



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN MÚSICA

TECNOLOGÍA MUSICAL

ESPACIO-TIMBRE: UNA FORMULACIÓN TEÓRICA EN LA ELECTROACÚSTICA MULTICANAL

TESIS

QUE PARA OPTAR POR EL GRADO DE
DOCTOR EN MÚSICA (TECNOLOGÍA MUSICAL)

PRESENTA:

EDMAR OLIVARES SORIA

TUTOR O TUTORES PRINCIPALES

ROBERTO MORALES MANZANARES (UNIVERSIDAD DE GUANAJUATO)

RODRIGO SIGAL SEFCHOVICH (ENES MORELIA, UNAM)

JUAN SEBASTIÁN LACH (ENES MORELIA, UNAM)

MIEMBROS DEL COMITÉ TUTOR

MANUEL ROCHA ITURBIDE (UAM LERMA)

ENRIQUE RUÍZ VELASCO SÁNCHEZ (IISUE UNAM)

EDUARDO MORALES MANZANARES (INAOE)



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Espacio-Timbre: una formulación teórica en la electroacústica multicanal

por

Edmar Olivares Soria

Mtro. en Tecnología Musical.
Lic. en Física y Matemáticas.

Tesis presentada para obtener el grado de

Doctor en Tecnología Musical

en la

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad de México. Enero, 2020

*A mi hijo Sayab.
A mis padres: María Teresa y Edmundo.*

Agradecimientos

Por ser maestros y verdaderos mentores en mi andar dentro de la música electroacústica: **Elizabeth Anderson, Roberto Morales Manzanares, Rodrigo Sigal, Manuel Rocha Iturbide, Esther Lamneck.**

Por su apoyo para con este trabajo de investigación y su guía en el mundo académico y del arte: **Juan Sebastián Lach, Enrique Ruíz Velasco, Eduardo Morales Manzanares, Luis Alfonso Estrada Rodríguez, Ricardo Dal Farra, Luz María Sánchez Cardona.**

Por compartir conmigo su gran experiencia y conocimiento en momentos específicos y en lugares específicos del mundo: **Beatriz Ferreira, Annete Vande Gorne, Joao Pedro Oliveira, Panayiotis Kokoras, Juan Pampín, Miller Puckette, Chris Chafe, Daniel Teruggi, Pablo Di Liscia, Diego Losa.**

Índice general

1. Percepción como fundamento de la experiencia de la espacialidad	1
1.1. Fundamentos generales sobre percepción.	2
1.1.1. Percepción y tiempos de reacción	6
1.2. Principios básicos de la Teoría Gestalt en la percepción	8
1.2.1. Analogía relacional	11
1.2.2. Sumario de definiciones y conceptos	12
1.3. Percepción estructural sonora	14
1.4. Análisis de escena auditiva	14
1.4.1. Modos de agrupamiento: simultáneo y secuencial.	17
1.4.2. Atributos	23
1.5. La teoría Gestalt de la música de James Tenney	24
1.6. Localización espacial del sonido	30
1.6.1. Difusión espacial y localización audible del sonido	33
1.6.2. Bandas direccionales y localización vertical	36
1.6.3. Propiedades acústicas de la espacialidad en la experiencia aural	38
1.6.4. Fuentes sonoras en movimiento y fuentes múltiples	42
2. Espacio y espacialidad	45
2.1. Construcción teórica del concepto de Espacio	46
2.1.1. Antecedentes sobre uso del espacio en la música	47
2.1.2. El concepto de espacio: espacio sonoro/musical y espacialidad	48
2.1.3. Espacio Interno - Espacio Externo y el problema de la disimilaridad serial/perceptual.	51
2.1.4. Espacio virtual, espacialidad e inmersividad	53

2.1.5.	Espacio virtual e inmersividad en la electroacústica multicanal	56
2.1.6.	Consideraciones generales sobre el espacio audible	58
2.1.7.	Diseño Serial vs Percepción	62
2.2.	Componentes primarios de la espacialidad	64
2.2.1.	Objetos sonoros fijos	67
2.2.2.	Trayectorias continuas u objetos sonoros móviles	69
2.2.3.	Superficies sonoras y involucimiento	71
2.2.4.	Observaciones finales sobre la descripción de espacialidad	72
3.	Espacio-Timbre: un modelo teórico.	74
3.1.	Perfil y Espacio Morfológico	76
3.2.	Perfil y Espacio Espectral	80
3.3.	Perfil y Espacio Espectromorfológico	81
3.3.1.	Distancia y medidas en el espacio espectromorfológico	82
3.4.	Espacio Timbre: construcción abstracta	84
3.4.1.	Donde la espacialidad y el timbre se unen.	87
3.5.	Ejemplos	89
3.5.1.	Ejemplo I	93
3.5.2.	Ejemplo II	94
3.5.3.	Ejemplo III	96
4.	Descripción espectromorfológica computacional y clasificación mediante aprendizaje máquina.	99
4.1.	Análisis primario de muestras	102
4.2.	Tratamiento de las series de tiempo y optimización preclasificatoria.	105
4.3.	Clasificación de series de tiempo mediante aprendizaje no supervisado.	111
4.4.	Formación de centroides mediante baricentros y bandas LB Keogh	118
4.5.	Clasificación de muestras de audio arbitrarias	125
4.6.	Conclusiones del capítulo	135
5.	Análisis de obra: <i>Vortex Flow</i>.	138
5.1.	Herramientas tecnológicas utilizadas para el diseño espacial	139

5.1.1.	VBAP	140
5.1.2.	Ambisonics	142
5.1.3.	Paneo multicanal por amplitud	142
5.1.4.	Microfonía multiarray	143
5.1.5.	Comparación de las herramientas.	145
5.2.	Materiales sonoros	147
5.3.	Diseño y construcción de la espacialidad.	149
5.3.1.	Sección I	150
5.3.2.	Sección II	152
5.3.3.	Sección III	152
5.4.	Comentarios finales	153
6.	Conclusiones	154
6.1.	Trabajo a futuro	156
7.	Apéndice	158
7.1.	Fundamentos de Sistemas Dinámicos	158
7.2.	Algoritmos computacionales.	160
7.2.1.	Dynamic Time Warping	160
7.2.2.	Alineación Global de Kernel k-Medios	162
7.2.3.	Bandas LB Keogh	164
7.2.4.	Métricas de similaridad para cadenas de caracteres.	165
7.3.	Entropía aproximada.	166
7.4.	Clasificación no supervisada utilizando medidas de tendencia central.	167
7.4.1.	Simulación Base	168
7.4.2.	Simulación 2	172
7.4.3.	simulación 3	175
7.4.4.	simulación 5	179
7.4.5.	simulación 6	183
7.4.6.	Simulaciones con el algoritmo propagación de afinidad	185
7.4.7.	Interpretación de los resultados	188

Índice de figuras

1-1. Representación diagramática de las TGU.	29
1-2. Sistema de coordenadas interaural polar (Letowski, 2001).	32
2-1. Representación gráfica del espacio audible y de la extensión del espacio virtual sonoro. . .	59
2-2. <i>Sonic Arts Research Centre</i> , Belfast.	59
2-4. <i>Birmingham Electroacoustic Sound Theatre</i> . BEAST.	60
2-3. Hexagram-UQAM ESPACE. Montreal.	60
2-5. CCRMA <i>3D Listening Room</i> . Stanford.	61
3-1. Espectrograma para muestra de audio I.	91
3-2. Espectrograma para muestra de audio II.	91
3-3. Forma de onda para muestra de audio I.	92
3-4. Forma de onda para muestra de audio II.	92
3-5. Series de tiempo del zcr para los objetos sonoros I y II.	96
3-6. Series de tiempo del centroide espectral para los objetos sonoros I y II.	98
4-1. Comparación de series de tiempo para RMSE de audio de corno francés. Nota A2 tocada en mezzo forte. . . .	108
4-2. Comparación de series de tiempo para RMSE de audio de corno francés. Nota A2 tocada en fortissimo. . . .	108
4-3. Comparación de series de tiempo para RMSE de audio de corno francés. Nota A3 tocada en trino forte. . . .	109
4-4. Identificación de onsets en la serie de tiempo.	110
4-5. Reconstrucción de la serie de tiempo a partir de la interpolación de onsets.	110
4-6. Cluster no. 2 de series de tiempo para la subbanda 4 (contraste espectral) de muestras de audio de flauta. . . .	113
4-7. Cluster no. 3 de series de tiempo para la subbanda 4 (ancho espectral) de muestras de audio de corno inglés. . .	114
4-8. Cluster no. 3 de series de tiempo para el zcr de muestras de audio de saxofón.	114

4-9.	Cluster no. 0 de series de tiempo para el zcr de muestras de audio de oboe.	115
4-10.	Cluster no. 5 de series de tiempo para el flujo espectral de muestras de audio de clarinete.	115
4-11.	Cluster no. 5 de series de tiempo para el zcr de muestras de audio de mandolina.	116
4-12.	Cluster no. 3 de series de tiempo para el flujo espectral de muestras de audio de guitarra.	116
4-13.	Cluster no. 2 de series de tiempo para el flujo espectral de muestras de audio de cello.	117
4-14.	Cluster no. 0 de series de tiempo para el flujo espectral de muestras de audio de banjo.	117
4-15.	Baricentro euclideano para cluster de flujo espectral de muestras de audio de violín.	120
4-16.	Centroide baricentro/bandas LB Keogh para subbanda 2 de ancho espectral de muestras de audio de clarinete .	121
4-17.	Centroide baricentro/bandas LB Keogh para subbanda 2 de ancho espectral de muestras de audio de contrabajo.	121
4-18.	Centroide baricentro/bandas LB Keogh para subbanda 2 de ancho espectral de muestras de audio de tuba.	122
4-19.	Centroide baricentro/bandas LB Keogh para zcr de muestras de audio de trombón.	122
4-20.	Centroide baricentro/bandas LB Keogh para zcr de muestras de audio de tuba.	123
4-21.	Centroide baricentro/bandas LB Keogh para zcr de muestras de audio de trompeta.	123
4-22.	Centroide baricentro/bandas LB Keogh para flujo espectral de muestras de audio de clarinete bajo.	124
4-23.	Centroide baricentro/bandas LB Keogh para flujo espectral de muestras de audio de clarinete banjo.	124
4-24.	Comparación de muestra 1 para baricentros de zcr.	127
4-25.	Comparación de muestra 1 para baricentros de flujo espectral.	127
4-26.	Comparación de muestra 2 para baricentros de zcr.	128
4-27.	Comparación de muestra 2 para baricentros de subbanda 2 de contraste espectral.	128
4-28.	Comparación de muestra 3 para baricentros de zcr.	129
4-29.	Comparación de muestra 3 para baricentros de subbanda 2 de contraste espectral.	129
4-30.	Comparación de muestra 4 para baricentros de flujo espectral.	130
4-31.	Comparación de muestra 4 para baricentros de zcr.	130
4-32.	Forma de onda de la muestra 1	132
4-33.	Espectrograma de la muestra 1.	132
4-34.	Cluster correspondiente al baricentro más cercano en zcr de la muestra 1	133
4-35.	Cluster clasificado interiormente mediante propagación de afinidad, correspondiente al baricentro más cercano en zcr de la muestra 1	133
4-36.	Histograma de representación de la muestra 1 para el baricentro más cercano en zcr	134
7-1.	Pseudocódigo para el algoritmo de DTW (Senin, 2009)	161

7-2.	Documentación para el algoritmo DTW en la librería tslearn.	162
7-3.	Pseudocódigo para el algoritmo de Kernel Global de K Medios (Tzortzis, 2008).	163
7-4.	Documentación para el algoritmo GAK de tslearn.	164
7-5.	Documentación para el algoritmo GAK de tslearn.	165
7-6.	Métrica Levinshtein	165
7-7.	Algoritmo para calcular la entropía aproximada. Fuente (Wikipedia)	166
7-8.	Número de componentes óptimo para PCA vs varianza explicada	169
7-9.	Método del codo, k-medios.	169
7-10.	Clustering de 10 etiquetas	170
7-11.	Clustering de 10 etiquetas	170
7-12.	Clustering de 10 etiquetas	171
7-13.	Número de clusters óptimos para k-medios	173
7-14.	Clustering para 8 etiquetas.	173
7-15.	Clustering para 8 etiquetas.	174
7-16.	Clustering para 8 etiquetas.	174
7-17.	Número de clusters óptimos para k-medios	175
7-18.	Clustering para 6 etiquetas.	175
7-19.	Clustering para 6 etiquetas.	176
7-20.	Clustering para 6 etiquetas.	176
7-21.	Matriz de correlación para los descriptores utilizados.	177
7-22.	Número de componentes óptimos para PCA.	177
7-23.	Número de clusters óptimos para k-medios	178
7-24.	Clustering para 10 etiquetas.	178
7-25.	Clustering para 10 etiquetas.	179
7-26.	Clustering para 10 etiquetas.	179
7-27.	Número de componentes óptimos para PCA.	180
7-28.	Número de clusters óptimos para k-medios	180
7-29.	Clustering para 10 etiquetas.	181
7-30.	Clustering para 10 etiquetas.	181
7-31.	Clustering para 10 etiquetas.	182
7-32.	Clustering para 11 etiquetas.	183

7-33. Clustering para 11 etiquetas.	184
7-34. Clustering para 11 etiquetas.	184
7-35. Comparación de ejecuciones del algoritmo: número de muestras vs clusters obtenidos . . .	186
7-36. Comparación de ejecuciones del algoritmo: número de muestras vs clusters obtenidos . . .	186
7-37. Clustering para 15 etiquetas.	187
7-38. Clustering para 15 etiquetas.	187
7-39. Clustering para 15 etiquetas.	188
7-40. Clustering para 8 etiquetas.	189
7-41. Clustering para 12 etiquetas.	190
7-42. Clustering para 15 etiquetas.	190

Prólogo

La composición electroacústica multicanal es en sí misma, una formulación estética cuyo desarrollo histórico ha estado íntimamente relacionado con la tecnología y con algunos campos del saber como la fenomenología, las ciencias cognitivas y por supuesto la acústica. Esta práctica (la música electroacústica multicanal) se posiciona irremediablemente como un punto de convergencia multidisciplinaria que además produce dinamismo sinérgico con el fin de desarrollar expresiones artísticas enmarcadas en el campo sonoro.

