersividad virtual. En este sentido, y de manera inversa, la capacidad de un espacio virtual de abstraer en mayor o menor grado al individuo del espacio físico conduce a la inmersividad. Lo anterior

tiene como implicación lógica un sentido de ida y vuelta. Por un lado, la inmersividad virtual deviene en espacio virtual y, por el otro, el espacio virtual produce inmersividad virtual.

Todo esto va a conjugarse en la idea de *obra inmersiva*, aquella que busca establecer experiencias estéticas que puedan generar experiencia de inmersividad en el espectador, teniendo la capacidad de expandir la estimulación sensorial no sólo en formato, sino en calidad de los elementos propios de la pieza y en el diseño de la misma, vista como un espacio virtual. Éste es el punto clave del razonamiento propuesto: *es posible generar obras capaces de dirigir al espectador hacia una experiencia de inmersividad si se les concibe y desarrolla a lo largo del proceso creativo como espacios virtuales en sí mismas*. Lo anterior concluye en una definición específica acerca de lo que se denominó *arte experiencial inmersivo*:

DEFINICIÓN 3.1. Al plantear la obra artística explícitamente como un espacio virtual en el que se privilegien la extensión de las experiencias sensoriales dirigidas a la búsqueda de la inmersividad, es posible hablar de un tipo muy particular de obra: arte experiencial inmersivo.

De este modo, la obra electroacústica concebida y desarrollada con el objetivo de detonar experiencias de inmersividad en el espectador sería una obra experiencial inmersiva. Para lograr esto, el marco teórico planteado en este libro podría resultar útil no sólo desde el punto de vista teórico sino incluso práctico.

2. Espacio virtual e inmersividad en la electroacústica multicanal

Algunos ejemplos de construcción intencional de espacio virtual se encuentran en los entornos desarrollados con tecnologías de realidad aumentada, realidad virtual y audio 3D (binauralidad o sistemas multicanal). Para el caso de la música electroacústica, el espacio virtual se construye a través de sistemas de audio que van desde las configuraciones binaurales, estéreo, multicanales 2D y multicanales 3D.

De acuerdo con el contexto expuesto en la sección anterior, es posible afirmar que la experiencia electroacústica multicanal se vuelve, en cierto modo, una creación deliberada de espacios virtuales cuyo fin deviene, en menor o mayor medida, en la búsqueda de la inmersividad. Por supuesto, ésta no es el objetivo primordial en todas las creaciones electroacústicas, incluso

puede no ser considerada por el compositor o compositora; sin embargo, el formato multicanal, por su propia naturaleza, plantea esta relación potencial inherente: la de la obra con la inmersividad.

Cuando la obra electroacústica busca intencionalmente generar una resolución espacial satisfactoria, es necesario que se establezcan métodos y procesos detallados que permitan al espectador obtener una experiencia de inmersividad convincente. Las metodologías, taxonomías y estrategias para lograr una sensación de inmersividad variarán de acuerdo con el creador, no obstante, al día de hoy la mayoría de ellas se desprenden o se formulan de manera heurística.

En esta sección se propone una metodología para el desarrollo de la inmersividad en la música electroacústica multicanal basada en la hipótesis de que el timbre y el espacio son parámetros indisolubles entre sí y que, por lo tanto, pueden ser considerados como un sólo elemento compositivo: *el Espacio-Timbre*. Esta premisa, fundamentada principalmente en todos los referentes descritos en los capítulos I y II, tiene como implicación directa lo siguiente:

Los objetos sonoros, de acuerdo con su perfil espectromorfológico, poseen una zona óptima de resolución/localización espacial como fuente virtual dentro de un sistema multicanal.

De este modo, se propone suerte de orquestación espacial de los objetos sonoros dentro de un marco de creación electroacústica multicanal basado en la hipótesis anterior. Este concepto será denominado a partir de este punto como *orquespectración espacial*.

OBSERVACIÓN 3.1.1. Como consecuencia directa de los resultados mostrados en el capítulo I, es posible afirmar que para los sonidos utilizados en los experimentos descritos, la espectromorfología de los objetos sonoros está asociada de manera directa y proporcional a su localización espacial. Dicho de otro modo, para los sonidos, configuraciones y metodologías descritas en los experimentos, *el timbre está asociado a una zona de resolución espacial óptima* tanto en el azimut como en el plano elevación, entendiéndose *timbre* tal y como se planteó en la definición 4.0.1, del capítulo IV.

Si bien los resultados psicoacústicos al respecto de la localización espacial y el contenido frecuencial de los sonidos han sido presentados para objetos sonoros simples, se hace la conjetura de que pueden ser extrapolados a sonidos mucho más complejos dependiendo de su

distribución de información espectral y, de este modo, se podrían utilizar como guía directa para toda una gama espectromorfológica mucho más compleja. Lo anterior ha sido probado heurísticamente a través de mi propia experiencia, con resultados bastante aceptables al respecto de los distintos tipos de localización espacial de objetos sonoros en obras electroacústicas multicanal.

Se mencionó con anterioridad que la inmersividad, en su aspecto más general, se considera compuesta por dos elementos primordiales: las variables intrapersonales del espectador y las características de la experiencia artística. En este sentido, la propuesta aquí planteada explora una metodología centrada en el diseño y construcción del segundo elemento con un objetivo paralelo: el desarrollo de una alta resolución de inmersividad en la creación electroacústica multicanal.

La experiencia de escucha de una obra electroacústica multicanal tiene lugar en un *espacio físico* específico: un recinto, un estudio o un espacio abierto, por ejemplo. Este recinto posee características estructurales y arquitectónicas particulares que, a su vez, definen y delimitan las propiedades acústicas del mismo. Como ejemplos de estas últimas se tienen: coeficiente de absorción, área total de absorción, tiempo de reverberación, campos sonoros, modos normales de vibración, etcétera. Ejemplos de características arquitectónicas son: diseño estructural del recinto y materiales de construcción (interior/exterior), entre otros.

Dado un recinto o espacio físico, la experiencia de escucha de una obra electroacústica multicanal requiere de la construcción de un *espacio virtual* que, en este caso, estará definido por la arquitectura del espacio físico en conjunto con las características del sistema multicanal. De este modo se establece una región delimitada dentro del *espacio físico*, delimitada también por componentes físicos (arquitectura del recinto y características del sistema multicanal). Se puede decir entonces que sus propiedades son de fisicalidad y son previas a la experiencia aural. El resultado de la conjunción de estas características (las de la arquitectura del recinto y del sistema multicanal) se nombrará *espacio audible*.

Dado que el sistema multicanal se insertó en el espacio físico, el primero hereda las propiedades acústicas del segundo aunque posee otras características propias; por ejemplo, el número de bocinas, la disposición de las bocinas en el espacio físico (la distancia y el ángulo entre ellas, la altura a la cual se colocan respecto del nivel del piso), el tipo y tamaño de bocinas y el diseño estructural del sistema en sí, el cual puede ser 2D (un solo nivel) o 3D (más de un nivel, con formas cúbicas, esféricas o semiesféricas).

Todas estas características, tanto las propiedades acústicas heredadas del espacio físico como las inherentes del sistema multicanal, conforman lo que se denominará *arquitectura del sistema multicanal*. Nótese que el espacio físico delimita, en gran medida, varias de las características que se denominaron como *inherentes* del sistema multicanal; sin embargo, por cuestiones de definición se consideran separadas de la arquitectura del mismo.

DEFINICIÓN 3.1.1. El *espacio audible* $[\Xi]$ es el espacio virtual que surge debido a la experiencia de fisicalidad delimitada por la arquitectura del espacio físico y la arquitectura del sistema multicanal.

La arquitectura del espacio físico será denotada mediante $A(\varphi)$ y la arquitectura del sistema multicanal por $A(\vartheta)$. La *arquitectura del espacio audible* será el término que haga referencia a la conjugación simultánea de $A(\varphi)$ y de $A(\vartheta)$.

En el momento en que el espectador se coloca dentro del espacio audible y ocurre la experiencia de escucha *in situ* de la obra, surge el *espacio virtual sonoro*.⁸

DEFINICIÓN 3.1.2. La experiencia perceptual aural *in situ* de la obra dentro de un espacio audible es la particularización del concepto de espacio virtual para el caso de la música electroacústica multicanal y será denominado *espacio virtual sonoro* o bien, espacialidad.

2.1 Consideraciones generales sobre el espacio audible

Cada espacio audible tendrá un *sweet spot* específico de acuerdo con su arquitectura y por, lo tanto, un área óptima de escucha que por lo general decae de manera proporcional mientras más alejado se encuentre el espectador de dicha ubicación. En la figura 6 se muestra la representación genérica de un espacio audible para una configuración usual octafónica. Nótese que la delimitación del espacio audible transita entre el perímetro octagonal y la periferia circular. Dependiendo de los procesos y herramientas compositivas, la delimitación del espacio vir-

⁸ La experiencia aural puede y es realizada múltiples veces por el o la compositora; es decir dentro del proceso compositivo, el creador intercambia continuamente su perspectiva de compositor a espectador.

tual sonoro puede extenderse más allá de dicha periferia del espacio audible; es decir, es posible crear objetos sonoros que sean percibidos con una localización fuera del espacio audible.

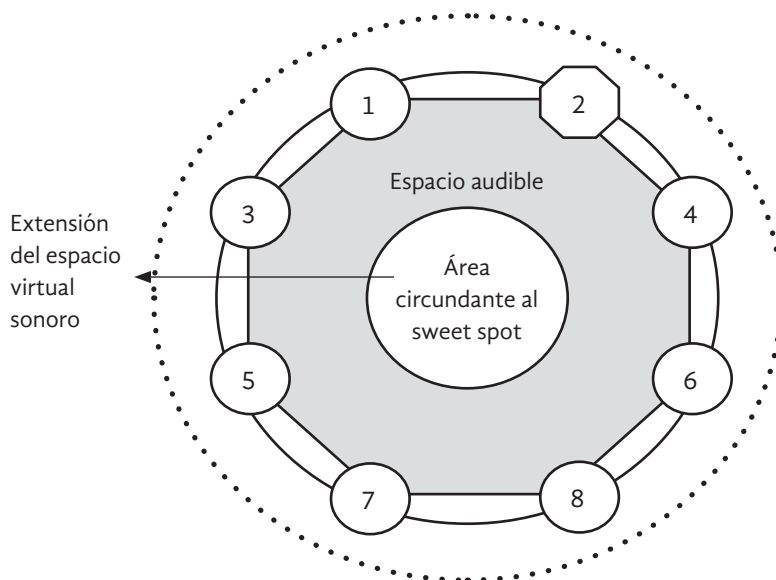


Figura 6 | Representación gráfica del espacio audible y de la extensión del espacio virtual sonoro.

En las figuras 6, 7, 8 y 9, se muestran ejemplos concretos de arquitecturas físicas-multicanal que generan espacios audibles 3D.

En el capítulo anterior se explicó que si la resolución del espacio virtual sonoro era lo suficientemente alta, entonces era posible que el espectador experimentara una abstracción del espacio físico para sentirse dentro del espacio virtual. Esta resolución dependerá básicamente de tres características primordiales. La primera está relacionada con la eficacia de la arquitectura del espacio audible, la segunda con la resolución espacial innata de la obra electroacústica y la tercera con la extrapolación o adaptación de esa resolución espacial de la obra al espacio audible. Las dos últimas características derivan de los conceptos de *espacio interno* y *espacio externo* definidos por Chion (2017), mientras que la primera puede considerarse como una medida cualitativa de la arquitectura del sistema multicanal.

¿Qué hace a una arquitectura multicanal más eficiente que otra? Definitivamente dependerá de distintas variables, pero entre las más importantes a considerar están la simetría del diseño estructural y la relación adecuada entre las dimensiones físicas del espacio audible y el



Figura 7 | Sonic Arts Research Centre, Belfast (<<https://presspack.rte.ie/2014/10/10/the-lyric-feature-126/>>).



Figura 8 | Birmingham Electroacoustic Sound Theatre (BEAST) (<<https://asmo23.wordpress.com/2008/03/22/behold-the-beast/>>).



Figura 9 | Hexagram-UQAM ESPACE, Montreal (<<https://getitinwriting.net/spat-revolution-helps-uqam-professor-bring-immersive-to-life/>>).



Figura 10 | CCRMA 3D Listening Room, Stanford (<<https://ccrma.stanford.edu/room-guides/listening-room>>).

número de bocinas (Rumsey, 2010). Por ejemplo, un espacio demasiado grande con un sistema multicanal de pocas bocinas difícilmente logrará arrojar una resolución espacial alta.

DEFINICIÓN 2.1.3. El grado de *immersividad* experimentado por el escucha es directamente proporcional a la eficacia de la arquitectura del espacio audible, a la resolución del espacio interno de la obra electroacústica y a la extrapolación del espacio interno en el espacio audible.

En este punto tiene sentido, entonces, referir como *diseño espacial multicanal* a las *estrategias, métodos y herramientas utilizadas para crear el espacio interno de una obra electroacústica*; más aún para poder lograr la inmersividad en un grado aceptable, partiendo del hecho de que uno de los objetivos primordiales es la búsqueda de la experiencia de la inmersividad en el espectador (obra experiencial inmersiva). De este modo y toda vez que se han planteado formalmente los términos adecuados para la comprensión del espacio según esta propuesta teórica particular, se procederá a abordar una metodología propia encaminada a investigar el problema del *diseño espacial multicanal* que permita contribuir positivamente a la inmersividad.

Si bien el espacio virtual sonoro constituye una conceptualización cuasi abstracta, para fines de la presente propuesta resulta útil asociarlo con una representación cuantificable como la representación tridimensional del sonido a partir de la conjunción del plano azimut, plano elevación y plano horizontal. De este modo, en las siguientes secciones, será posible realizar representaciones específicas de los elementos constitutivos del diseño espacial multicanal, el cual puede entenderse como la *composición del espacio* dentro de una obra de música electroacústica; esto es, la creación de espacios ficticios sonoros dinámicos y múltiples que surgen, se desarrollan y convergen dentro del espacio audible.

2.2. Diseño serial vs. percepción

Smalley (1997) introdujo el ahora recurrente término *espectromorfología*, mediante el cual caracteriza a los objetos sonoros a través de la interacción de dos componentes: el espectro sonoro y la manera en que dicho espectro es moldeado a través del tiempo en la dinámica (morfología). Smalley enfatiza, en su conocido artículo, que su propuesta está orientada hacia una explica-

ción teórica de la experiencia perceptual aural y desde el punto de vista compositivo de la forma y contenido.

Como se ha observado, el presente trabajo se fundamenta en gran medida en los aspectos perceptuales de la experiencia de escucha y es por eso que el concepto de *espectromorfología* resulta pertinente en este contexto. A partir de ese punto es necesario delinear de manera clara el concepto de *diseño serial* dentro del proceso compositivo.

Las concepciones de prediseño en el manejo de parámetros sonoro-musicales dentro del contexto de la música electroacústica son comunes y tienen bastante tiempo de ser utilizadas. Por prediseño se hace referencia a la manera en que se establece la evolución de uno o varios parámetros sonoro-musicales a través del tiempo de manera específica-predeterminada y que, por lo general, obedece a algún tipo de regla, proceso o fórmula matemática.⁹ Debido a que esta perspectiva se relaciona naturalmente (al menos desde un punto de vista metafórico) con las aproximaciones y procesos del serialismo integral, en algunas ocasiones dichas perspectivas se nombran o catalogan como *seriales* (Bates, 2012).

Una problemática recurrente de estas concepciones basadas en el *prediseño serial* es la difícil concordancia entre el resultado final percibido en la experiencia aural y el prediseño mismo; esto es, existe una diferencia potencial e innata entre el *diseño serial* y el *resultado final perceptual*. Lo anterior se vincula con la distinción hecha por Blauert (1969) entre *evento audible* y *evento sonoro*, donde el primero se refiere a la *percepción audible determinada por atributos espaciales y temporales* y el segundo se asocia al *aspecto físico del proceso de escucha*.

DEFINICIÓN 3.1.4. Dado cualquier proceso compositivo que implique el uso de uno o más parámetros sonoros, la *disimilitud serial-perceptual* será la diferencia existente entre el prediseño serial y la experiencia final aural percibida.

La *disimilitud serial-perceptual* se plantea como una medida cualitativa y hasta cierto punto heurística. El diseño y realización de pruebas perceptuales/psicoacústicas que puedan establecer una medida cuantitativa de la misma para diversos parámetros no resulta de ningún modo complicada, sobre todo si se consideran todos los resultados mostrados en el capítulo I. La importancia de establecer esta medida heurística de disimilitud para el presente trabajo ra-

⁹ Este aspecto se relaciona directamente con los diversos aspectos de la *composición algorítmica*.

dica en que aquí se puntualiza, como objetivo, que la composición del parámetro Espacio-Timbre esté en función de una validación perceptual acorde con el diseño serial. Si la disimilitud serial-perceptual no jugara un papel predominante, entonces nos encontraríamos frente a otro tipo de acercamiento estético más orientado, por ejemplo, hacia el arte procesual.

La disimilitud serial-perceptual, tal como se ha presentado, se desprende además de hechos psicoacústicos observados en distintas pruebas. De aquí se ha concluido, entre otras cosas, que “cambios en la parte física del sonido (incluyendo frecuencia, intensidad, fase y dirección) pueden no siempre generar una diferencia perceptual” (He, 2017). Cetta (2003) puntualiza, por su parte, la idea anterior, afirmando que dado que el espacio físico y el auditivo no son necesariamente idénticos; la posición del evento sonoro con respecto al evento auditivo¹⁰ no siempre coinciden y, por lo tanto, la localización espacial puede resultar ambigua.

3. Componentes primarios de la espacialidad

El diseño espacial multicanal dentro de la música electroacústica es un área amplia y bien documentada que ha evolucionado rápida e ingentemente desde los primeros días del *potentiométre d' espace* de Schaeffer y Pierre Henry. Es un campo que ofrece al compositor una múltiple paleta de recursos composicionales, estructurales, sensitivos y estéticos ya sea que se “busque colocar al escucha dentro de un ambiente de sala de concierto o dar la ilusión de traer a los músicos a su sala de escucha personal” (Rumsey, 2001). Una mezcla dinámica de ambas perspectivas juega un papel central, y en ocasiones hasta primordial, dentro de áreas como la música acusmática donde “el acercamiento espacio-forma es diferente de otras metodologías en el sentido que coloca al tiempo, al servicio del espacio” (Smalley, 2007).

En este sentido, el diseño espacial propuesto en este trabajo se asocia con la *composición del espacio* y está basado en tres elementos primarios:

1. Localización de objetos sonoros fijos.
2. Trayectorias de objetos sonoros móviles.
3. Envolvimiento sonoro.

¹⁰ Los términos evento sonoro y auditivo derivan de la terminología de Blauert (1997).

Cuando los elementos anteriores se mezclan adecuadamente en la creación de la obra, y si la resolución de cada una de ellos es adecuada, estas propiedades composicionales pueden ayudar a generar en el escucha la experiencia de *inmersión espacial* o *immersividad*, apoyada de los otros dos elementos descritos en la sección anterior. La decisión de fundamentar el diseño espacial en estos tres elementos surge después de realizar comparaciones y análisis de las distintas conceptualizaciones de espacialidad discutidas en el capítulo II, así como de comprobaciones prácticas dentro del proceso composicional por parte del autor de este libro.

Ya se ha analizado el problema de la diferencia entre el diseño serial y el resultado final percibido, la cual se definió como disimilitud serial-perceptual. Este concepto cobra especial relevancia al momento del diseño espacial debido a que en muchas ocasiones esta disimilitud puede resultar bastante considerable; más aún, según lo propuesto en este trabajo, cada objeto sonoro, de acuerdo con su perfil espectromorfológico, tendrá una zona de localización espacial óptima medianamente definida, en la cual se tendrá una mayor resolución de localización espacial perceptual.

En este punto es necesario establecer una clarificación en el uso del término *objeto sonoro* con el fin de evitar confusiones en secciones posteriores.

DEFINICIÓN 3.2.1. Un *objeto sonoro* es cualquier elemento o conjunto de elementos que son considerados como una sola unidad cohesiva ya sea desde el prediseño serial o en la experiencia perceptual aural final.

Considerado desde su cualidad física como fuente virtual (vibración producida por el sistema multicanal), éste radica en el espacio audible. Cuando se le considera desde el proceso de percepción auditiva (con base en los principios Gestalt y del ASA y en sus perfiles espaciales y espectromorfológicos), se puede afirmar que radica en el espacio virtual sonoro. El objeto sonoro se denotará como:

$$\mathcal{O} = \mathcal{O}_{spat} \cup \mathcal{O}_{spmor}$$

donde \mathcal{O}_{spat} representa las propiedades de espacialidad aural (localización espacial) y \mathcal{O}_{spmor} representa aquellas propiedades espectromorfológicas de dicho objeto sonoro; ambas desde las perspectivas simultáneas acústico-perceptuales.

La definición anterior hace una distinción explícita del objeto sonoro dependiendo de la perspectiva desde la cual se analice. Si se le observa como una vibración acústica desplazándose a través del espacio físico, entonces se afirma que radica en el espacio audible. En el caso de que exista un proceso perceptual de asociación con sus perfiles espectromorfológicos y espaciales, se considera que radica en el espacio virtual sonoro. En ambos casos se le denomina *objeto sonoro*. La definición anterior puede considerarse descriptiva y, tomando esto en cuenta, se hace aquí la aclaración de que en el capítulo IV se definirá el objeto sonoro de manera más detallada y abstracta. Mientras tanto, debe tomarse en cuenta que la definición anterior toma en cuenta las propiedades espacio-tímbricas desde una perspectiva acústica y espectromorfológica simultáneamente. En este sentido, dado que se asume que el Espacio-Timbre es un solo parámetro, un objeto sonoro estará definido, precisamente, por sus características conjuntas de espacio (o localización espacial) y timbre.

La razón de lo anterior radica en lo que se dijo en la definición 3.1.2 y es el hecho de que el espacio audible se convierte en espacio virtual sonoro cuando ocurre la experiencia aural perceptual *in situ*. Esto implica que existe una relación inherente entre el espacio audible y el espacio virtual sonoro y que la focalización de alguno de ellos depende de la perspectiva particular de observación.

CONJETURA 3.2.1. El objeto sonoro es equivalente al torrente auditivo tal y como está descrito en la definición 1.4.2.

CONJETURA 3.2.2. Las cualidades acústicas del objeto sonoro que en el proceso perceptual permiten que sea aprehendido como una unidad, están dadas por los factores del agrupamiento simultáneo y secuencial del ASA.

CONJETURA 3.2.3. El objeto sonoro considerado desde su cualidad perceptual y dotado de una función composicional o musical, es equivalente a la TGU definida por Tenney.

Las conjeturas anteriores establecen, entonces, mediante conexiones conceptuales, una definición robusta de objeto sonoro: como ente aural dotado de propiedades acústico-perceptuales (objeto sonoro como torrente auditivo) y como ente aural con función composicional/musical (TGU). Se establece que el objeto sonoro posee de manera innata estas dos características simultáneamente y que es sólo cuando se enfoca la atención en alguna de ellas cuando se tornan individuales.

De modo similar, es necesario considerar un concepto de *espacio* que conjugue a la vez las características físicas y las experienciales; es decir, que sea la concepción sincrónica de espacio audible y espacio virtual sonoro:

DEFINICIÓN 3.2.2. Se define el *espacio aural* $\mathfrak{S}(a)$ como la caracterización simultánea del espacio audible y del espacio virtual sonoro con focalización en la localización espacial. Esto es, la consideración general de espacialidad desde el prediseño serial y también desde la percepción final.

En otras palabras, el espacio aural es *la focalización conjunta del prediseño serial y la percepción final, con respecto a los atributos de la espacialidad*. Es importante hacer esta distinción ya que más adelante se hablará de espacio espectromorfológico que será, por lo tanto, la focalización conjunta del prediseño serial y la percepción final con respecto a los atributos de la espectromorfología.

Nótese que con la definición anterior se podrán desarrollar definiciones y resultados que contemplen tanto las características acústicas como las perceptuales respecto del proceso compositivo espacial, el cual es el objeto de las siguientes secciones.

DEFINICIÓN 3.2.3. Dado un espacio aural $\mathfrak{S}(a)$, un elemento $\mathcal{O}_{spat} \in \mathfrak{S}(a)$ será precisamente la descripción de localización espacial de un objeto sonoro \mathcal{O} . La duración temporal de dicho elemento estará dividida en dos: duración acústica y duración perceptual, y serán denotadas como $|\mathcal{O}_{spat}|_a = n_a$ y $|\mathcal{O}_{spat}|_p = n_p$ respectivamente. Cuando no se especifique otra cosa, la duración temporal hará referencia a la duración perceptual y se denotará simplemente como $|\mathcal{O}_{spat}| = n$.

Para fines generales, es posible asumir que la duración de un objeto sonoro coincide con la duración de su representación espacial y de la representación espectromorfológica. No obstante, derivado de lo planteado en el capítulo I, es claro que dichas duraciones pueden diferir en casos muy concretos; por ejemplo cuando los objetos sonoros se desplazan a través de trayectorias bajo condiciones particulares. A partir de este momento se asumirá, por facilidad de notación, que las duraciones son las mismas y que se puede hablar de *duración* del objeto sonoro en general: $|\mathcal{O}| = |\mathcal{O}_{spat}| = |\mathcal{O}_{spmor}| = n$.

OBSERVACIÓN 3.2.1. La *espacialidad virtual* entendida desde la electroacústica multicanal es la experiencia aural que surge al momento de experimentar el fenómeno perceptual-cognitivo de la escucha de la obra. Por esta razón, la espacialidad virtual dependerá del espacio audible y de la *conciencia audible espacial*.

La observación anterior hace hincapié en el hecho de que los *componentes primarios de la espacialidad* están relacionados de manera directa con la *conciencia audible espacial de ubicuidad* y con la *conciencia audible espacial de ambiente o difusión*. Por lo tanto, cuando más adelante se construya la representación abstracta de la *conciencia audible espacial*, se sobreentenderá que se refiere a la espacialidad y su descripción a través de los componentes primarios que se desarrollarán a continuación.

3.1. Objetos sonoros fijos

La idea de *objeto sonoro fijo* corresponde a aquel cuya posición espacial dentro del espacio aural no sólo no cambia durante su duración total, sino que puede ser perceptualmente localizable en un punto o un área específica delimitada dentro de dicho espacio. Recuérdese que, según lo descrito en la sección 3.2 del capítulo II, el sistema de coordenadas a utilizar para la descripción de localización espacial, será el *interaural polar*.

DEFINICIÓN 3.2.4. Dado un espacio aural $\mathfrak{S}(a)$ y $\mathcal{O}_{spat} \in \mathfrak{S}(a)$, para un objeto sonoro \mathcal{O} con una duración total $|\mathcal{O}| = n$, la representación de su ubicación espacial en un instante de tiempo t , con respecto al sistema de coordenadas interaural polar, estará dada por:

$$\gamma(\mathcal{O}_{spat}, t) = (\alpha(t), \beta(t), \theta(t), \phi(t), d(t))$$

Se dice que \mathcal{O}_{spat} es un *objeto sonoro fijo* si su posición dentro de $\mathfrak{S}(a)$ es constante a lo largo de su duración total:

$$\gamma(\mathcal{O}_{spat}, t) = \gamma(\mathcal{O}_{spat})$$

para todo $t \in [0, n]$.

La percepción de la espacialidad depende de la construcción de la experiencia aural, por un lado, y la abstracción individual subjetiva de cada espectador, por el otro. Tomando en cuenta esto último, es posible que la experiencia de la misma varíe en cada individuo. Así, se tiene que, para el caso del creador, esta espacialidad es una sola y por lo tanto se puede considerar constante. El objetivo del diseño espacial tal y como ha sido propuesto es, entonces, la creación de espacios ficticios dentro del espacio aural de modo tal que puedan dirigir la experiencia de la escucha a la inmersividad.

Para construir dichos espacios ficticios se plantea que el elemento básico y primario es el concepto de *distancia espacial*. Para el caso de fuentes fijas, dicho elemento se denominará *distancia espacial lineal*.

Supóngase que se coloca una única fuente fija $\mathcal{O}_{spat} \in \mathfrak{S}(a)$ dentro del espacio aural, con localización $\gamma(\mathcal{O}_{spat}, t) = (\alpha(t), \beta(t), \theta(t), \phi(t), d(t))$. Si bien los componentes de este sistema de coordenadas generan una sensación de distancia básica entre el escucha (en este caso el compositor) y la fuente sonora dentro del espacio aural, la percepción de la localización espacial de dicha fuente única será débil debido a que no existe un punto de referencia contra el cual contrastar dicha ubicación espacial.

Las escenas auditivas cotidianas muy rara vez están compuestas por una única fuente sonora y, por lo tanto, puede afirmarse que esta situación extraordinaria no es natural desde el punto de vista ecológico. Esto no implica que el escucha no pueda localizar una fuente única dentro de una escena auditiva dada (sobre todo porque en tal situación el proceso sensorial está directamente asociado a las ITD y a las ILD), lo que se afirma aquí es que una fuente virtual única dentro de un espacio audible (en un contexto electroacústico) no es suficiente por sí misma para establecer una resolución deseable de distancia y, por lo tanto, de espacio composicional en su nivel más básico, lo cual es además consistente con lo que se planteó en el capítulo I respecto de la localización espacial relativa. De esta forma queda justificado lo siguiente:

DEFINICIÓN 3.2.5. La *construcción mínima de distancia espacial* dentro de un espacio aural está conformada por dos fuentes sonoras simultáneas en localizaciones espaciales distintas y el escucha mismo; a ésta se le denotará como *distancia espacial lineal*. La magnitud de esta distancia espacial lineal dependerá de todos los factores de percepción descritos anteriormente.

De la definición anterior se sigue que, al momento de que al escucha le son presentados dos objetos sonoros simultáneos de localización espacial distinta, será posible establecer una sensación de distancia mucho más definida debido a que una de ellas se tomará como punto de referencia; luego se resolverá perceptualmente la distancia de ese punto de referencia con el otro objeto sonoro y, finalmente, se estableció una resolución de distancia respecto a él mismo. Todo ello, por supuesto, no en ese orden estricto desde el punto de vista cognitivo.

Recordando lo descrito sobre la *disimilitud serial-perceptual*, es necesario recalcar que *la distancia espacial lineal* (así como todos los conceptos relacionados que se desarrollarán enseguida) depende de una validación perceptual; esto es, *no basta con que exista una distancia entre dos objetos sonoros en el diseño composicional si dicha distancia no es correspondiente en la percepción real dentro del espacio aural*.

Cabe señalar que la capacidad de colocar objetos sonoros en localizaciones específicas en el espacio aural por parte del compositor se tiene como dada. Por supuesto, esta capacidad dependerá de las herramientas tecnológicas utilizadas, el método de espacialización y la capacidad del propio compositor para manejar tanto la espectromorfología de los objetos sonoros como el diseño espacial mismo (desde el punto de vista técnico). Por esa razón, esos detalles no se discutirán en este trabajo.

3.2. Trayectorias continuas u objetos sonoros móviles

Cuando a los objetos sonoros se les integra una característica de movimiento dentro del espacio aural, es posible referirse a ellos como *objetos sonoros móviles*. El camino o *recorrido* que realizan dichos objetos a través del espacio aural se denominará *trayectoria espacial móvil* o simplemente *trayectoria espacial*.

Las trayectorias espaciales, como recurso estético, pueden añadir dinamismo tanto en el nivel general (*i.e.* en la experiencia de la obra misma) como en el local (en la percepción espacial), apoyando (cuando son utilizadas correctamente) al desarrollo de la construcción de la espacialidad y por ende de la inmersividad.

DEFINICIÓN 3.2.6. Dado un espacio aural $\mathfrak{S}(a)$, y \mathcal{O} un objeto sonoro tal que $\mathcal{O}_{\text{spat}} \in \mathfrak{S}(a)$ con una duración total $|\mathcal{O}| = n$, se dice que \mathcal{O} es un *objeto sonoro móvil* si su localización es-

pacial dentro de $\mathfrak{S}(a)$ cambia a través del intervalo temporal de su duración total. La *trayectoria espacial* asociada a dicho objeto sonoro móvil estará dada por la parametrización:

$$\begin{aligned}
 I &= [0, n] : \rightarrow \mathbb{R}^5 \\
 t &: \rightarrow (\alpha(t), \beta(t), \theta(t), \phi(t), d(t))
 \end{aligned}$$

para $t \in I \subset \mathbb{R}^+$.

Cada trayectoria, tal y como es representada en la definición anterior, posee las propiedades de *velocidad* y *aceleración*. Estas propiedades delimitan el carácter distintivo espacial de los objetos sonoros desde el punto de vista composicional, en el sentido de que no sólo la posición espacial afectará la percepción de dicho objeto sonoro sino que la aceleración y la velocidad del mismo permitirán, en mayor o menor medida, desarrollar gestos o frases composicionales respecto del parámetro *Espacio-Timbre*.

El problema principal de las trayectorias no se limita a la deformación de la *continuidad* al momento de la traslación del espacio interno al espacio externo, sino que incluye la continuidad de la trayectoria en sí misma dentro del espacio interno. Dicha continuidad depende de diversas variables entre las que destacan la espectromorfología del objeto sonoro y la aceleración-velocidad de las mismas, pero también están directamente relacionadas con las propiedades perceptuales que se introdujeron en el capítulo I para la sección de las fuentes móviles. En este sentido, se puede hacer uso de dichos resultados experimentales con el fin de apoyar la construcción de la continuidad en las trayectorias de los objetos sonoros. Dicho lo anterior, es necesario aclarar lo que se entenderá formalmente por continuidad.

DEFINICIÓN 3.2.7. Dado un espacio aural, la *continuidad* de una trayectoria espacial asociada a un objeto sonoro móvil cualquiera es la percepción ininterrumpida de dicha trayectoria en su totalidad, a lo largo de su duración temporal.

Con la definición anterior en mente, puede resultar más o menos claro porque la *continuidad* de las trayectorias constituye uno de los parámetros más susceptibles a la deformación en la traslación espacio interno-espacio externo. Una de las principales razones es que las dimensiones y forma del espacio audible pueden cambiar dramáticamente al momento de dicha trasla-

ción. El estudio (lugar donde el o la compositor/a crea la obra y que en tal momento no es otra cosa que el recinto o espacio físico) puede considerarse un ambiente con condiciones más o menos ideales y por lo tanto el espacio aural se presenta estable. Las dimensiones de dicho estudio, y por lo tanto del espacio aural, son constantes y en general no son demasiado grandes. Esta dimensionalidad moldea, permea y afecta la construcción del espacio interno del mismo modo que las herramientas tecnológicas, los procesos de espacialización y las aptitudes del o la compositora. En este sentido puede afirmarse que el espacio interno de cada obra es una suerte de reflejo o condicionamiento del espacio físico donde fue creada.

Si la *continuidad* es uno de los parámetros críticos al momento de la traslación espacio interno-espacio externo, entonces, una apropiada percepción de ésta resulta fundamental para la coherencia y cohesión de los gestos musicales-sonoros. Así, la continuidad en las trayectorias espaciales es también uno de los parámetros más delicados al momento de diseñar experiencias espaciales inmersivas, ya que debido a su propia naturaleza física y psicoacústica, comprende elementos altamente susceptibles a la aparición de la discontinuidad.

Es posible afirmar que la mayor parte del tiempo resulta deseable diseñar trayectorias espaciales sonoras que sean capaces de permear una sensación convincente de *continuidad* en el escucha para la mayor cantidad de posibles tipos de objetos sonoros y, más aún, para crear escenarios donde múltiples trayectorias espaciales se encuentren simultáneamente, pero de un modo tal que cada una de ellas pueda ser escuchada de una manera específica e individual sin una pérdida demasiado significativa de coherencia perceptual tanto en el nivel local como en el general.

Por supuesto, esta *continuidad perceptual* de las trayectorias espaciales depende de diversas variables: extrapolación de la distancia y configuración del sistema multicanal llevado del estudio al recinto de concierto, métodos de paneo, contenido espectral de los objetos sonoros, velocidad y aceleración de los objetos sonoros, etcétera. Más aún, un desarrollo profundo de la continuidad no depende únicamente de estas variables por separado, sino de una relación dinámica y cambiante entre subconjuntos de éstas.