Dentro de la música electroacústica resaltan dos parámetros primordiales en el lenguaje estético de esta expresión artística: *espacio* y *timbre*. La exploración tímbrica se ha posicionado como un elemento predominante en la música de la segunda mitad del siglo XX y es posible afirmar que en cierto modo, ha dirigido a su vez, la composición del parámetro del espacio dentro de la electroacústica multicanal.

El espacio como elemento composicional con una identidad individual, ha sido desarrollado históricamente de manera más bien fragmentada y no sería exagerado afirmar que no existe hoy en día una consolidación homogénea respecto al mismo; ni desde un planteamiento teórico ni desde una formulación explicativa fenomenológica. Parece ser que el desarrollo más notable se ha concentrado en el diseño de herramientas digitales de *espacialización* y en las consecuencias directas en el uso de estas. En este sentido, si bien existen numerosos compositores que desde la segunda mitad del siglo pasado han abordado y desarrollado diversas conceptualizaciones del espacio, no han sido del todo desarrollados, marcos teóricos que aborden el uso del espacio como elemento composicional en la música electroacústica desde una mirada integral compuesta por elementos perceptuales, acústicos y teóricos,

Actualmente el espacio como parámetro dentro de la música electroacústica es considerado algo casi intrínseco e inherente a casi cualquier obra que sea presentada en formato multicanal, sin embargo, si bien existen numerosos recursos académicos realizados por compositores que describen sistemáticamente el parámetro del espacio dentro de la composición, estos recursos no siempre son llevados a la práctica al momento de la creación artística. Podría afirmarse que en no pocos casos, el nivel de detalle y exploración es más superficial en el caso del espacio que en el caso de otros parámetros como el timbre, el ritmo o la dinámica.

Hoy día parece existir una marcada tendencia a diseñar la espacialización de modo serial¹, (en el sentido de Stockhausen) de forma tal que el diseño espacial es frecuentemente basado casi por completo en el uso procesual de herramientas digitales de espacialización como *plug ins* y *VST's* y utilizando infinidad de modelos y fuentes de información que después son traducidos en forma de datos a parámetros asociados a la posición espacial y movimiento de objetos sonoros. Si bien esta práctica aporta mucho a la creación de la obra, deriva en una problemática concreta y muy común; el cuestionamiento de ¿qué tanto de lo que se diseña de manera *serial* sobre la espacialización es percibido en la realidad por el espectador? En otras palabras, ¿que tan alta es la resolución espacial de la obra electroacústica de acuerdo al prediseño serial?

¹El uso del término serial tiene el sentido descrito en (Bates 2009) donde se hace referencia a un diseño y manejo de los parámetros composicionales (y en particular de l espacio) enfocado al uso de procesos o heurísticas que están en cierto grado disociadas del resultado final perceptual.

Este cuestionamiento plantea una perspectiva clave alterna hacia el diseño espacial que está relacionada de manera conjunta con el diseño espectromorfológico dentro del proceso compositivo de la obra electroacústica. Dicho cuestionamiento tiene que ver con lo que se denominará *disimilitud serial-perceptual* el cual trata precisamente sobre la diferencia que existe entre el prediseño serial de los parámetros (con un énfasis en el timbre y el espacio) y el resultado final percibido de dicho diseño.

Derivado de la problemática anterior surge entonces el planteamiento propositivo de que una comprensión alternativa al proceso compositivo electroacústico multicanal puede estar basada en el uso informado de la acústica y la psicoacústica para la organización del material sonoro tomando como referente su perfil espectromorfológico en relación con la capacidad humana de localización espacial del sonido.

Dentro del marco teórico propuesto en este trabajo está entonces la hipótesis de que el espacio y el timbre están inexorablemente ligados de modo tal que puedan ser considerados como un sólo atributo compositivo: *espacio-timbre*. Esta perspectiva está basada sobre todo en el planteamiento de una estructura teórica que explique dicho fenómeno (el espacio timbre) a través de la integración en un sólo corpus, de elementos teóricos perceptuales, acústicos y psicoacústicos .

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*.

Posteriormente se propone una caracterización general a nivel teórico-matemático para cada uno de dichos componentes. Esta caracterización está basada en el lenguaje de sistemas dinámicos desde la mirada de la topología. Dicha herramienta matemática resulta por demás útil para establecer una estructura teórica capaz de *organizar* el concepto de *timbre*, de *espacio* y de *espacio-timbre* de una manera lo más general y formal posible.

La presente propuesta puede ser entendida entonces, como un modelo teórico original que describe el concepto de *espacio-timbre* desde una perspectiva abstracto-fenomenológica construida a partir de tres aspectos primordiales:

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

Los dos primeros aspectos anteriores se entienden por lo tanto como la referenciación acústico-perceptual de los diversos métodos, estrategias y procesos compositivos en torno al uso del timbre y el espacio que son bien conocidos dentro de la música electroacústica; i.e representan una formalización o *explicación acústico-perceptual informada* de dichos elementos dentro del proceso creativo.

Estos elementos como ya se dijo, son bastante bien conocidos y manejados por los creadores relacionados al formato multicanal, sin embargo, no existen hasta este punto, demasiadas formulaciones teóricas que expliquen el uso de estos elementos (los procesos y técnicas relacionadas al espacio y el timbre) mediante resultados perceptuales y acústicos y que además, los agrupen en un modelo como el aquí propuesto.

El tercer elemento; la representación abstracto-matemática, es la estructuración de los dos elementos anteriores mediante sistemas dinámicos. Estos objetos matemáticos son a juicio del autor de este trabajo, una de las herramientas más adecuadas para dicha representación abstracta debido a su flexibilidad para modelar y representar el comportamiento de cualquier objeto a través del tiempo, que es precisamente el objetivo de esta sección: representar de manera abstracta los espacios matemáticos correspondientes al timbre (perfil espectromorfológico) y la inmersividad (espacio audible-espacio virtual sonoro-espacio aural) a través del tiempo. Toda vez que se han construido estos espacios matemáticos por separado, haciendo uso de la *propiedad universal del producto de conjuntos* se obtendrá la representación matemática del *espacio-timbre*.

Los dos primeros capítulos de este trabajo contienen los resultados acústico-perceptuales acerca de la audición y la localización espacial, así como del fenómeno de percepción en general. Dentro de cada uno de estos capítulos existe una sección final en la cual se establece a modo de definiciones, observaciones

y conjeturas, los resultados más relevantes para el presente trabajo los cuales son propuestas originales que funcionan dentro del corpus de la presente investigación para cada capítulo en particular.

En el primer capítulo se describirán diversos enfoques relacionados a 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 correspondiente procesos cognitivos a través de herramientas como el análisis de escena auditiva, la psicoacústica del sonido espacial y la teoría de análisis musical de James Tenney.

En el capítulo dos se planteará 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 *immersividad* y se desarrollarán los *componentes primarios de la immersividad*. Se definirá de igual modo, el *perfil espectromorfológico* para concluir con la representación matemática de todos los elementos y del llamado *espacio-timbre*.

El capítulo tres corresponde a la 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 a este espacio de referencia y asociarla por grado de similaridad, a una espectromorfología de algún instrumento musical convencional de modo tal que dicha muestra puede ser tratada en cierto grado, de acuerdo a los principios de orquestación y acústica conocidos para tal instrumento; esto es lo que se nombrará *orquespectración virtual*.

El capítulo final es un análisis sobre el diseño espacial desarrollado en la obra *Vortex Flow* (2018). Este análisis incluye una descripción de la manera en que fueron aplicados de manera concreta, diversos resultados y conceptos expuestos en esta investigación. Del mismo modo, describe los elementos y condiciones que detonaron la necesidad de profundizar, contrastar y explorar nuevos elementos teórico-prácticos orientados sobre todo al uso de clasificación no supervisada para clustering de muestras y aplicación de principios psicoacústicos y acústicos para el diseño espacial, guiando por lo tanto, a modo de laboratorio de prueba, error y comprobación, la línea general de investigación del presente trabajo.

Espacio-Timbre: una formulación teórica en la electroacústica multicanal

por

Edmar Olivares Soria

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 la acústica y la percepción. El presente trabajo plantea un modelo original teórico que hace uso de resultados acústicos y psicoacústicos relacionados a los campos de la percepción sonora y localización espacial del sonido. Se construye además, una definición de la espacialidad desde una perspectiva fenomenológica-descriptiva. Todo lo anterior es posteriormente estructurado dentro de un modelo matemático abstracto 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 *immersividad* en el proceso composicional como fin último de la experiencia aural de la obra. Finalmente, se presenta 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), planteando con ello el concepto de *orquespectración virtual*.

Capítulo 1

Percepción como fundamento de la experiencia de la espacialidad

El papel de la percepción sensorial (y en particular de la percepción sonora) ha sido someramente abordado de manera formal y extensa en la literatura de la música electroacústica. Si bien los compositores electroacústicos conocen perfectamente diversos fundamentos acústicos/psicoacústicos que explican y desarrollan la sensación de espacialidad en la experiencia aural de una obra, hasta ahora no se ha prestado la suficiente atención a la construcción de una taxonomía o estructura que incorpore dichos elementos teóricos (aquellos asociados a la percepción).

El objetivo principal de este trabajo, tal y como se mencionó en la parte introductoria, es el de construir un modelo teórico que explique el concepto de *espacio-tímbre* basado en tres bloques fundamentales: la percepción (o experiencia aural), 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 establecerá todo lo necesario para desarrollar la parte asociada a la percepción. Se detallarán primeramente, las características básicas del fenómeno de la percepción sensorial analizado desde enfoques sumamente particulares que permitirán posteriormente, construir conceptos y definiciones categóricas acerca de la percepción auditiva. Lo anterior también permitirá definir en capítulos subsecuentes, los conceptos de espacio, espacialidad e inmersividad y permitirá dotar al presente trabajo, de un fundamento formal respecto a la formulación de definiciones y conceptos originales propuestos.

1.1. Fundamentos generales sobre percepción.

El fenómeno de la percepción sensorial implica una reconstrucción del mundo externo mediante un proceso cognitivo que tiene multiplicidad de aristas al momento de ser analizado. Existen numerosas posturas que intentan explicar dicho proceso las cuales incluyen campos como la filosofía de la mente, el idealismo objetivo-subjetivo, las neurociencias, la psicología cognitiva, etc. La percepción sensorial es entonces, una puerta para interactuar 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*. Dicha aseveración encierra de manera sintética una definición de este fenómeno de acuerdo a 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 de dos elementos primordiales: el proceso de representación y aquello que va a ser representado. En otras palabras para extraer información del mundo sensible (en el sentido platónico), 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 incertidumbre¹ 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 a las representaciones mentales finales. Sin embargo, dichas variaciones permanecen debajo de un nivel de tolerancia lo suficientemente funcional 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

¹Incertidumbre entendida aquí desde su acepción proveniente de la física

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 y cuyas características, al menos en una primera etapa, son 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 la percepción que pudiera denominarse como *pasiva* ya que los estímulos sensoriales son señales de información 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 posible diferenciación entre esta *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*.

Por supuesto esta idea de *percepción activa* planteada por Plumbley (2006), deriva directamente del campo de la *percepción máquina*; una subárea de la Inteligencia Artificial la cual trata acerca del estudio de la percepción humana adaptada o implementada en sistemas computacionales. Al respecto de la percepción activa dentro del campo de la percepción máquina, Bajcsy (1988) afirma que *es exploratoria, verificativa, exploradora; la percepción no cae simplemente sobre los sensores como la lluvia en el piso*.

Si bien la discusión de percepción pasiva vs percepción activa es pertinente, parece ser no del todo correcta en el sentido de colocarlas como elementos contrapuestos o quasiexcluyentes. Una mirada alternativa pudiera plantearse como que la percepción pasiva es 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 a un primer nivel básico. En la siguiente etapa, este flujo de información produce una respuesta fisiológica cuyo objetivo puede estar orientado a obtener mayor información acerca del objeto que produce la señal de información o acerca del ambiente así como planear una reacción acorde al estímulo dependiendo de su naturaleza (reacción ecológica)². Dicha respuesta o reacción está direccionada por múltiples factores pero de ese conjunto destacan los condicionamientos evolutivos (estímulos asociados con la supervivencia como amenazas, ausencia o presencia de comida, identificación de posibles parejas reproductivas, etc.) y asociación directa o indirecta sobre estímulos percibidos en el pasado.

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

Al respecto de la percepción relacionada directamente 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 es un proceso que contiene a la percepción pasiva y ésta (la percepción activa) es el fundamento a su vez del proceso de aprendizaje, entendiendo aprendizaje 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 puede ser dividida por etapas.

¿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 por dicho organismo. Dado que la realidad ontológica o la formulación epistemológica de ellos no nos atañe en esta discusión, su existencia es considerada a priori.