Entre los retos principales al momento de utilizar trayectorias sonoras en una obra se encuentra la *variación de la velocidad de las fuentes sonoras virtuales del espacio interno al espacio externo*. Cuando el compositor trabaja el espacio interno, lo hace por lo regular en un ambiente controlado y cercano a las condiciones ideales (relativamente buena acústica, equipo adecuado, aislamiento sonoro, etcétera). Al momento de interpretar la obra en el recinto y ajustarla al

espacio externo, el compositor pierde control de varios parámetros; entre ellos el cambio de la velocidad percibida de los objetos sonoros que realizan trayectorias espaciales, las cuales, dependiendo de las dimensiones del recinto, se pueden incrementar sustancialmente. Por tanto, la percepción de la continuidad y la localización dinámica pueden llegar a ser considerablemente afectadas. Una solución utilizada para este problema es la de aplicar tiempos de retardo específicos en las salidas de los canales de la configuración multicanal de acuerdo con la perspectiva panorámica del recinto. La otra se debe a la tradición de las escuelas francesa e inglesa de espacializar en vivo (*live diffusion*) de modo tal que la obra sea interpretada de acuerdo con cada recinto; es decir, trasladar el espacio interno al espacio externo.

Cabe señalar que estos dos primeros componentes corresponden de manera directa a lo que se denominó *conciencia audible espacial de ubicuidad* en la definición 2.5 del capítulo II.

3.3. Superficies sonoras y envolvimiento

El último componente primario de la inmersividad dentro del espacio aural que se presenta en este trabajo es el de las *superficies sonoras* y el *envolvimiento*.

Las superficies sonoras (de dos dimensiones o bidimensionales) resultan de la distribución de un objeto sonoro (por lo regular textural, más que puntual o percusivo) en una cantidad dada de salidas dentro del espacio aural de modo tal que, al contrario de la localización puntual fija, dicho objeto pueda ser percibido como difundido o extendido de manera plana a lo largo de ciertas áreas del espacio aural (o incluso a lo largo del espacio aural en su totalidad) de acuerdo a lo que se denominó *conciencia audible espacial de ambiente* (véase definición 2.5 del capítulo II).

La posibilidad de percibir estas superficies dentro de un espacio aural 2D, no sólo horizontalmente sino incluso con ciertos grados de inclinación, implica un especial manejo del perfil espectromorfológico y de la distribución espacial de los objetos sonoros que está directamente ligada a la capacidad de poder crear sensación de altura en sistemas 2D. Por supuesto, para sistemas multicanal 3D la sensación de altura se puede lograr de manera natural.

Esta capacidad de generar sensación de altura en los sistemas 2D implica además que se puedan crear superficies sonoras tridimensionales; es decir, superficies sonoras que puedan ser percibidas con extensión volumétrica. De nuevo, para sistemas multicanal 3D, estas super-

ficies tridimensionales se pueden lograr de manera más sencilla debido a la naturaleza inherente del espacio aural que resulta de dicha configuración.

DEFINICIÓN 3.2.8. La sensación de envolvimiento como componente de la conciencia audible espacial de ambiente, está directamente relacionada a la percepción de superficies o volúmenes sonoros.

Si se piensan las trayectorias espaciales como curvas paramétricas y por lo tanto como objetos abstractos de una dimensión, entonces las superficies sonoras pueden ser vistas como objetos abstractos dentro del espacio aural de dos y tres dimensiones: superficies y volúmenes de forma respectiva.

DEFINICIÓN 3.2.9. Dado un espacio aural $\mathfrak{S}(a)$ de dimensión k , con $k \in [2,3]$, una *superficie sonora* es un objeto sonoro distribuido de manera más o menos homogénea sobre una extensión del espacio aural. Si la superficie se diseña y se percibe plana, (*i.e.* tiene una frontera serial-perceptual plana), se dice que es bidimensional. Si la superficie se diseña y se percibe con extensión volumétrica (frontera serial-perceptual volumétrica), se dice que es tridimensional.

Cabe recalcar que los objetos sonoros que mejor logran disminuir la disimilitud serial-perceptual de superficie (bi o tridimensional) son aquellos cuyo perfil espectromorfológico está orientado hacia la textura.

4. Observaciones finales sobre la descripción de espacialidad

OBSERVACIÓN 3.2.2. Las superficies o volúmenes texturales son un componente de la conciencia audible espacial del ambiente.

OBSERVACIÓN 3.2.3. La *espaciosidad* puede construirse mediante la combinación dinámica principalmente de los objetos fijos y de las superficies o volúmenes sonoros.

OBSERVACIÓN 3.2.4. La *espaciosidad* puede construirse mediante la construcción de superficies o volúmenes cuyo material único sea la reverberación; es decir, la distribución de

reverberación (mediante algún proceso digital o de microfónica) a través del espacio aural como una superficie o volumen.

CONJETURA 3.2.4. La inmersividad como experiencia perceptual aural, puede construirse mediante tres componentes generales:

- Características estructurales. Tal y como se plantearon en la definición 3.1.3 del capítulo III.
- Características composicionales-procesuales. Las que corresponden precisamente a lo definido como los componentes primarios de la espacialidad en conjunción con las formulaciones acústicas, psicoacústicas y perceptuales de los perfiles espectromorfológicos de los objetos sonoros.
- Características subjetivas. Las que corresponden a las condiciones perceptual-cognitivas intrapersonales tal y como fueron descritas en la definición 2.3. del capítulo II.

A continuación se enlistan una serie de observaciones que están enfocadas a conjugar los elementos mostrados hasta el momento, al respecto del ASA, el Gestalt y la especialidad. Estas observaciones provienen además de toda una serie de experimentaciones y heurísticas personales y están encaminadas a formular ideas o propuestas que pudieran ser útiles al momento de diseñar espacialidad en una obra.

OBSERVACIÓN 3.2.5. A partir de los conceptos de *preatención* (o *atención pasiva*), el proceso *bottom-up*, y la diferenciación fondo-figura de la Gestalt, es posible diseñar objetos sonoros que detonen protagonismo momentáneo en una escena dada, sobre todo si se plantean con duración corta para posiciones fijas o de corta a media para trayectorias móviles; para estas últimas la velocidad de la trayectoria debería ser considerable, con el fin de detonar atención selectiva ecológica. Estos objetos sonoros pueden ser útiles para dirigir la atención de manera súbita hacia ciertas zonas del espacio audible o para plantear distancias virtuales.

OBSERVACIÓN 3.2.6. A partir de los conceptos de los *tiempos de reacción* y *tiempo simple de reacción perceptuales* (visuales y sonoros), y teniendo claro que la mayoría de los estudios realizados tienen que ver con estímulos ecológicos (asociados al peligro), es posible en primera instancia, continuar el planteamiento de la observación anterior y en segunda, diseñar elementos visuales que complementan los objetos sonoros tanto como estímulo sensorial

como para procesos de localización en el espacio audible. Estos elementos visuales pueden transitar desde simples estímulos hasta geometrías o creaciones visuales complejas y, por lo tanto, desde lo meramente sonoro apoyado por lo visual hasta la llamada música visual.

OBSERVACIÓN 3.2.7. Haciendo uso de los componentes estructurales representativos para la descripción de la memoria, del *sistema de procesamiento de la información* de Atkinson, es posible organizar y diseñar objetos sonoros que tengan una función específica para cada caso. Por ejemplo, el componente *registro sensorial* correspondería a los estímulos que se plantearon en las observaciones anteriores. Objetos sonoros asociados al *almacenamiento a corto plazo* tendrían que ver con cualidades tímbricas, las cuales pueden ir desde lo completamente reconocible hasta un cierto grado de reconocimiento (sin perder completamente esta asociación). Para la espacialidad, estos objetos sonoros podrían estar asociados a ubicaciones fijas dentro del espacio audible, o bien, a repeticiones de trayectorias móviles (tanto en posición como en energía/comportamiento y velocidad). Todo lo anterior estaría también asociado a la *atención paramétrica* en el sentido de Tenney. Los objetos sonoros asociados al *almacenamiento a largo plazo* tendrían relación con objetos con un alto grado de reconocimiento tímbrico y/o el uso de *leitmotifs* para ambos parámetros (Espacio-Timbre).

OBSERVACIÓN 3.2.8. Considerando que el concepto de *torrente auditivo* se puede asociar al TGU de Tenney, es posible entender los objetos sonoros (tal y como se presentan en esta obra) como equivalentes a ambos. Más aún, de acuerdo con lo planteado en esta obra, un objeto sonoro tiene no sólo una representación tímbrica sino también espacial; de manera inversa, las TGU de Tenney y los torrentes auditivos pueden ser replanteados en términos de lo espacial-tímbrico. En este sentido, un *clang* en el sentido de Tenney puede ser redefinido como un *clang espacio-tímbrico* y un *torrente auditivo* como un *torrente aural espacio-tímbrico*.

OBSERVACIÓN 3.2.9. A partir de la observación anterior, puede resultar bastante útil considerar la cohesión (en términos de las TGU) y los múltiples principios de segregación y agrupamiento del ASA con el fin de diseñar objetos sonoros que puedan ser percibidos en un lugar único dentro del espacio audible (tal y como se hace de acuerdo a los principios básicos de la mezcla estéreo). Al contemplar esos elementos en la creación de los objetos sonoros será posible, en gran medida, lograr escenas espaciales con mayor coherencia y estructura virtual de modo tal que el efecto de enmascaramiento pueda ser disminuido a niveles óptimos.

OBSERVACIÓN 3.2.10. En relación con lo anterior, es posible utilizar los principios gestalt para poder complementar el diseño de objetos sonoros con un alto grado de cohesión espacial-tímbrica. No sólo eso, los principios Gestalt pueden resultar bastante útiles al momento de generar escenas con múltiples objetos sonoros de los cuales se busca generar alguna especie de agrupamiento. Aplicando por ejemplo el factor de similaridad (objetos sonoros con similaridad espectromorfológica de acuerdo con lo descrito en este capítulo) en conjunto con el factor de proximidad (objetos sonoros con ubicación espacial cercana), será posible plantear agrupaciones de objetos sonoros que tengan un comportamiento espacial grupal en cierta medida unificado; ello puede resultar bastante útil al momento de diseñar trayectorias sonoras, superficies y/o volúmenes espaciales.

OBSERVACIÓN 3.2.11. La ley de la buena configuración de la Gestalt podría ser directamente aplicada al ámbito sonoro con el conocido efecto *ilusión voz-a-canción*, el efecto básico de los batimientos binaurales y en el espacial multicanal con el efecto Hass o la localización espacial a partir del paneo por intensidad, por mencionar algunos ejemplos. Sin embargo, para el caso de objetos sonoros (espacio-tímbricos), dicha ley puede utilizarse mediante pequeños desplazamientos en el tiempo de objetos sonoros altamente similares (filtrados para diferentes bandas frecuenciales) y ubicados en distintas posiciones espaciales con el fin de generar un sólo objeto sonoro resultante. Lo anterior sonará bastante familiar para el compositor o compositora, de acuerdo con las técnicas de espacialización tímbrica y espacialización por bandas frecuenciales.

OBSERVACIÓN 3.2.12. El factor de la buena curva (en conjunto con las distintas nociones de *continuidad* previamente planteadas) puede aplicarse para los casos en los cuales se busca diseñar trayectorias espaciales múltiples unidas y contiguas para que, a pesar de contener objetos o agrupaciones de objetos sonoros disimilares tímbricamente, dichas trayectorias distintas puedan percibirse en el espacio audible como una sola trayectoria que va evolucionando en función del timbre.

Cabe señalar, a modo de comentario final, que resulta bastante útil el considerar todos los resultados experimentales mostrados al respecto de la escucha espacial (localización de dirección y distancia). Con ello se puede diseñar la espacialidad de los objetos sonoros. Por supuesto, esto no significa que se deban restringir los mismos a dichas consideraciones, más bien se trata de construirlos o transformarlos con los énfasis espectrales correspondientes. ■ ■ ■



Capítulo IV. Espacio-Timbre: un modelo teórico.

El término Espacio-Timbre, tal y como se ha planteado a lo largo de este trabajo, se refiere a la relación natural e inherente que existe entre dos características fundamentales de los objetos sonoros dentro del campo de la composición electroacústica multicanal: el timbre y el espacio. Como se ha explicado, la percepción, segregación, agrupación y localización de las fuentes sonoras no es un proceso sencillo sino más bien no lineal y complejo.

Bajo la formulación de los resultados acústicos y psicoacústicos, el modelo aquí planteado tiene el objetivo principal de explicar (mediante dichos resultados y conceptos) diversas formulaciones conocidas dentro de la composición electroacústica multicanal derivadas de la experiencia práctica. Más aún, como se verá posteriormente, el modelo se propone incluso bajo una formulación matemática abstracta basada en sistemas, cuyo fin es justificar lo más formalmente posible la construcción teórica de aquello que podría denominarse Espacio-Timbre y mostrar un ejemplo de cómo puede utilizarse el lenguaje matemático teórico para la representación de conceptos.

Si bien el concepto de timbre es históricamente conflictivo debido a la falta de convención para definirlo, engloba de manera más o menos eficiente las características específicas que hacen que un escucha pueda, en cierto modo, distinguir una fuente sonora de otra. A lo largo de este capítulo el lector observará la construcción de una definición del objeto sonoro derivada del concepto de espectromorfología de Dennis Smalley; se trata de una definición com-

puesta por dos bloques fundamentales: la parte morfológica y la parte espectral. A su vez, cada uno de estos bloques poseerá dos aspectos individuales: el carácter acústico y el perceptual. Esto último dará sentido al concepto de *disimilitud serial-perceptual* propuesto en la definición 3.1.4 del capítulo III.

Antes de continuar, es necesario establecer la definición de timbre de acuerdo con el contexto del presente trabajo. Aunque esta definición se presenta en esta sección; se debe aclarar que los elementos que la componen (los perfiles morfológicos y espectrales con las características acústicas y perceptuales) se desarrollarán detalladamente a lo largo del presente capítulo.

DEFINICIÓN 4.0.1. El *timbre* es la representación conjunta de un objeto sonoro de sus perfiles morfológicos y espectrales simultáneamente con sus características acústicas y perceptuales.

De este modo, Espacio-Timbre es el término que denotará la relación que existe para cada objeto sonoro, entre sus propiedades de ubicuidad espacial y su perfil espectralmorfológico, ello incluye de manera simultánea el carácter acústico y perceptual de dichas propiedades. El Espacio-Timbre es por lo tanto, un concepto derivado de implicaciones acústicas, psicoacústicas y composicionales, cuyo objetivo es establecer un marco específico (una alternativa más a las ya existentes) tanto de términos como de categorización para la creación electroacústica multicanal.

Los perfiles morfológicos y espectrales, tanto en la parte acústica como en la parte perceptual, son representados de manera cuantitativa a través de los denominados *descriptores de audio*, que son de uso común en el área del MIR (*Music Information Retrieval*). Dependiendo del contexto, estos descriptores son clasificados de acuerdo con su nivel de abstracción (bajo/alto nivel), a su validez temporal o su descripción de dinámica.

Como se menciona en Gomez y Herrera (2004), y en Pohle (2005), para el caso de nivel de abstracción existen dos tipos: bajo y alto. Los descriptores de bajo nivel son calculados directamente de la señal de audio o de la representación FFT/STFT y son almacenados como valores flotantes. Por su parte, los descriptores de alto nivel requieren un proceso de inferencia y clasificación y los resultados corresponden a variables categóricas o etiquetas.

De acuerdo con la información que arrojan, los descriptores pueden ser clasificados del siguiente modo (Bogdanov *et al.*, 2013), (Termens, 2009):

- Espectrales/tímbricos: contraste espectral, flujo espectral, centroide espectral, inarmonicidad, coeficientes cepstrales Mel (MFFC), medida de contenido de alta frecuencia (HFC), bandas Bark, etcétera.
- Temporales: tasa de cruce cero (ZCR), medidas de volumen (*loudness*), energía media cuadrática (RMSE), etcétera.
- Rítmicos: intervalo de *onset*, histograma de *beat*, espectro de *beat*, etcétera.
- Tonales: perfil de clase altura armónica (HPCP), desviación igual temperada, razón de energía no temperada a temperada, fuerza diatónica, centroide de octava, rugosidad tonal, etcétera.

En el contexto de este trabajo, los descriptores espectrales, tímbricos y tonales tendrán la función de representar el *perfil espectral*, mientras que los temporales y rítmicos lo harán para el *perfil morfológico*. Ambos perfiles serán desarrollados detalladamente en las siguientes secciones.

1. Perfil y espacio morfológico

En esta sección se establecerá la idea de *perfil morfológico*, el cual funcionará como un concepto teórico con aplicación práctica directa para describir y caracterizar la morfología de un objeto sonoro; esto es, la forma en la que el contenido espectral es moldeado a lo largo del tiempo a través de la duración total de dicho objeto. Las siguientes definiciones tienen por objetivo construir la caracterización morfológica de un objeto sonoro de manera perceptual y acústica permitiendo, de este modo, construir la caracterización completa de \mathcal{O}_{morf}

DEFINICIÓN 4.1.1. Dado un objeto sonoro cualquiera \mathcal{O} , el contorno morfológico acústico

$$\mathfrak{G}_{morf}^{ac} \mathcal{O}$$

es la representación de la envolvente del sonido ligada a la representación físico-acústica y se denotará así:

$$\mathfrak{G}_{morf}^{ac} \mathcal{O} \simeq \bigcup_{i=0}^n \delta(\mathcal{O})_i^{ac}$$

donde lo anterior puede leerse como “el contorno morfológico es equivalente a la influencia conjunta de los descriptores acústicos $\delta(\mathcal{O})_i^{ac}$ asociados a dicha representación”.

El *contorno morfológico perceptual*

$$\mathfrak{G}_{morf}^{pr} \mathcal{O}$$

es la representación de la envolvente del sonido que muestra la relación existente entre la frecuencia y la intensidad del sonido de acuerdo con las características específicas perceptuales del oído humano; se denotará del siguiente modo:

$$\mathfrak{G}_{morf}^{pr} \mathcal{O} \simeq \bigcup_i \delta(\mathcal{O})_i^{pr}$$

donde lo anterior puede leerse como “el contorno morfológico perceptual es equivalente a la influencia conjunta de los descriptores $\delta(\mathcal{O})_i^{pr}$ asociados a dicha representación”.

El *perfil morfológico*

$$\mathfrak{B}_{morf}$$

se definirá entonces como la combinación cualitativa y cuantitativa del contorno morfológico acústico y del contorno morfológico perceptual.

$$\mathfrak{B}_{morph}(\mathcal{O}) \simeq \mathfrak{G}_{morf}^{ac} \mathcal{O} \cup \mathfrak{G}_{morf}^{pr} \mathcal{O}$$

De acuerdo con la definición anterior y a lo discutido previamente, se puede observar que tres ejemplos de *descriptores morfológicos acústicos* son el ZCR, el RMSE y la *transformada de Hilbert*, mientras que ejemplos de *descriptores morfológicos perceptuales* serían los llamados filtros de ponderación A, B y C.

Cabe señalar que esta clasificación no es completamente definitoria ya que tanto en la parte morfológica como en la parte espectral del sonido, los descriptores acústicos contribuyen en cierta medida a la percepción aurál y, de manera inversa, los descriptores perceptuales contribuyen a la parte físico-acústica. Sin embargo, se hace esta diferenciación para establecer una

distinción categórica, conceptual y práctica al momento de enunciar este modelo teórico, así como los subsiguientes ejemplos del capítulo.

Una consecuencia directa del concepto de *perfil morfológico* es la idea de *espacio morfológico*, el cual, dentro de este modelo, no es otra cosa que el ordenamiento de los descriptores como vectores de algún subconjunto de \mathbb{R}^n y que encuentra relación con lo que se mencionó en el capítulo anterior sobre el concepto de espacio como representación de parámetros sonoro/composicionales. Esta conceptualización permitirá establecer, de manera formal, el concepto de *espacio espectralmorfológico* y además facilitará el formular casos particulares de acuerdo con los descriptores utilizados.

El *espacio morfológico* es, entonces, una construcción arbitraria y particular acorde con el número de descriptores que se deseen utilizar para casos específicos; es decir, éste se construye luego de decidir cuántos y cuáles descriptores acústico-perceptuales se usarán en una situación particular. Antes de establecer formalmente dicha definición es necesario tener presente el concepto de *duración total* de un objeto sonoro tal y como se presentó en la definición 3.2.4 del capítulo III.

El concepto de duración total es de importancia relevante ya que los descriptores morfológicos de los objetos sonoros pueden formularse de dos formas muy generales:

- El descriptor se caracteriza mediante una cantidad o cuantificación numérica constante para toda la duración total.
- El descriptor se caracteriza como una función paramétrica a través del tiempo de la duración total.

En el primer caso se tiene, por ejemplo, la práctica común en el MIR (*Music Information Retrieval*) de obtener medidas de tendencia central y dispersión para formular el comportamiento general de los descriptores. Por supuesto, la desventaja de esta metodología radica en que se puede perder una gran cantidad de información acerca del comportamiento de la morfología del objeto sonoro a analizar, al hacer una reducción tan considerable. De hecho, como se verá más adelante, esta es la razón principal por la cual se propone la metodología descrita en este capítulo, basada en el análisis de series de tiempo.

Ahora bien, se debe clarificar, en este punto, la diferencia entre *espacio* y *parametrización de espacio*. Por ejemplo, el *espacio morfológico* es el espacio abstracto donde radica la caracteriza-

ción morfológica del objeto sonoro. La *parametrización* del *espacio morfológico* será, por otro lado, la descripción vectorial de dicho espacio en función de los descriptores acústico-perceptuales asociados al objeto sonoro. De manera análoga, el *espacio espectral* es el espacio abstracto donde radica la caracterización espectral del objeto sonoro, mientras que la parametrización de éste será la representación vectorial de dicho espacio en función de los descriptores acústico-perceptuales asociados al objeto sonoro. Lo anterior funciona exactamente igual para el espacio espectralmorfológico.

Para construir una parametrización de un *espacio morfológico* primero deberá elegirse el conjunto de descriptores morfológicos acústico/perceptuales y, posteriormente, a partir de estos conjuntos, describir dicho espacio. En este sentido, *el espacio morfológico* depende de la elección previa del conjunto de descriptores.

DEFINICIÓN 4.1.2. Sea n el número de descriptores morfológicos acústicos y k el número de descriptores perceptuales elegidos ambos de antemano. Entonces, el espacio morfológico para esta construcción en particular estará dado por:

$$\mathcal{M}_I \subset \mathbb{R}^n \times \mathbb{R}^k$$

Para cada objeto sonoro que se desee describir en el espacio anterior, sea $I \subset \mathbb{R}^+$ un conjunto indexador de tiempo asociado a la duración total de dicho objeto sonoro que se esté representando, es decir $|I| = |\mathcal{O}|$. Entonces, dado un instante de tiempo $t \in I$, un elemento o vector $m \in \mathcal{M}$ estará representado por la siguiente parametrización:

$$\begin{aligned} I &: \rightarrow \mathcal{M}_I \\ t &: \rightarrow m(t) \end{aligned}$$

donde:

$$m(t) = [m_0^{ac}(t), \dots, m_n^{ac}(t), m_0^{pr}(t), \dots, m_k^{ac}(t)]$$

y $m_i \in \mathfrak{C}_{morf}^{ac} \mathcal{O}$ para $i \in [0, n]$ y $m_j \in \mathfrak{C}_{morf}^{pr} \mathcal{O}$ para $j \in [0, k]$. Por lo tanto, todo objeto sonoro asociado a dicho espacio estará representado y definido por dicho vector.

Nótese que la definición anterior incluye los casos en los que el perfil morfológico está descrito tanto por una cuantificación numérica (como la media, varianza, etcétera), como por valores en series de tiempo. La primer situación puede considerarse como un caso particular de la segunda; cuando para todo instante de tiempo el valor del descriptor es el mismo: $m_i^{ac}(t) = \mu$ para todo $t \in I$.

La diferencia primordial entre el *perfil morfológico* y el *espacio morfológico* es que el primero describe las características morfológicas acústico-perceptuales de los *objetos sonoros* de acuerdo con los descriptores elegidos; mientras tanto, el segundo establece la organización *vectorial* de dichos descriptores dentro de un espacio subconjunto de $\mathbb{R}^n \times \mathbb{R}^k$. El vector $m(t)$ puede entonces pensarse como la caracterización numérica de un objeto sonoro en relación a los descriptores elegidos a lo largo de un intervalo de tiempo.

La forma en que dicha caracterización toma un valor concreto dependerá del referente numérico que se le asigne a cada descriptor, el cual, como ya se explicó, puede ser constante o paramétrico a través del tiempo. Supóngase, para fines de ejemplificación, un caso trivial en el que el *espacio morfológico* está dado por:

$$\mathcal{M}_I \subset \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$$

Lo anterior significa que dicho espacio está construido a partir de un descriptor acústico y uno perceptual. Supóngase además que dichos descriptores son el *ZCR* y el *RMSE* respectivamente, y que el *ZCR* está caracterizado por una cantidad numérica fija dada por la *media aritmética*, mientras que el *RMSE* está caracterizado por una parametrización en serie de tiempo. Entonces, cualquier objeto sonoro estará representado por estos dos descriptores del siguiente modo:

$$m(t) = [m_0^a(t), m_0^p(t)]_0^{t=n} = (\mu, m_0^p(t))_0^{t=n}$$

Por su parte, el perfil morfológico de ese objeto sonoro estará descrito por la serie de tiempo:

$$[(\mu, m_0^{pr}(0)), (\mu, m_0^{pr}(1)), \dots, (\mu, m_0^{pr}(n))]$$

Nótese que el procedimiento general para construir el *espacio morfológico* puede resumirse en los siguientes pasos:

1. Se definen los descriptores a utilizar tanto acústicos como perceptuales y se construye la dimensión del espacio como la suma del número de descriptores acústicos más los descriptores perceptuales.
2. Se establece cuáles descriptores serán caracterizados por una cantidad numérica fija y cuáles por una parametrización; ambos a lo largo de la duración total del objeto sonoro. Nótese en este punto que el espacio como tal no depende del objeto sonoro sino de los descriptores a utilizar y que, una vez construido el espacio, cualquier objeto sonoro puede ser analizado mediante esta estructura. De lo anterior se desprende que la duración total del objeto sonoro no afecta la construcción del espacio sino que la misma se adapta dependiendo de éste.
3. Cada vector $m(t)$ estará asociado a un objeto sonoro particular y este vector arrojará la descripción numérica o la parametrización de los descriptores seleccionados.

Resulta claro en este punto que existirán tantas parametrizaciones de espacios morfológicos como combinaciones posibles de descriptores morfológicos acústicos-perceptuales. Nótese además que, al definir la parametrización del espacio morfológico de una manera tan general, es posible incorporar sin ningún problema cualquier descriptor que pudiera ser desarrollado en el futuro. No sólo eso, es necesario enfatizar la diferencia entre espacio y su parametrización. Al respecto del espacio morfológico tenemos:

DEFINICIÓN 4.1.3. El espacio morfológico \mathcal{M} es el conjunto que contiene la representación morfológica de cualquier objeto sonoro \mathcal{O} desde el punto de vista conceptual; lo anterior significa que $\mathcal{O}_{morf} \in \mathcal{M}$, es la notación correspondiente para tal efecto. \mathcal{M}_I por su parte será, como ya se describió anteriormente, la caracterización vectorial de la morfología acústico-perceptual de dicho objeto sonoro.

Perfil y espacio espectral

En esta sección se definirá el *perfil espectral* de manera completamente análoga a la forma en que se definió el *perfil morfológico*; es decir, estableciendo un componente físico-acústico y otro perceptual.

DEFINICIÓN 4.2.1. Dado un objeto sonoro cualquiera \mathcal{O} , el *contenido espectral acústico*:

$$\mathfrak{M}_{spectra} \mathcal{O}$$

es la representación de dicha propiedad (el espectro del sonido) ligada a la representación físico-acústica:

$$\mathfrak{M}_{spectra}^{ac} \mathcal{O} \simeq \cup_i \omega(\mathcal{O})_i^{ac}$$

donde lo anterior puede leerse como “el contenido espectral acústico es equivalente a la influencia conjunta de los descriptores acústicos $\omega(\mathcal{O})_i^{ac}$ asociados a dicha representación”.
El contenido espectral perceptual

$$\mathfrak{M}_{spectra}^{ac} \mathcal{O}$$

de un objeto sonoro, es la representación de dicha propiedad (el espectro del sonido) ligada a la representación perceptual:

$$\mathfrak{M}_{spectra}^{pr} \mathcal{O} \simeq \cup_i \omega(\mathcal{O})_i^{pr}$$

donde lo anterior puede leerse como “el contorno morfológico perceptual es equivalente a la influencia conjunta de los descriptores perceptuales $\omega(\mathcal{O})_i^{pr}$ asociados a dicha representación”.

El perfil espectral

$$\mathfrak{P}_{spectra}$$

es la combinación cualitativa y cuantitativa del contenido espectral acústico y del contenido espectral perceptual:

$$\mathfrak{M}_{spectra}(\mathcal{O}) \simeq \mathfrak{M}_{spectra}^{ac} \mathcal{O} \cup \mathfrak{M}_{spectra}^{pr} \mathcal{O}$$

De manera análoga a la forma en que se construyó el espacio morfológico en la sección anterior, se propone aquí la definición de *espacio espectral*:

DEFINICIÓN 4.2.2. Sea n el número de descriptores espectrales acústicos y k el número de descriptores espectrales perceptuales elegidos ambos de antemano. La parametrización del espacio espectral para esta construcción en particular estará dado por:

$$\mathcal{E}_I \subset \mathbb{R}^n \times \mathbb{R}^k$$

Para cada objeto sonoro que se desee describir en el espacio anterior, sea $I \subset \mathbb{R}^+$ un conjunto indexador de tiempo asociado a la duración total de dicho objeto sonoro que se esté representando, es decir $|I| = |\mathcal{O}|$. Entonces, dado un instante de tiempo $t \in I$, un elemento o vector $e \in \mathcal{E}$ estará representado por la siguiente parametrización:

$$\begin{aligned} I &\rightarrow \mathcal{E}_I \\ t &\rightarrow e(t) \end{aligned}$$

donde:

$$e(t) = [e_0^{ac}(t), \dots, e_n^{ac}(t), e_0^{pr}(t), \dots, e_k^{pr}(t)]$$

y $e_i^{ac}(t) \in \mathfrak{B}_{spectra}^{ac} \mathcal{O}$, para $i \in [0, n]$, y $e_j^{pr}(t) \in \mathfrak{B}_{spectra}^{pr} \mathcal{O}$ para $j \in [0, k]$. Por lo tanto, todo objeto sonoro asociado a dicho espacio estará representado y definido por dicho vector.

DEFINICIÓN 4.2.3. El espacio espectral \mathcal{E} es el conjunto que contiene la representación espectral de cualquier objeto sonoro \mathcal{O} desde el punto de vista conceptual; lo anterior significa que $\mathcal{O}_{spectra} \in \mathcal{E}$, es la notación correspondiente para tal efecto. \mathcal{E}_I por su parte será, como ya se describió anteriormente, la caracterización vectorial del espectro acústico-perceptual de dicho objeto sonoro.

Perfil y espacio spectromorfológico

De lo expuesto en las dos secciones anteriores se deriva directamente la definición de lo que se propone como *perfil spectromorfológico*:

DEFINICIÓN 4.3.1. Dado un objeto sonoro cualquiera \mathcal{O} , el *perfil espectromorfológico* es la descripción cualitativa y cuantitativa del perfil morfológico y el perfil espectral de manera conjunta.

$$B_{\text{morf}}^{\text{spectra}}(\mathcal{O}) \simeq \mathfrak{B}_{\text{morph}}(\mathcal{O}) \cup \mathfrak{B}_{\text{spectra}}(\mathcal{O})$$

Antes de cerrar esta sección cabe mencionar algunos puntos importantes. Nótese que al establecer una definición tan general de los perfiles morfológico y espectral, tal y como se hizo anteriormente, es posible que dicha concepción se ajuste a los descriptores que se quieran utilizar y no a la inversa. En otras palabras, no es la elección predefinida de los descriptores la que delimita la definición. Lo anterior también se ve reflejado en el hecho de que para cada componente de cada perfil, el número de descriptores es distinto: n, m, k, l .

La formulación consecuente resumida es que *todo objeto sonoro puede ser descrito y representado por sus cualidades morfológicas y espectrales desde el punto de vista físico-acústico y perceptual simultáneamente*. Esto permite una definición abstracta y general que abarque cualquier objeto sonoro y, a la vez, un marco de referencia práctico y concreto para realizar esta representación mediante descriptores acústico-perceptuales que puedan ser directamente incluidos en algún modelo computacional.

Se establece a continuación, también de manera análoga a la forma en que se construyeron el espacio morfológico y el espectral, la definición de *espacio espectromorfológico*; en este caso la construcción es más sencilla puesto que dicho espacio se propone como el producto cruz de los dos anteriores.

DEFINICIÓN 4.3.2. Sean \mathcal{M} y \mathcal{E} los espacios morfológico y espectral con dimensiones $(n_1 + k_1)$ y $(n_2 + k_2)$ respectivamente, entonces, la parametrización del *espacio espectromorfológico* estará dado por:

$$S_I = \mathcal{M}_I \times \mathcal{E}_I$$

donde cada elemento $s \in S_I$ será un vector de dimensión $n = n_1 + k_1 + n_2 + k_2$ y cada objeto sonoro representado en este espacio estará caracterizado por los descriptores morfológico-

cos y espectrales elegidos, así como por su caracterización numérica. Del mismo modo que en los casos anteriores, cada vector estará representado por una parametrización que para este caso estará dada por:

$$\begin{aligned} I &: \longrightarrow S_I \\ t &\longrightarrow s(t) \end{aligned}$$

donde:

$$s(t) = [s_0(t), \dots, s_n^*(t)]$$

DEFINICIÓN 4.3.3. El espacio espectromorfológico $S = \mathcal{M} \times \mathcal{E}$ es el conjunto que contiene la representación espectromorfológica de cualquier objeto sonoro \mathcal{O} desde el punto de vista conceptual; lo anterior significa que

$$\mathcal{O}_{specM} = (\mathcal{O}_{morf}, \mathcal{O}_{spectra}) \in S$$

es la notación correspondiente para tal efecto. S_I por su parte será, como ya se describió anteriormente, la caracterización vectorial de la espectromorfología acústico-perceptual de dicho objeto sonoro.

Con todas las definiciones anteriores ya establecidas, nótese que el objeto sonoro \mathcal{O} puede entonces pensarse como la conjunción de sus propiedades espaciales ($\mathcal{O}_{spat} \in \mathfrak{S}(a)$) y sus propiedades espectromorfológicas ($\mathcal{O}_{specM} \in S$). Dicho de otro modo:

$$\mathcal{O} = (\mathcal{O}_{spat}, \mathcal{O}_{spectrM})$$

lo que es equivalente a decir que el Espacio-Timbre es el producto cruz del espacio aural con el espacio espectromorfológico:

$$\mathfrak{S}(a) \times S$$

DEFINICIÓN 4.3.4. El Espacio-Timbre puede ser representado como el producto cruz del espacio aural con el espacio espectromorfológico. De este modo, el objeto sonoro se repre-

senta a partir de la conjunción de sus propiedades acústico-perceptuales espaciales y espectromorfológicas respectivamente. De lo anterior se sobreentiende que la función proyección de cada coordenada del Espacio-Timbre es la representación correspondiente del objeto sonoro en términos de espacio-espectromorfología y así, la primer proyección corresponde a la *representación espacial* del objeto sonoro:

$$\begin{aligned} \pi_1 : \mathfrak{S}(a) \times S &\longrightarrow \mathfrak{S}(a) \\ \mathcal{O} = (\mathcal{O}_{spat}, \mathcal{O}_{specM}) &\longrightarrow \mathcal{O}_{spat} \end{aligned}$$

mientras que la segunda proyección corresponde a la *representación espectromorfológica* del objeto sonoro:

$$\begin{aligned} \pi_2 : \mathfrak{S}(a) \times S &\longrightarrow S \\ \mathcal{O} = (\mathcal{O}_{spat}, \mathcal{O}_{specM}) &\longrightarrow \mathcal{O}_{specM} \end{aligned}$$

Esta proposición será presentada de manera más formal en la sección de sistemas dinámicos.

1.1. Distancia y medidas en el espacio espectromorfológico

Hasta este punto se ha construido el espacio espectromorfológico como el producto cruz del espacio morfológico y del espacio espectral, siendo todos ellos subconjuntos de \mathbb{R}^n para alguna dimensión n . Esto tiene como consecuencia directa de que dichos espacios heredan diversas propiedades de \mathbb{R}^n , en particular la propiedad de ser espacios métricos cuando se consideran las métricas usuales como la euclidiana, la de Minkowsky o la Manhattan. En este sentido, la cuestión principal al respecto de la definición del espacio espectromorfológico radica en el hecho de poder establecer distancias entre dos vectores cualesquiera y, más aún, puntualizar el significado de dicha distancia.