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 sino el de establecer una definición de dicho fenómeno que funcione acordemente para fundamentar las características de éste aplicadas al sonido espacial, se resume a continuación todo lo descrito anteriormente:

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 afectores. 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 pero al mismo tiempo está delimitada por ellos. Esta primer etapa de conocimiento asocia los estímulos sensoriales con referentes aprendidos y vivencias experimentadas dirigiendo al proceso cognitivo de este modo, a las primeras formulaciones de significación.

La *Teoría de la atención de la integración de características* (Treisman, 1980), ha sido utilizada y desarrollada principalmente para estímulos visuales, sin embargo, su aplicación a los demás exeroceptores no es extraña. Bajo esta perspectiva se considera que los procesos perceptuales que ocurren en un intervalo entre 200 y 250 milisegundos (y que además son considerados como intrínsecos y fuera de control), son denominados como *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 percepción activa-pasiva debería resultar clara en este punto.

Si la percepción sensorial del mundo externo puede entonces ser entendida de acuerdo a la definición anterior, la siguiente pregunta lógica podría ser: ¿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).

Procesos Top-Down y Bottom-Up

Top-down y *bottom-up* son estrategias de procesamiento de información características de las ciencias de la información 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 en 1969 para automatizar ciertos indicadores del New York Times. El éxito administrativo e ingenieril del proyecto hicieron que el enfoque *top-down* se esparciera por IBM y por el resto de la industria de los computadores. Por su parte, Niklaus Wirth, que entre sus logros está el desarrollo del lenguaje de programación Pascal, publicó en 1971 el artículo *Program Development by Stepwise Refinement*, que funcionó como fuente de influencia importante para 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 (la percepción y 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), donde 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* el cual es también conocido como *data-driven processing* y que implica que el proceso de percepción comienza con el estímulo externo mismo 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 a un nivel primario por los exteroceptores.
2. Procesamiento de dicha información al nivel más básico sin crear relaciones semánticas cognitivas.

El proceso Top-Down 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 que son 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 con el fin de 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) y más aún, dichos principios también pueden ser utilizados para explicar y categorizar diversas ideas propuestas por Tenney acerca de la percepción musical.

1.1.1. Percepción y tiempos de reacción

Uno de los primeros modelos en establecer paralelismos entre las computadoras y el pensamiento humano dentro de la teoría psicológica es el modelo de memoria del *sistema de procesamiento de la infor-*

mación de Attkins (1968). 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 o decaimiento. La analogía con la computación, es que esta etapa puede ser comparada con los dispositivos de entrada de información como mouse, teclado, etc.
- **Almacenamiento a corto plazo.** (STS, *short time store*). Puede ser considerada como la *memoria de trabajo* del individuo; de hecho 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 además no depende necesariamente del tipo de entrada sensorial ocurrida en el SR por lo que 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* (Attkins 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 vista relativo) aunque no de manera definitiva pues 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 al campo de la percepción y sobre todo dentro del intervalo de SR y STS, permitiendo con ello establecer una caracterización cada vez más definida acerca de la etapa de *preatención*. Si bien el tiempo de reacción ha sido estudiado con mayor profundidad para el campo visual al igual que sucede con lo expuesto en las secciones anteriores, su aplicación para otros sentidos, y en particular para el campo audible, sigue siendo 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 tal y como es presentado por Shelton (2010).

³El 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.

Definición 1.1.2 ■ *El tiempo de reacción es el tiempo transcurrido entre la ocurrencia de un estímulo sensorial y la subsecuente respuesta de comportamiento.*

- *El tiempo simple de reacción es el tiempo requerido para un observador de 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 audibles, 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 particulares de los experimentos realizados, sin embargo, el punto principal de esta discusión es el de concebir el 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, por lo que la categoría de cual es percibido antes no tiene importancia para este trabajo.

1.2. 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 y que marcó un referente importantísimo en la historia del desarrollo de la psicología. El objeto primordial de estudio de la Gestalt trata sobre los mecanismos mediante los cuales tendemos a agrupar perceptualmente objetos externos (primordialmente visuales) en estructuras de mayor dimensión organizacional de acuerdo a las relaciones existentes de dichos objetos con sus alrededores. Esta tendencia natural de agrupamiento posiciona a la Gestalt como una teoría holística en cierto modo y hasta cierto punto sistémica en el sentido de Bertalanffy, pero 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 diferente a una colección de sensaciones, piezas o sumatoria de elementos.

En el trabajo de 1920 de Kohler, *Physical Gestalten*, el autor introduce el término *Gestalten*, el cual puede ser definido como *procesos* o fenómenos que revelan propiedades que son más que la suma de sus partes. El fenómeno perceptual se entiende en este contexto como un fenómeno emergente de la interacción de los elementos individuales o partes, donde el tipo más importante de interacción es nuestra tendencia de agrupar dichos elementos de acuerdo a ciertos principios. Kohler (1920) resalta por ejemplo, que *la función específica de una parte individual es percibida de acuerdo a su ubicación relativa a otras partes*.

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 de otros al mismo tiempo* (Wertheimer, 1923).

Uno de los aspectos primarios dentro de la teoría de la 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 se asume que 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 a continuación tal y como son presentados en (Brownie, 2006):

- **Factor de proximidad.** Este principio establece que los objetos que están localizados cercanos unos con otros serán percibidos como que están asociados entre ellos, i.e. pertenecientes a un mismo grupo o a partes de un todo más grande.
- **Factor de similitud.** Este principio afirma que los objetos que comparten propiedades similares se asumen 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 es presentada una serie de elementos complejos, tendemos a percibirlos como partes de un todo más simple* (van Wagner et al. 2006).
- **Factor de dirección** o ley de la continuidad. Afirma que cuando un conjunto de objetos acomodados de modo tal que hagan 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 son percibidos como agregados o anexos.
- **Factor de cerradura**. Los objetos que están cercanos entre sí, son percibidos como parte 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 puedan aplicar de manera simultánea 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 definitorios y que su jerarquía y aplicación es totalmente variable dependiendo tanto del individuo que percibe como del contexto.

1.2.1. Analogía relacional

La *analogía relacional* puede considerarse como un subcampo de estudio de las ciencias cognitivas y en particular del área que se enfoca en el estudio de la *similaridad* y la *analogía* como componentes estructurales del razonamiento tales como 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 que *puede ser vista como un tipo de razonamiento donde el conocimiento es transferido de una situación inicial (llamada **fuer**te) 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) enlista 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 a (Holyoak y Thagard 1995), existen al menos tres criterios específicos bajo los cuales es posible considerar cuándo dos situaciones son 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 de analogía se han realizado como implementaciones computacionales para el reconocimiento automático de elementos visuales, así como modelado computacional de la percepción y aprendizaje humano (sobre todo de nuevo, en el campo visual).

La esencia primordial de la analogía relacional para este trabajo, radica en la formulación concreta del estudio de cómo absorbemos nuevos conceptos con base en 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 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 que por lo general es a través precisamente de este tipo de asociaciones preestablecidas; este problema es central en *la escucha reducida* de Schaeffer, *la emancipación de la disonancia* de Schoenberg y en *la función musical de cualquier sonido* de James Tenney por mencionar algunos. 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.

1.2.2. 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 del autor de este trabajo 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. En este sentido, se pretende alejarse del concepto de *conjetura matemática* el cual posee una carga semántica mucho más rigurosa y profunda.

Definición 1.2.1 Una *escena perceptual* es una ubicación espacio-temporal de duración Δt , en la que un individuo recibe un *torrente multisensorial* a través de sus exteroceptores. 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 primer 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 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 primer 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; esto es, el 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).

1.3. Percepción estructural sonora

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 fundacionales 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 al timbre y al espacio.

Como se mencionó de igual modo, el objetivo de estas secciones no fue el de realizar un análisis profundo y comparativo sobre las distintas teorías de percepción sino únicamente el de documentar los referentes y antecedentes formales que justifiquen las definiciones utilizadas en el presente trabajo.

A continuación se desarrollarán los principios acerca de la percepción sensorial auditiva que fundamentarán el concepto de espacialidad e inmersión 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.

Este capítulo 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 claramente los diversos aspectos, mecanismos y escenarios que constituyen los numerosos procesos de percepción del sonido en su nivel más 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), la que hace al ASA una herramienta 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.

A continuación se desarrollarán los conceptos e ideas clave de la teoría de análisis musical de James Tenney (la cual está basada directamente en los principios Gestalt y está relacionada de manera indirecta con los resultados del ASA).

1.4. Análisis de escena auditiva

El sonido proveniente del entorno y que es captado por nuestros oídos, es 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 a 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 & Woszczyk 2004).

El ASA (Auditory Scene Analysis) es un modelo psicofísico que constituye la base teórica acerca de la investigación de la percepción auditiva. Este modelo fue desarrollado por Albert Stanley Bregman en 1990 en su famoso trabajo *Auditory Scene Analysis: the Perceptual Organization of Sound*. Esta investigación significó un importante referente y base para posteriores desarrollos en los campos de la percepción del habla, sistemas auditivos en humanos y animales, prótesis auditivas y sobresalientemente en el campo de la modelación computacional con el desarrollo de toda una línea de investigación; el *análisis de escena auditiva computacional* (Computational Auditory Scene Analysis, CASA).

De acuerdo a 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 todas 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 importantísima 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, en la de construir representaciones de los sonidos individuales y mediante ellas formar abstracciones relacionadas a los eventos que produjeron dichos sonidos. Además implica entre otras cosas:

- *Poder distinguir un mismo sonido o timbre en diferentes entornos o ambientes*. Entornos distintos poseen diferentes conjuntos de componentes frecuenciales por lo que 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 y de manera inversa, también se previene el caso en el que una palabra única sea

percibida como haber sido hablada 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.

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 agrupa ciertos subconjuntos espectrales como un sólo evento acústico si la diferencia cuantitativa de dichas evidencias es muy pequeña.

El ASA trabaja precisamente mediante una ponderación dinámica de estas *evidencias físico-perceptuales* que incluyen aspectos acústicos como: contenido espectral, localización espacial, sincronía de onsets, morfología de la envolvente de amplitud, etc. De hecho, parece existir fuerte evidencia de que la tendencia primaria del sistema auditivo al percibir una masa compleja de componentes espectrales, es 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 & Woszczyk, 2004).

Bregman (1993) define ciertos procesos básicos utilizados por el escucha para descomponer mezclas sonoras complejas provenientes de cualquier escena audible. Estos procesos están diferenciados en una característica primaria: la presencia o ausencia de *esquemas previamente aprendidos*. En el caso en que la escucha se dé 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 hace uso de un análisis básico de la escena audible de acuerdo a las características acústicas más primarias y genéricas; este proceso es denominado *análisis de escena audible primitivo*.

El *reconocimiento automático* implica una 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 y de este modo segrega dicho esquema de la mezcla sonora. Según Bregman (1993), este proceso es de hecho 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 que en este caso el escucha está concentrado en aislar de manera intencional un esquema de una mezcla compleja sonora. Es decir, dado 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 lo que para poder segregar eventos sonoros específicos de una escena audible (o integrarlos con otros), el ASA hace uso únicamente de las características acústicas más primarias las cuales son descritas en la secciones siguientes.

Por supuesto 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, sino que el proceso de segregación y/o integración se da de una manera condicionada a dichos esquemas.

1.4.1. Modos de agrupamiento: simultáneo y secuencial.

Las escenas audibles que ocurren en cualquier entorno presentan una multiplicidad de fenómenos acústicos que deben ser analizadas por el ASA y en cada caso, de acuerdo a la ponderación de las evidencias físico-perceptuales, realizar alguna de las dos acciones antes mencionadas: segregar o agrupar. Este proceso ocurre para eventos que suceden *simultáneamente* o *sucesivamente* a través del tiempo en la escena audible. Para el primer caso, es decir, cuando existe *agrupamiento simultáneo* es posible poner como ejemplo, la situación en la que tres personas están hablando 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 por supuesto características propias como altura y timbre.

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

Vale la pena hacer notar y recalcar que en cierto modo, el desarrollo teórico del ASA está basado en la premisa de que el sistema auditivo ha evolucionado, se ha adaptado y desarrollado a través de la observación y asimilación de las regularidades y detección de patrones en el entorno, asociando la información sensorial detectada a la *causa más probable* de acuerdo al patrón observado. Lo anterior es denominado *Principio de validez ecológica* (Bregman & Woszyk, 2004), y coincide naturalmente, con el concepto de *reacción ecológica* de Shepard (1961) planteado en la sección anterior.

Agrupamiento simultáneo

Como se mencionó anteriormente, el agrupamiento simultáneo implica, dada una escena audible, la acción de fusionar distintos subconjuntos de componentes frecuenciales que suceden en un mismo instante de tiempo y asociarlos 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 a ciertas circunstancias y restricciones específicas (asociadas a las evidencias físico-perceptuales) funcionan en conjunto (a modo de una ecuación de combinación no lineal) para realizar dicha tarea:

Definición 1.4.1 *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 está basado en la cualidad física de los *sonidos armónicos* (a veces también llamados *periódicos*) cuyo contenido espectral es una serie de armónicos que son 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, etc.

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. Esta es una de las observaciones concluyentes más notables de la teoría del ASA y debe notarse que lógicamente 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 no. 18 mostrada en (Bregman & 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. De hecho, esta observación deriva directamente de nuevo,

de una funcionalidad ecológica ya que en cualquier entorno es altamente improbable que un conjunto de componentes frecuenciales que son múltiplos de una fundamental, provengan de fuentes distintas.

Asincronía de *onsets* y *offsets*.

Esta característica es de suma importancia para el autor y de hecho la presenta en (Bregman 1993) como la primera de 4 *regularidades ambientales* que conforman la explicación fundamental de la teoría del ASA; regularidades utilizadas por el sistema auditivo para resolver precisamente escenas audibles de manera genérica. Esta *regularidad* o característica podría definirse como sigue:

Observación 1.4.1 *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 tomados 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 reverberantes debido a que el detalle granular de los onsets-offsets de diversos sonidos es difuminado por el efecto de la reverberación.

Independencia de envolventes

Observación 1.4.2 *Dada una escena audible, 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 hace una anotación respecto a la evidente similitud de 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.* De hecho 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 1.4.3 *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 de ASA se hace el comentario de 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 en la sección 1.3.3, sonidos con cierta espectromorfología tienen por lo regular zonas restringidas de localización espacial en el azimuth y el plano elevación. Un ejemplo comunmente 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 azimuth 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 está enfocado a sonidos cuyo contenido espectral se encuentra más o menos bien delimitado en bandas definidas (sonidos, los cuales por lo general están 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 sólo objeto perceptual, mientras que sonidos separados en el espectro serán percibidos como individuales.

Agrupamiento secuencial

El agrupamiento secuencial que realiza el ASA tiene que ver 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 son llamadas *torrentes auditivos*.

Definición 1.4.2 *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 al evento. La causa física es denominada **evento acústico** y las unidades auditivas físicas son denominadas sonido.*

De la definición anterior se nota inmediatamente la diferenciación 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 al agrupamiento secuencial es el llamado *efecto streaming* (Bregman 1990) y que permite la introducción de un concepto de suma importancia: la *distancia acústica*. Si bien Bregman no clarifica una definición específica de dicho concepto, sí que 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 & Woszczyk, 2004). De acuerdo a 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 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.** Es decir, la morfología de la amplitud del contenido espectral. Esta característica está asociada 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 confirma después de cierto lapso de tiempo de que dichos estímulos siguen siendo escuchados de manera más o menos continua. De hecho se ha calculado que la percepción separada de dos tonos en distintas regiones frecuenciales se construye en a lo más 4 segundos, y que además puede tomar hasta otros 4 segundos para que esta predisposición desaparezca.

Al igual que en el agrupamiento simultáneo, los factores anteriores no son linealmente jerarquizables y muchas veces de hecho sucede una *competencia* entre ellos. Más aún, de acuerdo a los autores, los últimos tres principios son factores más bien débiles y que pocas veces resultan relevantes por sí mismos, aunque en conjunto con otros factores pueden ayudar al ASA a segregar o agrupar correctamente los estímulos perceptuales externos.

Existe una característica de suma importancia para el agrupamiento secuencial que sirve como recurso utilizado por el ASA; **la suavidad de cambio**⁴ en algunas de las propiedades acústico-perceptuales de las fuentes sonoras. Dicha característica tiene su fundamento en 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 & Woszczyk, 2004).

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

1.4.2. Atributos

Una característica importante para un funcionamiento eficiente de la teoría del ASA, es el hecho de que para poder crear representaciones mentales de torrentes auditivos bien definidos, tendemos a adjudicar características particulares a dichos eventos de acuerdo a preconcepciones que nos son familiares, i.e. utilizamos en gran medida el *proceso de analogía* (como se describió en el capítulo anterior) para poder asociar propiedades específicas a 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 supuesto este cluster variará de individuo a individuo y a través de 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) es mediante 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 es el hecho de 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 en cierto modo, una suerte de pegamento que ejerce cohesión sobre estas propiedades o atributos para mantenerlos unidos y permitir que con base en ello podamos abstraer eventos individuales y por lo tanto identificar las fuentes separadas que corresponden a ellos.

Las propiedades o atributos de los objetos son abstracciones mentales cuya realidad ontológica reside en el proceso cognitivo; son en cierto modo, etiquetas que colocamos en los objetos del mundo externo para poder describirlos y de este modo realizar representaciones mentales de ellos. Esto significa que *las propiedades necesitan pertenecer a algo* (Bregman 1990). En este sentido tal y como ejemplifica el autor, el *stream* o torrente, nos permite realizar agrupaciones de estas propiedades de modo tal que podamos 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 cerca y la otra baja y lejana. Es sólo por el hecho de que los atributos de *cercanía* y de *alto* son agrupados como propiedades de un *stream* y *lejanía* y *bajo* como propiedades de otro, que podemos experimentar la unicidad de dos sonidos individuales más que una masa amorfa de cuatro atributos. Lo anterior puede ser enunciado del modo siguiente:

Definición 1.4.3 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 llamada heurística *old-plus-new*, la cual ha probado ser una de las herramientas fundamentales dentro del ASA para poder resolver escenas auditivas de manera aceptablemente correcta. De manera muy general, la heurística *old-plus-new* 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.

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 que lo anterior puede ser resuelto precisamente de ese modo: como un sonido que continua pero acompañado de los componentes 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 está en cierto modo relacionada con el factor de dirección de la gestalt.

1.5. La teoría Gestalt de la música de James Tenney

James Tenney (1934 - 2006) fue uno de los compositores y teóricos de la música contemporánea, más importantes del siglo XX. Alumno de Cage, de Ruggles y de Varesse, su obra transita diversos estilos que abarcan la música algorítmica, el minimalismo, la música por computadora, el spectralismo 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 y expuesta por primera vez en su tesis de maestría *Meta-Hodos* de 1961.

Catorce años después Tenney publicaría *META Meta+Hodos*, documento que representaría un *intento de organizar ciertas ideas presentadas por primera vez en Meta+Hodos, incorporando reflexiones y revisiones que han emergido desde entonces* (Tenney, 1975). A partir de entonces, profundiza en el tema

con la colaboración de Larry Polansky, desarrollando material teórico de gran importancia y del cual se destaca para propósitos del presente trabajo, los conceptos de métricas morfológicas y la percepción jerárquica gestalt musical.

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

Sustituyendo y extrapolarando parámetros musicales (expuestos en la teoría de Tenney) por parámetros sonoros relacionados al campo de la electroacústica, se planteará un modelo con el fin de establecer una opción de generalización que funcione como marco teórico de análisis y composición. Se hace un énfasis particular a la formalización del parámetro de espacialidad dentro de la música electroacústica y de todos sus elementos constitutivos de acuerdo a la propuesta particular del autor del presente trabajo.

La complejidad sonora y los tone clusters.

Una de las primeras afirmaciones del *Meta+Hodos* es el hecho de 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 por parte del espectador, demandando nuevos niveles de organización y asimilación sonora.

Tenney diferencia 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 el segundo con el agrupamiento secuencial del ASA.

Para el caso de la complejidad sonora, el autor puntualiza sobre el espectro sonoro continuo sobre el 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 está directamente relacionada al 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 puntualiza el cada vez más común 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 a los parámetros tradicionales (atonal, atemático, etc.) 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. En *Estilo y la Idea* Schoenberg (1950) afirma que:

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 básicamente que todos los diferentes tipos de sonidos (específicamente los descritos de acuerdo a la escala de complejidad tímbrica) tienen la misma potencialidad de ser usados como materiales elementales de construcción musical (Tenney, 1961). La consecuencia inmediata del principio anterior de acuerdo a Tenney, es el hecho de 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.

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

- la extensión o rango de dicho parámetro 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 dicho parámetro a través del tiempo. Esto es, qué tanto se introducen variaciones del rango del parámetro dentro de intervalos de tiempo específicos.

El timbre es el parámetro que pudiera ser considerado como aquel cuyo rango se ha extendido mucho más, 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 es un referente distintivo de la música del siglo XX. Las exploraciones composicionales respecto a cambios de densidad, dinámica y timbre en intervalos relativamente cortos de tiempo, son particularmente resaltables y su uso se ha establecido de una manera más o menos sólida como material compositivo y método de tratamiento de los materiales sonoros que integran una obra.

Las exploraciones sonoras que actúan de acuerdo al 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 en la mayoría de las veces. El mayor obstáculo que se presenta en tal situación es quizá 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).

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 a un contexto socio-cultural, una de las soluciones para generar las nuevas configuraciones de escucha que se mencionaron con anterioridad es mediante 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, 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. 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, está atado 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 ser 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 elementos menos importantes que componen una estructura, una sección o una obra completa en sí misma para un determinado lapso tiempo de percepción; formulación que es posible asociar (a modo de clarificación al menos) con los conceptos de figura y fondo de un paisaje sonoro de acuerdo a Moles (1991) e incluso desde el punto de vista de la Gestalt. De este modo, la atención textural está asociada precisamente a la percepción de una estructura de cierto grado de complejidad dentro de una obra (o la obra misma) dejando en segundo plano los elementos individuales o aislados que son constitutivos a la misma. Mientras más atención se dirige a uno o varios de estos elementos de menor importancia, se pierde comprensión de las estructuras fundamentales de mayor escala de la obra misma.

Por otro lado, la atención paramétrica, 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.

Unidades temporales Gestalt

Para Tenney la música puede ser considerada como una red jerárquica ordenada de sonidos, motivos, pasajes, frases, secciones, etc. (Tenney, 1980). Dentro del contexto teórico de Tenney, la percepción de tales elementos sonoros como entidades individuales musicales en el tiempo es de importancia vital. En este sentido, es posible identificar estos elementos como configuraciones de sonidos que son perceptualmente identificadas como entidades individuales y que están compuestas por sus elementos constitutivos; cualesquiera que sean. Esta percepción temporalmente-delimitada de dichos sonidos o configuraciones de sonidos, implica de manera natural una aprehensión conceptual jerárquica en el sentido de que 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 sucesiva-

mente. Tenney denomina a dichas configuraciones: *Temporal Gestalt-Units* (TGUs).

Las TGUs son temporalmente-delimitadas; esto significa que es posible percibir además esta individualidad respecto a otras TGs 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, cada TGU es internamente cohesiva y externamente segregada. La figura 1 ilustra esta idea.

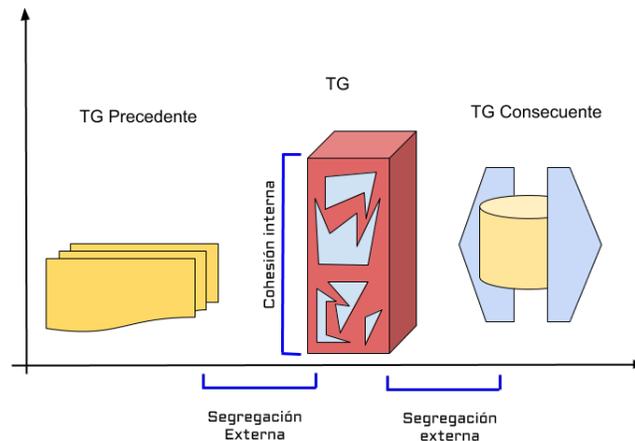


Figura 1-1: Representación diagramática de las TGU.

Tenney define tres niveles principales de jerarquía⁵. Dichos niveles tienen que ver con la organización de las TGs de acuerdo a la escala de tiempo considerada al momento de ser asimiladas. El nivel primario y fundamental de dicha jerarquía es el que 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 mencionó con anterioridad, de menor magnitud que el *clang*. En términos de Kohler, el *clang* representaría un *starke Gestalt* y la serie un *schwache Gestalt* (Kohler 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

⁵Tenney 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 TGs

a la obra misma, donde la percepción temporal está claramente delimitada por el inicio y el final de la música correspondiente a dicha obra.

1.6. Localización espacial del sonido

El problema de la localización espacial del sonido por el oído humano es un área de estudio bastante amplia y que ha logrado deducir y obtener resultados importantes a lo largo del tiempo de constante desarrollo que ha supuesto dicho campo del conocimiento. Sin embargo, hasta ahora no se tienen resultados completamente homogéneos capaces de resolver y contestar la interrogante de los mecanismos completos a través de los cuales el oído humano es capaz de realizar 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 acerca de la temática de la localización espacial audible ya que existe bastante literatura que aborda dicho objetivo de manera amplia y refinada. Lo único que se pretende es, exponer los conceptos y resultados más sobresalientes que sirvan para establecer los elementos constitutivos del modelo teórico propuesto.

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: evaluación de la distancia de la fuente sonora al espectador, dirección de dónde proviene la fuente sonora y 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 a factores individuales del escucha que Letowski (2001) enuncia y que se plantean aquí como definición debido a que se hará uso de estos más adelante.

Definición 1.6.1 *Se denominarán condiciones perceptual-cognitivas intrapersonales del escu-*

cha, 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: estatus fisiológico del sentido de escucha del espectador, experiencia audible, conocimiento de estrategias de escucha, familiaridad con el ambiente circundante y grado de involucramiento con la actividad de la escucha

Por supuesto, como es bien sabido, la capacidad de ubicar la dirección de la fuente, (la cual también es denominada *localización audible*) es un problema que es abordado en dos secciones: ubicación horizontal y elevación. El primero de ellos es el que ha tenido un desarrollo más considerable y los diversos resultados obtenidos ha 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 1.6.2 *Se denominará en este trabajo, **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***

De este modo, es posible por lo tanto hablar de *localización espacial* como la capacidad perceptual de ubicar una fuente en un espacio de manera general. Al respecto, se consideran por lo general, dos tipos de evaluación de la localización espacial: *relativa* o *discriminatoria* y *absoluta* o *categorica*. El primero de estos tipos tiene que ver 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 lo más preciso 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, es precisamente que (tal y como se presenta en la definición 2.1.5) 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 se toma como convención para la localización de las fuentes sonoras por parte del escucha, el sistema de coordenadas esféricas. En este sentido, se tienen tres componentes que el oído debe resolver: dirección en el plano horizontal (azimuth), dirección en el plano medio (elevación) y la distancia. Como es bien sabido, 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 2-2): *ángulo lateral* α , *ángulo de creciente* β , el azimuth θ , el ángulo de elevación ϕ y la distancia a la fuente sonora d .