Dada una parametrización del espacio morfológico \mathcal{M}_f junto con la métrica euclidiana usual y dos vectores $m(a), m(b) \in \mathcal{M}_p$ entonces la distancia $d[m(a), m(b)]$ tendrá el significado asociado a la caracterización numérica de cada descriptor. Ejemplificando lo anterior, supón-

gase que el espacio mencionado está construido a partir de dos descriptores en particular: RMSE y energía. Supóngase además que el RMSE es caracterizado por la media aritmética mientras que la energía es caracterizada por el coeficiente de variación. Entonces, el espacio morfológico resultante es un subconjunto de \mathbb{R}^2 tal que todo vector será de la forma:

$$m(t) = [m_0(t), m_1(t)]$$

y representará objetos sonoros de acuerdo con los descriptores establecidos. De este modo, para cada objeto sonoro, la primera coordenada de este vector será el valor asociado a la media aritmética del RMSE y la segunda será el valor del coeficiente de variación de la energía. Para estas caracterizaciones numéricas es claro que el espacio morfológico no sólo es un subconjunto de \mathcal{M}_1 sino de $[0, 1] \times [0, 1]$. De este modo, para cada instante de tiempo t , dados dos vectores $m(a), m(b) \in \mathcal{M}$, la distancia entre ellos representará un nivel de similaridad de los objetos sonoros de acuerdo con los descriptores y las caracterizaciones numéricas elegidas. Se propone ahora la definición formal de *caracterización numérica*.

DEFINICIÓN 4.3.5. Sea $I \subset \mathbb{R}^n$ la parametrización de un espacio morfológico, espectral o espectralmorfológico. Para cada descriptor que conforma dicho espacio, se denomina *caracterización numérica del descriptor* a la unidad de medida con la que es analizada la propiedad acústico-perceptual desde lo espectral, morfológico o espectralmorfológico del objeto sonoro. Dicha caracterización puede ser de dos tipos: fija o paramétrica. En el caso de que la caracterización sea paramétrica, se considerará como una serie de tiempo.

Resta la definición formal de distancia entre dos vectores dentro de un espacio morfológico, espectral o espectralmorfológico. Nótese que para el caso de que la caracterización sea paramétrica es necesario establecer una forma de medir la distancia entre las series de tiempo. Una de las técnicas más populares para realizar lo anterior es el *dynamic time warping*, que funciona sobre todo para series de tiempo que son de distinta longitud. De nuevo, dada la generalidad del modelo, desde el punto de vista conceptual, el método elegido para medir la distancia entre las series de tiempo es irrelevante; esto cobrará importancia al momento de realizar la aplicación práctica concreta. Recuérdese que el objetivo primordial de este modelo es establecer un marco teórico formal que permita formular la idea de distancia entre objetos

sonoros dentro de algún espacio morfológico, espectral o espectromorfológico y el significado de la misma. Así, que dos objetos sonoros dentro de un espacio dado estén más o menos separados, implicará un grado de similaridad o disimilaridad respectivamente. Esta referencia permitirá tener un panorama de organización de los mismos de acuerdo con dicho grado relativo de similaridad.¹¹

DEFINICIÓN 4.3.6. Sea $X \subset \mathbb{R}^n$ un espacio morfológico, espectral o espectromorfológico definido con n descriptores, junto con una métrica $d: X \times X \rightarrow \mathbb{R}_{\geq 0}$ donde además, cada uno de los n descriptores tiene definida una caracterización numérica específica. Entonces, dados dos objetos sonoros \mathcal{O}_a y \mathcal{O}_b y sus vectores correspondientes $s(a), s(b) \in X$ asociados, la distancia $d[s(a), s(b)]$ representará el grado de similaridad entre dichos objetos sonoros de acuerdo con los descriptores y la caracterización numérica establecidos.

2. Espacio-Timbre: construcción abstracta

Se presenta ahora la construcción del concepto del Espacio-Timbre utilizando sistemas dinámicos. Esta forma de definir dichos conceptos permitirá una contribución teórica formal haciendo uso de una herramienta matemática sólida, capaz de establecer un modelo lo suficientemente general como para englobar y explicar la mayoría de los casos particulares.

La dinámica de sistemas puede entenderse, en términos muy generales, como la descripción de la evolución de un fenómeno natural (social o de cualquier otro tipo), de un objeto abstracto, o bien, de un sistema a través del tiempo, dadas ciertas condiciones iniciales. Estas últimas se relacionan, usualmente, con el punto sobre el cual se elige aplicar la *dinámica* del sistema, así como con restricciones específicas sobre dicha dinámica. En el presente contexto, cada punto del sistema corresponderá a un objeto sonoro específico y, por lo tanto, el sistema dinámico será la representación de la evolución de dicho objeto sonoro a través del tiempo; dada la gran generalidad de este modelo, la representación no es para un punto único o específico sino para cualquiera. En esta característica radica la utilidad del sistema dinámico como

¹¹ El lector puede encontrar similitudes con las *métricas morfológicas* propuestas por Polansky (1996); sin embargo, la diferencia principal con lo propuesto en este trabajo radica en el hecho de que aquí se enfoca la atención a las cualidades espectromorfológicas de los objetos sonoros.

herramienta para representar de manera general un concepto teórico como lo es el de la conciencia espacial audible o el perfil espectromorfológico. Se comienza esta sección estableciendo la caracterización (mediante sistemas dinámicos) de la *conciencia audible espacial*, tal y como fue descrita en la sección anterior.

De acuerdo a la definición 3.2.2 del capítulo III, dado un espacio aural $\mathfrak{S}(a)$, un objeto sonoro tiene una representación de localización espacial dada por

$$\gamma(\mathcal{O}_{spat}, t) = (\alpha(t), \beta(t), \theta(t), \phi(t), d(t))$$

para cada instante de tiempo t . Si se define un conjunto indexador $I \subset \mathbb{R}^+$ para la representación del tiempo, es posible construir un sistema dinámico asociado al concepto de conciencia audible espacial del modo siguiente:

DEFINICIÓN 4.4.1. Sea $\mathfrak{S}(a)$ un espacio aural específico e $I \subset \mathbb{R}^+$ un conjunto indexador de tiempo. Entonces el sistema dinámico asociado a la *conciencia audible espacial* o simplemente *espacialidad*, estará dado por la tripleta $(\mathfrak{S}(a), I, \vartheta)$, donde:

$$\begin{aligned} \vartheta : I \times \mathfrak{S}(a) &\longrightarrow \mathfrak{S}(a) \\ (t, \mathcal{O}_{spat}) &\longrightarrow \tau(t, \mathcal{O}_{spat}) \end{aligned}$$

Sea $n \in I$, y $\mathcal{O}_{spat} \in \mathfrak{S}(a)$, entonces, se denominará *recorrido finito* a la sección de la órbita correspondiente al tiempo n :

$$\Upsilon[\mathcal{O}_{spat}](n) = \{ \tau(t, \mathcal{O}_{spat}) \mid t = \overline{0, n} \}$$

Es posible abreviar el operador evolución cuando se hace referencia a un punto específico del sistema, en este caso, un objeto sonoro $\mathcal{O}_{spat} \in \mathfrak{S}(a)$; por lo que dicha función puede expresarse como:¹²

$$\tau_{spat}(t) : I \rightarrow \mathfrak{S}(a)$$

¹²Nótese que cuando el contexto es claro, por abuso de notación se sustituye *spat* en vez de \mathcal{O}_{spat}

Haciendo uso de esta definición del espacio audible mediante sistemas dinámicos, es posible caracterizar los componentes de dicho espacio de manera particular para cada caso: objetos sonoros fijos, trayectorias espaciales y superficies.

PROPOSICIÓN 4.4.1. Sea $(\mathfrak{S}(a), I, \tau)$ el sistema dinámico asociado a la conciencia audible espacial, entonces, cualquier objeto sonoro fijo puede ser entendido como un punto fijo de dicho sistema.

DEMOSTRACIÓN. Sea $\mathcal{O}_{spat} \in \mathfrak{S}(a)$. Dado que el objetivo sonoro es fijo, esto significa que su posición de localización espacial es constante a lo largo de toda su duración:

$$\gamma(\mathcal{O}_{spat}, t) = (\alpha(t), \beta(t), \theta(t), \Phi(t), d(t)) = (\alpha, \beta, \theta, \Phi) = \gamma(\mathcal{O}_{spat})$$

para todo $t \in [0, n]$. Supóngase por otro lado que el objeto sonoro no es punto fijo del sistema, entonces existe al menos un punto de la órbita de dicho objeto tal que la posición espacial es distinta de la posición inicial, i.e.

$$\exists t^* \in [0, n] \text{ tal que, } \tau_{spat}(t^*) \neq \mathcal{O}_{spat}$$

Lo anterior implicaría que para ese instante t^* :

$$\gamma(\mathcal{O}_{spat}, t^*) \neq \gamma(\mathcal{O}_{spat}, t)$$

Contradiciendo el supuesto de que el objeto sonoro era fijo.

DEFINICIÓN 4.4.2. Sea $(\mathfrak{S}(a), I, \tau)$ el sistema dinámico asociado a un espacio audible. Dado $\mathcal{O}_{spat} \in \mathfrak{S}(a)$ y $n \in I \subset \mathbb{R}$, entonces, la trayectoria espacial continua de cualquier objeto sonoro asociado a dicha representación estará dada por el recorrido finito de la órbita del sistema dinámico correspondiente:

$$\Upsilon[\mathcal{O}_{spat}](n) = \{\tau(t, \mathcal{O}_{spat}) \mid t = 0, 1, \dots, n\}$$

en donde $|\mathcal{O}_{spat}| = n$.

Para el caso del *espacio spectromorfológico* $S \subset \mathbb{R}^k$, es necesario recordar que, la dimensión k dependía de la morfología y del número de descriptores utilizados para construir el *perfil spectromorfológico* del objeto sonoro en cuestión. En este sentido, es posible también definir este espacio mediante un sistema dinámico tal y como se presenta a continuación:

DEFINICIÓN 4.4.3. Sea S el *espacio spectromorfológico* e $I \subset \mathbb{R}^+$ un conjunto indexador de tiempo. Entonces, el *sistema dinámico spectromorfológico* asociado a S , estará dado por la tripleta (S, I, ψ) , donde:

$$\begin{aligned} \psi: I \times S &\longrightarrow S \\ (t, \mathcal{O}_{specM}) &\longrightarrow \psi(t, \mathcal{O}_{specM}) \end{aligned}$$

Sea $n \in I$, y $\mathcal{O}_{specM} \in S$, entonces, se denominará *recorrido finito* a la sección de la órbita correspondiente al intervalo de tiempo n :

$$Y_{specM}(n) = \{\psi(t, \mathcal{O}_{specM}) \mid t = 0, 1, \dots, n\}$$

De manera análoga a como se hizo con el sistema dinámico asociado a la espacialidad, es posible abreviar el operador evolución cuando se hace referencia a un punto específico del sistema, en este caso $\mathcal{O}_{specM} \in S$. Por esta razón, dicha función puede expresarse como:

$$\psi_{specM}: I \longrightarrow S$$

PROPOSICIÓN 4.4.2. Dado un *sistema dinámico spectromorfológico* por la tripleta (S, I, ψ) , los *objetos spectromorfológicamente estáticos* son asociados a los puntos fijos de dicho sistema.

La prueba de esta proposición es exactamente análoga a la mostrada para el caso del sistema dinámico de la espacialidad.

3. Donde la espacialidad y el timbre se unen

Uno de los principales intereses de este trabajo de investigación era el de plantear la conjetura de que existe una relación indisoluble entre la espacialidad y el perfil spectromorfológico de

cada objeto sonoro dentro del proceso compositivo de la música electroacústica. Se ha vislumbrado que esta conjetura es cierta hasta cierto grado, a partir de los resultados expuestos en el capítulo I. Estos resultados han sido comprobados para fuentes sonoras muy simples (tonos sinusoidales y ruido) por lo que no es posible extrapolar los mismos a sonidos más complejos de manera indiscriminada y directa. Un camino posible para obtener resultados acerca de la relación Espacio-Timbre con fuentes sonoras más complejas desde la parte psicoacústica consiste en plantear un *corpus* teórico formal que permita fundamentar de manera directa dichas comprobaciones prácticas. Al distinguir como problemática esta relación entre el espacio y el timbre, se propuso como solución el concepto Espacio-Timbre.

En este apartado se establecerá dicha correspondencia a través de la representación realizada mediante sistemas dinámicos. Como se mencionó anteriormente, cada sistema dinámico (en cualquiera de los dos espacios, el de la conciencia audible espacial o el del perfil espectromorfológico) puede pensarse como un elemento matemático que modela el comportamiento de la cualidad del espacio que actúa sobre el objeto sonoro. Por ejemplo, una trayectoria espacial que recorre el arreglo multicanal en forma cíclica puede ser modelada por un sistema dinámico específico, el cual puede actuar sobre un número cualquiera de objetos sonoros. De igual modo, una transformación espectromorfológica puede ser modelada mediante un sistema dinámico el cual, a su vez, podrá ser aplicado a un sinnúmero de objetos sonoros.

Bajo esta lógica, tiene más sentido hablar de una *colección de sistemas dinámicos para el espacio audible* y una *colección de sistemas dinámicos para el espacio espectromorfológico*.

DEFINICIÓN 4.4.4. Sea Π la colección de los sistemas dinámicos que actúan sobre el espacio aural y Ψ la colección de los sistemas dinámicos que actúan sobre el espacio espectromorfológico.

Para poder establecer entonces el concepto Espacio-Timbre de manera matemática, se hará uso de la propiedad universal del producto de conjuntos y el procedimiento será básicamente el de establecer las relaciones necesarias entre elementos genéricos Π y Ψ de modo tal que mediante dicha propiedad se genere un mapeo tal que conjugue ambas características simultáneamente.

PROPOSICIÓN 4.4.5. Sean $(\mathfrak{S}(a), I, \tau) \in \Pi$ un sistema dinámico asociado a la conciencia audible espacial y $(S, I, \psi) \in \Psi$ un sistema dinámico asociado al perfil espectromorfológico. Existe un sistema dinámico $\mathcal{E} = (\mathfrak{S}(a) \times S, I, \zeta)$ tal que:

$$\zeta = (\tau, \psi)$$

A \mathcal{E} se le denominará *sistema dinámico del Espacio-Timbre*.

DEMOSTRACIÓN. Sea $\mathcal{O} = (\mathcal{O}_{spat}, \mathcal{O}_{spmor})$ un objeto sonoro con representaciones espacio-espectromorfológicas dadas por los sistemas dinámicos correspondientes:

$$\tau_{spat}: I \rightarrow \mathfrak{S}(a)$$

$$\psi_{specM}: I \rightarrow S$$

Para I un conjunto indexador de tiempo. Por la *propiedad universal de productos de conjuntos* existe una única función $\zeta: I \rightarrow \mathfrak{S}(a) \times S$ tal que el siguiente diagrama conmuta:

$$\begin{array}{ccc}
 & & \\
 & \swarrow & \searrow \\
 & \mathfrak{S}(a) \times S & \\
 & \swarrow & \searrow \\
 \tau_{spat} & & (\tau\psi)S \\
 & \swarrow & \searrow \\
 & \mathfrak{S}(a) & \\
 & \swarrow & \searrow \\
 & I & \\
 \psi_{specM} & &
 \end{array}$$

De este modo es posible construir un nuevo sistema dinámico $\mathcal{E} = (\mathfrak{S}(a) \times S, I, \zeta)$, tal que:

$$\zeta = (\tau_{spat}, \psi_{specM}): I \rightarrow \mathfrak{S}(a) \times S$$

donde:

$$\zeta(t) = (\tau_{spat}(t), \psi_{specM}(t))$$

De lo anterior se observa que \mathcal{E} tiene como espacio de estados al producto de los estados de fase de los sistemas dinámicos del espacio aural y espectromorfológicos; de este modo, todo elemento del estado de fase de \mathcal{E} es una descripción espacio-espectromorfológica de

un objeto sonoro \mathcal{O} y por lo tanto dicho sistema expresará simbólicamente el parámetro único del *Espacio-Timbre*.

El resultado anterior implica que se tiene un objeto teórico lo suficientemente general como para poder englobar casi cualquier descripción, proceso o acercamiento que involucre una relación entre la espacialidad y el timbre dentro de la música electroacústica. La utilidad de haber modelado las definiciones de una manera tan abstracta, a partir de sistemas dinámicos, radica en que la definición no dependerá de una visión o un caso particular sino que es expandible a virtualmente cualquier caso.

4. Ejemplos

Las secciones anteriores mostraron cómo se construyeron, en la presente propuesta, los conceptos de *espacio morfológico*, *espacio espectral* y la consecuente formación del *espacio espectral-morfológico* a partir de los dos anteriores. Se mostró además la construcción formal del *Espacio-Timbre* haciendo uso de sistemas dinámicos con el fin de establecer dicho concepto de una manera formal y general.

A continuación se mostrarán algunos ejemplos concretos en los que se definirán dichos espacios de acuerdo con parámetros espectralmorfológicos específicos, con el fin de que se pueda obtener un panorama más claro de la metodología conceptual propuesta y, sobre todo, de una de las mayores aportaciones del presente trabajo: la generalización. Una vez planteado el modelo conceptual-abstracto, una gran cantidad de casos particulares pueden ser abordados y explicados a través del mismo. Una manera alternativa y un tanto informal de enfatizar la importancia de la generalización, consiste en decir que, en realidad, se pueden construir tantos *espacios espectralmorfológicos* como combinaciones de descriptores se tengan.

Desde el punto de vista conceptual es en cierto modo irrelevante el conjunto de descriptores que se considere para conformar los distintos espacios morfológicos y espectrales o espectralmorfológicos. Se menciona de nuevo el hecho de que la construcción de dichos espacios (esto es, la elección de la combinación de los descriptores a utilizar) dependerá completamente de las necesidades y/o intereses particulares de cada caso.

La pregunta que surge como consecuencia de lo anterior es: ¿cuáles descriptores conformarán espacios más adecuados para fines de análisis y/o creación? La respuesta a la cuestión se

encuentra en función de los objetivos y metas de cada caso particular, pero también de las propiedades que cada descriptor revele al respecto del sonido. Como se verá en la siguiente sección, cada descriptor posee características específicas que pueden o no ser útiles dependiendo de: *a)* qué otros descriptores se usen en combinación con los primeros y *b)* el contexto que se plantee para dicho uso. En resumen, es necesario enfatizar dos puntos importantes:

- El modelo plantea una definición general de espacios morfológicos y espectrales de tal forma que es posible construir cualquier caso particular a partir de dicha conceptualización abstracta.
- La elección de los descriptores a utilizar (y por lo tanto de la construcción de los casos particulares de dichos espacios) dependerá de los fines y objetivos particulares de análisis/creación, así como de la combinación sinérgica de dichos descriptores.

Enseguida se presentan ejemplos concretos para casos específicos de combinaciones de descriptores. Para todos los ejemplos propuestos se utilizarán los mismos dos objetos sonoros con el fin de mostrar la diversidad de representaciones de estos a partir de la construcción de los espacios espectromorfológicos con base en las distintas combinaciones de los descriptores y caracterizaciones numéricas elegidas.

Sean $\mathcal{O}(a)$, $\mathcal{O}(b)$ los objetos sonoros mencionados y que están asociados a dos muestras de audio cuyos espectrogramas y formas de onda se muestran en las figuras 11 a 14. Los dos objetos sonoros corresponden a muestras de golpes de madera y cristal (respectivamente), ambas procesadas digitalmente mediante resonadores y filtros con transformación de altura.

Todos los valores de las caracterizaciones numéricas de los descriptores de dichos objetos sonoros fueron obtenidos utilizando la librería *Librosa* (McFee *et al.*, 2015) dentro del lenguaje de programación Python.

4.1. Ejemplo I

Considérense como descriptores morfológicos: *a)* ZCR y *b)* Energía. Supóngase que para el ZCR se considera como caracterización numérica, el *coeficiente de variación*, mientras que para la energía se considera la *entropía*. Supóngase además que se considera un solo descriptor espectral: *coeficiente de tonalidad espectral*, donde dicho descriptor tiene como caracterización nu-

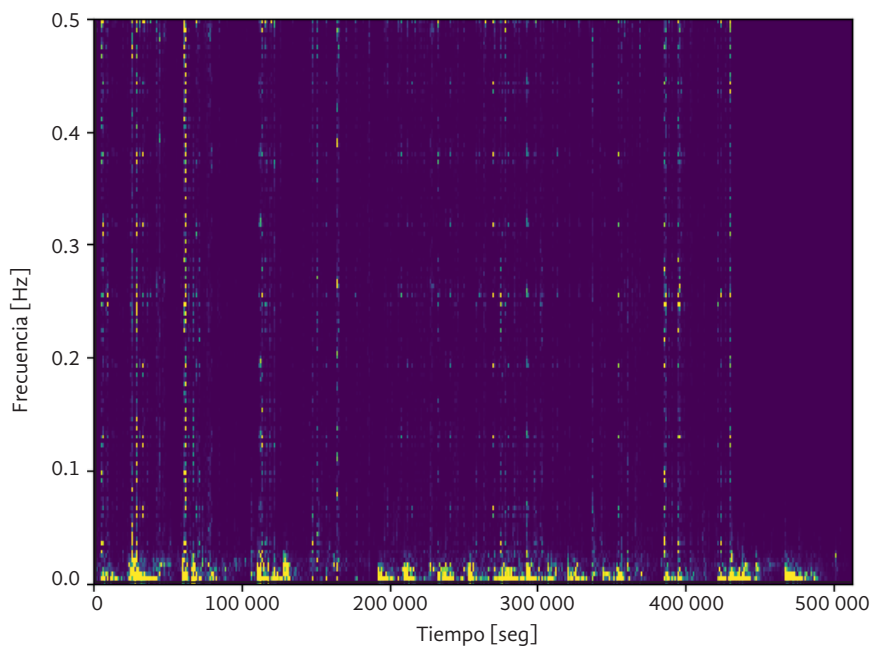


Figura 11 | Espectrograma para muestra de audio I.

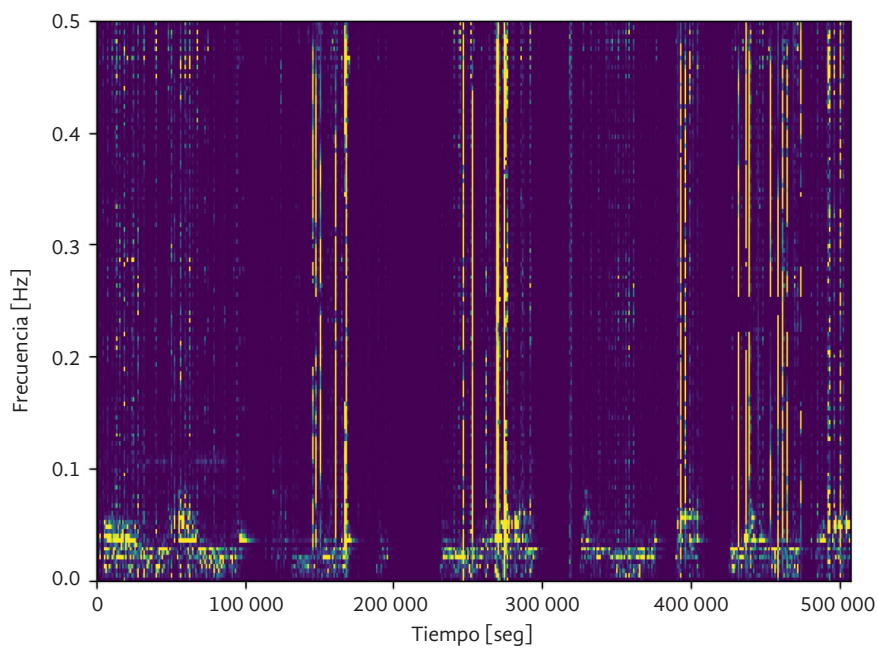


Figura 12 | Espectrograma para muestra de audio II.

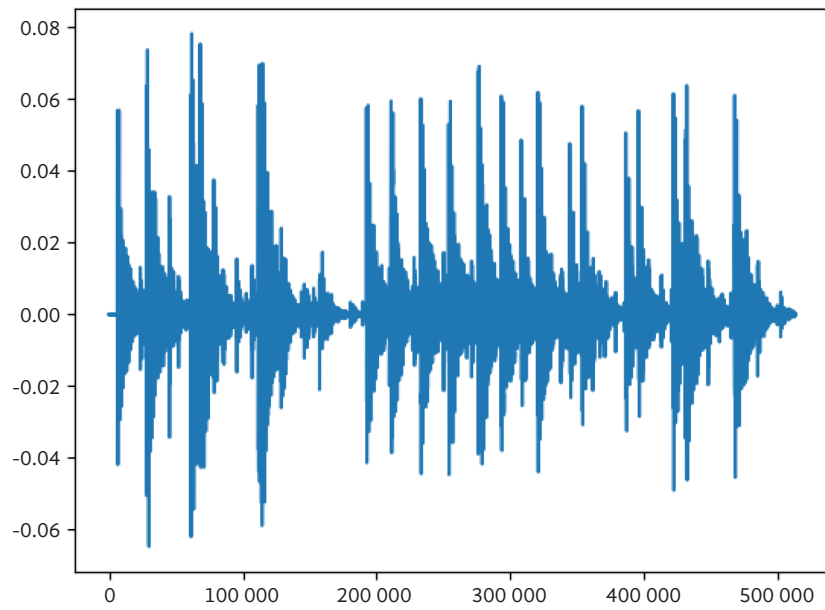


Figura 13 | Forma de onda para muestra de audio I.

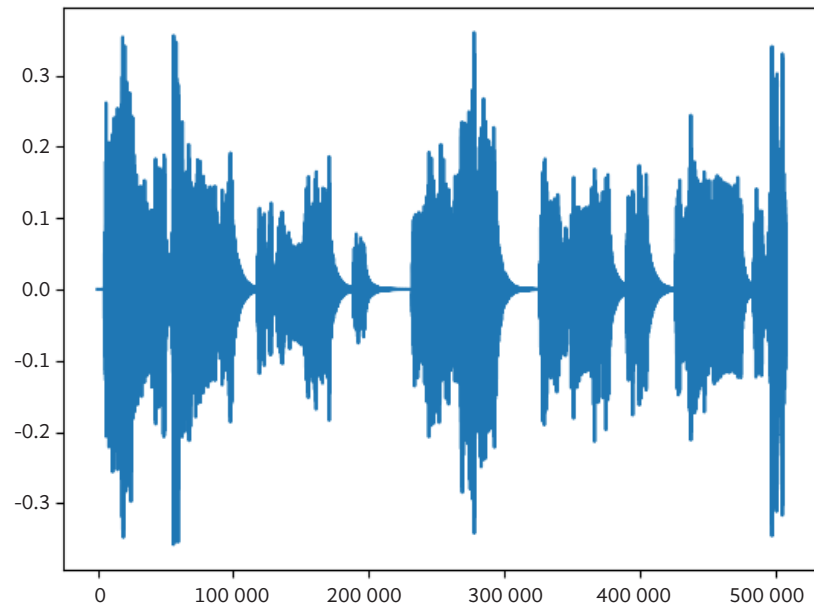


Figura 14 | Forma de onda para muestra de audio II.

mérica la *entropía* (los valores obtenidos se muestran en los vectores correspondientes). Supóngase que se elige la métrica euclideana como medida de distancia entre vectores. Entonces, la parametrización del espacio espectromorfológico estará dado por:

$$S = \mathcal{M} \times \mathcal{E} \subset [0, 1]^2 \times [0, 1]^2 \subset \mathbb{R}^4$$

Para el espacio morfológico \mathcal{M} los vectores $m(a)$ y $m(b)$ asociados a los objetos sonoros $\mathcal{O}(a)$ y $\mathcal{O}(b)$ respectivamente, tienen los valores:

$$\begin{aligned} m(a) &= [m(a)_0, m(a)_1] = [2.131538, 0.002888036] \\ m(b) &= [m(b)_0, m(b)_1] = [0.656021, 0.008939] \end{aligned}$$

donde para cada vector, la primera coordenada m_0 corresponde al valor del *coeficiente de variación* del ZCR, mientras que la segunda coordenada m_1 es el valor de la *entropía* de la energía, para cada uno de los objetos sonoros en cuestión.

De manera análoga, para el espacio espectral, los vectores asociados a los objetos sonoros estarán dados por:

$$\begin{aligned} e(a) &= [e(a)_0, e(a)_1] = [0.02223, 0.1222] \\ e(b) &= [e(b)_0, e(b)_1] = [0.0143, 0.4522] \end{aligned}$$

De lo anterior se observa que los vectores dentro del espacio espectromorfológico están dados por:

$$\begin{aligned} s(a) &= [m(a)_0, m(a)_1, e(a)_0, e(a)_1] = [0.02223, 0.1222, 0.33, 0.567] \\ s(b) &= [m(b)_0, m(b)_1, e(b)_0, e(b)_1] = [0.0143, 0.4522, 0.1223, 0.13345] \end{aligned}$$

Y de este modo, la distancia dentro del espacio espectromorfológico entre dichos vectores estará dada por:

$$\begin{aligned} d[(a), s(b)] &= (|m(a)_0 - m(b)_0|^2 + |m(a)_1 - m(b)_1|^2 + |e(a)_0 \\ &\quad - e(b)_0|^2 + |e(a)_1 - e(b)_1|^2)^{1/2} = 0.53375928387192 \end{aligned}$$

La interpretación de esta distancia entre los objetos sonoros es que, *dados los descriptores elegidos y las caracterizaciones numéricas*, los objetos sonoros muestran ese nivel de similaridad denotado por dicha distancia.

Tal y como se mencionó al inicio, las distintas posibles construcciones de espacios espectromorfológicos son bastante numerosas y dependen, ciertamente, de las combinaciones de descriptores elegidos para conformar la dimensión de cada espacio (espectral y morfológico) y de las caracterizaciones numéricas. Para clarificar esta idea, hágase un ejercicio mental y del ejemplo que se ha mostrado en esta sección, supóngase caracterizaciones numéricas distintas para cada uno de los descriptores. Esto afectará natural y directamente al valor de la distancia entre los objetos sonoros y no sólo eso, como implicación conceptual, se tendrá un espacio espectromorfológico distinto. Entonces, se reitera lo dicho con anterioridad: el espacio espectromorfológico depende en su construcción de la elección de los descriptores y de las caracterizaciones numéricas para cada descriptor.

4.2. Ejemplo II

Considérense para este segundo ejemplo como descriptores morfológicos: *a)* ZCR, *b)* energía y *c)* envolvente de Hilbert. Supóngase que para todos los descriptores se tiene como caracterización la entropía. Supóngase por otro lado, que se consideran como descriptores espectrales:

- Coeficiente de tonalidad espectral.
- Centroide espectral.
- Relevo espectral.

Supóngase que para los dos primeros descriptores espectrales se considera como caracterización numérica el *coeficiente de variación*, mientras que para el tercero se considera la varianza. Entonces, el espacio espectromorfológico estará dado por:

$$S = \mathcal{M} \times \mathcal{E} \subset [0, 1]^3 \times [0, 1]^3 \subset \mathbb{R}^6$$

Para el espacio morfológico \mathcal{M} los vectores $m(a)$ y $m(b)$ asociados a los objetos sonoros $\mathcal{O}(a)$ y $\mathcal{O}(b)$ respectivamente, tienen los valores:

$$\begin{aligned}
 m(a) &= [m(a)_0, m(a)_1, m(a)_2] = [0.02223, 0.1222, 0.33456] \\
 m(b) &= [m(b)_0, m(b)_1, m(b)_2] = [0.0143, 0.4522, 0.298845]
 \end{aligned}$$

donde para cada vector, la primer coordenada m_0 corresponde al valor de la entropía del *coeficiente de variación* del ZCR, la segunda coordenada m_1 es el valor de la *entropía* de la energía de los objetos sonoros en cuestión y la tercera la correspondiente a la entropía de la envolvente de Hilbert.

De manera análoga, para el espacio espectral, los vectores asociados a los objetos sonoros estarán dados por:

$$\begin{aligned}
 e(a) &= [e(a)_0, e(a)_1, e(a)_2] = [0.63324, 0.19282, 0.055572] \\
 e(b) &= [e(b)_0, e(b)_1, e(b)_2] = [0.32338, 0.30548214, 0.0880419]
 \end{aligned}$$

Y de este modo, la distancia dentro del espacio espectralmorfológico entre los vectores asociados a los objetos sonoros, estará dada por:

$$\begin{aligned}
 d[s(a), s(b)] &= (|m(a)_0 - m(b)_0|^2 + |m(a)_1 - m(b)_1|^2 + |m(a)_2 - m(b)_2|^2 + |e(a)_0 - e(b)_0|^2 + |e(a)_1 - e(b)_1|^2 + |e(a)_2 - e(b)_2|^2)^{1/2}
 \end{aligned}$$

4.3. Ejemplo III

Para este último ejemplo, considérese como descriptor morfológico único el ZCR. Supóngase que ahora dicho descriptor tiene una caracterización paramétrica dada por la serie de tiempo obtenida del análisis FFT, que para los objetos sonoros utilizados se puede visualizar en la figura siguiente.

Las figuras mencionadas representan el valor del ZCR para cada *frame* correspondiente a lo largo de la duración total de cada objeto sonoro. De este modo se tiene la siguiente parametrización para cada objeto sonoro:

$$\begin{aligned}
 I &\longrightarrow \mathcal{M} \\
 t &\longrightarrow m(t)
 \end{aligned}$$

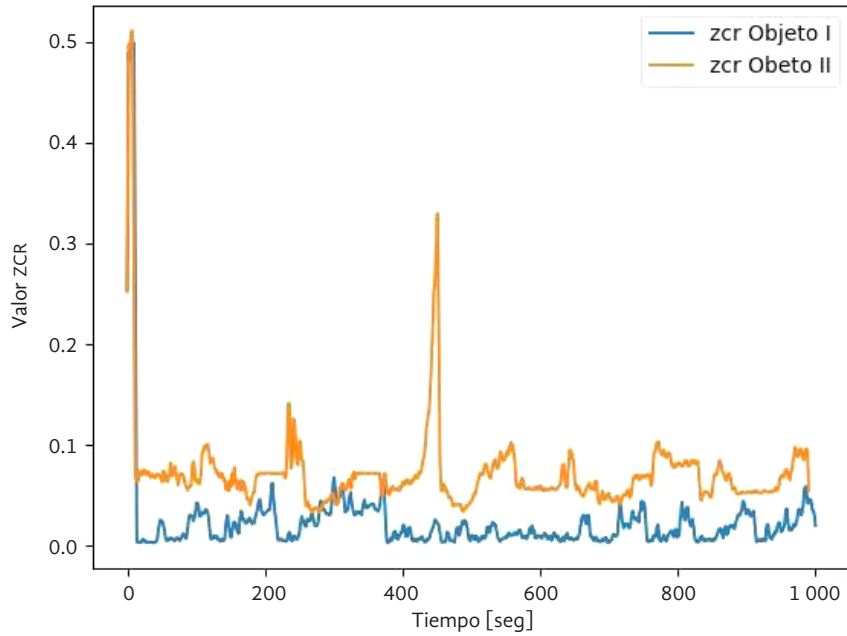


Figura 15 | Series de tiempo del ZCR para los objetos sonoros I y II.

Sean entonces, $m(a, t)$, $m(b, t)$ los vectores del espacio morfológico asociados a los objetos sonoros $\mathcal{O}(a)$ y $\mathcal{O}(b)$ respectivamente:

$$m(a, t) = [m(a, 0), m(a, 1), m(a, 2), \dots, m(a, k)] = [0.2345, 0.123, 0.321, \dots, 0.8596]$$

$$m_1(t) = [m(b, 0), m(b, 1), m(b, 2), \dots, m(b, l)] = [0.2345, 0.123, 0.321, \dots, 0.8596]$$

Si se desea calcular la distancia $d[m(a), m(b)]$ entre estos dos vectores de manera práctica, uno de los métodos más útiles es el llamado *dynamic time warping* o DTW, el cual permite estimar niveles de disimilaridad entre series de tiempo incluso si estas poseen tamaños distintos y que puede ser obtenido a través de cualquier paquetería computacional como PYPY¹³ para Python. Supóngase ahora que se tienen como descriptores espectrales los siguientes:

- Coeficiente de tonalidad espectral.
- Centroide espectral.

¹³ <<https://pypi.org/project/dtw-python/>>.

Tales que el *coeficiente de tonalidad espectral* tiene una caracterización numérica fija dada por el *coeficiente de variación*, mientras que el *centroide espectral* tiene una caracterización numérica paramétrica. En este caso, es posible considerar ambos descriptores con una representación paramétrica, sólo que para el primer descriptor, el valor del mismo será constante en cualquier instante de tiempo:

$$\begin{aligned} I &\rightarrow \mathcal{E} \\ t &\rightarrow e(t) \end{aligned}$$

donde:

$$\begin{aligned} e(a, t) &= [e(a, 0), e(a, 1), e(a, 2), \dots, e(a, k)] = [0.2345, 0.2345, 0.2345, \dots, 0.2345] \\ e(b, t) &= [e(b, 0), e(b, 1), e(b, 1), \dots, e(b, l)] = [0.2345, 0.123, 0.321, \dots, 0.8596] \end{aligned}$$

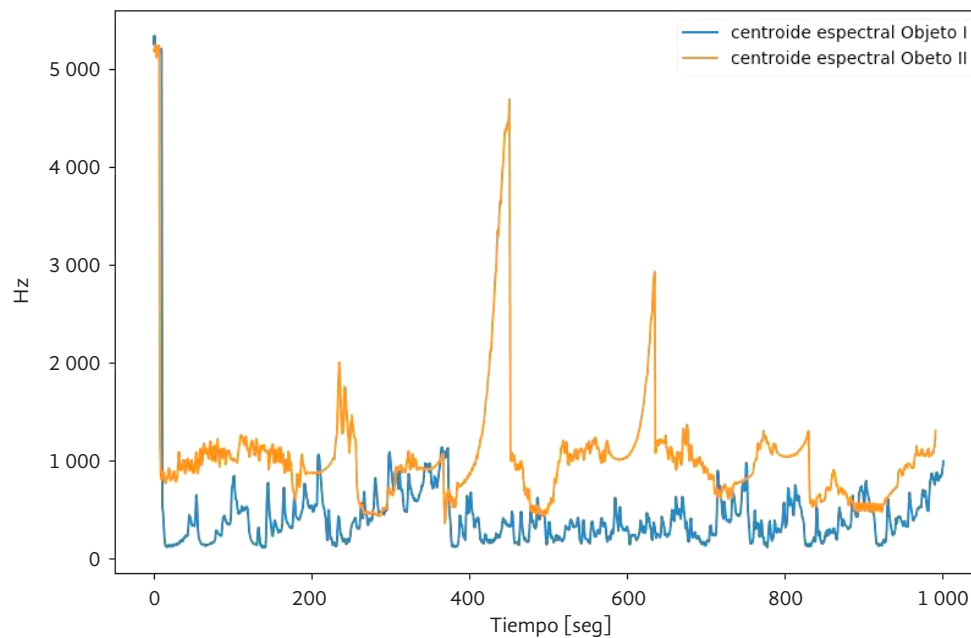


Figura 16 | Series de tiempo del centroide espectral para los objetos sonoros I y II.

Bajo esta lógica, para calcular una distancia entre ambas series de tiempo se puede proceder del mismo modo se mencionó en el caso anterior; esto es, utilizando DTW.

4.4. Ejemplo IV. Metodología computacional básica mediante aprendizaje máquina

En esta sección se describirá una aplicación concreta computacional que se desarrolló con el fin de realizar dos tareas específicas: *a*) descripción espectromorfológica y *b*) clasificación mediante aprendizaje no supervisado dado un conjunto de muestras de audio. Esta aplicación, más allá de ser un estudio profundo sobre implementación de técnicas de aprendizaje máquina (desde la perspectiva del estado del arte), es una manera específica y concreta que pretende ejemplificar las posibilidades prácticas del modelo general a partir del uso de herramientas básicas de aprendizaje máquina. La implementación (y contribución de esta sección) se diferencia de las metodologías convencionales del aprendizaje no supervisado, en el hecho de que el análisis es realizado a partir de la comparación de series de tiempo y, por lo tanto, se incluyen métricas y metodologías que no son comunes en el contexto mencionado. Como se verá más adelante, la justificación para realizar los análisis a partir de la comparación de series de tiempo en lugar de utilizar medidas de tendencia central o de dispersión es que por su propia naturaleza, estas últimas tienden a generar sobreestimación en lo local, además de presentar en no pocas ocasiones distribuciones no normales, conduciendo a la probabilidad de ambigüedad respecto del significado de los valores obtenidos; además, la información derivada de los descriptores no se presenta como una distribución normal. Por supuesto, esta metodología tiene la desventaja de que implica un gasto computacional considerablemente más alto con respecto a la comparación básica de las medidas de tendencia central y de dispersión; sin embargo, para casos en donde la complejidad espectromorfológica es alta y sobre todo dinámica a través del tiempo, dicha metodología computacional puede resultar bastante útil para extraer información más precisa acerca de las características de los objetos sonoros.

A partir de un conjunto de muestras de audio (de instrumentos musicales convencionales tocados a diferentes alturas y con distintas dinámicas, todos ellos propiamente etiquetados de acuerdo con tres parámetros específicos: *a*) instrumento, *b*) nota o altura y *c*) dinámica),¹⁴ se obtienen diversos descriptores acústicos para establecer perfiles espectromorfológicos específicos (tal y como se hizo en la sección anterior) y poder definir posteriormente *clusters* y centroides mediante los algoritmos simples de *alineación global de Kernel k-Medios* y *K-Vecindades*.

¹⁴ Las muestras de audio fueron obtenidas de la página web <philarmonia.co.uk/explore/soundsamples>.

Toda vez que se han obtenido estos centroides, cualquier conjunto de muestras de audio puede ser comparado y/o clasificado (de acuerdo con un perfil espectromorfológico específico), ya sea por grados de similaridad utilizando algoritmos como el *dynamic time warping* (DTW) o mediante aprendizaje supervisado, utilizando como etiquetas los centroides anteriores. De este modo será posible asociar muestras de audio cualquiera a un Espacio-Timbre conocido y así trabajar composicionalmente dichas muestras de acuerdo con los principios de orquestación ya conocidos para los instrumentos musicales utilizados en el análisis primario.

Como se mencionó con anterioridad, la metodología está basada en el análisis y clasificación mediante la comparación de las series de tiempo y haciendo uso de técnicas como el *dynamic time warping* (DTW). Más aún, los centroides son obtenidos mediante una hibridación de herramientas algorítmicas. Dado un *cluster* específico, se obtiene su respectivo *baricentro*, para luego calcular una banda alrededor de él mediante el método de I.B. Keogh.

Lo anterior permite obtener superficies como centroides que se encuentran posicionadas alrededor del baricentro, con lo que los perfiles espectrales tienen un rango de acción no lineal (es decir no es una serie única de tiempo sino una región alrededor de dicha serie representada por el baricentro).

Se muestran aquí las rutinas escritas en Python para que el lector interesado pueda realizar las tareas antes mencionadas de la manera más directa y automática posible y con un fin más bien didáctico. Con estas rutinas, el usuario puede analizar cualquier número de muestras de audio eligiendo un conjunto de descriptores espectrales y de envolvente predefinida de acuerdo con sus intereses particulares o tomando como guía los resultados obtenidos con los conjuntos utilizados en las simulaciones aquí mostradas.

Las rutinas propuestas contienen un número específico de descriptores; sin embargo, dada la forma en la que están escritas, cualquier descriptor puede ser añadido a futuro de manera directa. El objetivo es entonces, el de establecer, en primer lugar, un *espacio espectromorfológico de referencia* derivado del análisis de muestras de audio de instrumentos musicales convencionales.

Como se mencionó con anterioridad, los centroides se definen como zonas o *bandas* alrededor de los baricentros calculados para cada *cluster*, con ello se construye el *espacio espectromorfológico de referencia*. Una vez obtenido el espacio anterior, es posible analizar un conjunto arbitrario de muestras de audio y clasificarlo (o asociarlo a través de una simple relación métrica de cercanía) mediante aprendizaje supervisado (de acuerdo con las etiquetas obtenidas en

la parte inicial), asociando cada muestra nueva a los perfiles espectromorfológicos de instrumentos musicales convencionales del *espacio espectromorfológico de referencia*.

Lo anterior permite, por lo tanto, establecer puntos referenciales de acuerdo con los principios de orquestación conocidos para los instrumentos musicales especificados y obtener un panorama de información referencial para dichas muestras arbitrarias de audio. Antes de continuar es necesario establecer ciertas observaciones respecto de las restricciones, limitaciones y enfoques de la metodología aquí planteada.

OBSERVACIÓN 4.0.1. Las asociaciones de los perfiles espectrales de las muestras de audio arbitrarias con las de los instrumentos musicales son relativas. Al momento de establecer que una muestra de audio arbitraria está relacionada con el perfil espectromorfológico de un instrumento en particular, no se debe nunca perder de vista los descriptores que han sido utilizados y tener siempre presente que dicha asociación dependerá de tales descriptores, así como del método de comparación utilizado (métrica de similaridad o aprendizaje supervisado). No se está afirmando aquí que los principios de orquestación pueden ser aplicados de manera directa e indiscriminada a dichas muestras, sino que estos sirven como guías referenciales para el proceso de organización tímbrica de las muestras de audio arbitrarias.

OBSERVACIÓN 4.0.2. El espacio espectromorfológico de referencia puede ser ampliado de manera directa. Dada la metodología anterior, es posible incluir tanto muestras de audio de diversos instrumentos musicales no integrados previamente como nuevos descriptores, y construir de este modo distintos espacios espectromorfológicos de referencia.

OBSERVACIÓN 4.0.3. Si bien la metodología aquí planteada puede ser directamente aplicada al proceso compositivo, el objetivo de ésta no es el de presentarla como una estrategia y/o herramienta inmediata y de uso directo para el compositor. Esta metodología propone una manera alternativa y original de realizar análisis de audio mediante series de tiempo y clasificación no supervisada. Su propósito principal radica en el desarrollo de un método computacional para la obtención de perfiles espectromorfológicos y no en la presentación de una solución para la composición asistida por computadora. Sin embargo, se derivan de manera natural y directa ciertas soluciones aplicables al proceso compositivo tales como la asociación con los principios orquestales de los instrumentos convencionales o la clasificación de muestras de audio de acuerdo con ciertos perfiles espectromorfológicos con fines de organización tímbrica.

Metodología

1. Obtención de los perfiles espectromorfológicos de las muestras de audio mediante los descriptores morfológicos y espectrales, tanto acústicos como perceptuales.
2. Organización de la información anterior en series de tiempo.
3. Acortamiento proporcional de las series de tiempo de modo tal que conserven su forma original pero con una longitud mucho menor.
4. Obtención de *clusters* de las series de tiempo acortadas mediante alineación global de Kernel por K-Medios para cada descriptor y para cada instrumento musical.
5. Para cada *cluster* anterior, cálculo del baricentro euclidiano.
6. Para cada baricentro obtenido, cálculo de la banda LB Keogh para definir el centroide representativo específico.
7. Utilización de los centroides anteriores para comparar mediante DTW el grado de similitud con muestras de audio arbitrarias.
8. Asociación de las muestras de audio arbitrarias al centroide más cercano y sumariación de las características del *cluster* correspondiente de acuerdo con los tres parámetros etiquetados en las muestras de audio: *a*) instrumento, *b*) nota o altura y *c*) dinámica.

Análisis primario de muestras

Para el análisis realizado en este apartado se utilizó la base de datos alojada en <philharmonia.co.uk/explore/sound_amp>, la cual cuenta con 27 100 muestras de audio pertenecientes a los siguientes instrumentos musicales: cello, contrabajo, clarinete, banjo, clarinete bajo, fagot, flauta, corno francés, guitarra, mandolina, oboe, saxofón, trombón, trompeta, tuba, viola y violín. Estas muestras incluyen notas en diferentes alturas tocadas con distintas dinámicas y algunas otras con motivos melódicos. Todas las muestras están etiquetadas en el formato: x_y_z , donde x es el instrumento, y es la nota o altura y z es la dinámica.

Dado que para los fines del presente análisis era necesario descartar todos aquellos casos que presentaran motivos melódicos y conservar únicamente aquellas muestras que presentaran notas únicas, se escribió una rutina que identificara de manera automática las muestras con esta característica (motivos melódicos), con el fin de descartarlas del conjunto final. Lo anterior se realizó utilizando análisis de envolvente mediante la transformada de Hilbert e

identificando aquellos casos en los que la envolvente presenta variaciones significativas a través del tiempo. La rutina consta básicamente de los siguientes pasos:

1. Cargar muestras de audio. Dada la ruta de acceso de la carpeta raíz donde se encuentran todas las muestras de audio (organizadas por subcarpetas), la rutina carga las muestras y las organiza automáticamente arrojando dos arreglos; el primero correspondiente a los valores flotantes y el segundo a los nombres de las muestras (extraídos también automáticamente de los archivos originales).
2. Para cada muestra, localizar el *onset* e identificar la envolvente del mismo.
3. Seleccionar todas aquellas muestras que correspondan a un impulso sonoro único (es decir muestras que no sean motivos melódicos por ejemplo).
4. Para todas las muestras seleccionadas, trasladar en el tiempo, el *onset* de cada una de ellas a un mismo *frame* con el fin de normalizar el inicio de las mismas.

A continuación se muestran los pasos anteriores escritos como subrutinas en Python. Paso 1:

```
import librosa
import numpy as np
import os, os.path

def loadAll(root_path):
    subdirs_array = [os.path.join(root_path,name) for name in os.listdir(root_path) if name[:2]!='.']
    f_names = files_name(root_path)
    print(subdirs_array)
    data_sound = [load_soundsX(path_x)[0] for path_x in subdirs_array]
    data_names = [load_soundsX(path_x)[1] for path_x in subdirs_array]
    return data_sound, data_names

def load_soundsX(path_x):
    files_array = files_path(path_x)
    f_names = files_name(path_x)
    librosa_array=[librosa.load(y, sr = sample_rate, duration=sample_duration, mono=True)[0] for y in files_array]
    return librosa_array, f_names

def files_name(path_x):
    temp_array = [name for name in os.listdir(path_x) if os.path.isfile(os.path.join(path_x,name)) if name[:2]!='.']
    return temp_array

def files_path(path_x):
    temp_array = [os.path.join(path_x,name) for name in os.listdir(path_x) if os.path.isfile(os.path.join(path_x,name))]
    return temp_array
```

Paso 2-4:

```
#Clasificación automática de muestras de un sólo onset. #Ejemplo:
#data_OneShot, data_OneShotNames = oneShotClassification(data_sound, data_names)

def oneShotClassification(data_input, data_names):
    onset_data = [onset_extract(x) for x in data_input]
    oneShot_index = [[[i,j] for j in range(len(onset_data[i])) if len(onset_data[i][j])<2] for i
    in range(len(onset_data))]
    data_namesOneShot = [[index_place(data_names,i) for i in j] for j in oneShot_index]
    data_soundsOneShot = [[index_place(data_input,i) for i in j] for j in oneShot_index]
    return data_soundsOneShot, data_namesOneShot

def onset_extract(sample_array):
    onset_array = [Onset_detect(x) for x in sample_array]
    return onset_array

def Onset_detect(signal):
    onset_frames = librosa.onset.onset_detect(signal, sr=22050, wait=1, pre_avg=1, post_avg=1, pre_max=1, post_max=1)
    return onset_frames
```

Del proceso anterior se obtuvo un conjunto final de 13 533 muestras con las características antes mencionadas, éste fue el conjunto de muestras con el cual se trabajó de manera final. Para todas esas muestras se calcularon descriptores acústicos utilizando la librería *librosa* de Python (McFee *et al.*, 2019). Los descriptores calculados para cada una de las muestras fueron: ZCR, RMSE, ancho de banda espectral, flujo espectral y contraste espectral.

Para fines de organización de la información anterior, cada una de las seis subbandas del *contraste espectral* se trataron como un descriptor individual, arrojando en total 10 descriptores para cada muestra de audio; dicha información fue almacenada en formato de series de tiempo.

Tratamiento de las series de tiempo y optimización preclasificatoria

Tal y como se mencionó al inicio de esta sección, gran parte de la información arrojada por los descriptores presenta distribuciones no normales; por ello, el uso de medidas de tendencia

central o dispersión, como la media o la desviación estándar podrían fácilmente generar resultados espurios o fuera de contexto sobre todo para muestras de audio con una complejidad espectromorfológica considerable y dinámica a través del tiempo. Con el fin de solucionar esta situación se decidió realizar el análisis y el procesamiento de la información a partir de las series de tiempo y mediante grados de similaridad entre ellas haciendo uso de algoritmos como el *dynamic time warping* o *DTW*.

Debido a que el análisis y el proceso de clasificación con aprendizaje no supervisado para series de tiempo es mucho más caro computacionalmente que la clasificación con medidas de tendencia central, era crítico incrementar la eficiencia de dicho proceso lo más posible. Para ello se diseñó un método simple que permite representar las series de tiempo con un número menor de puntos pero manteniendo, de manera bastante aceptable, la forma de la envolvente original (y por lo tanto la información relevante al respecto). Este método reduce el tamaño de la serie original a cualesquiera divisores enteros de la longitud original de dicha serie pero, para fines de aplicación práctica, se observó que la reducción más eficiente está entre la mitad y la tercera parte de la longitud original. La conclusión anterior se obtuvo a partir de una comparación directa de la forma de las envolventes original y recortada, al hacer un escalamiento lineal de la envolvente original en el rango de tiempo de la envolvente recortada. Lo anterior se resume en los siguientes pasos:

1. Cargar la serie de tiempo.
2. Calcular la longitud de la serie de tiempo.
3. Definir el rango del acortamiento como la razón entre la longitud de la serie y el número de intervalos en los que se desea recortar la señal o *chunks*.
4. Dividir la señal mediante intervalos de longitud igual al parámetro *chunks*.
5. Para cada intervalo calcular el máximo local y reconstruir la señal a partir de dichos valores. Como método alternativo se puede usar el valor medio de cada intervalo en lugar del máximo.

Se muestran a continuación las rutinas escritas en Python correspondientes al proceso anterior. Obsérvese que se tiene la alternativa de transformar la señal recortada con una transformada de Hilbert y suavizada mediante un filtro *butterwooth* tipo A con el fin de aumentar o disminuir el grado de detalle de la misma.

```

#Ejemplo:
#data_OneShot, data_OneShotNames = oneShotClassification(data_sound, data_names)

from scipy.signal import hilbert, chirp
from scipy.signal import butter, filtfilt

def chopAll(array_input, chunk):
    chop_out = [[chop_array(x, chunk) for x in array_input[i]] for i in range(len(array_input))]
    return chop_out

def chop_array(data_array, chunk):
    #factor_mtx = factorSet(data_array, -2)
    #factorsMtx = [int(len(x)/y) for x,y in zip(data_array, factor_mtx)]
    #chop_array = [chop(z,y) for z,y in zip(data_array, factorsMtx)]
    chop_data = [chop(z, chunk) for z in data_array]
    return chop_data

def chop(signal, chunks):
    signal_length = len(signal)
    chunk_range = int(signal_length/chunks)

    if signal_length%2 == 0:
        signal = signal[:-1]
        signal_length = len(signal)
    chunk_array = [np.max(signal[i*chunks:(i+1)*chunks]) for i in range(chunk_range)]
    #chunk_array = [np.mean(signal[i*chunks:(i+1)*chunks]) for i in range(chunk_range)]
    #chunk_filt = hilbert_env(chunk_array, rolloff)
    return chunk_array

def hilbert_env(data, rolloff):
    analytic_signal = hilbert(data)
    amp_env = np.abs(analytic_signal)
    amp_env = FilteredSignal(amp_env, rolloff)
    return amp_env

def FilteredSignal(signal, cutoff):
    B, A = butter(1, cutoff, btype='low')
    filtered_signal = filtfilt(B, A, signal, axis=0)
    return np.array(filtered_signal)

```

Una vez que se ha aplicado el proceso anterior a todas las series de tiempo de los descriptores utilizados, el siguiente paso consiste en comprobar que las series de tiempo acortadas mantienen la forma original de tal modo que la información primaria no se pierda y continúe

siendo significativa. Para ello se hizo un escalamiento al intervalo de tiempo de las series acortadas para cada una de las series originales. En este sentido, las series originales se compararon con las series acortadas dentro del mismo intervalo de tiempo mediante una interpolación lineal. Las figuras 17, 18 y 19 muestran ejemplos del proceso anterior. La gráfica superior muestra la serie de tiempo original, la gráfica siguiente (hacia abajo) representa la serie de tiempo recortada y finalmente, la gráfica inferior muestra a la serie de tiempo recortada junto con la gráfica original interpolada en el intervalo de tiempo de la serie recortada con el fin de comparar la forma de ambas.

Existen casos en los cuales las series de tiempo acortadas pueden presentar variaciones más significativas con la forma de la serie original. Como proceso alternativo al anterior, es posible reconstruir la señal mediante la identificación de *onsets* y la reconstrucción de la serie mediante interpolación lineal y un filtrado *butterwooth* tipo A. En la figura 20 en la que se pueden observar los *onsets* identificados en la serie de tiempo original mientras que la figura 21 muestra la reconstrucción de la serie de tiempo recortada a partir de la interpolación de dichos valores.

Clasificación de series de tiempo mediante aprendizaje no supervisado

Dado que se tiene un conjunto de 13 533 de muestras de audio y 10 descriptores, el total de las series de tiempo obtenidas como fuente de información es de 135 330. Recuérdese que estas muestras de audio están organizadas en 19 distintos instrumentos musicales convencionales.

La metodología aquí propuesta tiene el objetivo de establecer *clusters* para cada instrumento musical y para cada descriptor con el objetivo de definir perfiles espectromorfológicos específicos que permitan generar puntos de referencia para futuras comparaciones con sonidos arbitrarios.

El algoritmo utilizado para construir los *clusters* fue el de Global Kernel K-Means, con alineación basada en *DTW* implementado en la librería *tslearn* de Python (Tavenard *et al.*, 2017). El proceso se puede resumir del siguiente modo:

1. Elección del conjunto de las series de tiempo correspondientes a un descriptor particular para las muestras de audio correspondientes a un instrumento particular.

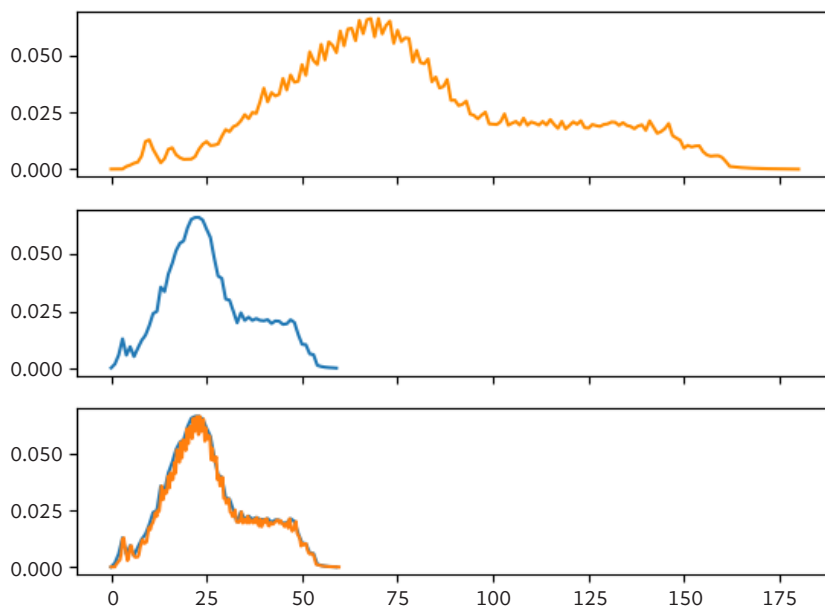


Figura 17 | Comparación de series de tiempo para RMSE de audio de corno francés. Nota A2 tocada en *mezzo forte*.

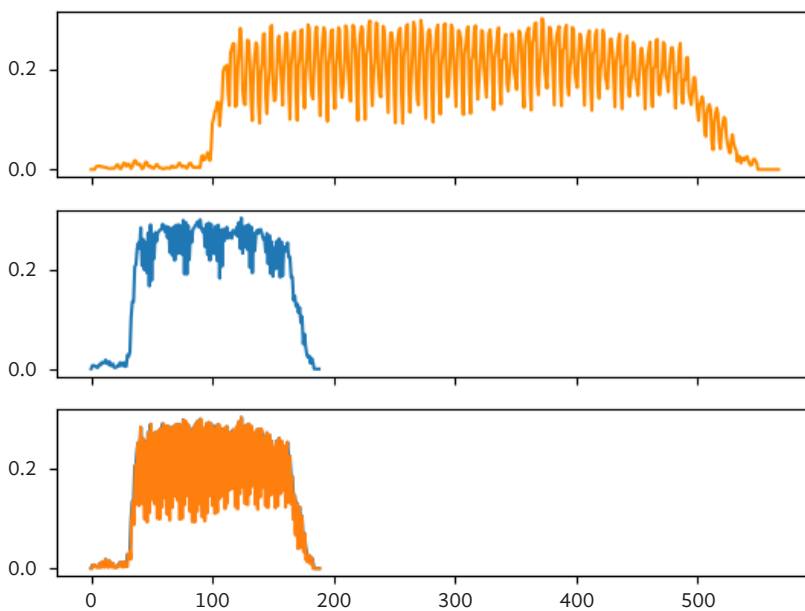


Figura 18 | Comparación de series de tiempo para RMSE de audio de corno francés. Nota A2 tocada en *fortissimo*.

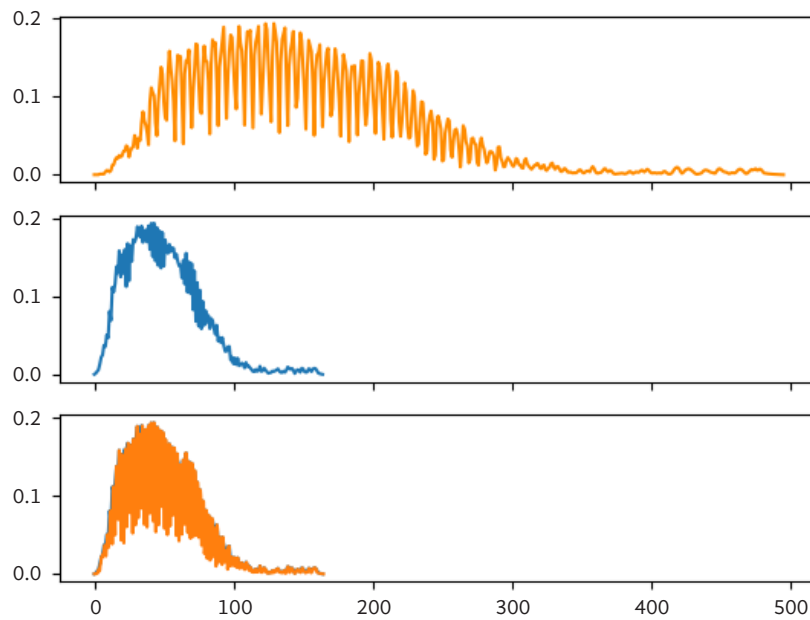


Figura 19 | Comparación de series de tiempo para RMSE de audio de corno francés. Nota A3 tocada en trino fuerte.

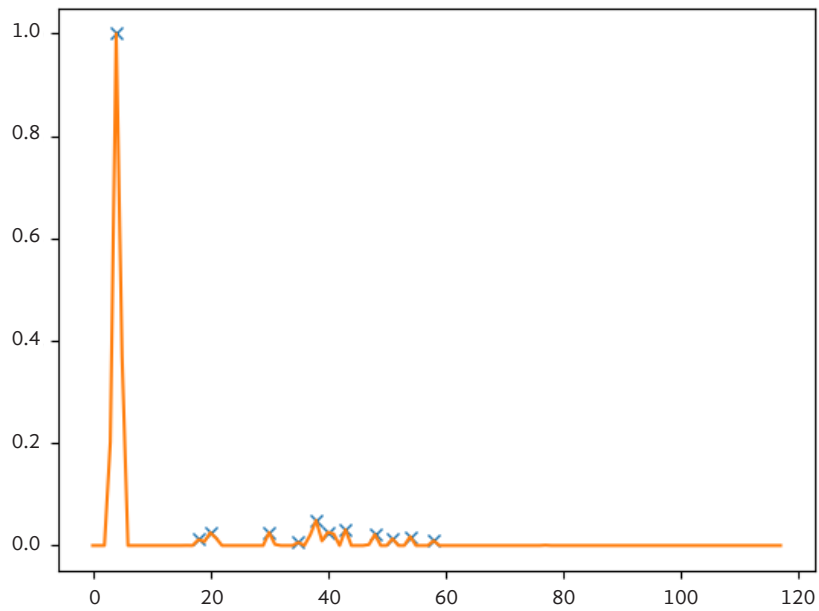


Figura 20 | Identificación de *onsets* en la serie de tiempo.

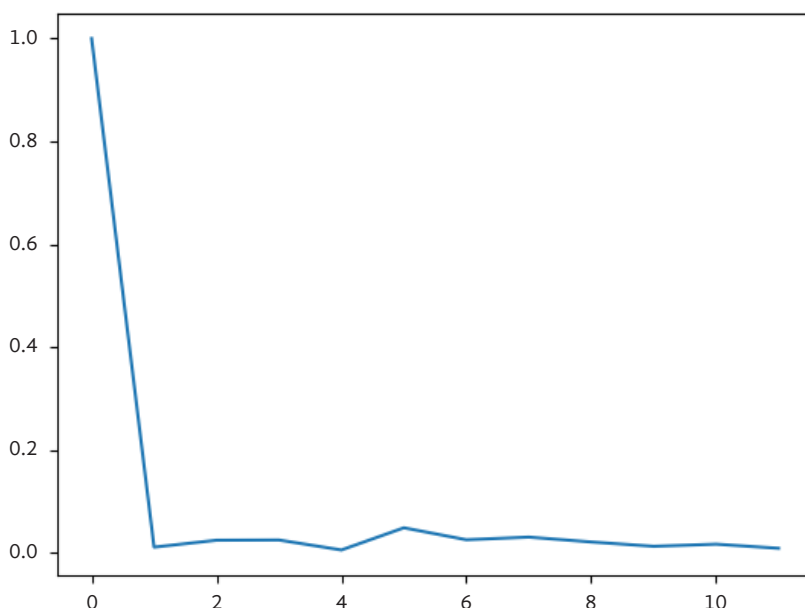


Figura 21 | Reconstrucción de la serie de tiempo a partir de la interpolación de *onsets*.

2. Aplicación de formato particular de dichas series de acuerdo con los requisitos de la librería *tslearn* y reemplazo de cualquier posible valor *nan* por 0.
3. Selección del número de *clusters* para realizar la clasificación.
4. Obtención de los *clusters*.
5. Repetición de todo el proceso para todos los descriptores.
6. Agrupamiento de las muestras en *arrays* de acuerdo con los *clusters* obtenidos.

El algoritmo GAK ofrece la ventaja de poder comparar series de tiempo que no son de la misma longitud debido precisamente a que usa DTW para realizar el proceso. En este sentido, el algoritmo alinea las series de tiempo, las compara de acuerdo con su forma (envolvente) y posteriormente realiza la clasificación basada en K-Medios. Por supuesto, éste es un algoritmo básico de toda la gama de posibilidades del aprendizaje máquina y no se plantea como una implementación computacional compleja (al respecto de dicho contexto); al contrario, se presenta como una de las primeras y más naturales opciones al momento de trabajar con series de tiempo. En este sentido, comparaciones de eficiencia con otros métodos no tienen mucho sentido para este ejemplo, pues no se busca realizar un análisis profundo sobre herramientas de

aprendizaje máquina sino mostrar al lector o lectora cómo se puede aplicar el modelo teórico a un caso práctico concreto.

#Formateo de las series de tiempo para la librería tslearn y ejecución del algoritmo de clasificación GAK.

#Ejemplo. Clustering de 10 elementos para cada instrumento musical y para un descriptor en particular; a saber, el zcr (las series de tiempo re

```
#zcr_formatData = time_seriesFormat(zcr_data)
#gak_km = GlobalAlignmentKernelKMeans(n_clusters=10)
#zcr_labels = gak_km.fit_predict(zcr_formatData)
#zcr_grouped = data_Group(zcr_data,grouping_labels(zcr_labels))
```

```
from tslearn.clustering import GlobalAlignmentKernelKMeans as GAKM
from tslearn.utils import to_time_series_dataset

def time_seriesFormat(series_set):
    out_array = [to_time_series_dataset(i) for i in series_set]
    return out_array

def checkNaN():
    nanArray = []
    for i in input_array:
        dummy = [[x if np.isnan(x)==False else 0 for x in z] for z in i]
        nanArray.append(i)
    return nanArray

def grouping_labels(labels_array):
    max_a = np.max(labels_array)
    min_a = np.min(labels_array)
    range_a = list(range(min_a,max_a+1))
    out = [np.where(labels_array == i)[0] for i in range_a]
    return out

def data_Group(data_array, index_array):
    data_np = np.array(data_array)
    index_st = [np.array(i) for i in index_array]
    data_grouped = [data_np[i] for i in index_st]
    return data_grouped
```

```

def plot3D_array(array_data, index_data):
    fig = plt.figure()
    ax = fig.gca(projection='3d')
    for i,j in zip(index_data, range(len(index_data))):
        ax.plot3D(np.full(len(array_data[i]),j), np.arange(len(array_data[i])),array_data[i])
    
```

Las figuras 22 a la 30 muestran ejemplos de *clusters* obtenidos con el proceso anterior para distintos instrumentos y distintos descriptores.

La clasificación posee un alto grado de precisión aunque, como puede observarse en alguno de los ejemplos, están presentes ciertos elementos que pueden ser considerados *outliers* o mal clasificados. El camino a seguir para esta situación es el estándar dentro de cualquier proceso de aprendizaje máquina: realizar pruebas con distintos algoritmos de clasificación y/o con diferentes condiciones iniciales. Las restricciones para el caso particular de clasificación mediante series de tiempo es que la mayoría de los algoritmos requieren que las series sean de igual longitud. Lo anterior implica que las muestras de audio deberán normalizarse de algún modo, lo que puede derivar en un grado de pérdida o sobreajuste de información.

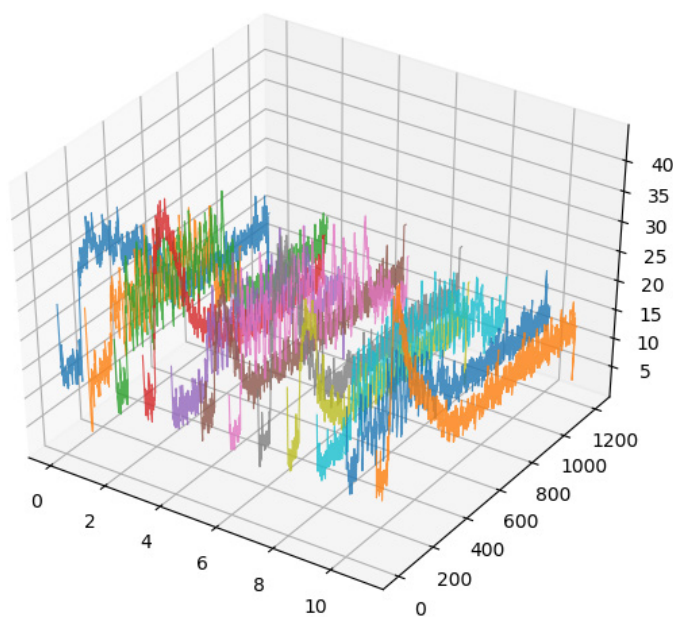


Figura 22 | Cluster no. 2 de series de tiempo para la sub-banda 4 (contraste espectral) de muestras de audio de flauta.

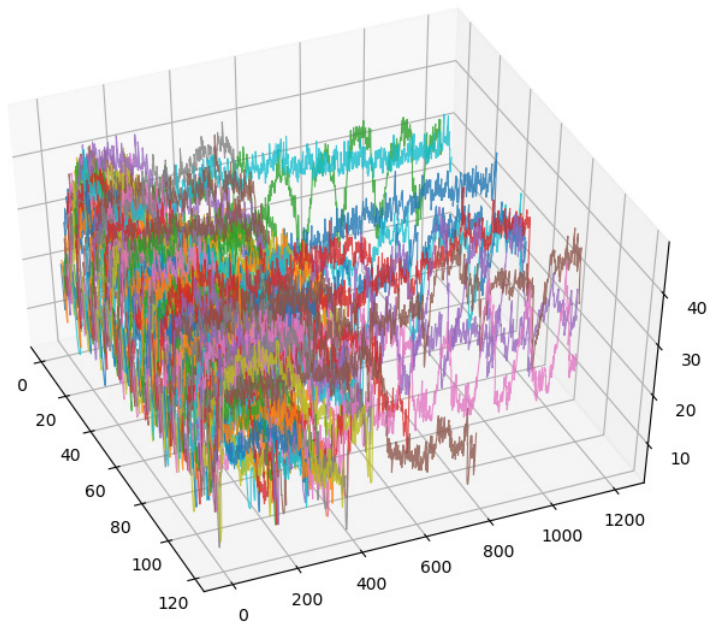


Figura 23 | Cluster no. 3 de series de tiempo para la sub-banda 4 (ancho espectral) de muestras de audio de corno inglés.