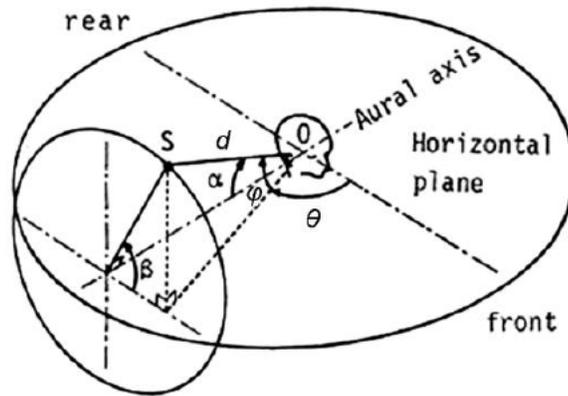


Figura 1-2: Sistema de coordenadas interaural polar (Letowski, 2001).

A lo largo del presente trabajo se utilizará como referencia el *sistema de coordenadas interaural polar* ya que como bien describe Letowski (2001), este sistema presenta la ventaja principal de que la longitud de arco entre dos ángulos laterales es independiente de la elevación, además del hecho de que a través del uso de este sistema, la descripción de la ubicación espacial es más adecuada para el *espacio interno*⁶ formado dentro del sistema multicanal. Además 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).

Se comenzó esta sección planteando el concepto de conciencia audible espacial y se mencionó que este era el concepto a partir del cual se desprendían de manera sino formal, al menos sí indirecta, los demás elementos que conforman la estructura de la localización de la escucha espacial. Esta conciencia está por supuesto 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.

⁶Véase la sección 2.1.2

- 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 implican una desviación de entre $\pm 20^\circ$, $\pm 25^\circ$.
- El *error de confusión* es la discrepancia entre la localización percibida y la 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 visceversa (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.

1.6.1. 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 y es el hecho de que éstas por lo general se dividen en dos categorías: fuente y ambiente. Las primeras son entidades relativamente discretas y localizables, mientras que las segundas poseen un carácter difusivo y de difícil localización. La descripción anterior, aunque muy general y por supuesto bastante conocida, establece una estructura primaria para dividir conceptualmente el fenómeno de la localización espacial y la experiencia de inmersividad.

En esta sección se enlistarán los aspectos y conceptos más importantes acerca de la escucha espacial en relación a los objetivos del presente trabajo. Existe muchísima literatura que aborda la escucha 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.

Del mismo modo, existen diferencias sustanciales en la localización espacial del sonido entre escenarios de campo abierto y recintos cerrados las cuales de nuevo, están descritas de manera detallada en las referencias antes mencionadas. La convención usual es que en campos abiertos, la localización espacial es 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.

Respecto a las fuentes sonoras en sí, es bien sabido que la mayoría de ellas 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 a la categoría de fuentes y la localización espacial de las mismas en el plano horizontal, es por demás conocido que se tienen primordialmente dos escenarios: campos sonoros libres y campos sonoros reverberantes y que uno de los mecanismos analíticos básicos para estudiar la escucha espacial de esta índole, está compuesto por lo propuesto en 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 dicho fenómeno en situaciones ideales compuestas por una sola fuente sonora.

Más aún, es conocido el hecho de que 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 ente los dos oídos (típicamente entre 22 a 23 cm) y se considera usualmente con un valor alrededor de los 1500 Hz* (He J. 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*. Por supuesto 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 IPD⁷ para frecuencias por debajo de los 1400 Hz y para frecuencias más altas mediante IID*.

Debe notarse que al igual que como ocurre con la Teoría del ASA, las conclusiones anteriores así como 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 resultados se han planteado conjeturas bastante formales para establecer la validez de las mismas para el caso de fuentes sonoras más complejas.

De acuerdo a diversos investigadores tales como (Blauert 1997), (Begault 2000), (Algazi y Duda 2011) y (Xie 2003), en general se acepta que los siguientes mecanismos son los *más significativos para la localización del sonido* (He J. 2017):

⁷IPD o *Interaural phase difference* es el término utilizado en cierta literatura como sustitutivo de las IID, para el caso de tonos puros.

- ITD.
- ILD.
- 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 tienen que ver 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 conocidas 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 J. 2017).

Anteriormente, 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 superpuestos*, 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 en cierto modo 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 apunta Blauert (1997) 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 ($\pm 90^\circ$ azimuth) 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 por ejemplo del frente horizontal).

Como bien retoma Cetta (2003), la localización espacial de fuentes sonoras de banda estrecha presentan además una anomalía particular, y es el hecho que para ciertos casos, dicha fuente es percibida en una posición simétrica respecto al eje interaural. Lo anterior es resultado del hecho 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, algunos conjeturas sugieren que debido precisamente 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).

1.6.2. Bandas direccionales y localización vertical

El mecanismo de bandas direccionales para resolver la ubicación espacial vertical de ciertas fuentes sonoras radica de manera única en el contenido espectral de las mismas. 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 fenomenológico. 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 hacia una representación mental del espacio. Rusconi et al. (2005) mostraron que de hecho existe *una naturaleza integral del procesamiento de los estímulos audibles tanto espectrales como espaciales*. Se cita a continuación, explícitamente 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) demuestra que la dirección de una imagen sonora para una banda de ruido de $\frac{1}{3}$ de octava, es una función únicamente del centro de la frecuencia y no depende en el 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 arriba de 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.

1.6.3. Propiedades acústicas de la espacialidad en la experiencia aural

Según diversos estudios, 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 envolvimiento*. Lo anterior plantea de entrada una discusión sobre una posible dirección o perspectiva de las características puntuales al momento de realizar creación electroacústica multicanal y sobre todo, para 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 al tamaño del recinto y al sentido de externalización del sonido por parte del escucha (Rumsey, 2001).

Por otro lado, el término *envolvimiento*, está ligado precisamente 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 en cierto grado con el coeficiente interaural de correlación cruzada (IACC), 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* y 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.

Este factor (la naturalidad) es de una importancia bastante peculiar 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. Tales situaciones se presentan de manera más inmediata cuando dichos elementos (objetos o escenas sonoras) son percibidos con un comportamiento inusual o incluso contrario a la experiencia aural ecológica usual, llegando a ser este factor (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).

Una pregunta lógica pudiera surgir entonces en este punto: ¿porqué es importante considerar todos estos factores acústico, 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 sino 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. Se agrupan a modo de definición todos los conceptos antes descritos.

Definición 1.6.3 *La conciencia audible espacial puede dividirse en dos subgrupos:*

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

La primera englobará 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: 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 1.6.1 *Derivado de lo anterior, el concepto de **naturalidad** puede entenderse tanto a nivel grupal (conciencia audible espacial) como local de cada uno de los componentes; por ejemplo, naturalidad de la distancia o naturalidad de la espaciosidad.*

A continuación se enlistan diversos resultados obtenidos de acuerdo a 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 a (Kuhn, 1977), la ITD se puede calcular mediante la fórmula:

$$ITD = \frac{ar}{c} \sin(\theta)$$

para $a = 3$ para frecuencias por debajo de los 500 Hz, $a = 2$ para 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's 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's 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's, existe una incertidumbre considerable de localización para el rango frecuencial $[1, 1.6]$ KHz para sonidos con desviaciones de hasta $\pm 30^\circ, 40^\circ$ con respecto a 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's 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's 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's son particularmente activas para la localización izquierda-derecha para sonidos únicamente por debajo de los 800 Hz. Por su parte, las ITD's 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° ó 2° para la posición frontal y del orden de 6° ó 7° para ubicaciones por detrás.
10. El MAA vertical para la posición frontal en el plano medio está entre 3° y 9° .

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's 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° .

Indicios dinámicos

1. Los movimientos de la cabeza causan cambios momentarios 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 600-800 ms pueden ser bastante bien localizados 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 3 ms a 1 s, disminuye el error promedio de localización

de 10° a 20°.

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 1 ms, 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 (5-2000 ms) con tiempos de ataque de 5 ms son los más sencillos para localizar en espacios cerrados.

1.6.4. 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, ingenieriles y por supuesto científicas.

Como se verá en la sección 2.1.5, 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 y 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*⁸, 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

⁸La velocidad angular es aquella a la cual el sonido rota alrededor del escucha, mientras que la velocidad radial es la que se refiere a la velocidad con la que la fuente sonora se aleja o se acerca en referencia al escucha (Letowski, 2012)

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 de hecho 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 Perrontt y Musicant (1977) de las cuales se obtuvieron las primeras conclusiones al respecto de las fuentes móviles: las posiciones de los *onsets* y los *offsets* son desplazadas primordialmente en la dirección del movimiento. Esto es recopilado por Getzmann y Lewald (2007):

La magnitud del desplazamiento del onset se incrementa con la velocidad de la fuente sonora (entre $90^\circ/\text{seg}$ y $600^\circ/\text{seg}$)... 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 estar directamente relacionado con el *efecto Frhöllich* 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 primera 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 esas etapas: al inicio, durante o al final, o incluso durante todas ellas.

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; de hasta 1.5 a 3.0 veces más grande para $\pm 60^\circ$ a lo largo del azimuth.
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 period 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 es del orden 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 2

Espacio y espacialidad

El *Espacio-Timbre* tal y como se propone en el presente trabajo, será construido a través de dos etapas básicas. La primera será la formulación abstracta y generalizada de las características de *espacialidad* y de *timbre* de un objeto sonoro cualquiera; ambas considerando los aspectos acústicos y perceptuales. Una vez que se han construido estas definiciones, se establecerá el concepto de *espacio aural* y *espacio espectromorfológico* mediante sistemas dinámicos. Este capítulo corresponde a la construcción de la primer parte; el de la formalización del concepto de *espacio* y de *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 espectromorfológico* es el objeto abstracto que representa las características espectromorfológicas 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 espectromorfológico*. De este modo se concluye el capítulo con una definición formal de *Espacio-Timbre* que abarca una concepción teórica y generalizada que permite que cualquier aplicación práctica y concreta pueda ser explicada y desarrollada a través de dicho modelo.

Cabe mencionar que la idea de establecer esta relación indisoluble entre las características espaciales y tímbricas puede parecer demasiado obvia si se adopta la premisa de que en la composición de la música electroacústica y la música acusmática se tiene consciencia constante de dicha relación. Sin embargo,

tomando en cuenta que no existe una convención o unificación de términos y conceptos respecto al espacio y al timbre dentro de la música electroacústica, se observa que existen pocos trabajos académicos que aborden esta temática (la relación espacio-timbre) de manera explícita.

Se tienen antecedentes de esto por supuesto 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 azimuth en el plano medio*.

Se comenzará realizando 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 plantearán los fundamentos teóricos de la concepción de espacio a utilizar dentro del contexto de música electroacústica, para continuar estableciendo los elementos primarios que conforman a juicio del autor, la caracterización primordial de la espacialidad en la música electroacústica. Finalmente, como se mencionó anteriormente, se establecerán los componentes matemáticos tanto del *espacio aural* como del *espacio espectromorfológico* para concluir con la derivación generalizada de *Espacio-Timbre*.

2.1. Construcción teórica del concepto de Espacio

A lo largo de esta sección se construirá una definición original del concepto de *espacio* basada 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.

2.1.1. Antecedentes sobre uso del espacio en la música

A principios de los años 50's 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). Estas obras 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 establecieron 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), y cuyos antecedentes primarios pueden ser trazados desde 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, en esa misma época 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 Henry, los conceptos de spatial relief y sound projection. Xenakis exploraba de manera sistematizada, las posibilidades intrínsecas de la relación del sonido con el espacio arquitectónico (al igual que Varese) y desarrollaban los conceptos de contrapunto espacial y geometría espacial. Por su parte, Stockhausen sentaba además las bases de la llamada serialización del espacio en Gesang der Juunglinge (1956) y desarrollaba en formas alternativas las ideas de la música espacial de Henry Brant, con Gruppen (1957). Hacia 1966 Francois 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 establecía teóricamente los conceptos de espacio interno y espacio externo y resaltaba la necesidad de establecer una interrelación adecuada entre ambos desde el punto de vista compositivo-performático, siendo además el primero en usar el término spatialization, y que junto con Risset resaltaban las relaciones de la manipulación del espacio externo durante el performance de la música en vivo.

Por otro lado, en su teoría de la espectromorfología, Dennis Smalley, definiendo entre otros conceptos,

los de subrogación de primer y segundo orden, el gesto como movimiento energético, entre otros, establece además 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.

2.1.2. El concepto de espacio: espacio sonoro/musical y espacialidad

A lo largo de finales del siglo XIX y casi todo el siglo XX, se han ido desarrollando diversas concepciones de *espacio* en la música. Macedo (2015) por ejemplo, desarrolla mediante un revisionismo histórico, una taxonomía de cinco categorías: *metáfora*, *espacio acústico*, *espacialización sonora*, *referencia y localización*. Henriksen (2002) por su parte, recoge y estructura un análisis mediante seis tipos de espacios divididos en dos grandes categorías: el espacio musical en su base perceptual y el espacio musical en su perspectiva compositiva-de performance. La primera categoría consta de: *espacio intrínseco*, *espacio extrínseco* y *espacio espectral*, mientras que la segunda está compuesta 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 como 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. Esto muestra que como se mencionaba al inicio de este trabajo, no existe una homegenización teórica y conceptual sobre lo que significa *espacio* 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* de acuerdo 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 es la relacionada a la idea del *espacio de parámetros musicales* y la segunda la de 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 a la escucha aural.