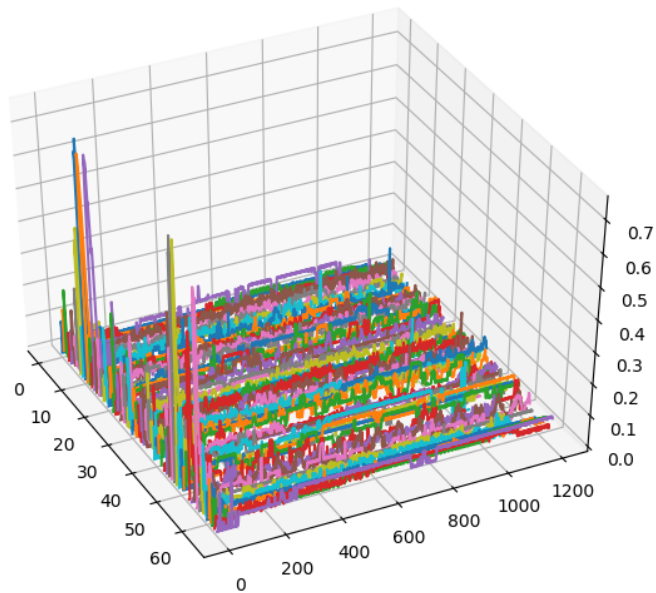


Figura 24 | Cluster no. 3 de series de tiempo para el ZCR de muestras de audio de saxofón.

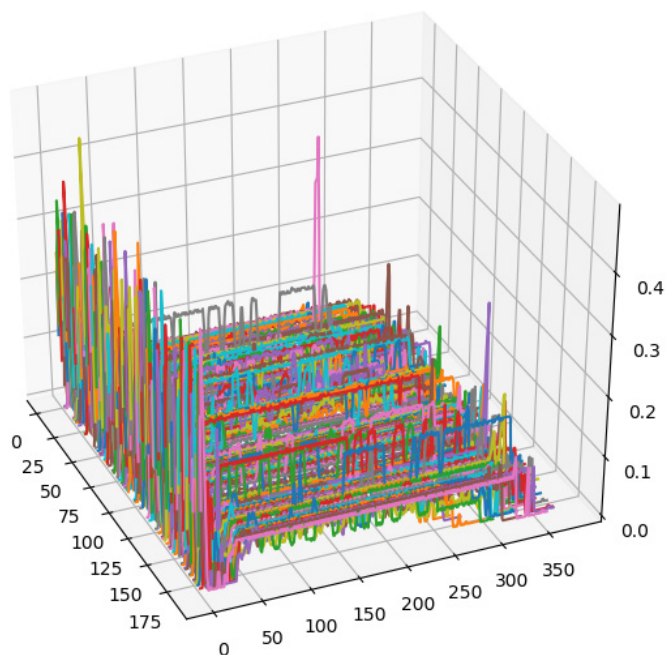


Figura 25 | Cluster no. 0 de series de tiempo para el zcr de muestras de audio de oboe.

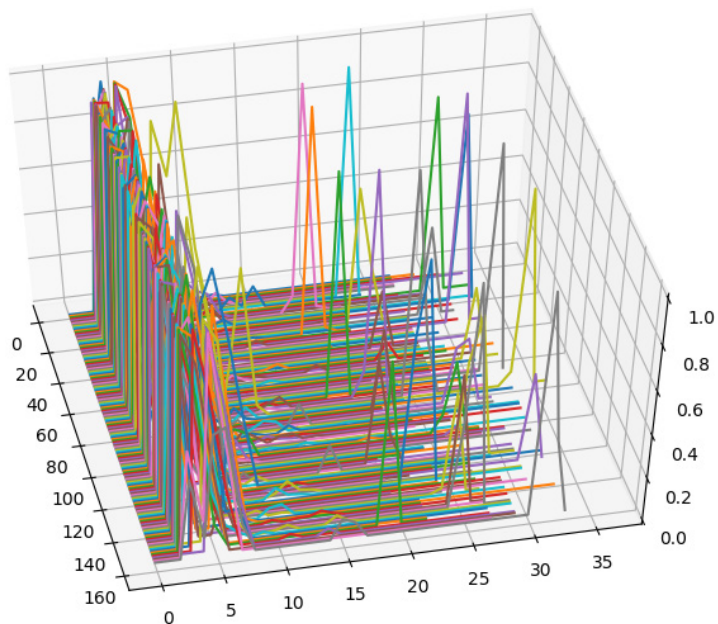


Figura 26 | Cluster no. 5 de series de tiempo para flujo espectral muestras audio de clarinete.

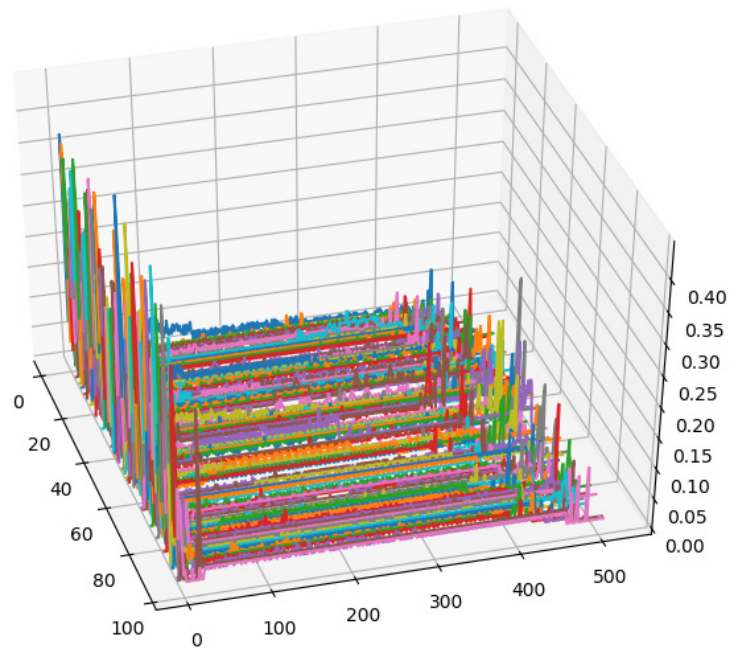


Figura 27 | Cluster no. 5 de series de tiempo para el zcr de muestras de audio de mandolina.

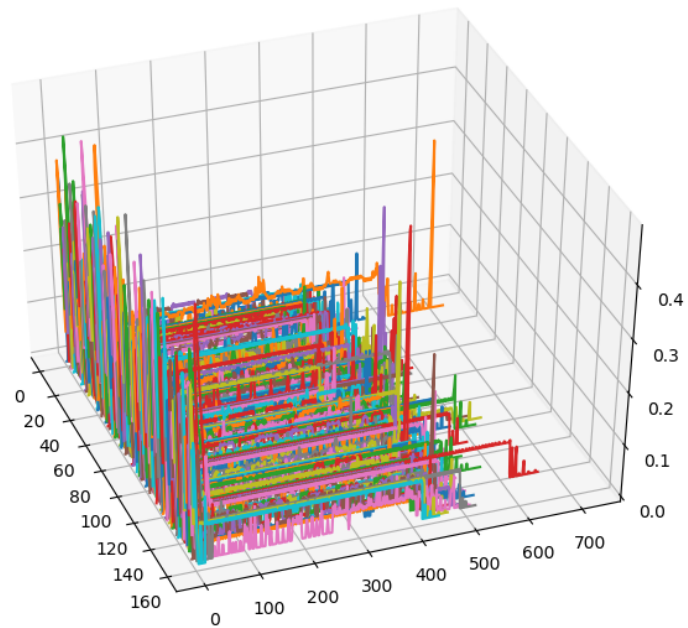


Figura 28 | Cluster no. 3 de series de tiempo para flujo espectral de muestras de audio de guitarra.

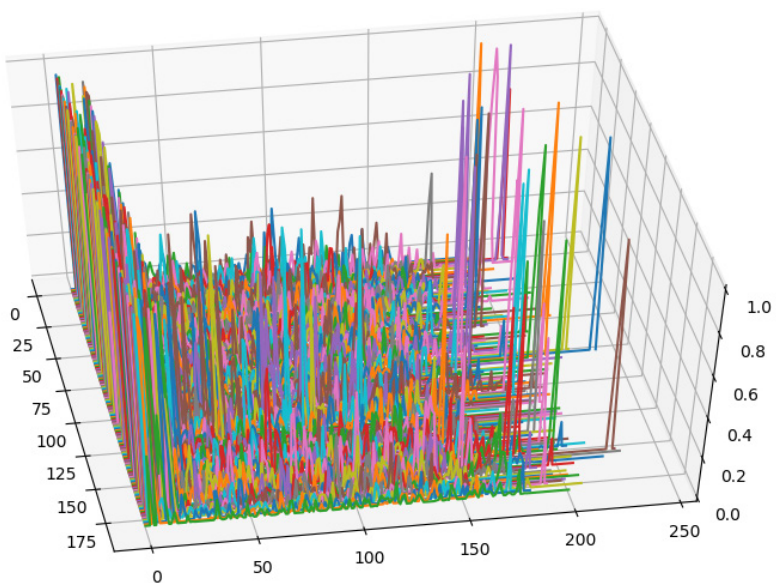


Figura 29 | Cluster no. 2 de series de tiempo para el flujo espectral de muestras de audio de cello.

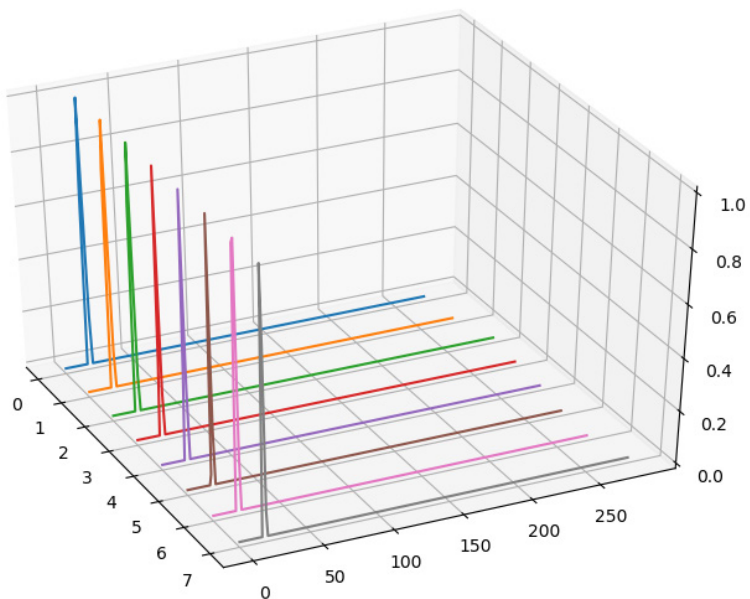


Figura 30 | Cluster no. 0 de series de tiempo para el flujo espectral de muestras de audio de banjo.

Formación de centroides mediante baricentros y bandas LB Keogh

Una vez que se obtuvieron los *clusters* anteriores, en la presente metodología se propone que se definan centroides de referencia para cada *cluster* mediante el cálculo de baricentros estándar euclidianos tal y como está implementado en la librería *tslearn*.¹⁵ Como su nombre lo indica, este método arroja una serie de tiempo que representa el *centro de masa* del *cluster* en cuestión que puede ser considerada como la serie representativa de éste.

En este punto, el propio baricentro podría ser considerado como el centroide de dicho *cluster*; sin embargo, dado que el objetivo final es que se comparen series de tiempo provenientes de muestras de audio arbitrarias, el centroide es definido como un área cuyo eje es el baricentro, mientras que sus límites inferior y superior están dados por el cálculo de las bandas LB Keogh. De este modo, al momento de realizar comparaciones de dicho centroide con muestras de audio arbitrarias haciendo uso de DTW, el cálculo se vuelve más eficiente y preciso. El método completo puede resumirse entonces del siguiente modo:

1. Calcular el baricentro euclideo para un *cluster* específico.
2. Dado el baricentro, calcular las bandas LB Keogh inferior y superior respecto al baricentro.
3. El área delimitada por dichas bandas se considera entonces como el centroide de dicho *cluster*.

A continuación se muestra la rutina en Python escrita para realizar el proceso anterior:

```
def bary_centers(data_array, index_array):
    data_np = np.array(data_array)
    index_st = [np.array(i) for i in index_array]
    bary_centers = [euclidean_barycenter(data_np[i]) for i in index_st]
    #bary_centers = [softdtw_barycenter(data_np[i], gamma=1., max_iter=100) for i in index_st]
    #bary_centers = [dtw_barycenter_averaging(data_np[i], max_iter=150, verbose=False) for i in index_st]
    return bary_centers
```

¹⁵ La librería *tslearn* también permite calcular los baricentros mediante el algoritmo conocido como *DTW barycenter averaging* o *DBA* (Petitjean *et al.*, 2011), el cual es una alternativa a los baricentros euclidianos. Sin embargo, dado que el *clustering* se realizó con alineación por DTW, el uso del método DBA podría resultar redundante en ciertos casos.

```

def lb_envelopes(bary_array, radius_s):
    out = [metrics.lb_envelope(i, radius=radius_s) for i in bary_array]
    return out

def plot3D_ts(data_array,alpha,line,env):
    fig = plt.figure()
    ax = plt.axes(projection='3d')
    env_down = env[0].ravel()
    env_up = env[1].ravel()

    for i,j in zip(data_array,range(len(data_array))):
        x_axis = np.full(len(i),j+1)
        y_axis = list(range(len(i)))
        #ax.plot3D(x_axis,y_axis,i,alpha=alpha,linewidth=line)
        ax.scatter(x_axis,y_axis,i,alpha=alpha,s=line)
        ax.plot3D(np.full(len(env_down),j+1),list(range(len(env_down.ravel()))),env_down,c='blue',alpha=0.5,linewidth=0.8)
        ax.plot3D(np.full(len(env_up),j+1),list(range(len(env_up.ravel()))),env_up,c='red',alpha=0.5,linewidth=0.8)

def plot2D_ts(data,env):
    fig = plt.figure()
    for i in data:
        plt.plot(i,alpha=0.6,linewidth=0.8)
    plt.plot(env[0],c='red',linewidth=1.4)
    plt.plot(env[1],c='blue',linewidth=1.4)

def plot3D_Env(bary,env):
    fig = plt.figure()
    ax = plt.axes(projection='3d')
    for i,j,k in zip(bary,range(len(bary)),env):
        env_down = k[0].ravel()
        env_up = k[1].ravel()
        x_axis = np.full(len(i),j+1)
        y_axis = list(range(len(i)))
        #ax.scatter(x_axis,y_axis,i,s=0.8,alpha=0.7)
        ax.plot3D(x_axis,y_axis,i.ravel())
        ax.plot3D(np.full(len(env_down),j+1),list(range(len(env_down.ravel()))),env_down,c='blue',alpha=0.8,linewidth=0.7)
        ax.plot3D(np.full(len(env_up),j+1),list(range(len(env_up.ravel()))),env_up,c='red',alpha=0.8,linewidth=0.7)

```

La figura 31 muestra un ejemplo de construcción del baricentro dado un *cluster* particular; en este caso para el flujo espectral correspondiente a las muestras de audio del violín.

En las figuras 32 a 39 se muestran ejemplos específicos de *clusters* definidos mediante baricentros euclidianos en conjuntos con bandas LB Keogh superior e inferior.

La ventaja de haber realizado el *clustering* para descriptores individuales y para muestras correspondientes a cada instrumento musical radica en que las muestras arbitrarias subse-

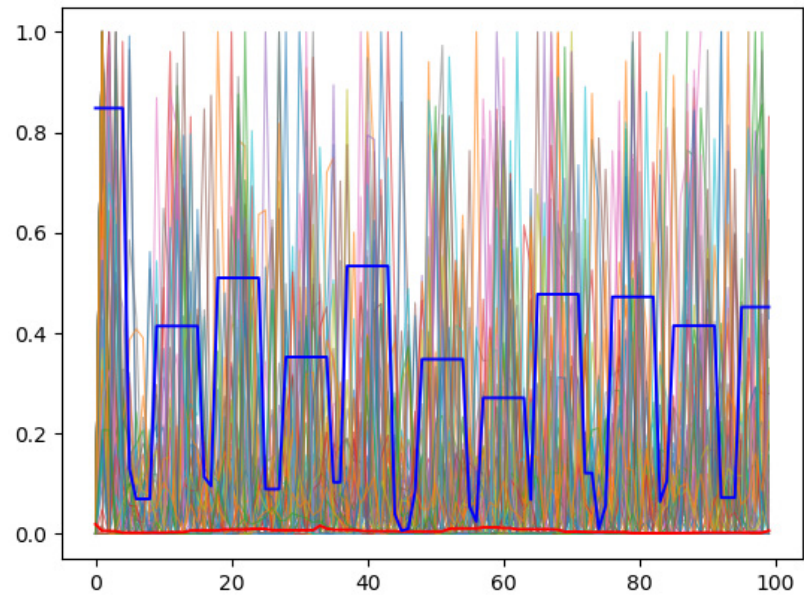


Figura 31 | Baricentro euclidiano para *cluster* de flujo espectral de muestras de audio de violín.

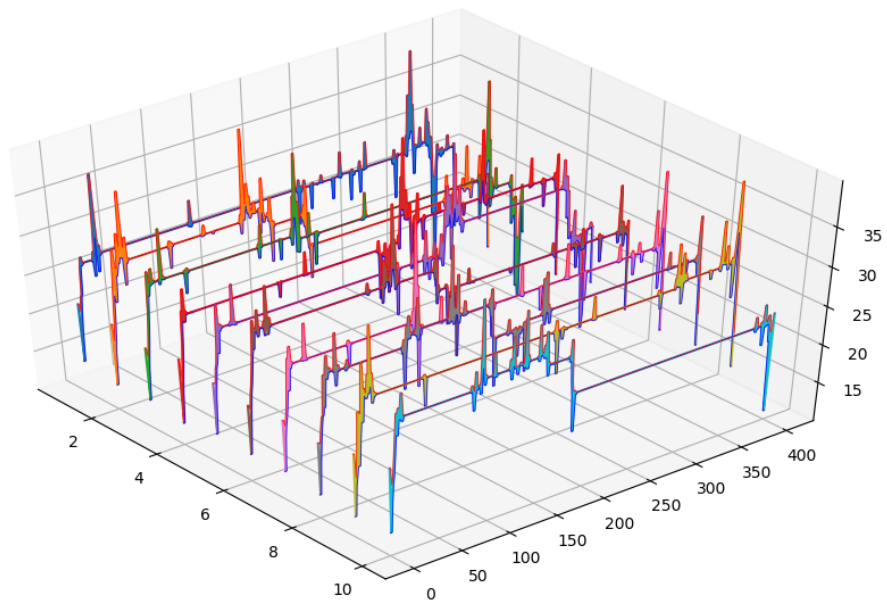


Figura 32 | Centroide baricentro/bandas LB Keogh para sub-banda 2 de ancho espectral de muestras de audio de clarinete.

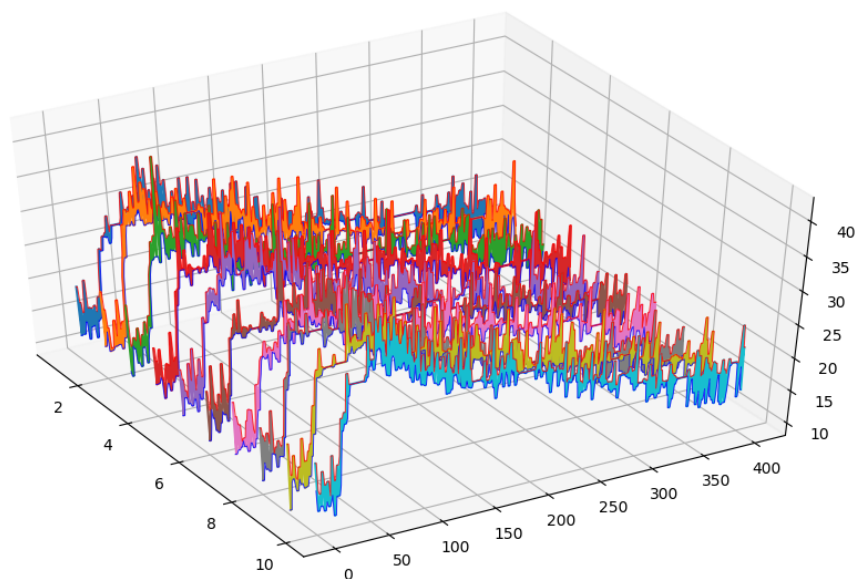


Figura 33 | Centroide baricentro/bandas LB Keogh para sub-banda 2 de ancho espectral de muestras de audio de contrabajo.

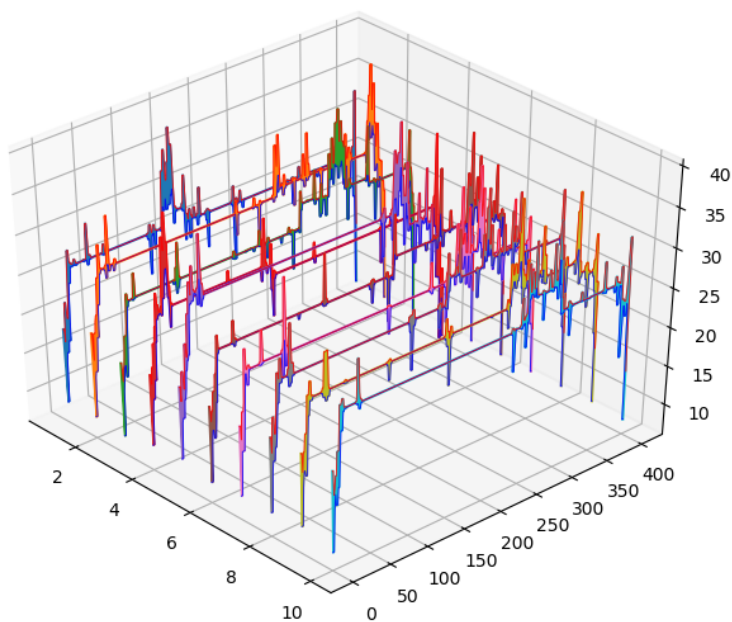


Figura 34 | Centroide baricentro/bandas LB Keogh para sub-banda 2 de ancho espectral de muestras de audio de tuba.

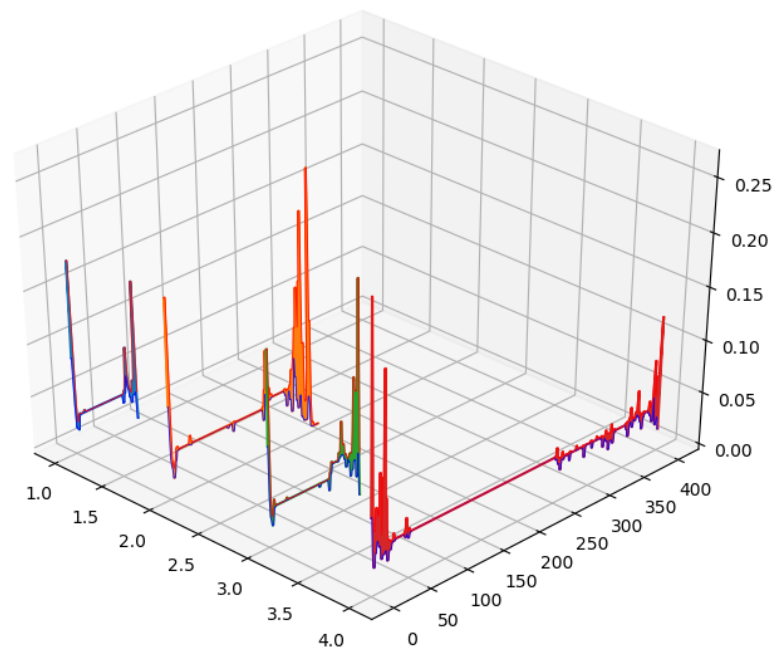


Figura 35 | Centroide baricentro/bandas LB Keogh para ZCR de muestras de audio de trombón.

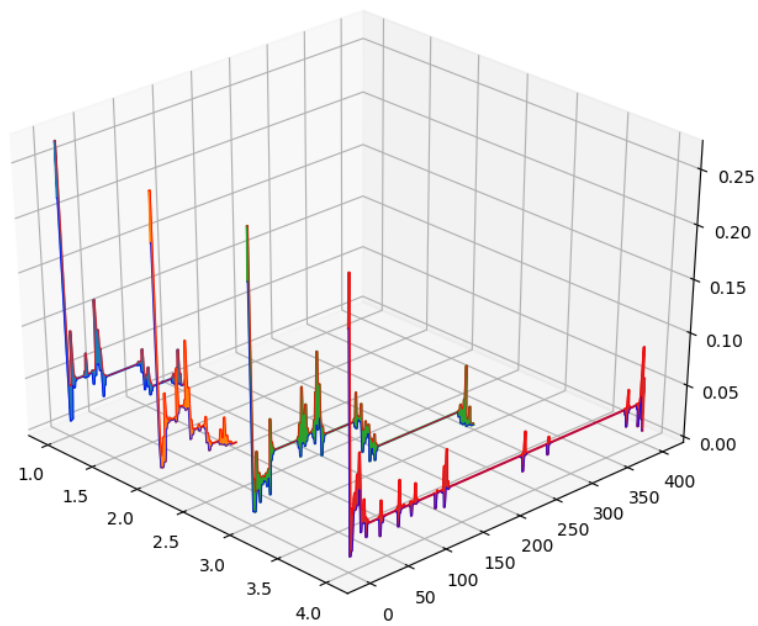


Figura 36 | Centroide baricentro/bandas LB Keogh para ZCR de muestras de audio de tuba.

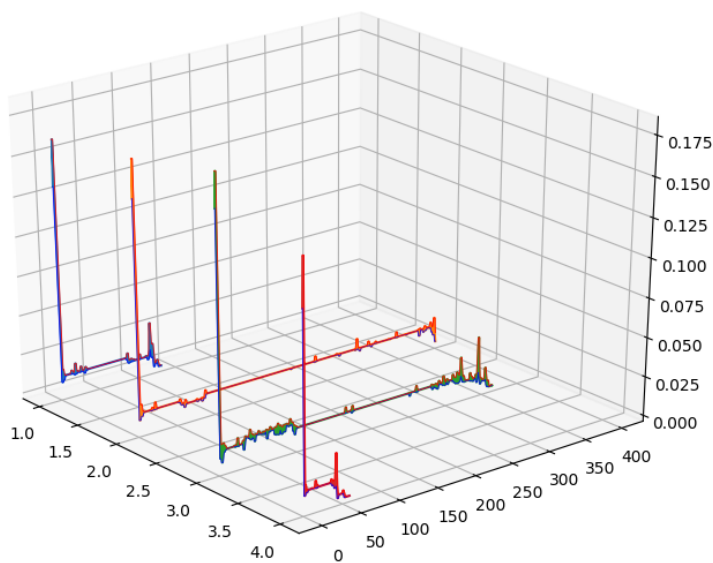


Figura 37 | Centroide baricentro/bandas LB Keogh para ZCR de muestras de audio de trompeta.

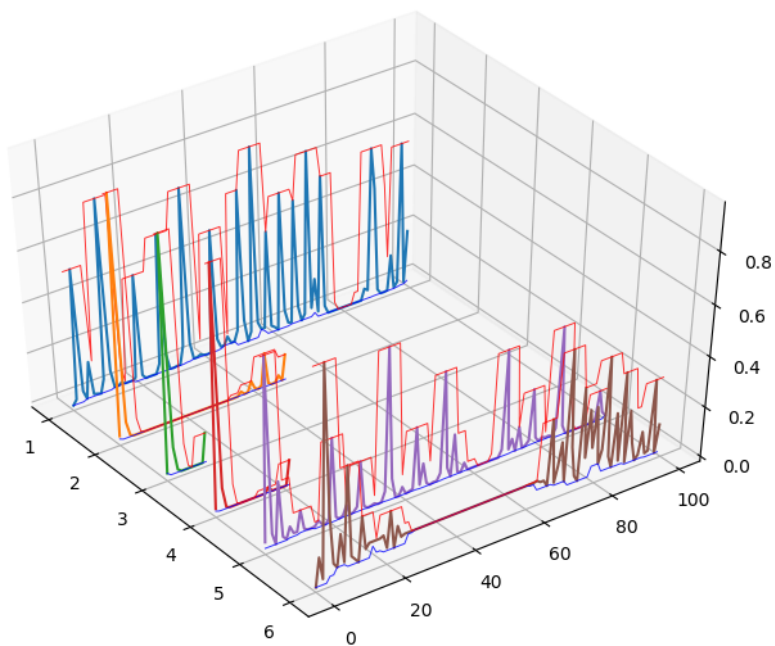


Figura 38 | Centroide baricentro/bandas LB Keogh para flujo espectral de muestras de audio de clarinete bajo.

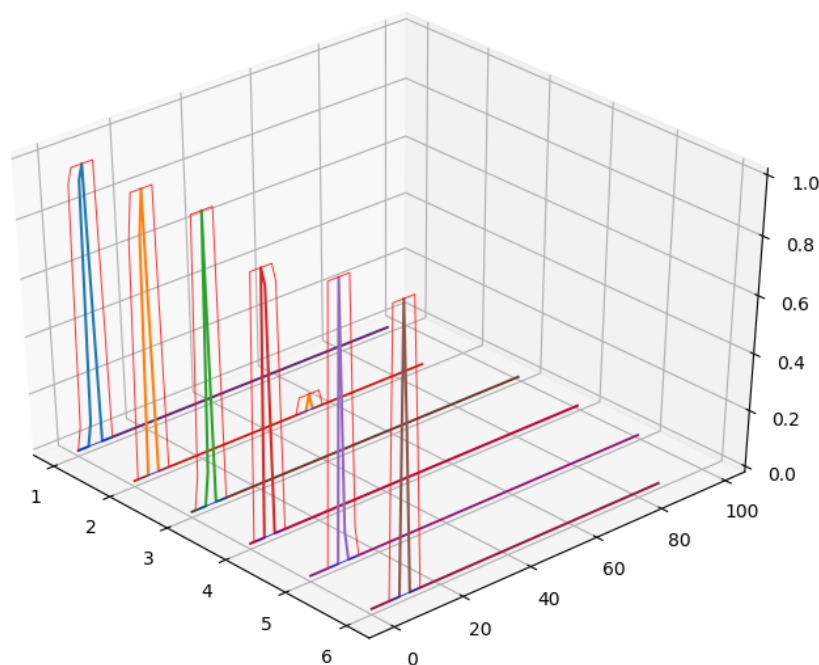


Figura 39 | Centroide baricentro/bandas LB Keogh para flujo espectral de muestras de audio de clarinete banjo.

cuentas pueden ser comparadas de acuerdo, precisamente, con descriptores particulares asociados a muestras de audio de instrumentos musicales conocidos, tocados con una dinámica específica, además del hecho de que esta comparación no está sujeta o delimitada a un conjunto preestablecido de tales descriptores, sino que el espacio espectromorfológico de referencia puede ser construido a partir de una elección arbitraria (justificada claro está, desde el punto de vista metodológico) de cualesquiera descriptores.

Clasificación de muestras de audio arbitrarias

Una vez que se obtuvieron los centroides para cada descriptor, tal y como se describió en la sección anterior, se construye ahora el *espacio espectromorfológico de referencia* y, con éste, una forma de establecer comparaciones específicas con muestras de audio arbitrarias.

El método propuesto para realizar dicha comparación consiste en extraer, para cada muestra de audio arbitraria, un conjunto de descriptores específicos (correspondientes a los obtenidos en el espacio espectromorfológico de referencia), realizar la comparación de similitud

mediante DTW y asociar dicho descriptor de esa muestra al *cluster* correspondiente. Esto permitirá tener un perfil de similaridad de la muestra arbitraria en relación al *cluster* en cuestión. Lo anterior se resume en los siguientes pasos:

1. Cargar la muestra de audio a clasificar junto con el array 2D de centroides (baricentros o bandas LB Keogh).
2. Convertir el array de baricentros a 1D con el fin de realizar la comparación de manera lineal con todos los baricentros (recuérdese que los baricentros están organizados por instrumentos musicales).
3. Calcular el grado de similaridad mediante DTW de la muestra a comparar con cada uno de los centroides.
4. Revisar valores *nan* o *inf* y solucionar posibles problemas de normalización.
5. Obtener *array* de índices derivado del *array* del paso 2 y formar el *array* conjunto del valor de similaridad DTW con dicho índice.
6. Ordenar el *array* anterior de menor a mayor de acuerdo con el valor de similaridad.
7. Obtener el elemento con el valor más pequeño y más grande que representarán los baricentros más similares y más disimilares respectivamente.
8. Repetir para todas las muestras a comparar.

Se muestra a continuación la rutina correspondiente a lo descrito previamente:

```

from tslearn.metrics import dtw as DTW

def dtw_comparison(bary_array, input_data):
    step = len(bary_array[0])
    bary_extend = [i.ravel() for i in bary_array.ravel()]
    bary_dtw = [DTW(input_data, i) for i in bary_extend]
    bary_dtw = [x if np.isnan(x) == False else -100 for x in bary_dtw]
    index_array = [[int(i/step), i%step] for i in range(len(bary_extend))]
    #lowBand_LB = [dtw(input_data, i[0]) for i in lb_array]
    #highBand_LB = [dtw(input_data, i[1]) for i in lb_array]
    joint = [[i, j] for i, j in zip(bary_dtw, index_array)]
    sorted_array = sorted(joint, key=itemgetter(0, 1))
    return [sorted_array[0], sorted_array[-1]]

def plot_comparison(stat_array, zcr_bary, feature_data):
    for element, j in zip(stat_array, feature_data):

```

```

fig = plt.figure()
ax = fig.gca()
bary_x = element[0][1]
bary_y = element[1][1]
plt.plot(zcr_bary[bary_x[0]][bary_x[1]],label='baricentro más cercano')
plt.plot(zcr_bary[bary_y[0]][bary_y[1]],label='baricentro más lejano')
plt.plot(j,label = 'descriptor del audio de prueba')
plt.legend(loc=2, prop={'size': 6})

```

Las figuras 40 a 47 ilustran además, la visualización del proceso anterior para distintas muestras de audio comparadas con distintos descriptores. Se presentan tres series de tiempo: la correspondiente al descriptor de la muestra a comparar, la del baricentro correspondiente más cercano y la del baricentro correspondiente más lejano. Las muestras presentadas corresponden a cuatro objetos sonoros: 1) fricción rápida con vidrio, 2) fricción rápida con madera, 3) *hit* procesado con *pitch* bajo sin reverberación, 4) *hit* procesado con *pitch* medio con reverberación.

Para concluir el proceso, es necesario establecer las relaciones de los baricentros con las muestras comparadas de acuerdo con las características de los audios recordando que las muestras de los instrumentos musicales están etiquetadas con tres parámetros: *a*) instrumento, *b*) nota (o altura) y *c*) dinámica.

El *cluster* correspondiente al baricentro más cercano de la muestra comparada tendrá asociado, por lo tanto, un conjunto de muestras específicas de los instrumentos musicales y, consecuentemente, será posible extraer el etiquetado correspondiente haciendo una minería sencilla de datos. Se propone aquí que para poder realizar un análisis final de toda la metodología expuesta, es necesaria una clasificación sobre dicho etiquetado:

1. Obtener los nombres de las muestras correspondientes al *cluster* del baricentro más cercano dada la muestra a comparar y organizarlos en un arreglo de tipo *string*.
2. Calcular matriz de combinaciones de pares de comparaciones entre entradas del arreglo anterior para la métrica Damerau.¹⁶

¹⁶ Esta métrica es utilizada para medir grados de similaridad entre cadenas de caracteres, la librería utilizada aquí para ese propósito fue *strsimpy* (Zhou Yang Luo, 2019).