En este trabajo, el modelo teórico de *Espacio-Timbre* está compuesto por la conjugación de esas dos perspectivas de espacio en el sentido de que mediante el desarrollo de una representación paramétrica de ciertos aspectos sonoros (espectromorfología y localización espacial) se construye una formulación fenomenológica 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. Autores como Helmholtz, Rieszler, Conrad, Stofft, Kagel, McDermott, Rochberg o Morgan, han contribuido al desarrollo de propuestas de formulaciones de espacios de parámetros sonoros.

El propio Wishart (1998) plantea un modelo tridimensional de *espacio sonoro*; los elementos de este espacio 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).

Para Clifton (1983) por ejemplo, el espacio se entiende más 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*.

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 hacia 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 junto con metamorfosis de sonido 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 auricular 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).

Por su parte, Francois Bayle propone tres tipos de *especies de espacios* definidos con base en aspectos perceptuales y elementos cognitivos de la escucha: *présentification*, *identification* e *interprétation*. La *présentification* tiene que ver con el proceso preatención y percepción pasiva descrito en el capítulo I, así como 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 *identification* tiene lugar un proceso cognitivo más complejo y 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 *interprétation*, el escucha realiza asociaciones sobre la espacialidad de acuerdo a 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* tiene que ver con un prediseño de la espacialización de acuerdo a 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 (tanto desde el punto de vista paramétrico como de la espacialidad) y 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 estos 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*.

Los antecedentes previamente descritos han sido reutilizados, mezclados y reformulados por los compositores de música electroacústica interesados en el componente espacial y para el caso específico de una clasificación sobre espacialidad desde la perspectiva meramente compositiva/performativa es posible diferenciar al menos tres formatos explícitos: música acusmática multicanal en formato fijo, música acusmática para difusión en vivo, música mixta (cinta e instrumento acústico con o sin procesamiento en tiempo real). De lo anterior se desprende una clasificación bastante 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, etc.)
- **Serialización y concepción geométrica del espacio.** Diseño predefinido de trayectorias espaciales y relaciones matemáticas del sonido con el espacio. (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 basados en la teoría de Smalley y en los conceptos de Chion; es decir, componer el parámetro espacial de acuerdo a las características espectralmorfológicas de cada objeto sonoro, generando así un espacio interno, para después adecuarlo o modificarlo de acuerdo al espacio externo en el performance de la obra; en este sentido el espacio está en cierto modo “sometido” a la funcionalidad tímbrica. De hecho, desde una perspectiva reduccionista, salvo la música espacial, todos los acercamientos compositivos descritos, crean espacios artificiales en el espacio interno de la obra (usando la terminología de Chion) a través de espacialización y/o localización mediante plug ins o 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).

2.1.3. Espacio Interno - Espacio Externo y el problema de la disimilaridad serial/perceptual.

Uno de los principales problemas con los que se enfrenta la creación electroacústica es sin duda, la traslación de la obra desde el estudio al recinto de concierto. Esta traslación no es lineal sino más bien compleja ya que cada recinto es por lo general completamente distinto en sus características arquitectónicas y acústicas.

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 concreta: 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. Esta traslación es en cierto modo una reinterpretación del espacio composicional trabajado en el estudio a las circunstancias específicas del performance en vivo de la obra en la sala de concierto, lo cual es realizado casi siempre mediante difusión por *faders* en mezcladoras multicanal.

Estableciendo analogías un tanto arbitrarias con el espacio interno y externo, Risset habla de *espacio ilusorio* y *espacio real* y de como estos dos tipos de espacios interactúan para desarrollar una experiencia final sobre el escucha. Para él, los sonidos pueden ser espacializados, i.e. *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* que se encuentran al 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 compositor. 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 concepto espacio interno-espacio externo, la traslación del primero al segundo implica un número circunstancias técnicas que afectan considerablemente la percepción de la obra al momento del performance siendo quizás una de las más importantes la distorsión de la imagen espacial. Esta distorsión por supuesto afecta la localización fija, el involucramiento sonoro y en particular la *percepción de la continuidad*; conceptos que se especificarán en las secciones siguientes.

Como se vió en el primer capítulo, desde el punto de vista perceptual, la continuidad sensorial es muy importante para establecer representaciones mentales coherentes y convincentes. Disrupciones en la continuidad provocan ambigüedades perceptuales que aon interpretadas como errores y más aún, alejan al espectador de la experiencia inmersiva, que es uno de los objetivos planteados en este trabajo de investigación para la creación electroacústica multicanal.

2.1.4. Espacio virtual, espacialidad e inmersividad

Para plantear un modelo teórico que aborde la problemática de lo que se denominó *Espacio-Timbre*, es necesario primero definir claramente el concepto de *espacio* en sí mismo. Como se mencionó con anterioridad, esta definición está orientada hacia una mirada fenomenológica que no intenta desarrollar una ontología del *espacio* sino un acercamiento epistemológico del mismo con el objetivo de estructurar este concepto hacia el campo de la electroacústica.

Una posible perspectiva de la idea de espacio (quizá las 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 *fenomenológico*, 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 desde la fisicalidad como desde la temporalidad del espacio mismo. A esta extensión se le puede denominar *dimensionalidad*.¹

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 sensible. En este sentido, se puede proponer como primera aproximación, que *el espacio fenomenológico es la acción de delimitar el espacio físico a través del fenómeno de la experiencia sensorial subjetivo*.

Dicho lo anterior, resulta necesario establecer ¿qué se entiende por *delimitación*? para posteriormente

¹Por supuesto, la discusión acerca de la relación espacio-tiempo desde la mirada fenomenológica es bastante conocida y dado que la intención de este apartado es establecer únicamente los precedentes formales para definir el concepto de *espacio*, dicha problemática se deja de lado.

enfaticar ¿qué elementos se consideran para establecer esa delimitación dentro de un marco específico para definir *espacialidad* desde la perspectiva del arte y en particular desde la perspectiva de la música electroacústica multicanal?

Delimitar en el sentido anteriormente expuesto, no significa otra cosa que establecer los límites del fenómeno de la percepción de la experiencia sensorial 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, a través 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 dimensionalidad y esta delimitación fenomenológica 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 un constructo fenomenológico o 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 sensorial 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 en la definición 2.3.1 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.

Se mencionaron en el párrafo anterior, como ejemplos de espacios físicos, recintos cerrados o lugares abiertos. Sin embargo, es posible que el observador experimente el fenómeno de la experiencia sensorial en otro tipo de espacios; espacios ficticios. Estos pueden entenderse como construcciones, reconstrucciones o incluso deconstrucciones ficticias de cualquier tipo de recinto o región los cuales pueden ser consecuencia de procesos naturales o a través de métodos 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 entonces que dicho individuo está dentro de un espacio ficticio o *virtual*.

Tómese como ejemplo más básico la experiencia de la realidad virtual. En el momento en que el individuo se coloca el dispositivo, sigue estando dentro de un espacio físico dado (un cuarto, una sala, un auditorio, etc.), sin embargo, en ese mismo instante, su experiencia fenomenológica se traslada hacia un espacio virtual que intencionalmente construye una región ficticia. En este sentido, el espacio virtual es construido como una ficción dentro de un espacio físico y por lo tanto, el espacio virtual es un espacio fenomenológico en sí mismo y por lo tanto posee una espacialidad particular.

Para estos espacios virtuales, el proceso perceptual del individuo para establecer la *espacialidad* se encuentra condicionado no sólo por las herramientas y métodos tecnológicos con los que se construye esta ficción sino por las herramientas, procesos y contextos tecnológico-situacionales a través de los cuales el observador experimenta dicha ficción; es decir, por las condiciones perceptual-cognitivas intrapersonales.

Claramente, todo *espacio virtual* se construye y existe dentro de un espacio físico específico. Como se verá más adelante, la capacidad de dicho espacio virtual, de abstraer en mayor o menor grado al individuo del espacio físico, conduce a la inmersividad.

Sin embargo, ¿qué sucede cuando se integra una experiencia artística a algún espacio específico? ¿la espacialidad se mantiene idéntica o existe una perturbación de la misma dentro del espacio físico?

Si el eje primordial sobre el cual se contruye la espacialidad es la experiencia sensorial del observador en un instante dado, entonces el elemento artístico claramente perturba la espacialidad original del espacio físico. De este modo es posible afirmar entonces, que al momento de dicha integración (la de la experiencia artística en el espacio físico), el observador se convierte en espectador ya que dicho observador se encuentra ahora dentro de una experiencia sensible muy específica, a saber: el acto artístico, sea del tipo que sea. En este sentido, el espectador es un observador, pero el observador no es siempre un espectador.

Pero más aún, nótese que el acto artístico ocurre en un espacio físico específico, por consecuencia y de acuerdo a lo que se expuso con anterioridad, es posible afirmar que el fenómeno de la experiencia de tal acto artístico genera un *espacio virtual*.

Haciendo hincapié en que la espacialidad es una abstracción perceptual individual subjetiva de cada espectador, entonces el grado de la perturbación antes mencionada discrepará de individuo a individuo, siendo esta variación como se dijo anteriormente, directamente dependiente de las condiciones perceptual-cognitivas intrapersonales y de las herramientas-procesos-contextos tecnológico-situacionales con las que

se construye dicha virtualidad. Para el caso de estas últimas y dentro del contexto de la experiencia artística como detonante del espacio virtual, se pueden considerar: formato de presentación, medio de transmisión, resolución de los elementos expuestos, calidad general de la experiencia, etc.

En este sentido, el espacio virtual creado a partir de la perturbación del acto artístico puede ser entendido como aquel que contiene las experiencias estético-sensoriales que experimenta el espectador debido a la contemplación de dicho acto artístico.

La hipótesis principal en este punto es que, si existe el balance adecuado dentro del *espacio virtual*, entre las condiciones perceptual-cognitivas intrapersonales y las características de la experiencia artística, entonces será posible que dicho espectador experimente un grado específico de *immersividad*, la cual puede entenderse como la sensación específica de experimentar un alto grado de abstracción del espacio físico real y encontrarse dentro de ese espacio virtual.

2.1.5. Espacio virtual e immersividad en la electroacústica multicanal

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

De acuerdo al 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 hacia la búsqueda de la immersividad. Por supuesto, ésta no es el objetivo primordial en todas las creaciones electroacústicas o incluso pudiera ni siquiera ser considerada por el compositor, sin embargo, el formato multicanal por su naturaleza propia, plantea esta relación potencial inherente; la de la obra con la immersividad.

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 immersividad convincente. Las metodologías, taxonomías y estrategias para lograr una sensación de immersividad variarán de creador en creador, sin embargo al día de hoy la mayoría de ellas, se desprenden o se formulan casi siempre de manera heurística.

En esta sección se propone una metodología del desarrollo de la immersividad 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, que los objetos sonoros de acuerdo a su perfil espectromorfológico poseen una zona óptima de resolución espacial como fuente virtual dentro de un sistema multicanal. De este modo, la presente propuesta plantea entonces, una suerte de orquestación espacial de los objetos sonoros dentro de un marco de creación electroacústica multicanal. Este concepto será denominada a partir de este punto como *orquespectración espacial*.

Observación 2.1.1 *Como consecuencia directa de los resultados mostrados en la sección 1.6.3 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 azimuth como en el plano elevación, entendiéndose timbre tal y como se planteó en la definición 2.0.1.*

Se mencionó con anterioridad que la inmersividad en su aspecto más general, puede ser vista como 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 específico: 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 unas características estructurales y arquitectónicas específicas 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, etc. Ejemplos de características arquitectónicas son: diseño estructural del recinto, materiales de construcción (interior/exterior), etc.

Entonces, 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, estableciendo de este modo, una región delimitada dentro del *espacio físico*. Esta región está delimitada por componentes físicos (arquitectura del recinto y características del sistema multicanal). Se puede decir entonces que sus propiedades son de fisicalidad y 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 encuentra insertado dentro del espacio físico, el primero hereda las propiedades acústicas del segundo pero además posee otras características propias como 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 al nivel del piso), el tipo y tamaño de bocinas que se use, y el diseño estructural del sistema en sí, el cual puede ser 2D (un sólo 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 la *arquitectura del sistema multicanal*. Nótese que en general 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 2.1.1 *El espacio audible Ξ 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 $\mathcal{A}(\varphi)$ y la arquitectura del sistema multicanal por $\mathcal{A}(\vartheta)$. La **arquitectura del espacio audible** será el término que haga referencia a la conjugación simultánea de $\mathcal{A}(\varphi)$ y de $\mathcal{A}(\vartheta)$.*

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*.²

Definición 2.1.2 *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 más brevemente, *espacialidad*.*

2.1.6. Consideraciones generales sobre el espacio audible

Cada espacio audible tendrá un *sweet spot* específico de acuerdo a 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 del *sweet spot*.

²Por supuesto, la experiencia aural puede y es realizada múltiples veces por el mismo compositor; es decir dentro del proceso composicional, el creador intercambia continuamente su perspectiva de compositor a espectador.

En la figura 3-1 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 composicionales, la delimitación del espacio virtual sonoro puede extenderse más allá de dicha periferia del espacio audible; es decir, es posible crear objetos sonoros que sean percibidos con una localización afuera del espacio audible.

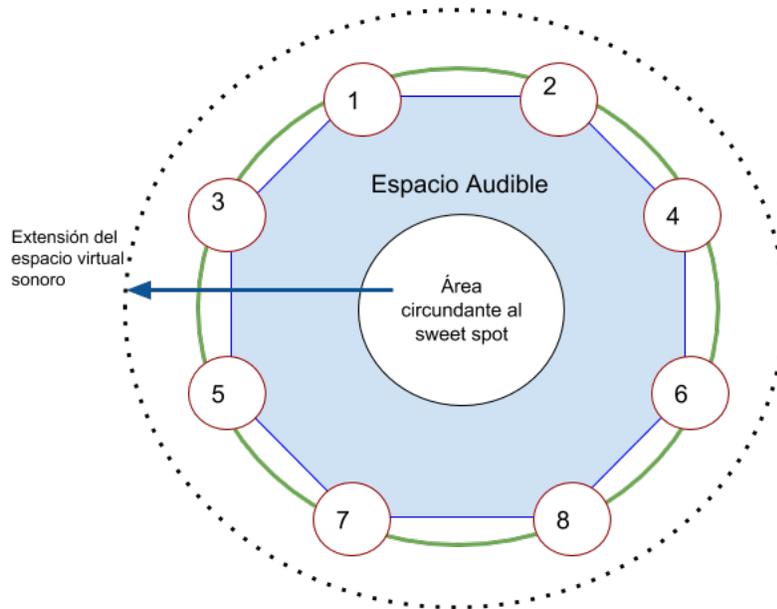


Figura 2-1: Representación gráfica del espacio audible y de la extensión del espacio virtual sonoro.

En las figuras 3-1, 3-1, 3-4 y 3-5, se muestran ejemplos concretos de arquitecturas físicas-multicanal que generan espacios audibles 3D.

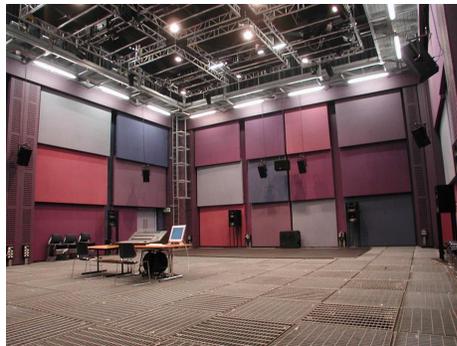


Figura 2-2: *Sonic Arts Research Centre*, Belfast.



Figura 2-4: *Birmingham Electroacoustic Sound Theatre*. BEAST.



Figura 2-3: *Hexagram-UQAM ESPACE*. Montreal.

Se mencionó en el capítulo anterior 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. Por supuesto, las dos últimas características están directamente derivadas 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



Figura 2-5: CCRMA *3D Listening Room*. Stanford.

la relación adecuada entre las dimensiones físicas del espacio audible y el número de bocinas (Rumsey). 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, referirse 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 immersividad en un grado lo suficientemente aceptable. De este modo y toda vez que se han planteado formalmente los términos adecuados para la comprensión del espacio según lo expuesto por esta propuesta teórica particular, se procederá a abordar en la siguiente sección, la metodología propia encaminada a investigar el problema del *diseño espacial multicanal* que permita contribuir positivamente a la immersividad.

Si bien el espacio virtual sonoro es 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 de los conocidos: plano azimuth, 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.

2.1.7. 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 (morfología). Smalley hace hincapié en su conocido artículo, que su propuesta está orientada hacia una explicación teórica de la experiencia perceptual aural y desde el punto de vista composicional de la forma y contenido.

Como se ha observado, el presente trabajo está fundamentado en gran medida en los aspectos perceptuales de la experiencia de escucha y es por eso que el término *espectromorfología* resulta tan idóneo 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 muy comunes y tienen ya bastante tiempo de ser utilizadas. Por prediseño se hace referencia a la manera de establecer la evolución de uno o varios parámetros sonoro-musicales a través del tiempo de manera específica y que por lo general obedece a algún tipo de regla o proceso. Esta perspectiva se relaciona naturalmente (al menos desde un punto de vista metafórico) con las aproximaciones y procesos del serialismo integral por lo que 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 que resulta complicado que el resultado final percibido en la experiencia aural concuerde con el prediseño mismo; i.e. existe una diferencia entre el *diseño serial* y el *resultado final perceptual*. Esto por supuesto está directamente relacionado con la distinción hecha por (Blauert 1969) entre *evento audible* y *evento sonoro*, donde el primero se refiere a *a la percepción audible determinada por atributos espaciales y temporales* y el segundo está asociado al *aspecto físico del proceso de escucha*.

Definición 2.1.4 *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 entonces como una medida cualitativa y hasta cierto punto heurística. Sin embargo, el diseño y realización de pruebas perceptuales 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 tiene implicaciones sustanciales para el presente trabajo ya que aquí se puntualiza como objetivo, que la composición del parámetro espacio esté en función de una validación perceptual acorde al diseño serial. Si la disimilitud serial-perceptual no jugara un papel predominante como en este caso, entonces nos encontraríamos frente a otro tipo de acercamiento estético más orientado hacia el arte procesual por ejemplo.

La *disimilitud serial-perceptual* tal y como ha sido presentada aquí se desprende además de hechos psicoacústicos observados en distintas pruebas de las que 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 J., 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 auditivo³ no siempre coinciden y por lo tanto, la localización espacial puede resultar ambigua.

2.2. 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 *potentiomètre d' espace* de Schaeffer y Henry. Es un campo que ofrece al compositor una múltiple paleta de recursos compositivos, 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 especiales 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 está directamente asociado a 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.**

Cuando los elementos anteriores se mezclan adecuadamente en la creación de obra y si la resolución de cada una de ellas es adecuada, estas propiedades compositivas 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 la sección 2.1.2 así como de comprobaciones prácticas en el proceso compositivo del autor de este trabajo de investigación.

En la sección anterior se analizó 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

³Los términos evento sonoro y auditivo son derivados de la terminología de Blauert (1997).

considerable; más aún, según lo propuesto en este trabajo, cada objeto sonoro de acuerdo a su perfil espectromorfológico tendrá una zona de localización espacial óptima medianamente definida, en la cual se tendrá una mayor resolución de localización espacial perceptual.

Dependiendo de la arquitectura del espacio audible, es posible caracterizar el diseño espacial mediante coordenadas esféricas o cilíndricas, pero de acuerdo a lo discutido en la sección 1.6, el sistema de coordenadas más adecuado es el *sistema de coordenadas interaural polar*.

Sistemas 3D con niveles de igual número de salidas también pueden ser descritos directamente mediante este tipo de sistemas de coordenadas como los que se mostraron en las figuras 2-2, 2-3, 2-4 y 2-5.

Para el caso de sistemas 3D con niveles y disposiciones delimitadas por formas semiesféricas o piramidales, su descripción espacial puede ser resuelta de mejor manera mediante coordenadas esféricas o incluso coordenadas elipsoidales.

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 2.2.1 *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 multi-canal), 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 a sus perfiles espaciales y espectromorfológicos), se puede afirmar que entonces radica en el espacio virtual sonoro.

La definición anterior hace una distinción explícita del objeto sonoro dependiendo 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 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 razón de lo anterior radica en lo que se dijo en la definición 2.1.2 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 2.2.1 *El objeto sonoro es equivalente al torrente auditivo tal y como está descrito en la definición 1.4.2.*

Conjetura 2.2.2 *Las cualidades acústicas del objeto sonoro que en el proceso perceptual permiten que se aprehendido como una unidad, están dadas por los factores del agrupamiento simultáneo y secuencial del ASA.*

Conjetura 2.2.3 *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 por separado cuando surgen como individuales.

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

Definición 2.2.2 *Se define el **espacio aural** $\mathfrak{S}(a)$, como la caracterización simultánea del espacio audible y del espacio virtual sonoro. Esto es, la consideración general de espacio desde el prediseño serial y también desde la percepción final.*

Nótese que con la definición anterior se podrán desarrollar de manera natural, definiciones y resultados que contemplen tanto las características acústicas así como las perceptuales respecto al proceso composicional espacial el cual es precisamente el objetivo de las siguientes secciones.

Definición 2.2.3 *Dado un espacio aural $\mathfrak{S}(a)$, un objeto sonoro dentro de él será denotado como $\mathcal{O} \in \mathfrak{S}(a)$.*

La duración temporal de dicho objeto sonoro estará dividida en dos: duración acústica y duración perceptual, y serán denotadas como $|\mathcal{O}|_a = n_a$ y $|\mathcal{O}|_p = n_p$ respectivamente.

Cuando no se especifique otra cosa, la duración temporal hará referencia a la duración perceptual y se denotará simplemente como $|\mathcal{O}| = n$

Observación 2.2.1 *La espacialidad virtual vista desde la electroacústica multicanal, es entonces, la experiencia aurál 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 es muy importante ya que entonces, los *componentes primarios de la espacialidad* están relacionados de manera directa a la *conciencia audible espacial de ubicuidad* y a 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 desarrollados a continuación.

2.2.1. Objetos sonoros fijos

La idea de *objeto sonoro fijo* corresponde a aquellos cuya posición espacial dentro del espacio aurál 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 de acuerdo a lo descrito en la sección 2.3, el sistema de coordenadas a utilizar para la descripción de localización espacial, será el *interaural polar*.

Definición 2.2.4 *Dado un espacio aurál $\mathfrak{S}(a)$, un objeto sonoro $\mathcal{O} \in \mathfrak{S}(a)$ con una duración total $|\mathcal{O}| = n$, la representación de su ubicación espacial en un instante de tiempo t , estará dada por:*

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

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

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

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

Se dijo con anterioridad que la espacialidad es una abstracción perceptual individual subjetiva de cada espectador y por lo tanto es posible que la experiencia de la misma varíe de individuo a individuo. Derivado de esto 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 más básico y primario sea el concepto de distancia espacial. Para el caso de fuentes fijas, dicho elemento se denotará como *distancia espacial lineal*.

Supóngase que se coloca una única fuente fija $\mathcal{O} \in \mathfrak{S}(a)$ dentro del espacio aural, con localización $\gamma(\mathcal{O}, t) = (\alpha(t), \beta(t), \theta(t), \phi(t), d(t))$. 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 es 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. Esto no implica por supuesto que el escucha no pueda localizar una fuente única dentro de una escena auditiva dada (más por el hecho de que en tal situación el proceso sensorial está directamente asociado a las ITD y a las ILD) pero lo que se afirma aquí es que una fuente virtual única dentro de un espacio audible (en un contexto electroacústico) no es suficiente por si misma para establecer una sensación de distancia y por lo tanto de espacio composicional en su nivel más básico. Lo anterior justifica la siguiente:

Definición 2.2.5 *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 diferencia en el perfil espectromorfológico de dichas fuentes, acentuarán la percepción de la distancia antes mencionada.*

De la definición anterior se desprende la premisa de que al momento de que al escucha le son presentados dos objetos sonoros simultáneos de localización espacial distinta, será posible entonces, establecer una sensación de distancia mucho más definida debido al hecho de que una de ellas se tomará como punto de referencia, para después resolver perceptualmente la distancia de ese punto de referencia al otro objeto sonoro y finalmente establecer una resolución de distancia respecto a él mismo; lo anterior por supuesto no en ese orden estricto.

Recordando lo descrito anteriormente respecto a la *disimilitud serial-perceptual*, es necesario recalcar que la *distancia espacial lineal* (así como todos los conceptos relacionados que se desarrollarán posteriormente) dependen de una validación perceptual; esto es, no basta con que exista una distancia entre

dos objetos sonoros en el diseño compositivo si dicha distancia no es correspondiente en la percepción real dentro del espacio aural.

Cabe señalar que en este punto, la capacidad de colocar objetos sonoros en localizaciones específicas en el espacio aural por parte del compositor, se tiene como dada. Por supuesto esa capacidad dependerá de las herramientas tecnológicas utilizadas, el método de espacialización y claro está, la capacidad del propio compositor. Por esa razón, esos detalles no se discutirán en este trabajo.

2.2.2. Trayectorias continuas u objetos sonoros móviles

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

Las trayectorias espaciales como recurso estético pueden añadir dinamismo tanto a nivel general (i.e. en la experiencia de la obra misma) como a nivel 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 2.2.6 *Dado un espacio aural $\mathfrak{S}(a)$, un objeto sonoro $\mathcal{O} \in \mathfrak{S}(a)$ con una duración total $|\mathcal{O}| = n$, se dice que es un **objeto sonoro móvil** si su localización 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:*

$$\begin{array}{rcl} I = [0, n] : & \longrightarrow & \mathbb{R}^3 \\ t & \longrightarrow & (\alpha(t), \beta(t), \theta(t), \phi(t), d(t)) \end{array}$$

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

Como se mencionó anteriormente, el problema principal de las trayectorias es el de la deformación de la *continuidad* al momento de la traslación del espacio interno al espacio externo por lo que es necesario establecer claramente que se entenderá por *continuidad*

Definición 2.2.7 *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.*

Con la definición anterior en mente puede resultar más o menos claro porque la *continuidad* de las trayectorias es 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 traslación. El estudio (lugar donde el compositor 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 por lo 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 compositor. 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 es fundamental para la coherencia y cohesión de los gestos musicales-sonoros,. En este sentido, la continuidad en las trayectorias espaciales es uno de los parámetros mas delicados al momento de diseñar experiencias espaciales inmersivas, ya que debido a su propia naturaleza física y psicoacústica, son elementos altamente susceptibles a la aparición de la discontinuidad.

Por lo tanto, 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 a nivel local como 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 panning, contenido espectral de los objetos sonoros, velocidad y aceleración de los objetos sonoros, etc. Más aún, un desarrollo profundo de la continuidad no es dependiente únicamente de estas variables por separado, sino de una relación dinámica y cambiante entre subconjuntos de éstas.

Uno de los retos principales al momento de utilizar trayectorias sonoras en una obra es 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, etc.). Al momento de interpretar la obra en el recinto y ajustarla al espacio externo, el compositor pierde control de varios parámetros; uno de ellos es el cambio de la velocidad percibida de los objetos sonoros que realizan trayectorias espaciales, los cuales, dependiendo de las dimensiones del recinto pueden incrementar sustancialmente de modo tal que 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 a la perspectiva panorámica del recinto. La otra radica precisamente en la tradición de la escuela francesa de espacializar en vivo (*live diffusion*) de modo tal que la obra sea interpretada de acuerdo al 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 1.6.3.