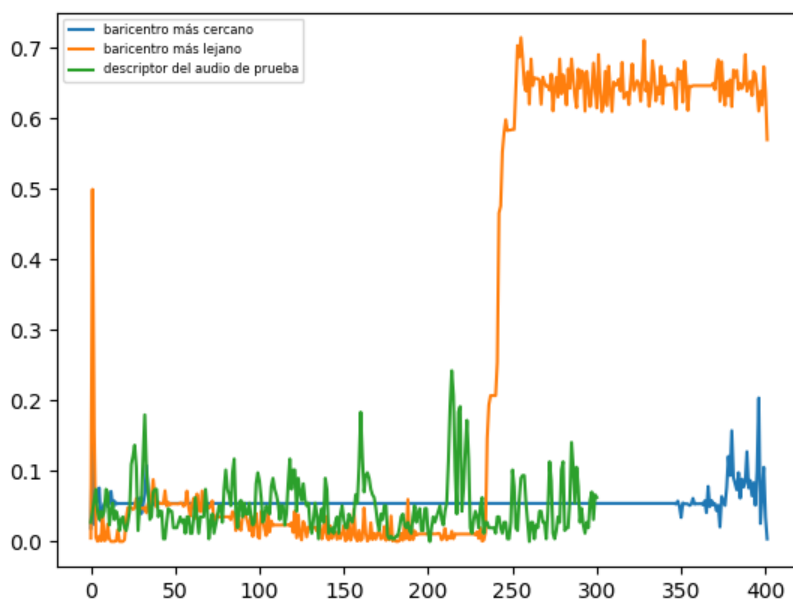


Figura 40 | Comparación de muestra 1 para baricentros de ZCR.

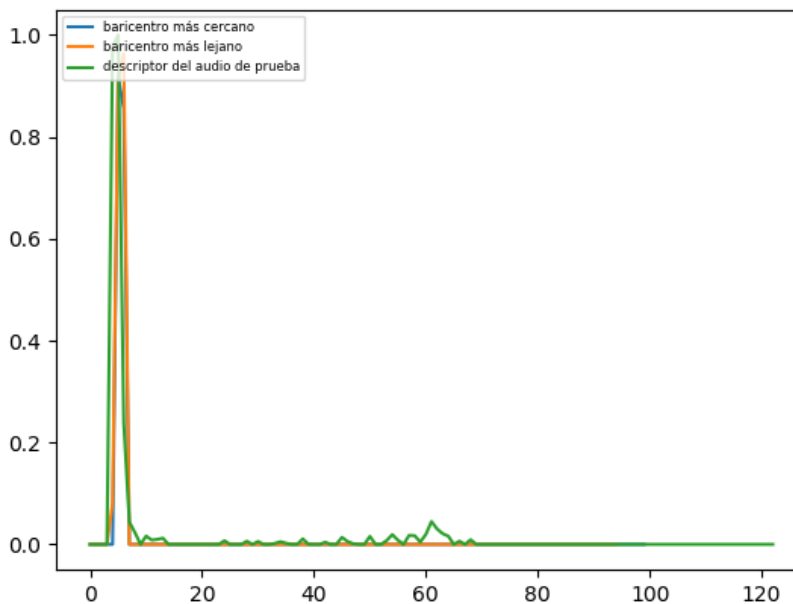


Figura 41 | Comparación de muestra 1 para baricentros de flujo espectral.

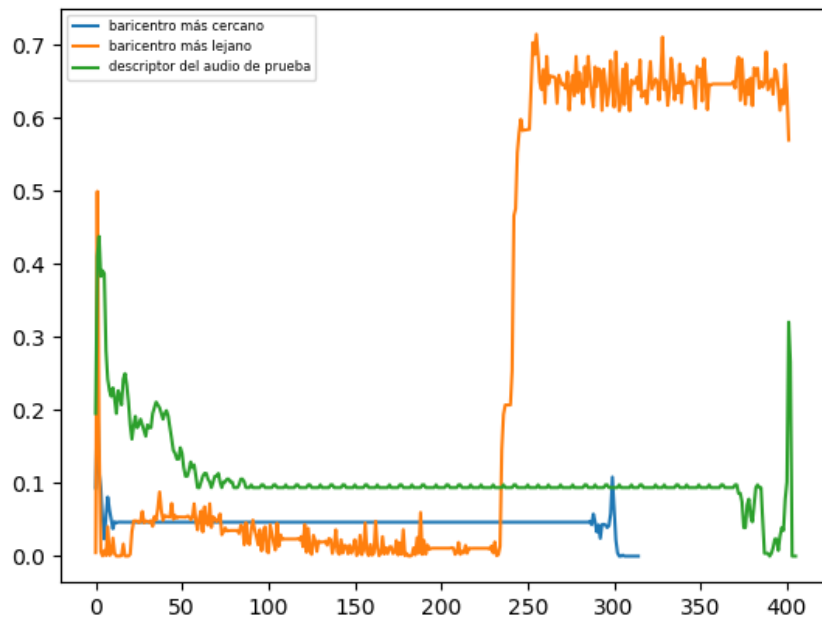


Figura 42 | Comparación de muestra 2 para baricentros de ZCR.

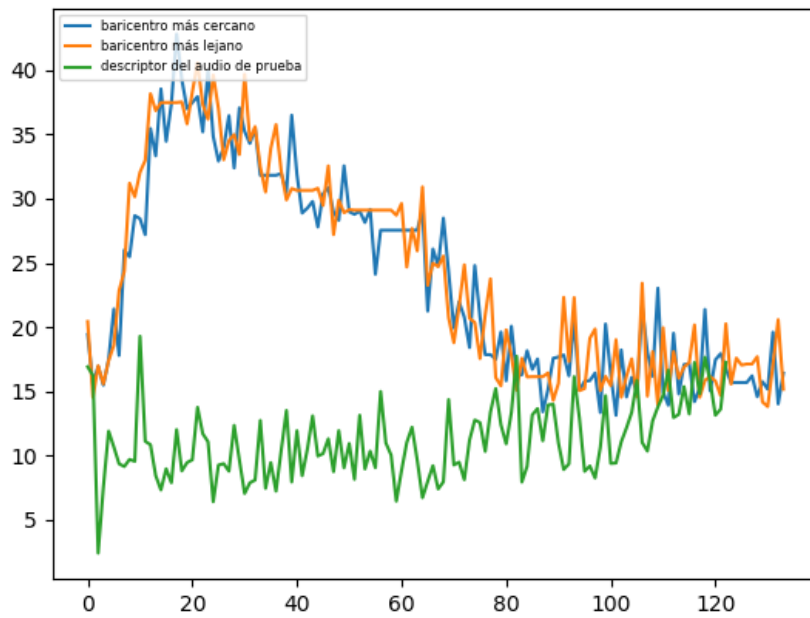


Figura 43 | Comparación de muestra 2 para baricentros de sub-banda 2 de contraste espectral.

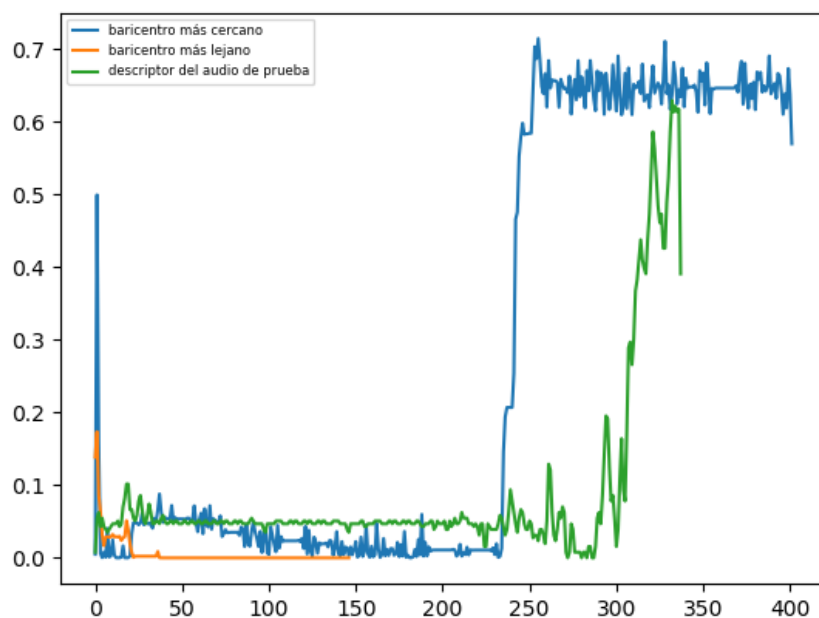


Figura 44 | Comparación de muestra 3 para baricentros de ZCR.

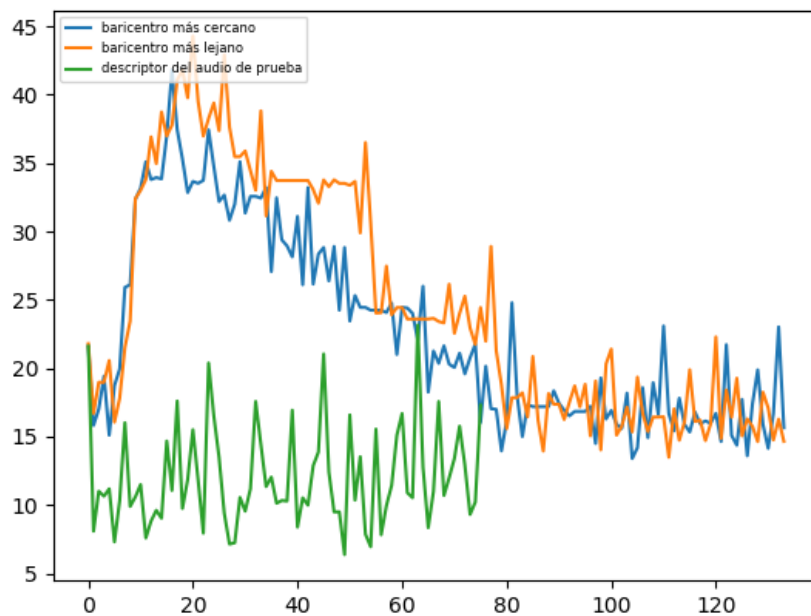


Figura 45 | Comparación de muestra 3 para baricentros de sub-banda 2 de contraste espectral.

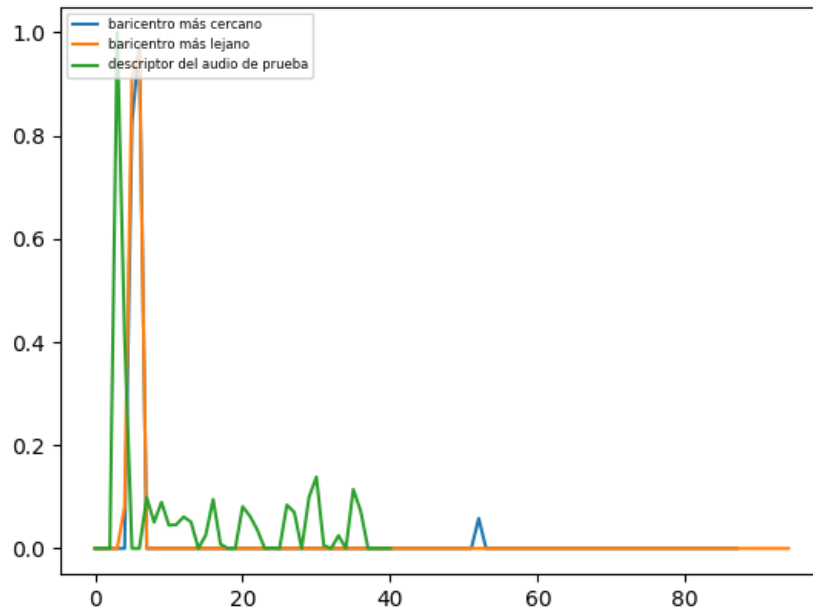


Figura 46 | Comparación de muestra 4 para baricentros de flujo espectral.

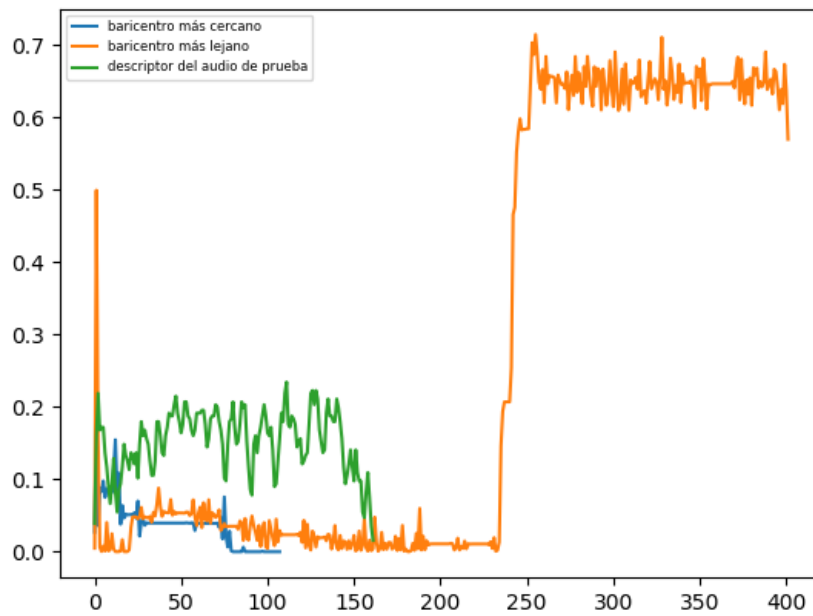


Figura 47 | Comparación de muestra 4 para baricentros de zcr.

3. Obtener *clusters* de clasificación no supervisada mediante el algoritmo de *propagación de afinidad* dada la matriz anterior.
4. Asociar los *clusters* a las etiquetas correspondientes y obtener histograma de los elementos de los *clusters*.
5. Dado el histograma anterior, es posible construir una asociación comparativa de la muestra de audio arbitraria con las muestras de audio de los instrumentos musicales, creando de este modo un perfil espectromorfológico asociado.

A continuación se muestra la rutina correspondiente al proceso anterior:

```

from strsimpy.damerau import Damerau
import sklearn

def string_cluster(string_array):
    damerau = Damerau()
    sim_array = -1*np.array([[damerau.distance(w1,w2) for w1 in string_array] for w2 in string_array])
    affprop = sklearn.cluster.AffinityPropagation(affinity="euclidean", damping=0.5)
    affprop.fit(lev_similarity)
    labels =affprop.labels_
    labels_x = np.unique(labels)
    group_data = data_Group(string_array,grouping_labels(labels))
    centers = [string_array[affprop.cluster_centers_indices_[i]] for i in labels_x]
    return group_data, centers

def hist_names(x_input,centers):
    z_temp = [np.full(len(x_input[i]),i) for i in range(len(x_input))]
    plt.hist(z)
    plt.legend(labels=centers,loc=2, prop={'size':6})
    
```

Las figuras 48 y 49 muestran la forma de onda y el espectrograma de la muestra comparada 1, la cual se utilizará como ejemplo único para finalizar esta sección.

La figura 50 muestra el *cluster* correspondiente al baricentro más similar para dicha muestra para el ZCR. La figura 51 muestra el mismo *cluster* después de hacer la clasificación mediante la métrica Damerau y el algoritmo de propagación de afinidad.

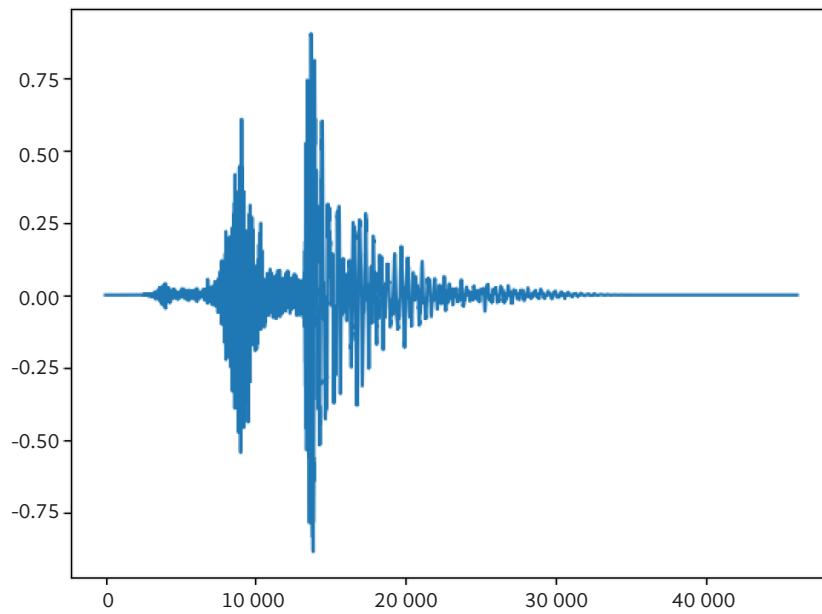


Figura 48 | Forma de onda de la muestra 1.

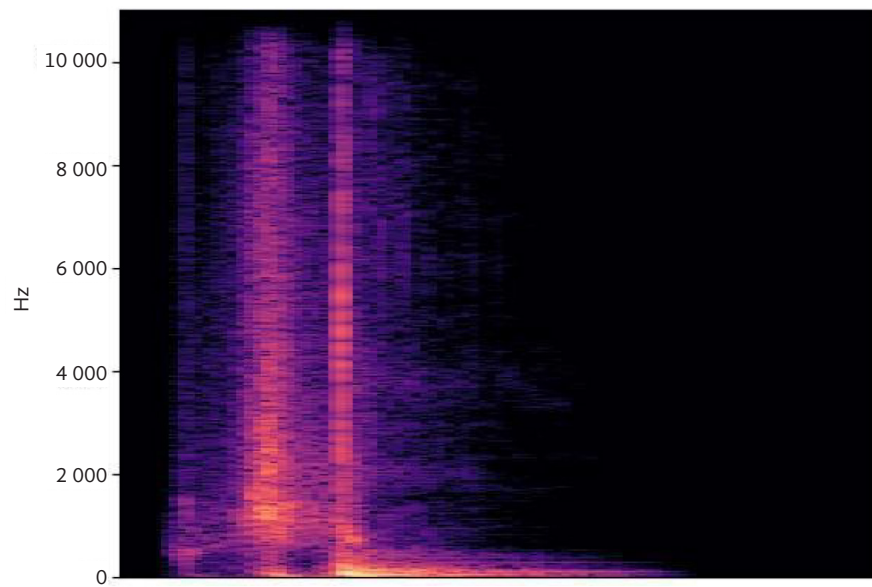


Figura 49 | Espectrograma de la muestra 1.

```
['bass-clarinet_Gs3_025_forte_normal',
 'bass-clarinet_C3_15_mezzo-forte_normal',
 'bass-clarinet_Gs4_1_fortissimo_normal',
 'bass-clarinet_A3_15_forte_normal',
 'bass-clarinet_Ds4_1_mezzo-forte_normal',
 'bass-clarinet_B2_1_piano_normal',
 'bass-clarinet_F4_phrase_forte_major-trill',
 'bass-clarinet_Gs2_1_pianissimo_normal',
 'bass-clarinet_Fs4_05_piano_normal',
 'bass-clarinet_G2_1_fortissimo_normal',
 'bass-clarinet_Fs5_05_pianissimo_normal',
 'bass-clarinet_C6_very-long_forte_normal',
 'bass-clarinet_E4_15_forte_normal',
 'bass-clarinet_Ds5_05_pianissimo_normal',
 'bass-clarinet_B4_05_forte_normal',
 'bass-clarinet_As4_025_pianissimo_normal',
 'bass-clarinet_Ds3_1_piano_normal',
 'bass-clarinet_A2_1_forte_normal',
 'bass-clarinet_E2_025_mezzo-forte_normal',
 'bass-clarinet_A3_05_fortissimo_normal']
```

Figura 50 | Cluster correspondiente al baricentro más cercano en ZCR de la muestra 1.

```
[array(['bass-clarinet_C3_15_mezzo-forte_normal',
 'bass-clarinet_Ds4_1_mezzo-forte_normal',
 'bass-clarinet_C6_very-long_forte_normal',
 'bass-clarinet_E2_025_mezzo-forte_normal'], dtype='<U41'),
 array(['bass-clarinet_F4_phrase_forte_major-trill'], dtype='<U41'),
 array(['bass-clarinet_Gs4_1_fortissimo_normal',
 'bass-clarinet_G2_1_fortissimo_normal',
 'bass-clarinet_A3_05_fortissimo_normal'], dtype='<U41'),
 array(['bass-clarinet_Gs3_025_forte_normal',
 'bass-clarinet_A3_15_forte_normal',
 'bass-clarinet_E4_15_forte_normal',
 'bass-clarinet_B4_05_forte_normal',
 'bass-clarinet_A2_1_forte_normal'], dtype='<U41'),
 array(['bass-clarinet_Gs2_1_pianissimo_normal',
 'bass-clarinet_Fs5_05_pianissimo_normal',
 'bass-clarinet_Ds5_05_pianissimo_normal',
 'bass-clarinet_As4_025_pianissimo_normal'], dtype='<U41'),
 array(['bass-clarinet_B2_1_piano_normal',
 'bass-clarinet_Fs4_05_piano_normal',
 'bass-clarinet_Ds3_1_piano_normal'], dtype='<U41')],
 ['bass-clarinet_C3_15_mezzo-forte_normal',
 'bass-clarinet_F4_phrase_forte_major-trill',
 'bass-clarinet_G2_1_fortissimo_normal',
 'bass-clarinet_E4_15_forte_normal',
 'bass-clarinet_Ds5_05_pianissimo_normal',
 'bass-clarinet_Ds3_1_piano_normal']])
```

Figura 51 | Cluster clasificado interiormente mediante propagación de afinidad, correspondiente al baricentro más cercano en ZCR de la muestra 1.

Finalmente, la figura 52 muestra el histograma correspondiente a dicho *cluster* clasificado de manera interna. Los *subclusters* del anterior están representados a su vez por un centroide que es el que se muestra en el histograma y que es el que se puede considerar como el representativo de dicho *subcluster*. En este sentido, es posible interpretar dicho histograma de la siguiente forma:

- 25 % del *cluster* está representado por el perfil espectromorfológico dado por el clarinete bajo tocado en E4 con una dinámica de *forte normal*.
- 20 % del *cluster* está representado por el perfil espectromorfológico dado por el clarinete bajo tocado en C3 con una dinámica *mezzo forte normal*.
- 20 % del *cluster* está representado por el perfil espectromorfológico dado por el clarinete bajo tocado en Ds5 con una dinámica *pianissimo normal*.

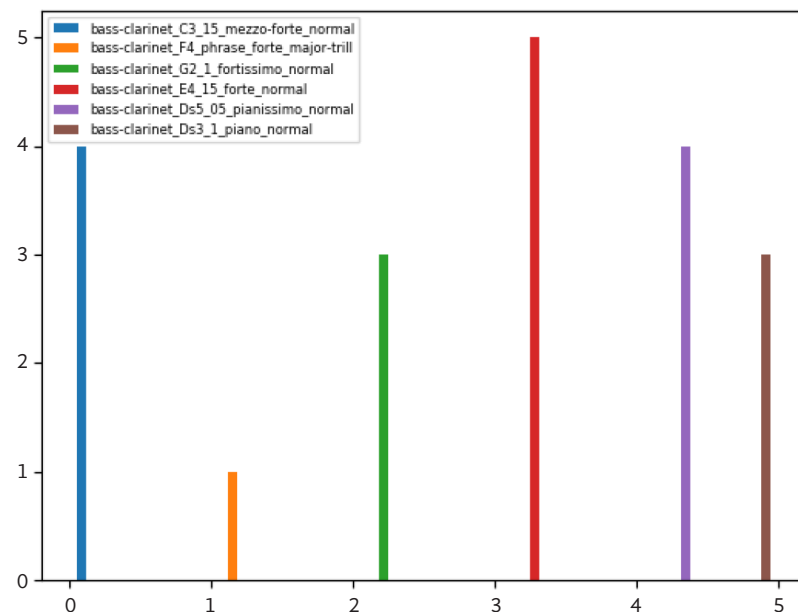


Figura 52 | Histograma de representación de la muestra 1 para el baricentro más cercano en ZCR.

Nótese que, en este caso, los *subclusters* del etiquetado tuvieron preponderancia (al momento de hacer la agrupación) sobre la dinámica. Una manera alternativa de solucionar lo anterior sería obtener descriptores de las cadenas de caracteres como longitud de la cadena y

número de ocurrencias de ciertos símbolos para realizar una clasificación no supervisada sobre ese conjunto de descriptores.

Se reitera el hecho de que la metodología anterior propone establecer un espacio espectromorfológico de referencia para comparar muestras de audio arbitrarias con perfiles espectromorfológicos de instrumentos musicales conocidos, permitiendo extrapolar aspectos de orquestación e información acústica conocida de dichos instrumentos a las muestras de audio arbitrarias. Por supuesto, podría establecerse una metodología más familiar, como la introducción del conjunto de muestras de audio y realizar clasificación mediante algún algoritmo de aprendizaje no supervisado dado un conjunto de descriptores específicos. Se exponen los resultados de dicho acercamiento utilizando los mismos algoritmos que en la sección anterior:

1. Cargar las muestras de audio.
2. Calcular para cada muestra, el conjunto de descriptores: ZCR, RMSE y flujo espectral.
3. Organizar la información en series de tiempo.
4. Convertir las series de tiempo al formato requerido por *tslearn*.
5. Obtener *clusters* de clasificación para cada descriptor mediante DTW GAK.
6. Para cada descriptor, agrupar las muestras de acuerdo con su *cluster* asignado.

Observaciones finales

El ejemplo descrito presentó una metodología de clasificación de muestras de audio basada en el análisis de perfiles espectromorfológicos derivados de ciertos descriptores mediante series de tiempo. A diferencia de la mayoría de los procedimientos de clasificación automática de audio (los cuales utilizan medidas de tendencia central), en esta metodología se propuso realizar el proceso mediante la comparación de series de tiempo utilizando el algoritmo DTW GAK. Derivado de lo anterior es posible enlistar las siguientes observaciones.

OBSERVACIÓN 4.4.1. El costo computacional de realizar el análisis a través de series de tiempo en comparación con caracterizaciones numéricas como la media o la desviación estándar es significativamente más elevado; sin embargo, dicho costo se justifica con el hecho de que es posible comparar muestras de audio de distinta longitud (debido a la ali-

neación hecha por DTW) y con un mayor nivel de precisión (ya que la comparación de similaridad se hace directamente de la serie de tiempo).

OBSERVACIÓN 4.4.2. Los resultados de clasificación utilizando DTW GAK presentaron una precisión significativamente mayor (tanto para las muestras de audio de los instrumentos musicales convencionales como para las muestras de audio arbitrarias) que aquellos obtenidos mediante caracterizaciones numéricas como la entropía aproximada o el coeficiente de variación.

OBSERVACIÓN 4.4.3. Los *outliers* que ocurren utilizando DTW GAK muestran una fuerte correlación con la diferencia de las longitudes de series de tiempo. Estos *outliers* ocurren para diferencias de longitudes de series de tiempo del doble o más.

OBSERVACIÓN 4.4.4. La metodología propuesta, basada en el establecimiento de centroides de clasificación a partir de las bandas LB Keogh y los baricentros euclidianos, no ha sido abordada de este modo en la literatura revisada al respecto de música contemporánea y música electroacústica; representa, por lo tanto, una propuesta del presente trabajo. Las ventajas de establecer los centroides de este modo son:

- Dado un *cluster*, el baricentro euclidiano funciona eficientemente como una serie de tiempo representativa de dicho *cluster*.
- Las bandas superior e inferior LB Keogh funcionan como límites de tolerancia para la comparación de muestras arbitrarias y permiten establecer criterios robustos de comparación de similaridad mediante DTW. Esto es, la comparación de la muestra mediante DTW puede hacerse respecto del baricentro o a alguna de las bandas LB Keogh, dependiendo del contexto específico, y con la seguridad de que el área establecida por estos tres componentes es la más óptima como centroide representativo del *cluster* específico.
- Los centroides pueden ser utilizados posteriormente como etiquetas para ser implementadas en procesos subsecuentes de aprendizaje supervisado.

OBSERVACIÓN 4.4.5. Es posible extender el método aquí propuesto y utilizar otros algoritmos de clasificación (manteniendo el paradigma de la comparación de series de tiempo) que permitan explorar mejoras en la precisión de la formación de los *clusters* tales como máquinas de soporte vectorial, árboles de decisión o algunos tipos de redes neuronales (en-caminadas al aprendizaje profundo). Solamente se debe tener especial cuidado en identi-

car los métodos que permitan analizar series de tiempo de longitudes distintas o, en su caso, normalizar todas las series de tiempo a una longitud fija.

OBSERVACIÓN 4.4.6. Al establecer clasificaciones de muestras de audio tal y como se expuso en el presente capítulo, ya sea en relación con un espacio espectromorfológico de referencia o de manera directa, es posible hablar de *orquespectración virtual* en el sentido de la organización espacial y espectromorfológica de los objetos sonoros (en un contexto compositivo) de acuerdo con su perfil espectromorfológico obtenido computacionalmente y tomando en cuenta los principios y características asociadas a dicho objeto sonoro desde la parte perceptual y acústica.

5. Aplicaciones computacionales actuales sobre espacio y timbre

Las aplicaciones de la inteligencia artificial al campo del sonido y la música han sido generalmente planteadas como metodologías asociadas a lo que se conoce como ingeniería de descriptores (*feature engineering*); a través de ella, las características o parámetros sonoros de bajo/alto nivel vinculados a la data de entrada (ZCR, RMSE, contraste espectral, MFCC, etcétera), son calculados de manera manual por el usuario para después alimentar con esta información algún algoritmo de clasificación supervisada o no supervisada. Esta perspectiva, asociada a las técnicas de aprendizaje automático denominado “tradicional”, ha rendido bastantes buenos resultados y ha permitido establecer metodologías alternativas en pos de una mejor eficiencia tanto de cálculo computacional como de precisión de resultados.

Entre estas alternativas actuales resaltan diversas técnicas de lo que se conoce como aprendizaje profundo (*deep learning*), tanto para aprendizaje no supervisado y aprendizaje supervisado, como para aprendizaje híbrido. Al día de hoy las técnicas de aprendizaje profundo que han sido desarrolladas con mayor énfasis para aplicaciones diversas incluyen (para modelos discriminativos profundos) las redes neuronales profundas (Deep Neural Networks, DNN), las redes neuronales de convolución (Convolutional Neural Networks, CNN) o las redes neuronales recurrentes (Recurrent Neural Networks, RNN). Por su parte, para los modelos generativos/no supervisados es posible destacar las máquinas restringidas de Boltzmann (Restricted Boltzmann Machine, RBM), las redes neuronales de creencia profundas (Deep Belief Networks, DBN), las máquinas profundas de Boltzmann (Deep Boltzmann Machines, DBM), los

autocodificadores en sus distintas versiones (*autoencoders*) o los modelos ocultos/jerárquicos de Markov (Hierarchical/Hidden Markov Models, HMM/HHMM).

Además, en los últimos años algunos autores han realizado diversos esfuerzos por establecer lo que se denomina *análisis matemático del aprendizaje profundo* (Berner *et al.*, 2021), el cual implica el uso de teoría matemática para formalizar resultados, definiciones y conceptos propios del aprendizaje profundo y de la denominada *teoría del aprendizaje*, a través de teoría matemática relacionada con la topología, la teoría de la medida, el análisis funcional y el análisis convexo. Asimismo, destaca el aprendizaje máquina en general y el reciente enlace interdisciplinar con la teoría matemática de categorías utilizada sobre todo como una propuesta de lenguaje unificador y de formalización.

Apenas en junio de este año, el equipo de Magenta, encargado de la parte de DSP (Digital Signal Processing) liberó su VST denominado: DDSP-VST: Neural Audio Synthesis for All, el cual supone el desarrollo de una herramienta computacional específica para procesamiento de audio en tiempo real basado en lo que se ha denominado *síntesis neuronal*. Este tipo de procesamiento fue denominada DDSP, y “combina la eficiencia y la interpretabilidad de los elementos clásicos del DSP (filtros, osciladores, reverberadores, etc.) con la expresividad del aprendizaje profundo” (Jurek *et al.*, 2022). El proceso general de esta técnica (la cual está optimizada para audio en tiempo real y de baja latencia) se compone de tres fases generales: *a*) extracción de descriptores, *b*) controles de predicción DSP y *c*) síntesis. La primera etapa hace uso de dos parámetros básicos, el RSM (como representación del volumen) y la altura (obtenida a través de una red neuronal previa entrenada específicamente para ello). La segunda etapa utiliza dos tipos de sintetizador para cada altura y amplitud que recibe: del tipo aditivo armónico y del tipo sustractivo de ruido.

Pons *et al.* (2017) propusieron adaptar un modelo CNN para realizar aprendizaje automático del timbre sonoro a partir de las representaciones de espectrogramas (para escalas logarítmicas y mel). En ese trabajo los autores plantean una metodología para extraer los contextos relevantes en el dominio tiempo-frecuencia y con ello realizar aprendizaje automático del timbre implementando técnicas de *domain-knowledge*. Obtienen resultados aceptables para casos que incluyen clasificación de voz y fonemas, reconocimiento de instrumentos musicales y *auto-tagging* musical.

Gabrielli *et al.* (2019) propusieron un autocodificador como propuesta de técnicas de hibridación en el contexto de síntesis de audio basado en aprendizaje de propiedades sonoras de

dos fuentes concretas, a partir de una exploración de las dependencias temporales de algunos descriptores semánticamente significativos. Su arquitectura logró modificar el timbre original (de las muestras base) manteniendo atributos esenciales (aprendidos en la etapa de entrenamiento) y preservando la envolvente de altura original.

Otro avance notable en el campo del aprendizaje profundo que puede tener aplicaciones potenciales importantes en el campo de la música electroacústica multicanal es el trabajo de Francl y Dermott (2022) al respecto de la localización espacial del sonido, con el objetivo de estudiar los procesos por medio de los cuales la percepción se adapta a entornos y ambientes reales. En este trabajo, los autores desarrollaron simuladores virtuales acústicos mediante BRIR (*binaural room impulse responses*), con el fin de entrenar distintas arquitecturas de redes neuronales implementadas con características propias de HRTF, pero con una especificidad particular (al respecto del entrenamiento de las redes): para cada uno de los aspectos generales de la localización espacial del sonido, como el preprocesamiento del estímulo mediante un modelo coclear, la integración de fuentes sonoras naturales con fuentes sonoras que corresponden a ruido de fondo, y la simulación de ambientes no naturales a partir de características como entornos anecoicos y fuentes sonoras no naturales.

En esta misma línea, Pujol *et al.* (2021) presentaron una arquitectura de aprendizaje profundo denominada *BeamLearning*, la cual es capaz de realizar localización espacial del sonido enfocada en lo angular (al respecto del azimut), a partir de *data* de presión sonora acústica multicanal. De acuerdo con los autores, esta perspectiva permite codificar información relevante contenida en las señales acústicas no procesadas captadas por arreglos de micrófonos; a partir de los resultados obtenidos, los beneficios más notables están orientados a tareas de localización de audio 2D en tiempo real para ambientes reverberantes y con ruido considerable.

Finalmente, aunque un poco menos reciente, resalta el trabajo de Gao y Grauman (2019), quienes desarrollaron un sistema de aprendizaje profundo (basado en redes de convolución) capaz de crear en el individuo la sensación de tridimensionalidad de un sonido que originalmente es monoaural, mediante el uso asistido de un video. De acuerdo con el autor, este sistema puede aprender a decodificar una señal monoaural en su contraparte binaural mediante la conjugación de la información visual del objeto y la información de la escena auditiva. Esto permite (a partir de la información visual) que el sistema pueda deducir la dirección de los sonidos y modificar las correspondientes diferencias interaurales en tiempo y amplitud para generar el efecto final de audio 3D. ■■■

Conclusiones

A

l establecer que el *Espacio-Timbre*, como relación inherente entre la espectromorfología y la localización espacial de los objetos sonoros, puede considerarse como un parámetro por sí mismo, se propuso un modelo teórico original a partir del cual se construyó y definió el concepto anterior desde tres perspectivas específicas: la perceptual, la acústica y la teórica.

El Espacio-Timbre como modelo se constituyó de dos elementos: la *espacialidad* y el *timbre*. Ambos se plantean, a su vez, como integrados por dos ejes fundamentales: la parte perceptual y la parte acústica. En este sentido, el capítulo I fundamentó estos dos aspectos sintetizando conceptos y herramientas teóricas específicas, así como resultados de pruebas experimentales relacionadas con la percepción sensorial.

Esta construcción dual (perceptual-acústica) quedó establecida en las formulaciones de los perfiles y espacios morfológicos y espectrales, los cuales a su vez formaron el perfil y espacio espectromorfológico de acuerdo con lo desarrollado en el capítulo IV. El modelo quedó posteriormente formalizado mediante el lenguaje de sistemas dinámicos haciendo uso de la propiedad universal del producto de conjuntos.

Debe entenderse este modelo, por lo tanto, como una construcción teórica que formula el Espacio-Timbre desde la parte acústica y la parte perceptual para luego inscribir dicho concepto dentro de una abstracción matemática basada en sistemas dinámicos. De este modo, es posible destacar las siguientes observaciones al respecto de lo desarrollado en esta obra:

- Se realizó una investigación puntual de los antecedentes y principios generales sobre percepción, psicoacústica y acústica, tales que permitieron desarrollar y fundamentar la construcción de los conceptos de *espectromorfología* y *espacialidad*.
- Se propuso una definición original de *espacialidad* a través de la idea de virtualidad y experiencia estética.
- Se estableció el concepto de *disimilitud serial-perceptual*.
- Se definió, desde una mirada particular, los *componentes primarios de la espacialidad*.
- Se construyeron teóricamente, de acuerdo con lo anterior, los conceptos de perfil y espacio morfológico, espectral y espectromorfológico.
- Se definió la forma de construcción métrica del espacio espectromorfológico.
- Se fundamentó el concepto de Espacio-Timbre con elementos teóricos abstractos bajo un modelo matemático basado en sistemas dinámicos, con el fin de establecer una estructura teórica de organización y síntesis de conceptos lo más general posible.
- Como consecuencia de los apartados previos, se estableció la propuesta teórica de que el timbre y el espacio pueden considerarse como un sólo parámetro sonoro dentro del contexto de la composición de la música electroacústica: el *Espacio-Timbre*.
- Se desarrolló una metodología computacional capaz de realizar análisis del perfil espectromorfológico de muestras de audio en series de tiempo, utilizando algoritmos de aprendizaje no supervisado para fines de clasificación y comparación.
- Se estableció el concepto de *orquespectración virtual*.

Este modelo presenta una mirada alternativa al proceso de comprensión y análisis de la música electroacústica, ya que plantea emancipar de forma estructurada y sistematizada al espacio como un parámetro primordial e indisoluble del timbre; además, establece las bases teóricas para lograr ese objetivo no sólo como una mera abstracción sino mediante un resultado práctico. En este punto se comprende la importancia de los referentes de percepción y psicoacústica desarrollados al principio de este trabajo de investigación.

La idea de la música electroacústica dentro de este modelo tiene un matiz primordial: el de la generación de una experiencia en el espectador de un espacio virtual inmersivo o *espacio ficticio*. Este planteamiento presupone una postura estética específica que absorbe categóricamente las construcciones compositivas dirigidas únicamente por exploraciones espectromorfológicas o conceptuales. En este sentido, absorber significa que la inmersividad es considerada de

forma general y, por lo tanto, el desarrollo de los demás parámetros son considerados como dados; “el parámetro espacio no es más un efecto en la música tonal, sino que el tono es sólo un efecto en la música espacial. Espacio como finalidad en la expresión musical” (Kupper, 1988).

El modelo, tal y como se planteó, posee un grado considerable de robustez y formalidad debido sobre todo a la estructuración de los elementos acústico-perceptuales y a la estructuración de las definiciones y conceptos propuestos a lo largo de la misma. Si bien los resultados mostrados en el capítulo I son evidencia suficiente de la hipótesis es decir, que en términos reales existe una relación inherente entre el perfil espectromorfológico y la posición espacial de cada objeto sonoro, la información no es lo suficientemente extensa como para afirmar con un alto grado de tolerancia que lo anterior se cumple para objetos sonoros distintos a los mencionados en las pruebas psicoacústicas: tonos puros, ruidos de banda corta, mezclas simples de tonos puros, etcétera. Sin embargo, este modelo plantea de manera natural una potencial metodología para desarrollar pruebas psicoacústicas enfocadas al análisis de fuentes sonoras asociadas al ámbito de la música electroacústica. Lo anterior utilizando, precisamente, la herramienta de clasificación aquí presentada.

Trabajo a futuro

Derivado de esta investigación, se identificaron tres puntos específicos que podrían ser desarrollados de manera inmediatamente posterior a este trabajo:

- Ampliar los resultados de estudios acústico-psicoacústicos para obtener una mejor descripción de las zonas óptimas de ubicación espacial para objetos sonoros de acuerdo con su perfil espectromorfológico. Lo anterior requiere de pruebas con metodología estadística especialmente diseñadas para el ámbito de creación electroacústica dentro de un sistema multicanal.
- Incorporar distintos algoritmos de aprendizaje no supervisado con el fin de elevar la precisión de la metodología computacional propuesta en el capítulo IV.
- Derivado de los dos puntos anteriores, será posible establecer asociaciones mucho más detalladas sobre *zonas óptimas* de localización espacial para objetos sonoros de acuerdo

con su perfil espectromorfológico. Este punto en particular resultaría de gran impacto para el proceso composicional y sobre todo para aquel asistido por computadora.

- Desarrollar estrategias de diseño espacial basadas en la aplicación de los principios Gestalt en conjunto con elementos puntuales del ASA y con lo que se describió como *componentes primarios de la espacialidad*. ■ ■ ■

Apéndice

Este apéndice contiene un desarrollo breve pero formal de las herramientas matemático-computacionales utilizadas en el presente trabajo de investigación, tanto en el nivel teórico como en el práctico. En el caso de la parte matemática, las secciones correspondientes contienen en su mayoría las definiciones de los objetos matemáticos relacionados con sistemas dinámicos y, salvo casos muy específicos y necesarios, algunos resultados derivados de proposiciones o teoremas. La parte de cómputo contiene los fundamentos teóricos primordiales de las técnicas utilizadas para el aprendizaje máquina y algunos de los algoritmos utilizados para el análisis de las series de tiempo.

Notación matemática

A lo largo de los capítulos que componen este libro se hace uso de la notación matemática convencional la cual puede ser consultada en cualquier libro de texto correspondiente. A continuación se muestran un par de tablas que presentan dicha notación:

Para el caso de las representaciones vectoriales y sucesiones, la notación correspondiente es:

Fundamentos de sistemas dinámicos

Los sistemas dinámicos son sistemas cuyos estados evolucionan a través del tiempo. La teoría matemática que los estudia es usada con bastante regularidad para modelar fenómenos que

<i>Símbolo</i>	<i>Nombre</i>	<i>Símbolo</i>	<i>Nombre</i>
+	Suma o adición	∈	Pertenece a
−	Resta o sustracción	∉	No pertenece a
±	Más menos	⊂	Contenido o inclusión
∓	Menos más	⊄	No contenido
×	Multiplicación ordinaria	∪	Unión entre conjuntos
÷	División	∩	Intersección entre conjuntos
:	Razón	∅, {}	Conjunto vacío
=	Igual a	⇒	Implica que
≈	Aproximado a	⇔	Sí sólo sí
≠	Diferente a	∃	Existe
≡	Idéntico a	∴	Por lo tanto
∝	Proporcional a	∵	Porque
<	Menor que	¬	Negación
≤	Menor o igual que	∠	Ángulo
>	mayor que	∠	Ángulo medido
≥	Mayor o igual que	!	Factorial
⊥	perpendicular a	%	Porcentaje
∥	Paralelo a	Δ	Incremento
~	Semejante a	√	Radical
∧	Conjunción (y)	∂	Derivada parcial
∨	Disyunción (o)	∑	Sumatoria
≈	Casi igual a	∫	Integral
≅	Aproximadamente ygal con	ℕ	Conjunto de números naturales
	Tal que (unicidad)	ℤ	Conjunto de números enteros
<	Precede	ℚ'	Conjunto de números irracionales
>	Sucede	ℚ	Conjunto de números racionales
∀	para todo	ℝ	Conjunto de números reales
∞	Infinito	ℂ	Conjunto de números complejos

Principales símbolos matemáticos.

<i>Notación</i>	<i>Se lee como</i>
$P(x_1, x_2, \dots, x_n)$	Punto P en el espacio de n dimensiones
\overline{AB}	Segmento dirigido \overline{AB}
$\vec{a}(x_1, x_2, \dots, x_n)$	Vector \vec{a} en el espacio de n dimensiones
\cdot	Producto escalar, producto punto o producto interno entre dos vectores
\times	Producto vectorial o producto cruz entre dos vectores
\hat{a}	Elemento inverso del elemento a
$\text{Comp}_{\vec{b}} \vec{a}$	Componente del vector \vec{a} sobre el vector \vec{b}
$\text{Proy}_{\vec{b}} \vec{a}$	Proyección del vector \vec{a} sobre el vector \vec{b}
$\vec{a} \cdot \vec{b}$	\vec{a} multiplicado escalarmente por \vec{b}
$\vec{a} \times \vec{b}$	\vec{a} multiplicado vectorialmente por \vec{b}
$\vec{a} \cdot \vec{b} \times \vec{c}$	Doble producto mixto de los vectores $\vec{a}, \vec{b}, \vec{c}$

cambian y se transforman a través del tiempo. Existen dos grandes tipos de sistemas dinámicos: discretos y continuos. Los primeros son modelados mediante ecuaciones en diferencias y los segundos a través de ecuaciones diferenciales; por supuesto, tiene que ver con el tipo de parametrización del tiempo. En esta sección se presentarán conceptos básicos de los sistemas dinámicos vistos desde su construcción abstracta.

DEFINICIÓN A1. Sea M un conjunto cualquiera, $\mathbb{I} \in \mathbb{R}^+$ para el caso continuo, o $\mathbb{I} \subset \mathbb{Z}^+$ para el caso discreto, entonces, un sistema dinámico está definido por la siguiente función:

$$\psi : \mathbb{I} \times M \longrightarrow M$$

tal que para cualquier $x \in M$:

- a) $\psi(0, x) = x$
- b) $\psi(t + s, x) = \psi(t, \psi(s, x))$ para $t, s \in \mathbb{I}$.

M es conocido como el espacio de estados o espacio fase y Φ como el operador evolución.

Obsérvese que debido a la restricción de que el indexador del tiempo es un subconjunto de \mathbb{R}^+ o de \mathbb{Z}^+ , los sistemas considerados están definidos para tiempos que van *hacia adelante*. $\psi(t, x) = x_t$ puede interpretarse como el estado que es alcanzado por el sistema después del intervalo de tiempo de longitud t . Cuando se considera la evolución del sistema para un punto específico x , el operador evolución puede representarse como:

$$\psi_x(t) : \mathbb{I} \rightarrow M$$

Una parte de gran interés para el análisis de los sistemas dinámicos es clasificar el comportamiento de un punto respecto del paso del tiempo. Se tienen tres formas generales: punto fijo, periodicidad y caos. A grandes rasgos, los puntos fijos son aquellos valores para los cuales el sistema dinámico permanece inalterable tras un periodo de tiempo considerable. La periodicidad denota que el sistema repite un comportamiento dado cada intervalo de tiempo específico y, finalmente, el caos denota impredecibilidad total dentro de un rango específico. Estos tres aspectos se presentan de acuerdo con los valores iniciales con los que se evalúe el sistema, también conocidos como condiciones iniciales.

DEFINICIÓN A2. Un punto $x \in M$ es un *punto periódico* de periodo n si se cumple que

$$\psi(n, x) = x$$

y $\psi(j, x) = x$ para todo $0 \leq j < n$. Si $n = 1$ entonces x es denominada *punto fijo*.

El conjunto que recolecta los valores o estados del sistema dado un punto inicial y un intervalo de tiempo específico es conocido como *órbita hacia adelante*. Existe una analogía para tiempos negativos la cual es llamada “órbita hacia atrás” pero requiere una condición especial de invertibilidad en el operador evolución. En este trabajo será suficiente la órbita hacia adelante por lo que a partir de este momento se hará referencia únicamente como órbita del sistema a dicho conjunto.

DEFINICIÓN A3. Dado un sistema dinámico, la *órbita hacia adelante* o simplemente *órbita*, es el conjunto:

$$\mathcal{O} = \{\pi(t, x) = x_t \in M \mid t \geq 0+\}$$

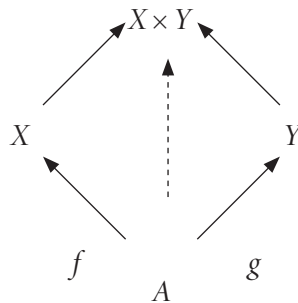
Nótese que en el caso en que el sistema sea discreto, $\pi(t, x) = x_t$ es el estado del sistema después de t iteraciones de la función π y por lo tanto la *órbita* será la colección de estados o valores de la función evolución iterada desde el tiempo inicial hasta el tiempo final.

Se presenta a continuación el resultado que se conoce como la propiedad universal del producto de conjuntos y se enuncia aquí tal y como está planteada en Spivak (2014).

PROPOSICIÓN A1. Sean X, Y dos conjuntos cualesquiera. Para cualquier conjunto A y funciones f, g tales que: $f: A \rightarrow X$ y $g: A \rightarrow Y$, existe una única función:

$$\langle f, g \rangle: A \rightarrow X \times Y$$

tal que el siguiente diagrama conmuta:



Se dice $\langle f, g \rangle: A \rightarrow X \times Y$ está inducida por f y por g y se cumple además que:

$$\langle f, g \rangle(a) = (f(a), g(a))$$

$$\pi_1 \circ \langle f, g \rangle = f$$

$$\pi_2 \circ \langle f, g \rangle = g$$

Para π_1, π_2 las funciones proyección con respecto a f y g .

La demostración de esta proposición está expuesta en la referencia antes mencionada por lo que se remite al lector a dicha lectura.

Algoritmos computacionales

En el presente trabajo se utilizaron algoritmos computacionales específicos para desarrollar la implementación computacional mostrada en el último capítulo. En esta sección se describen brevemente dichos algoritmos con el fin de contextualizar su aplicación y referenciar las fuentes correspondientes para el lector interesado.

Dynamic Time Warping (DTW)

El DTW es un algoritmo bastante popular que se usa en diversas áreas como un método de medida de similaridad entre series de tiempo cuya característica prominente es que “minimiza los efectos del cambio y distorsión en el tiempo al permitir una transformación elástica de la serie de tiempo con el fin de detectar formas similares pero con diferente fase” (Senin, 2009).

La resolución de dicho algoritmo se realiza mediante programación dinámica para obtener el alineamiento óptimo de las series de tiempo utilizando una matriz de costo.

Dadas dos series de tiempo X, Y , dicha matriz está dada por:

$$C_l \in \mathbb{R}^{N \times M}: c_{i,j} = \|x_i - y_j\|, i \in [1, N], j \in [1, M]$$

El pseudocódigo del algoritmo de acuerdo con Senin (2009), está dado por:

Como se mencionó en el capítulo correspondiente, la principal ventaja que presenta este algoritmo es que permite comparar series de tiempo de fase distinta y de distinta longitud, arrojando una medida de similaridad a partir de ello. Existen diversas librerías en Python que calculan este algoritmo, entre ellas destacan: *PyPi dtw* (Rouanet, 2014), *dtwdistance* (Meert, 2018), *dtw-python* (Giorgino, 2009) y la propia *tslearn* ya mencionada. Para esta última, la figura 7-2 muestra la documentación correspondiente.

Alineación global de Kernel K-Medios

Este algoritmo es una variación del popular K-Medios con la ventaja que es adaptable para procesar series de tiempo, “es una extensión del algoritmo de clasificación estándar de K-Me-

Algorithm 2.1 ACCUMULATEDCOSTMATRIX(X, Y, C)

```

1:  $n \leftarrow |X|$ 
2:  $m \leftarrow |Y|$ 
3:  $dtw[] \leftarrow new [n \times m]$ 
4:  $dtw(0, 0) \leftarrow 0$ 
5: for  $i = 1; i \leq n; j ++$  do
6:    $dtw(i, 1) \leftarrow dtw(i - 1, 1) + c(i, 1)$ 
7: end for
8: for  $j = 1; j \leq m; j ++$  do
9:    $dtw(1, j) \leftarrow dtw(1, j - 1) + c(1, j)$ 
10: end for
11: for  $i = 1; i \leq n; j ++$  do
12:   for  $j = 1; j \leq m; j ++$  do
13:      $dtw(i, j) \leftarrow c(i, j) + \min \{ dtw(i - 1, j); dtw(i, j - 1); dtw(i - 1, j - 1) \}$ 
14:   end for
15: end for
16: return  $dtw$ 

```

Figura 53 | Pseudocódigo para el algoritmo de DTW (Senin, 2009).

`tslearn.metrics.dtw(s1, s2, global_constraint=None, sakoe_chiba_radius=1, itakura_max_slope=2.0)` [source]

Compute Dynamic Time Warping (DTW) similarity measure between (possibly multidimensional) time series and return it.

DTW is computed as the Euclidean distance between aligned time series, i.e., if P is the optimal alignment path:

$$DTW(X, Y) = \sqrt{\sum_{(i,j) \in P} \|X_i - Y_j\|^2}$$

Note that this formula is still valid for the multivariate case.

It is not required that both time series share the same size, but they must be the same dimension. DTW was originally presented in [1].

Parameters: **s1**

A time series.

s2

Another time series.

global_constraint: {"itakura", "sakoe_chiba"} or None (default: None)

Global constraint to restrict admissible paths for DTW.

sakoe_chiba_radius: int (default: 1)

Radius to be used for Sakoe-Chiba band global constraint. Used only if `global_constraint="sakoe_chiba"`.

itakura_max_slope: float (default: 2.)

Maximum slope for the Itakura parallelogram constraint. Used only if `global_constraint="itakura"`.

Figura 54 | Documentación para el algoritmo DTW en la librería *tslearn* <<https://tslearn.readthedocs.io/en/stable/>>.

Algorithm outline: Global kernel k -means

Input: Kernel matrix K , Total number of clusters M
Output: Final clustering of the points C_1, C_2, \dots, C_M
 //There is no need to solve for 1 cluster as the solution is trivial and optimal. $C_1^* = X$

1. Solve all k -clustering problems for $k = 2$ to M
2. For each such problem run kernel k -means N times for $n = 1$ to N with input $(K, k, C_1^*, \dots, C_{k-1}^*, C_k = \{x_n\})$ and output $(C_1^n, \dots, C_k^n, E_k^n)$
3. Find $E_k^* = \min_n (E_k^n)$ and set C_1^*, \dots, C_k^* to the partitioning corresponding to E_k^* (this is the solution with k clusters).
4. Set $C_1 = C_1^*, \dots, C_M = C_M^*$ as output of the algorithm

Algorithm outline: Kernel k -Means

Input: Kernel matrix K , Number of clusters k , Initial clusters C_1, \dots, C_k
Output: Final clusters C_1, \dots, C_k , Clustering error E

1. For each point x_n and every cluster C_i compute $\|\phi(x_n) - m_i\|^2$ using (3)
2. Find $c^*(x_n) = \operatorname{argmin}_i (\|\phi(x_n) - m_i\|^2)$
3. Update clusters as $C_i = \{x_n | c^*(x_n) = i\}$
4. If not converged go to step 1 otherwise stop and return final clusters C_1, \dots, C_k and E calculated using (2).

Figura 55 | Pseudocódigo para el algoritmo de Kernel global de K-Medios (Tzortzis, 2008).

dios con la particularidad que es capaz de identificar *clusters* separables de manera no lineal” (Tzortzis, 2008).

Como se describe en Tzortzis (2008), el algoritmo “mapea puntos del espacio de datos entrantes a un espacio de propiedades de mayor dimensión a través del uso de una función Kernel optimizando el error de agrupamiento en el espacio de propiedades al localizar mínimos quasi óptimos”.

Este algoritmo fue adaptado en la librería *tslearn* para poder ser aplicable a series de tiempo, a través del `<métodotslearn.clustering.GlobalAlignmentKernelKMeans>`, el cual se muestra en la figura 56.

Bandas LB Keogh

El método LB Keogh es un algoritmo que permite establecer fronteras óptimas para poder agilizar la comparación de series de tiempo para distintas métricas y en particular para DTW. “Puede ser usada de igual modo para indizar bajo escalamiento uniforme y bajo otro tipo de

```
class tslearn.clustering. GlobalAlignmentKernelKMeans (n_clusters=3, max_iter=50, tol=1e-06, n_init=1,
sigma=1.0, n_jobs=None, verbose=False, random_state=None) \[source\]
Global Alignment Kernel K-means.

Parameters:
  n_clusters : int (default: 3)
    Number of clusters to form.

  max_iter : int (default: 50)
    Maximum number of iterations of the k-means algorithm for a single run.

  tol : float (default: 1e-6)
    Inertia variation threshold. If at some point, inertia varies less than this threshold between two consecutive iterations, the model is considered to have converged and the algorithm stops.

  n_init : int (default: 1)
    Number of time the k-means algorithm will be run with different centroid seeds. The final results will be the best output of n_init consecutive runs in terms of inertia.

  sigma : float or "auto" (default: "auto")
    Bandwidth parameter for the Global Alignment kernel. If set to 'auto', it is computed based on a sampling of the training set (cf tslearn.metrics.sigma\_gak)

  n_jobs : int or None, optional (default=None)
    The number of jobs to run in parallel for GAK cross-similarity matrix computations. None means 1 unless in a joblib.parallel\_backend context. -1 means using all processors. See scikit-learn's Glossary for more details.

  verbose : bool (default: False)
    Whether or not to print information about the inertia while learning the model.
```

Figura 56 | Documentación para el algoritmo GAK de tslearn.

distorsiones, puede ser usada para eficientizar el procesamiento de la transmisión de series de tiempo” (Ratanamahatana y Keogh, 2005).

De manera concreta, las bandas LB Keogh establecen dos cotas alrededor de la serie de tiempo analizada: una superior y otra inferior. Cualquier serie de tiempo puede ser entonces comparada contra esta banda utilizando algún algoritmo específico, luego de formar estas envolventes en las cuales las series de tiempo se encuentran inmersa, dependiendo del sesgo que se forma bajo todas las posibles trayectorias permitidas bajo la restricción global. El algoritmo puede plantearse de acuerdo con lo mostrado en las figuras 57 y 58.

Métricas de similitud para cadenas de caracteres

En la última sección del capítulo IV se utilizó la métrica Damerau-Levinshtein para poder clasificar las etiquetas de las muestras de audio mediante propagación de afinidad. Ésta es una

$$LB_Keogh_{\lambda}(Q, C) = \sum_{i=1}^n \begin{cases} (q_i - u_i)^2, & \text{if } q_i > u_i \\ (q_i - l_i)^2, & \text{if } q_i > l_i \\ 0 & \text{otherwise} \end{cases}$$

Figura 57 | Distancia LB Keogh planteada en términos de la distancia euclidiana (Li y Libin Yang, 2014).

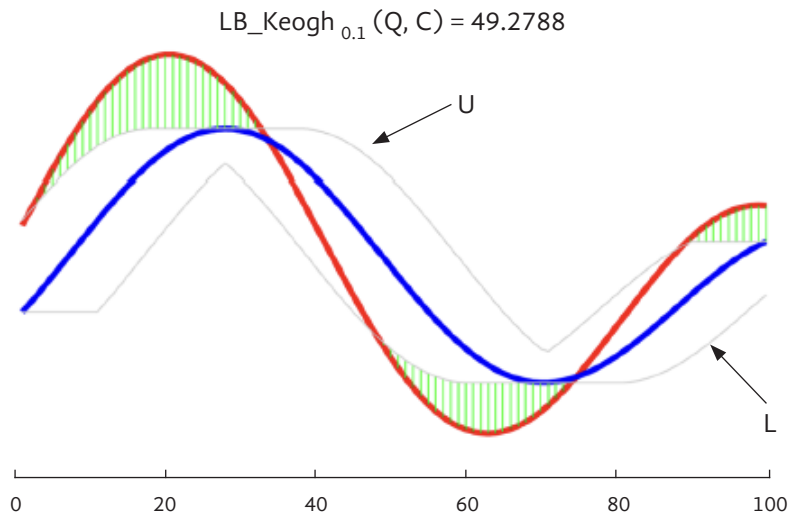


Figura 58 | Representación intuitiva de la distancia LB Keogh (Li y Libin Yang, 2014).

métrica particular, de una clase mayor nombrada Edit Distance, las cuales miden el grado de disimilaridad entre dos cadenas de caracteres mediante el cálculo del número de operaciones mínimas que se requieren para convertir una en otra.

La métrica Damerau-Levinshtein, es, a su vez, un subtipo de la métrica Levinshtein y, como mencionan Zhao C. y Sahni S. (2019), difiere de esta última en el sentido de que incluye operación de transposición y edición específica de caracteres (substitución, borrado e inserción) entre las operaciones permitidas al momento de hacer la comparación.

$$lev_{a,b}(i, c) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0 \\ \min, \begin{cases} lev_{a,b}(i-1, j) + 1 \\ lev_{a,b}(i, j-1) + 1 \\ lev_{a,b}(i-1, j-1) + 1_{(a,b \neq b_j)} \end{cases} & \text{otherwise} \end{cases}$$

Figura 59 | Métrica Levinshtein.

Entropía aproximada

Como alternativa a las medidas de tendencia central dentro del contexto de descriptores numéricos fijos para el análisis de muestras de audio, se plantearon dos valores: la *entropía aproximada* y el *coeficiente de variación*.

La *entropía aproximada*, junto con la *entropía muestral* son medidas utilizadas para calcular grados de predictibilidad, regularidad o complejidad dentro de series de tiempo y que reflejan, de manera general, la pérdida de información dentro de las mismas. Se denota como *ApEn* y “forma parte de procesos más generales como la tasa de entropía para la aproximación de cadenas de Markov para procesos” (Pincus, 1995). La figura 6-7 muestra el algoritmo correspondiente para calcular la *ApEn*.



Figura 60 | Algoritmo para calcular la entropía aproximada (fuente: Wikipedia).

Flood (2021) presenta EntropyHub, un *toolkit* de libre acceso que permite realizar cálculos complejos con series de tiempo, incluida la entropía aproximada, pero de manera mucho más general; permite además estimaciones de multiescala cruzada y bidimensional. Por su parte, Vallat (2018) desarrolló *Anthropy*, el cual es un paquete para Python con el cual se pueden realizar diversas estimaciones de complejidad en series de tiempo que también consideran la entropía aproximada. ■ ■ ■

Índice de figuras

- 33** 1. Principios Gestalt (Soegaard, 2020. Autor: Impronta. Copyright: CC BY-SA 3.0).
- 41** 2. Representación del funcionamiento del ASA.
- 48** 3. Flujo general de procesamiento del ASA.
- 55** 4. Representación diagramática de las TGU.
- 61** 5. Sistema de coordenadas interaural polar.
- 91** 6. Representación gráfica del espacio audible y de la extensión del espacio virtual sonoro.
- 92** 7. *Sonic Arts Research Centre*, Belfast.
- 92** 8. *Birmingham Electroacoustic Sound Theatre* (BEAST).
- 93** 9. Hexagram-UQAM ESPACE.
- 93** 10. CCRMA 3D *Listening Room*.
- 133** 11. Espectrograma para muestra de audio I.
- 133** 12. Espectrograma para muestra de audio II.
- 134** 13. Forma de onda para muestra de audio I.
- 134** 14. Forma de onda para muestra de audio II.
- 138** 15. Series de tiempo del ZCR para los objetos sonoros I y II.
- 139** 16. Series de tiempo del centroide espectral para los objetos sonoros I y II.
- 149** 17. Comparación de series de tiempo para RMSE de audio de corno francés. Nota A2 tocada en *mezzo forte*.
- 149** 18. Comparación de series de tiempo para RMSE de audio de corno francés. Nota A2 tocada en *fortissimo*.

- 150 19. Comparación de series de tiempo para RMSE de audio de corno francés.
Nota A3 tocada en *trino forte*.
- 150 20. Identificación de *onsets* en la serie de tiempo.
- 151 21. Reconstrucción de la serie de tiempo a partir de la interpolación de *onsets*.
- 154 22. *Cluster* no. 2 de series de tiempo para la sub-banda 4 (contraste espectral) de muestras de audio de flauta.
- 154 23. *Cluster* no. 3 de series de tiempo para la sub-banda 4 (ancho espectral) de muestras de audio de corno inglés.
- 155 24. Cluster no. 3 de series de tiempo para el ZCR de muestras de audio de saxofón.
- 155 25. *Cluster* no. 0 de series de tiempo para el ZCR de muestras de audio de oboe.
- 156 26. *Cluster* no.5 de series de tiempo para flujo espectral muestras audio de clarinete.
- 156 27. *Cluster* no. 5 de series de tiempo para el ZCR de muestras de audio de mandolina.
- 157 28. *Cluster* no.3 de series de tiempo para flujo espectral de muestras de audio de guitarra.
- 157 29. *Cluster* no. 2 de series de tiempo para el flujo espectral de muestras de audio de cello.
- 158 30. *Cluster* no. 0 de series de tiempo para el flujo espectral de muestras de audio de banjo.
- 161 31. Baricentro euclidiano para *cluster* de flujo espectral de muestras de audio de violín.
- 161 32. Centroide baricentro/bandas LB Keogh para sub-banda 2 de ancho espectral de muestras de audio de clarinete.
- 162 33. Centroide baricentro/bandas LB Keogh para sub-banda 2 de ancho espectral de muestras de audio de contrabajo.
- 162 34. Centroide baricentro/bandas LB Keogh para sub-banda 2 de ancho espectral de muestras de audio de tuba.
- 163 35. Centroide baricentro/bandas LB Keogh para ZCR de muestras de audio de trombón.
- 163 36. Centroide baricentro/bandas LB Keogh para ZCR de muestras de audio de tuba.
- 164 37. Centroide baricentro/bandas LB Keogh para ZCR de muestras de audio de trompeta.
- 164 38. Centroide baricentro/bandas LB Keogh para flujo espectral de muestras de audio de clarinete bajo.

- 165** 39. Centroide baricentro/bandas LB Keogh para flujo espectral de muestras de audio de clarinete banjo.
- 168** 40. Comparación de muestra 1 para baricentros de ZCR.
- 168** 41. Comparación de muestra 1 para baricentros de flujo espectral.
- 169** 42. Comparación de muestra 2 para baricentros de ZCR.
- 169** 43. Comparación de muestra 2 para baricentros de sub-banda 2 de contraste espectral.
- 170** 44. Comparación de muestra 3 para baricentros de ZCR.
- 170** 45. Comparación de muestra 3 para baricentros de sub-banda 2 de contraste espectral.
- 171** 46. Comparación de muestra 4 para baricentros de flujo espectral.
- 171** 47. Comparación de muestra 4 para baricentros de ZCR.
- 173** 48. Forma de onda de la muestra 1.
- 173** 49. Espectrograma de la muestra 1.
- 174** 50. *Cluster* correspondiente al baricentro más cercano en ZCR de la muestra 1.
- 174** 51. *Cluster* clasificado interiormente mediante propagación de afinidad, correspondiente al baricentro más cercano en ZCR de la muestra 1.
- 175** 52. Histograma de representación de la muestra 1 para el baricentro más cercano en ZCR.
- 191** 53. Pseudocódigo para el algoritmo de DTW (Senin, 2009).
- 191** 54. Documentación para el algoritmo DTW en la librería *tslearn*.
- 192** 55. Pseudocódigo para el algoritmo de Kernel global de K-Medios (Tzortzis, 2008).
- 193** 56. Documentación para el algoritmo GAK de *tslearn*.
- 194** 57. Distancia LB Keogh planteada en términos de la distancia euclidiana (Li y Libin Yang, 2014).
- 194** 58. Representación intuitiva de la distancia LB Keogh (Li y Libin Yang, 2014).
- 194** 59. Métrica Levinshtein.
- 195** 60. Algoritmo para calcular la entropía aproximada (fuente: Wikipedia).

Bibliografía

- Abellanas, Pedro (1975), *Elementos de matemáticas*, Madrid, Heraclio Fournier, S.A.
- Alameda-Pineda, Xavier y Radu Horaud (2014), "A Geometric Approach to Sound Source Localization from Audio", en *IEEE/ACM Transactions on Audio, Speech, and Language Processings*, vol. 22, núm. 6, pp. 1082-1095.
- Alexandridis, Anastasios; Anthony Griffin y Athanasios Mouchtaris (2013), "Directional Coding of Audio Using a Circular Microphone Array", en *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, disponible en <https://www.researchgate.net/profile/Athanasios-Mouchtaris/publication/259465647_Directional_coding_of_audio_using_a_circular_microphone_array/links/5582e7a408ae12bde6e63891/Directional-coding-of-audio-using-a-circular-microphone-array.pdf>.
- Alonso, Edith (2012), "El concepto de 'imagen-de-lo-sonoro' en la música acusmática según el compositor François Bayle", en *Escritura e Imagen*, vol. 9 (2013), pp. 101-124.
- Alpern, Adam (1995), "Techniques for Algorithmic Composition of Music", Hampshire College, disponible en <<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.9364&rep=rep1&type=pdf>>.
- Andrew, Francl y Josh H. McDermott (2022), "Deep Neural Network Models of Sound Localization Reveal How Perception is Adapted to Real-World Environments", en *Nature Human Behaviour*, vol. 6, núm. 1, pp. 111-133.
- Atick, Joseph (1992), "Could Information Theory Provide an Ecological Theory of Sensory Processing?", en *Network: Computational in Neural System*, vol. 3, núm. 2, pp. 213-251.
- Atkinson, Richard C. y Richard M. Shiffrin (1968), "Human Memory: A Proposed System and its Control Processes", en K.W. Spence y J.T. Spence (eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol. 2, Nueva York, Academic Press, pp. 89-195.
- Attneave, Fred (1954), "Some Informational Aspects of Visual Perception", en *Psychological Review*, vol. 61, núm. 3, pp. 183-193.

- Bajcsy, Ruzena (1988), "Active Perception", en *Proceedings of the IEEE*, vol. 76, núm. 8, pp. 966-1005.
- Barlow, Horace (1961), "The Coding of Sensory Messages Current Problems", en W.H. Thorpe y O.L. Zangwill (eds.), *Current Problems Animal Behaviour*, Cambridge, Cambridge University Press.
- Barlow, Horace (1990), "Conditions for Versatile Learning, Helmholtz's Unconscious Inference and the Task of Perception", en *Vision Research*, vol. 30, núm. 11, pp. 1561-1571.
- Barth, Friedrich G.; Patrizia Giampieri-Deutsch y Hans-Dieter Klein (eds.) (2012), *Sensory Perception: Mind and Matter*, Nueva York, Springer-Verlag/Wien.
- Bertin-Mahieux Thierry; Douglas Eck y Michael Mandel (2010), "Automatic Tagging of Audio: The State-of-the-Art", en W. Wang (ed.), *Machine Audition: Principles, Algorithms and Systems*, Hershey, Pennsylvania, IGI Publishing.
- Bértola, Elena de (2014), "On Space and Time in Music and the Visual Arts", en *Leonardo*, vol. 5, núm. 1 (invierno, 1972), pp. 27-30.
- Boden, Margaret (1999), "Computer Models of Creativity", en *Handbook of Creativity*, Cambridge, Cambridge University Press, pp. 351-372.
- Boon Jean Pierre, Olivier Decroly (1994), "Dynamical Systems Theory for Music Dynamics", en *Chaos*, vol. 5, núm. 3, Nueva York, Woodbury, pp. 501-508.
- Boone, Marinus M.; Edwin N.G. Verheijen y Peter F. van Tol (1995), "Spatial Sound-Field Reproduction by Wave-Field Synthesis", en *Journal of the Audio Engineering Society*, vol. 43, núm. 12, pp. 1003-1012.
- Boulanger-Lewandowski, Nicolas; Yoshua Bengio y Pascal Vincent (2013), "Audio Chord Recognition with Recurrent Neural Networks", disponible en <<https://archives.ismir.net/ismir2013/paper/000243.pdf>>.
- Bregman, Albert S. (1990), *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge, MA, The MIT Press, 1990 (hardcover)/1994 (paperback).
- Bregman, Albert S. (1993), "Auditory Scene Analysis: Hearing in Complex Environments", en Stephen McAdams y Emmanuel Bigand (eds.), *Thinking in Sound: The Cognitive Psychology of Human Auditions*, Oxford, Oxford University Press, pp. 10-36.
- Bregman, Albert S. y Pierre A. Ahad (1990), *Demonstrations to Accompany Bregman's Auditory Scene Analysis. The Perceptual Organization of Sound*, Department of Psychology Auditory Research Laboratory McGill University, disponible en <http://www.rctn.org/bruno/data/auditory_demonstrations/ASA-Demo-Booklet9V4.pdf>.
- Bregman, Albert S. y Wieslaw Woszczyk (2004), "Controlling the Perceptual Organization of Sound: Guidelines Derived from Principles of Auditory Scene Analysis (ASA)", en K. Greenebaum y R. Barzel (eds.), *Audio Anecdotes: Tools, Tips and Techniques for Digital Audio*, vol. 1, Natick, MA, A. K. Peters, pp. 35-64.
- Bregman, Albert S. y Wieslaw Woszczyk (2005), "Creating Mixtures: The Application of Auditory Scene Analysis (ASA) to Audio Recording", En K. Greenebaum Y R. Barzel (eds.), *Audio Anecdotes III: Tools, Tips and Techniques for Digital Audio*, vol. 1, Natick, MA, A. K. Peters.

- Bresson Jean, Carlos Agon y Marlon Schumacher (2007), “Représentation des données de contrôle pour la spatialisation dans openmusic”, disponible en <<http://articles.ircam.fr/textes/Bresson10a/index.pdf>>.
- Brochard, Renaud; Carolyn Drake, Marie Claire Botte y Stephen McAdams (1999), “Perceptual Organization of Complex Auditory Sequences: Effect of Number of Simultaneous Subsequences and Frequency Separation”, en *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25, núm. 6, pp. 1742-1759.
- Brown, Guy J. y Martin Cooke (1994), “Computational Auditory Scene Analysis”, en *Computer Speech and Language*, vol. 8, pp. 297-336, disponible en <<https://staffwww.dcs.shef.ac.uk/people/G.Brown/pdf/csl.pdf>>.
- Brownie, Barbara (2006), “Key Gestalt Theories and Principles”, disponible en <https://mafiadoc.com/key-gestalt-theories-and-principles_5a22fa711723ddd437f09d74.html>.
- Büchler, Michael (2002), “Algorithms for Sound Classification in Hearing Instruments”, tesis de doctorado en Ciencias Técnicas, Swiss Federal Institute of Technology Zurich.
- Cangea, Catalina; Charlie Chen, Drew Jaegle, Curtis Hawthorne e Ian Simon (2022), “Autoregressive Long-Context Music Generation with Perceiver AR”, disponible en <<https://magenta.tensorflow.org/perceiver-ar>>.
- Carlile, Simon y Johahn Leung (2016), “The Perception of Auditory Motion”, en *Trends in Hearing*, vol. 20, núm. 1, pp. 1-19, disponible en <https://www.researchgate.net/publication/301539178_The_Perception_of_Auditory_Motion>.
- Choi, Keunwoo; György Fazekas, Mark Sandler y Kyunghyun Cho (2018), “Transfer Learning for Music Classification and Regression Tasks”, Proceedings of the 18th ISMIR Conference, Suzhou, China, 23-27 de octubre, disponible en <<http://www.eecs.qmul.ac.uk/~gyorgyf/files/papers/choi2017ismir.pdf>>.
- Chowning, John M. (2004), “The Simulation of Moving Sound Sources”, en *Computer Music Journal*, vol. 1, núm. 3, pp. 48-52, disponible en <<https://www.jstor.org/stable/3679609>>.
- Clifton, Thomas (1983), *Music as Heard: A Study in Applied Phenomenology*, New Haven, Yale University Press.
- Cobos, Maximo; José J. López y Sascha Spors (2010), “A Sparsity-Based Approach to 3D Binaural Sound Synthesis Using Time-Frequency Array Processing”, en *EURASIP Journal on Advances in Signal Processing*, disponible en <<http://dx.doi.org/10.1155/2010/415840>>.
- Collins, Nick (2010), “Computational Analysis of Musical Influence: A Musicological Case Study Using Mir Tools”, en J. Stephen Downie y Remco C. Veltkamp (eds.), *Proceedings of the 11th International Society for Music Information Retrieval Conference*, agosto 9-13, Utrecht, Netherlands, IS-MIR, pp. 177-182.
- De Vries, J. (2000), *Elements of Topological Dynamics*, Dordrecht, Netherlands, Springer.
- Dinh Cong. Nguyen (1997), *Topological Dynamics of Random Dynamical Systems*, Oxford/Nueva York, Clarendon Press/Oxford University Press.

- Disbergen, Niels; Giancarlo Valente, Elia Formisano y Robert J. Zatorre (2018), "Assessing Top-Down and Bottom-Up Contributions to Auditory Stream Segregation and Integration With Polyphonic Music", en *Frontiers in Neuroscience*, vol. 12, article 121, pp. 1-16.
- Duero, Dante G. (2003), "La Gestalt como teoría de la percepción y como epistemología: aportes y desarrollos", disponible en <[http://psicologiamonserrat.zonalibre.org/Monserrat%20\(Gestalt\).pdf](http://psicologiamonserrat.zonalibre.org/Monserrat%20(Gestalt).pdf)>.
- Faller, Christof y Juha Merimaa (2004), "Source Localization in Complex Listening Situations: Selection of Binaural Cues Based on Interaural Coherence", en *Journal of the Acoustical Society of America*, vol. 116, núm. 5, pp. 3075-3089.
- Fay, Richard R. y Arthur N. Popper (2008), *Springer Handbook of Auditory Research, Series Editors*, Springer, Science+Business Media.
- Fazi, Filippo M. y Philip A. Nelson (2007), "The Ill-Conditioning Problem in Sound Field Reconstruction", en *123rd Audio Engineering Society Convention*, 5-8 de octubre, Nueva York, Audio Engineering Society.
- Finn, Jacobsen (2011), "The Sound Field in a Reverberation Room", tesis de doctorado, Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark.
- Flood, Matthew W. (2021), "EntropyHub: An Open-Source Toolkit for Entropic Time Series Analysis", en *PLoS ONE*, disponible en <<https://doi.org/10.1371/journal.pone.0259448>>.
- Fodor, Jerry (1983), *The Modularity of Mind: An Essay on Faculty Psychology*, Cambridge, MA, MIT Press.
- Francl, Andrew y Josh H. McDermott (2022), "Deep Neural Network Models of Sound Localization Reveal How Perception is Adapted to Real-World Environments", en *Nature Human Behaviour*, vol. 6, núm. 1, pp. 111-133.
- French, Robert M. (2002), "The Computational Modeling of Analogy-Making", en *Trends in Cognitive Sciences*, vol. 6, núm. 5, pp. 200-205.
- Fu, Zhouyu; Guojun Lu, Kai Ming Ting y Densheng Zhang (2011), *A Survey of Audio-Based Music Classification and Annotation*, en *Journal IEEE Transactions on Multimedia*, vol. 13, núm. 2, pp. 303-319.
- Gao, Ruohan y Kristen Grauman (2019), "2.5D Visual Sound", en *IEEE Conference on Computer Vision and Pattern Recognition*, disponible en <https://www.cs.utexas.edu/~grauman/papers/CVPR19_2.5d-visual-sound.pdf>.
- Gentner, Dedre (1983), "Structure-Mapping: A Theoretical Framework for Analogy", en *Cognitive Science*, vol. 7, núm. 2, pp. 155-170.
- Gentner, Dedre (1989), "The Mechanisms of Analogical Learning", en S. Vosniadou y A. Ortony (eds.), *Similarity and Analogical Reasoning*, Londres, Cambridge University Press, pp. 199-241.
- Getzmann, Stephan y Jörg Lewald (2007), "Localization of Moving Sound", en *Perception & Psychophysics*, vol. 69, núm. 6, pp. 1022-1034.
- Gibson, James Jerome (1966), *The Senses Considered as Perceptual Systems*, Boston, Houghton Mifflin.

- Gibson, James Jerome (1972), "A Theory of Direct Visual Perception", en A. Nöe y E. Thompson (eds.), *Vision and Mind: Selected Reading in the Philosophy of Perception*, Cambridge, MA, MIT Press, pp. 77-89.
- Giorgino, Toni (2009), "Computing and Visualizing Dynamic Time Warping Alignments", en *Journal of Statistical Software*, vol. 31, núm. 7, pp. 1-24.
- Gómez, Emilia y Perfecto Herrera (2004), "Automatic Extraction of Tonal Metadata from Polyphonic Audio Recordings", en *AES 25th International Conference*, 17-19 de junio, Londres.
- Gouyon, Fabien; Simon Dixon, Elias Pampalk y Gerhard Widmer (2004), "Evaluating Rhythmic Descriptors for Musical Genre Classification", en *AES 25th International Conference*, 17-19 de junio, Londres., disponible en <<https://ofai.at/papers/oefai-tr-2004-10.pdf>>.
- Gregory, Richard (1970), *The Intelligent Eye*, Londres, Wiedenfeld & Nicolson.
- Gregory, Richard (1974), *Concepts and Mechanisms of Perception*, Londres, Duckworth.
- Gregory, Richard (1997), "Knowledge in Perception and Illusion", *Philosophical Transactions of the Royal Society*, vol. 352, núm. 358, pp. 1121-1127.
- Grey, John (1976), "Multidimensional Perceptual Scaling of Musical Timbres", en *The Journal of the Acoustical Society of America*, vol. 61, núm. 5, pp. 1270-1277.
- Hafer, Ervin R., Samuel C. Carrier (1971), "Binaural Interaction in Low-Frequency Stimuli: The Inability to Trade Time and Intensity Completely", en *The Journal of the Acoustical Society of America*, vol. 51, núm. 6B, disponible en <<https://doi.org/10.1121/1.1913044>>.
- Harley, Maria Anna (1994a), *Space and Spatialization in Contemporary Music*, Montreal, McGill University.
- Harley, Maria Anna (1994b), "From Point to Sphere: Spatial Organization of Sound in Contemporary Music (after 1950)", en *Revue de Musique des Universités Canadiennes*, núm. 13, pp. 123-144, disponible en <<https://doi.org/10.7202/1014300ar>>.
- Helmholtz, Hermann von (1925), *Physiological Optics. Volume III. The Theory of the Perceptions of Vision*, Washington, D.C., Optical Society of America.
- Henriksen, Frank E. (2002), "Space in Electroacoustic Music: Composition, Performance and Perception of Musical Space", tesis de doctorado inédita, Londres, City University, Department of Music.
- Hernández-Olivan, Carlos; Ignacio Zay Pinilla, Carlos Hernández-López y José R. Beltrán (2021), "A Comparison of Deep Learning Methods for Timbre Analysis in Polyphonic Automatic Music Transcription", en *Electronics*, vol. 10, núm. 7, pp. 1-16.
- Herrera, Perfecto; Juan Pablo Bello, Gerhard Widmer y Mark Brian Sandler (2005), "SIMAC: Semantic Interaction with Music Audio Contents", en *The 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, Londres, IET, pp. 399-406.
- Hidaka, Takayuki; Toshiyuki Okano y Leo Beranek (1992), "Interaural Cross Correlation (IACC) as a Measure of Spaciousness and Envelopment in Concert Halls", en *Journal of the Acoustic Society of America*, vol. 92, núm. 4, 2492.

- Hofstadter, Douglas R. (1985), "Analogies and Roles in Human and Machine Thinking", en *Metamagical Themas: Questings for the Essence of Mind and Pattern*, Nueva York, Basic Books, pp. 547-604.
- Hosken, Daniel W. (2011), *An Introduction to Music Technology*, Londres, Routledge.
- Humphrey, Eric J.; Juan Pablo Bello y Yann LeCun (2013), "Feature Learning and Deep Architectures: New Directions for Music Informatics", en *Journal for Intelligent Information Systems*, vol. 41, núm. 3, pp. 461-481.
- Itoh, Motokuni, Kazuhiro Iida y Masayuki Morimoto (2006), "Individual Differences in Directional Bands", WESPAC IX, disponible en <<http://www.iida-lab.it-chiba.ac.jp/literature/International.Conference.Proceedings/24.Individual%20differences%20in%20directional%20bands.pdf>>.
- Juha, Merimaa (2006), *Analysis, Synthesis and Perception of Spatial Sound: Binaural Localization Modeling and Multichannel Loudspeaker Reproduction*, Helsinki, University of Technology/Laboratory of Acoustics and Audio Signal Processing.
- Jurek, Megan; Nikhil Bhanu, Wilson Zhao, Søren Nielsen, Sehmon Burnam y Jesse Engel (2022), "DDSP-VST: Neural Audio Synthesis for All", disponible en <<https://magenta.tensorflow.org/ddsp-vst-blog>>.
- Kendall, Gary (2010), "Spatial Perception and Cognition in Multichannel Audio for Electroacoustic Music", en *Organised Sound*, vol. 15, núm. 3, Cambridge University Press, pp. 228-238.
- Kendall, Gary y Mauricio Ardila (2009), "The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata", en R. Kronland-Martinet, S. Ystad y K. Jensen (eds.), *Computer Music Modeling and Retrieval. Sense of Sounds. CMMR 2007. Lecture Notes in Computer Services*, vol. 4969, Berlin, Springer Verlag.
- Keogh, Eamon y Ann Ratanamahatana (2005), "Exact Indexing of Dynamic Time Warping", en *Knowledge and Information Systems*, vol. 7, núm. 3, pp. 358-386.
- Kim, Teajun; Jongpil Lee y Juhan Nam (2018), *Sample-Level CNN Architectures for Music Autotagging Using Raw Waveforms*, disponible en <<https://arxiv.org/pdf/1710.10451.pdf>>.
- Kofka, Kurt (1922), "Perception: An introduction to the Gestalt-Theorie", en *Psychological Bulletin*, vol. 19, núm. 10, pp. 531-585.
- Kofka, Kurt (1935), *Principles of Gestalt Psychology*, Londres, Lund Humphries.
- Kollmeier, Birger; Georg Klump, Volker Hohmann, Ulrike Langemann, Manfred Mauermann, Stefan Uppenkamp y Jesko Verhey (2007), *Hearing-From Sensory Processing to Perception*, Berlín, Springer.
- Kootstra, Gert, Niklas Bergström y Danica Kragic (2012), *Gestalt Principles for Attention and Segmentation in Natural and Artificial Vision Systems*, Swedish Foundation for Strategic Research, disponible en <https://www.ais.uni-bonn.de/~holz/spme/04_kootstra_gestalt_principles.pdf>.
- Kowalczyk, Konrad; Oliver Thiergart, Maja Taseska et al. (2015), "Parametric Spatial Sound Processing: A Flexible and Efficient Solution to Sound Scene Acquisition, Modification, and Reproduction", en *IEEE Signal Processing Magazine*, vol. 32, núm. 2, pp. 31-42.
- Krizhevsky, Alex; Ilya Sutskever y Geoffrey Hinton (2012), "ImageNet Classification with Deep Convolutional Neural Networks", en *Advances in Neural Information Processing Systems*, vol. 5, Lake

- Tahoe, disponible en <<https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>>.
- Letowski, Tomasz y Szymon Letowski (2012), "Auditory Spatial Perception: Auditory Localization", en *Army Research Laboratory*, mayo, disponible en <<https://apps.dtic.mil/sti/pdfs/ADA563540.pdf>>.
- Li, Hailin; Chonghui Guo y Wangren Qiu (2011), "Similarity Measure Based on Piecewise Linear Approximation and Derivate Dynamic Time Warping for Time Series Mining", en *Expert Systems with Applications*, vol. 38, núm. 12, pp. 14732-14743.
- Li, Hailin y Libing Yang (2014), "Extensions and Relationships of Some Existing Lower-Bound Functions for Dynamic Time Warping", en *Journal of Intelligent Information Systems*, vol. 43, núm. 1, pp. 59-79.
- Lindsay, Peter y Donald Norman (1972), *Human Information Processing*, Nueva York, Academic Press.
- Macedo, Frederico (2015), *Investigating Sound in Space: Five Meanings of Space in Music and Sound Art*, en *Organised Sound*, vol. 20, núm. especial 2, Cambridge University Press, pp. 241-248.
- Marcus, Lauks (2010), *Computer-Aided Algorithmic Composition*, Saarbrücken, VDM Verlag Dr. Müller.
- McDonald, Kyle (2017), "Neural Nets for Generating Music", disponible en <<https://medium.com/artists-and-machine-intelligence/neural-nets-for-generating-music-f46dffac21c0>>.
- Morimoto, Masayuki (2001), "The Contribution of Two Ears to the Perception of Vertical Angle in Sagittal Planes", en *The Journal of the Acoustical Society of America*, vol. 109, núm. 4, pp. 1596-1603.
- Morimoto, Masayuki y Yoichi Ando (1980), "On the Simulation of Sound Localization", en *Journal of the Acoustical Society of Japan*, vol. 1, num. 3, pp. 167-174.
- Mountain, Rosemary (2007), "Auditory Scene Analysis and Electroacoustics", disponible en <<http://www.ems-network.org/ems09/papers/mountain.pdf>>.
- Müller, Alexander y Rudolf Rabenstein (2009), "Physical Modeling for Spatial Sound Synthesis", disponible en <https://dafx09.comopolimi.it/proceedings/papers/paper_54.pdf>.
- Müsseler, Jochen y Gisa Aschersleben y W. Prinz (1998), "Localizing the First Position of a Moving Stimulus: The Fröhlich Effect and an Attention-Shifting Explanation", en *Perception & Psychoacoustics*, vol. 60, núm. 4, pp. 683-695.
- Nakashima, H. y T. Mukai (2005), "3D Sound Source Localization System Based on Learning of Binaural Hearing", en *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 3534-3539.
- Nikunen, Joonas y Tuomas Virtanen (2014a), "Multichannel Audio Separation by Direction of Arrival Based Spatial Covariance Model and Non-Negative Matrix Factorization", en *IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, pp. 6677-6681 disponible en DOI <10.1109/ICASSP.2014.6854892>.
- Normandeau, Robert (2009), "Timbre Spatialisation: The Medium is the Space", en *Organised Sound*, vol. 14, núm. 3, Cambridge University Press, pp. 277-285.

- Normandeau, Robert (2010), "A Revision of the TARTYP, published by Pierre Schaeffer", en *Proceedings of the Seventh Electroacoustic Music Studies Network Conference*, disponible en <<https://docplayer.net/46175174-Robert-normandeau-a-revision-of-the-tartyp-published-by-pierre-schaeffer-faculte-de-musique-universite-de-montreal.html>>.
- Oramas, Sergio; Oriol Nieto, Francesco Barbieri y Xavier Serra (2017), "Multi-Label Music Genre Classification from Audio, Text and Images Using Deep Features", en *Proceedings of the 18th ISMIR Conference*, Suzhou, China, 23-27 de octubre, disponible en <<https://zenodo.org/record/1417427#.Y0s-PbvMLIU>>.
- Parry, R. Mitchell e Irfan Essa (2006), "Estimating the Spatial Position of Spectral Components in Audio", en J. Rosca, D. Erdogmus, J.C. Príncipe y S. Haykin, S. (eds), *Independent Component Analysis and Blind Signal Separation*, ICA 2006, Lecture Notes in Computer Science, vol 3889, pp. 666-673, Berlín, Springer Heidelberg, disponible en <https://doi.org/10.1007/11679363_83>.
- Pavlidis, Despoina; Anthony Griffin, Matthieu Puigt y Athanasios Mouchtaris (2013), "Real-Time Multiple Sound Source Localization and Counting Using a Circular Microphone Array", en *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, núm. 10, pp. 2193-2206.
- Payri, Blas y José Luis Miralles Bono (2007), "Auditory Scene Analysis and Sound Source Coherence as a Frame For The Perceptual Study of Electroacoustic Music Language", en *Electroacoustic Music Studies Network*, Universidad Politécnica de Valencia, Campus de Gandía, disponible en <http://www.ems-network.org/IMG/pdf_PayriEMS07.pdf>.
- Plenge, G. (1974), "On the Differences between Localization and Lateralization", en *The Journal of the Acoustical Society of America*, vol. 56, núm. 3, pp. 944-951.
- Perrott, D. R. y Alan Musican (1977), "Minimum Auditory Movement Angle: Binaural Localization of Moving Sound Sources", en *The Journal of the Acoustical Society of America*, vol. 62, núm. 6, pp. 1463-1466.
- Pincus, Steve (1995), "Approximate Entropy (ApEn) as a Complexity Measure", en *Chaos*, vol. 5, núm. 1, Woodbury, Nueva York, pp. 110-117.
- Plumbley, Mark y Samer Abdallah (2007), "Information Theory and Sensory Perception", en J.A. Bryant, M.A. Atherton y M.W. Collins (eds.), *Design and Information in Biology*, cap. 7, Southampton, UK, WIT Press.
- Pohle, Tim; Dominik Schnitzer, Markus Schedl y Peter Knees (2009), "On Rhythm and General Music Similarity", en *Proceedings 10th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 525-530, disponible en <<https://zenodo.org/record/1418229#.Y0tjhbvMLIU>>.
- Pons, Jordi; Olga Slizovskaia, Rong Gong, Emilia Gómez y Xavier Serra (2017), "Timbre Analysis of Music Audio Signals with Convolutional Neural Networks", en *Proceedings 25th European Signal Processing Conference*, Kos, Greece, pp. 2813-2817, disponible en <<https://eurasip.org/Proceedings/Eusipco/Eusipco2017/papers/1570347061.pdf>>.
- Pressnitzer, Daniel, Alain de Cheveigné, Stephen McAdams y Lionel Collet (2005), *Auditory Signal Processing. Physiology, Psychoacoustics, and Models*, Nueva York, Springer.

- Pulkki, V. (1997), "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", en *Journal of the Audio Engineering Society*, vol. 45, núm. 6, pp. 456-466.
- Pulkki, Ville (2001), *Spatial Sound Generation and Perception by Amplitude Panning Techniques*, Finlandia, Helsinki University of Technology.
- Pulkki, Ville, Symeon Delikaris-Manias y Archontis Politis (2018), *Parametric Time-Frequency Domain Spatial Audio*, Hoboken, N.J., John Wiley & Sons Ltd.
- Rath, Toni M. y R. Manmatha (2002), "Lower-Bounding of Dynamic Time Warping Distances for Multivariate Time Series", Center for Intelligent Information Retrieval University of Massachusetts, disponible en <https://works.bepress.com/r_manmatha/19/>.
- Reck Miranda, Eduardo (2002), *Computer Sound Design*, Oxford, Focal Press.
- Reuter, Christoph y Saleh Siddiq (2017), "The Colourful Life of Timbre Spaces-Timbre Concepts from Early Ideas to Meta-Timbre Space and Beyond", disponible en <<https://www.researchgate.net/publication/318467904>>.
- Rogers, Elliott (1968), "Simple Visual and Simple Auditory Reaction Time: A Comparison", en *Psychonomic Science*, vol. 10, disponible en <<https://link.springer.com/content/pdf/10.3758/BF03331548.pdf>>, pp. 335-336.
- Rumsey, Francis (2001), *Spatial Audio*, Oxford, Focal Press.
- Rumsey, Francis y Tim McCormick (2009), *Sound and Recording*, 6a. ed., Oxford, Focal Press.
- Sandvad, J. (1996), "Dynamic Aspects of Auditory Virtual Environments", en *Audio Engineering Society Reprints. AES 100th Convention*, vol. 2, pp. 54-60.
- Sawada, Hiroshi; Shoko Araki, Ryo Mukai y Shoji Makino (2007), "Grouping Separated Frequency Components by Estimating Propagation Model Parameters in Frequency-Domain Blind Source Separation", en *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, núm. 5, pp. 1592-1604.
- Schaeffer, Pierre y Guy Reibel (1967), *Solfège de L'objet sonore*, reedición 1998, París, Coedition Ina-Publications.
- Scheinerman, R. Edward (1996), *Dynamical Systems*, Department of Mathematical Sciences the Johns Hopkins University, disponible en <<http://cmc.de.iscte.pt/Scheinerman-Book.pdf>>.
- Schumacher, Marlon y Jean Bressons (2010), "Spatial Sound Synthesis in Computer Aided Composition", en *Organised Sound*, vol. 15, núm. 3, pp. 271-289.
- Schumacher, Federico y Claudio Fuentes (2017), "Space-Emotion in Acousmatic Music", en *Organised Sound*, vol. 22, núm. 3, pp. 394-405.
- Senin, Pavel (2009), "Dynamic Time Warping Algorithm Review", Information and Computer Science Department, University of Hawaii at Manoa Honolulu, disponible en <https://senin.github.io/assets/pubs/senin_dtw_litreview_2008.pdf>.
- Shepard, Roger N. (1984), "Ecological Constraints on Internal Representation: Resonant Kinematics of Perceiving, Imagining, Thinking and Dreaming", en *Psychological Review*, vol. 91, núm. 4, pp. 417-447.

- Shelton, Jose y Gideon P. Kumar (2010), "Comparison Between Auditory and Visual Simple Reaction Times", en *Neuroscience and Medicine*, vol. 1, núm. 1, septiembre, pp. 30-32.
- Shimada, Shoji; Nobuo Hayashi y Shinji Hayashi (1994), "A Clustering Method for Sound Localization Transfer Functions", en *Journal of the Audio Engineering Society*, vol. 42, núm. 7-8, pp. 577-584.
- Shin-ichi, Sato y Ando Yoichi (2002), "Apparent Source Width (ASW) of Complex Noises in Relation to the Interaural Cross-correlation Function", en *Journal of Temporal Design in Architecture and the Environment*, vol. 2, núm. 1, pp. 29-32.
- Smalley, Dennis (1997), "Spectromorphology: Explaining Sound-Shapes", en *Organised Sound*, vol. 2, núm. 2, Cambridge University Press, pp. 107-126.
- Smalley, Dennis (2007), "Space-Form and the Acousmatic Image", en *Organised Sound*, vol. 12, núm. 1, Cambridge University Press, pp. 35-58.
- Soria, Edmar (2014), "Espacialización Multicanal Algorítmica: un modelo teórico con implementaciones en Supercollider", tesis de maestría en Música, México, UNAM.
- Soria, Edmar (2020), "Espacio, virtualidad e inmersividad como experiencia estética: un análisis desde la percepción sensorial", en *Espacio inmersividad. Miradas desde la transversalidad filosofía-arte-ciencia-tecnología*, Ciudad de México, Juan Pablos/UAM Lerma.
- Spickler, Don y Jennifer Bergner (2011), *The Mathematics Behind Anamorphosis*, Salisbury, Salisbury University.
- Spivak, David (2014), *Category Theory for Sciences*, Cambridge, MA, MIT Press.
- Stefani, Ewan y Karen Lauke (2010), "Music, Space and Theatre: Site-Specific Approaches to Multichannel Spatialisation", en *Organised Sound*, vol. 15, núm. 3, diciembre, pp. 251-259.
- Sternberg, Shlomo (2000), *Dynamical Systems*, Nueva York, Dover Publications.
- Suzuki, Yōiti; Hae-Young Kim, Shouichi Takane y Toshio Sone (1998), "A Modeling of Distance Perception Based on Auditory Parallax Model (Abstract)", en *The Journal of the Acoustical Society of America*, vol. 103, núm. 5, 3083, disponible en <<https://doi.org/10.1121/1.422911>>.
- Tashev, Ivan (2009), *Sound Capture and Processing: Practical Approaches*, Nueva York, John Wiley & Sons Inc.
- Taney, James (1988), *Meta-Hodos and Meta Meta-Hodos: A Phenomenology of 20th Century Musical Materials and an Approach to the Study of Form*, Frog Peak Music, 2a. ed., Oakland, California, Frog Peak Music.
- Termens, Enric Gaus i (2009), "Audio Content Processing for Automatic Music Genre Classification: Descriptors, Databases and Classifiers", tesis de doctorado, Barcelona, Department of Information and Communication Technologies, Universitat Pompeu Fabra, disponible en <<https://www.tdx.cat/bitstream/handle/10803/7559/tegt.pdf?sequence=1>>.
- Theile, Günther (1980), "On the Localisation in the Superimposed Soundfield", tesis de doctorado, Technische Universität Berlin, disponible en <https://hauptmikrofon.de/theile/1980-2_Diss_Theile_english.pdf>.
- Theile, Günther, Helmut Wittek (2004), Wave Field Synthesis: A Promising Spatial Audio Rendering Concept, en *Acoustical Science and Technology*, vol. 25, núm. 6, pp. 393-399.

- Treisman, Anne M. y Garry Gelade (1980), "A Feature-Integration Theory of Attention", en *Cognitive Psychology*, vol. 12, núm. 1, pp. 97-136.
- Tzortzis, Grigorios y Aristidis Likas (2008), "The Global Kernel k-Means Clustering Algorithm", en *IEEE International Joint Conference on Neural Networks (World Congress on Computational Intelligence)*, disponible en <https://www.researchgate.net/publication/224330932_The_Global_Kernel_k-Means_Clustering_Algorithm>.
- Ulyanov, Dmitri y Vadim Lebedev (2017), "Audio Texture Synthesis and Style Transfer", disponible en <<https://dmitryulyanov.github.io/audio-texture-synthesis-and-style-transfer/>>.
- Usman Liaquat, Muhammad; Hafiz Suliman Munawar, Amna Rahman, Zakria Qadir, Abbas Z. Kouzani y M. A. Parvez Mahmud (2021), "Localization of Sound Sources: A Systematic Review", en *Energies*, vol. 14, núm. 13, 3910, MDPI-Energies.
- Van de Gorne, Annette (1998), "Space, Sound and Acousmatic Music", en *Academia*, disponible en <https://www.academia.edu/9065377/Space_sound_and_acousmatic_music>.
- Vilkamo, Juha y Symeon Delikaris-Manias (2015), "Perceptual Reproduction of Spatial Sound Using Loudspeaker-Signal-Domain Parametrization", en *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, núm. 10, pp. 1660-1669.
- Wang, DeLiang, Guy J. Brown (2005), *Computational Auditory Scene Analysis*, Nueva York, John Wiley & Sons, Inc.
- Wang, Yuxuan *et al.* (2017), "Tacotron: Towards End-to-End Speech Synthesis", en *Interspeech*, disponible en <https://www.isca-speech.org/archive_v0/Interspeech_2017/pdfs/1452.PDF>.
- Warren, Richard M. (2008), *Auditory Perception An Analysis and Synthesis*, Cambridge, Cambridge University Press
- Watanabe, Satoshi (1960), "Information-Theoretical Aspects of Inductive and Deductive Inference", en *IBM Journal of Research and Development*, vol. 4, núm. 2, pp. 208-231.
- Wightman, Frederic L. y Doris J. Kistler (1997), "Monaural Sound Localization Revisited", en *The Journal of the Acoustical Society of America*, vol. 101, núm. 2, pp. 1050-1063, disponible en <<https://doi.org/10.1121/1.418029>>.
- Wightman, Frederic L. y Doris J. Kistler (1999), "Resolution of Front-Back Ambiguity in Spatial Hearing by Listener and Source Movement", en *The Journal of the Acoustical Society of America*, vol. 105, núm. 5, pp. 2841-2851, disponible en <<https://doi.org/10.1121/1.426899>>.
- Windsor, W. Luke (1997), "Frequency Structure in Electroacoustic Music: Ideology, Function and Perception", en *Organised Sound*, vol. 2, núm. 2, pp. 77-82.
- Xie, Bosun (2013), *Head-Related Transfer Function and Virtual Auditory Display*, Florida, J. Ross Publishing.
- Xie, Bosun (2000), "Principle and Misunderstanding on Virtual Surround Sound", en *Audio Engineering (in Chinese)*, vol. 2, pp. 8-13.
- Xie, Bosun (2002), "Interchannel Time Difference and Stereophonic Sound Image Localization", en *Acta Acústica*.

- Xie, Bosun y X. F. Xie (1996), "Analyse and Sound Image Localization Experiment on Multi-Channel Plannar Surround Sound System", en *Chinese Journal of Acoustics*, vol. 15, núm. 1, pp. 52-64.
- Xie, Bosun y Yu Guangzheng (2021), "Psychoacoustic Principle, Methods, and Problems with Perceived Distance Control in Spatial Audio", en *Applied Sciences*, vol. 11, núm. 23, 11242, MDPI.
- Yagi, Yasuo; Kerry L. Coburn, Kristi M. Estes y James E. Arruda (1999), "Effects of Aerobic Exercise and Gender on Visual and Auditory P300, Reaction Time, and Accuracy", en *European Journal of Applied Physiology and Occupational Physiology*, vol. 80, núm. 5, pp. 402-408.
- Yalta, Nelson; Kazuhiro Nakadai y Tetsuya Ogata (2017), "Sound Source Localization Using Deep Learning Models", en *Journal of Robotics and Mechatronics*, vol. 29, núm. 1, pp. 37-48.
- Zahorik, Pavel; Douglas S. Brungart y Adelbert W. Bronkhorst (2021), "Auditory Distance Perception in Humans: A Summary of Past and Present Research", en *Acta Acustica united with Acustic*, vol. 91, núm. 3 (2005), pp. 409-420.
- Zelli, Bijan (2010), "Musique Acousmatique and Imaginary Spaces", Canadian Electroacoustic Community, disponible en <http://www.bijanzelli.com/Musique%20Acousmatique%20and%20Imaginary%20Spaces_econtact13.3.pdf>.
- Zhao, Chunchun y Sartaj Sahni (2019), "String Correction Using the Damerau-Levenshtein Distance", en *BMC Bioinformatics*, vol. 20, suplemento 11, 277, disponible en <<https://doi.org/10.1186/s12859-019-2819-0>>.

Espacio-Timbre.
Hacia una teoría interdisciplinar
en la música electroacústica
se terminó en enero de 2023
en Juan Pablos Editor, S.A. de C.V.
2a. Cerrada de Belisario Domínguez 19
Col. del Carmen, Alcaldía Coyoacán
04100, Ciudad de México
<juanpabloseditor@gmail.com>

Libro electrónico



La composición electroacústica multicanal hace uso enfático de dos parámetros en particular: el timbre y el espacio. Si bien éstos han sido bastante explorados de manera heurística y práctica, no existen demasiados modelos estructurados que expliquen dichos elementos composicionales de manera conjunta e integral desde los fundamentos de la acústica, la psicoacústica y la percepción. Se plantea aquí la hipótesis fundamental de que el espacio y el timbre están inexorablemente ligados de modo tal que, dentro del contexto específico perceptual-conceptual-teórico, pueden ser considerados como un sólo atributo composicional: el Espacio-Timbre. En este sentido, el presente libro plantea, por lo tanto, un modelo teórico original que hace uso de resultados específicos provenientes de estas áreas de estudio, para construir y proponer una definición original de la espacialidad y del timbre, vistos como un fenómeno integrado, desde una perspectiva teórico-abstracta y conceptual. Todo lo anterior es posteriormente estructurado dentro de un modelo matemático basado en el lenguaje de sistemas dinámicos el cual, debido a su generalidad, permitirá describir y explicar cualquier caso particular referente a la relación Espacio-Timbre. El desarrollo de los elementos del Espacio-Timbre están guiados, por lo tanto, bajo esta perspectiva, por la búsqueda del desarrollo de la inmersividad en el proceso composicional como fin último de la experiencia aural de la obra.

DOCUMENTOS

DAH

Los documentos del Departamento de Artes y Humanidades (DAH) exploran las prácticas, metodologías, historias, teorías, curadurías y arqueologías del arte digital. Estos documentos buscan aportar elementos a la arena de discusión en torno a las humanidades y las artes digitales. Escritos y/o coordinados por investigadores y creadores del DAH, cada uno de estos títulos explora distintas vías del arte de nuestro tiempo.

Edmar Olivares Soria

Obtuvo un doctorado y una maestría en Tecnología Musical por la UNAM, una licenciatura en Matemáticas en el Instituto Politécnico Nacional y otra maestría en Economía Aplicada por la UNAM. Actualmente realiza un posdoctorado en Inteligencia Artificial (II-MAS-UNAM). Es profesor investigador en el Departamento de Artes y Humanidades en la UAM, Unidad Lerma, México, donde también se desempeña como jefe del área de investigación PiATS (Práctica como investigación en el Arte, Transdisciplina y Sonido). Ganador del Concurso Acousmonium INA GRM-Fonoteca Nacional 2016 (Francia-México), del concurso SONOM 2014 (Festival Internacional de Arte Sonoro), finalista del “Concours International de Composition Electroacoustique SIME 2018” (Lille, Francia) y becario de los Programas Creadores Escénicos 2018 (Fonca, México), Resiliencias Sonoras-Composición Electroacústica 2020 (UNAM México) y ECOS SONOROS 2021 (México), ha realizado residencias artísticas en dos de los estudios de música acusmática más importantes del mundo: Musique & Recherches (Bélgica) y el INA GRM (París, Francia). Ha realizado residencias de composición de música electroacústica y arte digital en DXArts (El Centro de Artes Digitales y Medios Experimentales, Universidad de Washington), el Conservatorio de Belo Horizonte-Brasil (bajo la dirección del compositor Joao Pedro Oliveira), y el CMMAs (Morelia, México). Ha sido reconocido internacionalmente mediante comisiones oficiales para composiciones y performances multimedia por el Difrazzoni Festival 2016 (Florencia, Italia), Multiphonies GRM 2017 (Francia) y New York University Music Ensemble (2017), así como en Bienales de Arte Digital y sus obras han sido presentadas en diversos festivales nacionales e internacionales. Como investigador académico, actualmente pertenece al Sistema Nacional de Investigadores (SNI-Conacyt) y sus áreas de trabajo incluyen la espacialidad/localización del sonido, visualización/sonificación de datos, cómputo, matemáticas y tópicos de inteligencia artificial aplicados al arte digital y la música y tiene distintas publicaciones al respecto. Es director y fundador del Coloquio Internacional Espacio Inmersividad que cuenta hasta ahora con tres emisiones (2018-2019-2020) y de Desfases, Festival Inmersivo de Producción Multimedia (2021-2022).