2.2.3. 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) son resultado de distribuir 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 en la definición 2.3.3 como *conciencia audible espacial de ambiente*.

El poder percibir estas superficies dentro de un espacio aural 2D, no sólo horizontalmente sino incluso con ciertos grados de inclinación, implican un especial manejo del perfil espectromorfológico y de la distribución espacial de los objetos sonoros que está directamente ligado 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; i.e. superficies sonoras que pueden ser percibidas con extensión volumétrica. De nuevo, para sistemas multicanal 3D, estas superficies tridimensionales se pueden lograr de manera más sencilla debido a la naturaleza inherente del espacio aural que resulta de dicha

configuración.

Definición 2.2.8 *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 como se mencionó anteriormente, las superficies sonoras pueden ser vistas como objetos abstractos dentro del espacio aural de dos y tres dimensiones: superficies y volúmenes, correspondientemente.

Definición 2.2.9 *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, en oposición a los objetos sonoros puntuales o de perfil morfológico percusivo. Para el último caso, si dicho objeto es distribuido a través de varias salidas del sistema multicanal, la frontera serial-perceptual será más bien una localización hacia el interior del espacio aural (dependiendo de qué salidas se usen claro está).

2.2.4. Observaciones finales sobre la descripción de espacialidad

Observación 2.2.2 *Las superficies o volúmenes texturales son un componente de la conciencia audible espacial de ambiente.*

Observación 2.2.3 *La **espaciosidad** puede construirse mediante la combinación dinámica principalmente de los objetos fijos y de las superficies o volúmenes sonoros.*

Observación 2.2.4 *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 microfónica⁴) a través del espacio aural como una superficie o volumen.*

⁴Véase el capítulo V.

Conjetura 2.2.4 *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.*
- **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 2.3.1.*

Capítulo 3

Espacio-Timbre: un modelo teórico.

El término *Espacio-timbre* tal y como se ha planteado a lo largo de esta 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 vió en los capítulos y las secciones anteriores, 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 plantea incluso bajo una formulación matemática abstracta basada en sistemas dinámicos con el fin de justificar lo más formalmente posible la construcción teórica de aquello que podría denominarse *espacio-timbre*.

Si bien el concepto de timbre es históricamente conflictivo debido a la falta de convención para definirlo, es un término que 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 como se construirá una definición del objeto sonoro derivada del concepto de espectromorfología de Dennis Smalley; esta definición propuesta, está de manera lógica, compuesta 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 2.1.4.

Antes de continuar, es necesario establecer la definición de timbre de acuerdo al contexto del pre-

sente trabajo. Esta definición se presenta en esta sección pero se hace la aclaración que los elementos que la componen (los perfiles morfológicos y espectrales con los caracteres acústicos y perceptuales) se desarrollarán detalladamente a lo largo del presente capítulo.

Definición 3.0.1 *El timbre es la conjunción representativa de un objeto sonoro de sus perfiles morfológicos y espectrales simultáneamente con sus caracteres acústicos y perceptuales.*

De este modo, *Espacio-Timbre* es el término que denotará la relación indisoluble que existe para cada objeto sonoro, entre sus propiedades de ubicuidad espacial y su perfil espectromorfológico y que 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* que son de uso común en el área del MIR (*Music Information Retrieval*). Dependiendo del contexto estos descriptores son clasificados de acuerdo a su nivel de abstracción, a su validez temporal o su descripción de dinámica.

Como se menciona en (Gomez & 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 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 a 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 ceptrales Mel (MFFC), medida de contenido de alta frecuencia (HFC), bandas Bark, etc.
- Temporales: tasa de cruce cero (ZCR), medidas de volumen (loudness), energía media cuadrática (rmse), etc.
- Rítmicos: intervalo de inter noiset, histograma de beat, espectro de beat, etc.

- Tonales: perfil de clase altura armónica (HPCP), desviación igual temperada, razón de energía no temperada a temperada, fuerza diatónica, centroide de octava, rugosidad tonal, etc.

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 son desarrollados detalladamente en las siguientes secciones.

3.1. 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 durante la duración total de dicho objeto.

Definición 3.1.1 *Dado un objeto sonoro cualquiera $\mathcal{O} \in \mathfrak{S}(a)$*

El contorno morfológico acústico $\mathfrak{C}_{morph}^a \mathcal{O}$ es la representación de la envolvente del sonido ligada a la representación físico-acústica:

$$\mathfrak{C}_{morph}^a \mathcal{O} \simeq \bigcup_i \{\delta(\mathcal{O})_i^a\}$$

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

- *El contorno morfológico perceptual $\mathfrak{C}_{morph}^p \mathcal{O}$ es la representación de la envolvente del sonido que muestra la relación existente entre la frecuencia y la intensidad del sonido (en el sentido de las curvas isofónicas) de acuerdo a las características específicas perceptuales del oído humano.*

$$\mathfrak{C}_{morph}^p \mathcal{O} \simeq \bigcup_j \{\delta(\mathcal{O})_j^p\}$$

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

- *El perfil morfológico \mathfrak{P}_{morph} es la combinación cualitativa y cuantitativa del contorno morfológico*

acústico y del contorno morfológico perceptual.

$$\mathfrak{F}_{morph}(\mathcal{O}) \simeq \mathfrak{e}_{morph}^a \cup \mathfrak{e}_{morph}^p$$

De acuerdo a la definición anterior y a lo discutido previamente, se observa 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 por supuesto tanto en la parte morfológica como en la parte espectral del sonido, los descriptores acústicos contribuyen en cierta medida a la percepción aural y de manera inversa, los descriptores perceptuales contribuyen a la parte físico-acústica. Sin embargo se hace esta diferenciación a modo de establecer una distinción categórica, conceptual y práctica al momento de establecer este modelo teórico así como las implementaciones prácticas del capítulo IV.

Como consecuencia directa del concepto de *perfil morfológico*, se deriva la idea de *espacio morfológico* que dentro de este modelo no es otra cosa que el ordenamiento de los descriptores como vectores de algún subconjunto de \mathbb{R}^n . Esta conceptualización permitirá establecer más adelante de manera formal, el concepto de *espacio espectromorfológico* además de que permitirá formular casos particulares de acuerdo a los descriptores utilizados.

El *espacio morfológico* es entonces, una construcción arbitraria y particular de acuerdo al número de descriptores que se deseen utilizar para casos específicos; es decir, éste (el espacio morfológico) se construye a partir 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 2.2.3.¹

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.

¹Al respecto de este punto, recuérdese que existe una diferencia entre el inicio u *onset* acústico de una señal de audio y su inicio o percepción del mismo. *El onset perceptual de un tono musical puede definirse como el momento en el tiempo en el cual el estímulo es percibido por primera vez* (Vosch & Rasch 1981).

En el primer caso se tiene por ejemplo, la práctica común en MIR 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 el capítulo IV basada en el análisis de series de tiempo.

Para construir un *espacio morfológico* deberá primero 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 3.1.2 Sean 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:

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

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 $t \in \mathcal{I}$, un elemento o vector $\vec{m} \in \mathcal{M}$ estará representado por la siguiente parametrización:

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

donde:

$$\vec{m}(t) = (m_0^a(t), \dots, m_n^a(t), m_0^p(t), \dots, m_k^p(t))$$

y $m_i^a \in \mathfrak{C}_{morf}^a(\mathcal{O})$ para $i \in [0, n]$ y $m_j^p \in \mathfrak{C}_{morf}^p(\mathcal{O})$ 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 esté descrito tanto por una cuantificación numérica (como la media, varianza, etc.) como por series de tiempo ya que la primer situación puede considerarse como un caso particular del segundo; cuando para todo instante de tiempo el valor del descriptor es el mismo: $m_i^a(t) = \mu$ 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 a los descrip-

tores elegidos, mientras que el segundo establece la organización *vectorial* de dichos descriptores dentro de un espacio subconjunto de $\mathbb{R}^n \times \mathbb{R}^k$. El vector \vec{m} 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 vió 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 $\mathcal{M} \subset \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$; i.e. 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 ² 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 obtenida a través del análisis FFT. Entonces, cualquier objeto sonoro estará representado por estos dos descriptores del siguiente modo:

$$\vec{m}(t) = (m_0^a, m_0^p(t)) = (\mu, m_0^p(t))$$

Entonces, el perfil morfológico de ese objeto sonoro estará descrito por la serie de tiempo:

$$[(\mu, m_0^p(0)), (\mu, m_0^p(1)), \dots, (\mu, m_0^p(n))]$$

Nótese que el procedimiento general para construir el *espacio morfológico*³ 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 $\vec{m}(t)$ estará asociado a un objeto sonoro particular y este vector arrojará la descripción

²Véase el apéndice en la sección..

³Y que de hecho será totalmente análogo para el caso del espacio espectral y por lo tanto, del espacio espectralmorfológico.

numérica o la parametrización de los descriptores seleccionados.

Resulta claro en este punto entonces, que existirán tantos espacios morfológicos como combinaciones posibles de descriptores morfológicos acústicos-perceptuales. Nótese además, que al definir el espacio morfológico de una manera tan general, es posible incorporar sin ningún problema, cualquier descriptor que pudiera ser desarrollado a futuro.

3.2. Perfil y Espacio Espectral

En esta sección se definirá el *perfil espectral* de manera completamente análoga a como se definió el *perfil morfológico*, i.e. estableciendo un componente físico-acústico y otro perceptual.

Definición 3.2.1 *Dado un objeto sonoro cualquiera $\mathcal{O} \in \mathfrak{S}(a)$:*

- *El contenido espectral acústico $\mathfrak{W}_{spectra}^a \mathcal{O}$ es la representación de dicha propiedad (el espectro del sonido) ligada a la representación físico-acústica:*

$$\mathfrak{W}_{spectra}^a \mathcal{O} \simeq \bigcup_i \{\omega(\mathcal{O})_i^a\}$$

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

- *El contenido espectral perceptual $\mathfrak{W}_{spectra}^p \mathcal{O}$ es la representación de dicha propiedad (el espectro del sonido) ligada a la representación perceptual:*

$$\mathfrak{W}_{spectra}^p \mathcal{O} \simeq \bigcup_j \{\omega(\mathcal{O})_j^p\}$$

donde lo anterior puede leerse como que, el contorno morfológico perceptual es equivalente a la influencia conjunta de los descriptores $\omega(\mathcal{O})_i^p$, asociados a dicho prediseño.

- *El perfil espectral $\mathfrak{F}_{spectra}$ es la combinación cualitativa y cuantitativa del contenido espectral acústico y del contenido espectral perceptual.*

$$\mathfrak{F}_{spectra}(\mathcal{O}) \simeq \mathfrak{W}_{spectra}^a \mathcal{O} \cup \mathfrak{W}_{spectra}^p \mathcal{O}$$

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

Definición 3.2.2 Sean 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:

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

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 $\vec{e} \in \mathcal{E}$ estará representado por la siguiente parametrización:

$$\begin{aligned} \psi: I &\longrightarrow \mathcal{E} \\ t &\longrightarrow \vec{e}(t) \end{aligned}$$

donde:

$$\vec{e}(t) = (e_0^a(t), \dots, e_n^a(t), e_0^p(t), \dots, e_k^p(t))$$

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

3.3. Perfil y Espacio Espectromorfológico

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

Definición 3.3.1 Dado un objeto sonoro cualquiera $\mathcal{O} \in \mathfrak{S}(a)$, el **perfil espectromorfológico** es la descripción cualitativa y cuantitativa del perfil morfológico y el perfil espectral de manera conjunta.

$$\mathfrak{P}_{spectra}^{morph}(\mathcal{O}) \simeq \mathfrak{P}_{morph}(\mathcal{O}) \cup \mathfrak{W}_{spectra}(\mathcal{O})$$

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; i.e 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 tener una definición abstracta y general que abarca cualquier objeto sonoro y a la vez, tener un marco de referencia práctico y concreto para realizar esta representación mediante descriptores acústico-perceptuales que puedan ser directamente incluidos en algún modelo computacional.

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

Definición 3.3.2 Sean \mathcal{M} y \mathcal{E} los espacios morfológico y espectral con dimensiones $n_1 + k_1$ y $n_2 + k_2$ respectivamente, entonces, el **espacio espectromorfológico** estará dado por:

$$\mathcal{S} = \mathcal{M} \times \mathcal{E}$$

donde cada elemento $\vec{s} \in \mathcal{S}$, 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 morfológicos 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 particular estará dada por:

$$\begin{aligned} \psi : I &\longrightarrow \mathcal{S} \\ t &\longrightarrow \vec{s}(t) \end{aligned}$$

donde:

$$\vec{s}(t) = [s_0(t), \dots, s_{n^*}(t)]$$

3.3.1. 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 cualquiera, esto tiene como consecuencia directa que dichos espacios heredan diversas propiedades de \mathbb{R}^n y en particular, la propiedad de ser espacios métricos cuando se consideran las métricas usuales como la euclídeana, la métrica de Minkowsky o la Manhattan.⁴

⁴Véase el apéndice sección..

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.

Dado por ejemplo un espacio morfológico \mathcal{M} junto con la métrica euclideana usual y dos vectores $\vec{m}^0, \vec{m}^1 \in \mathcal{M}$, entonces la distancia $d(\vec{m}^0, \vec{m}^1) = \| \vec{m}_0 - \vec{m}_1 \|$ tendrá el significado asociado a la caracterización numérica de cada descriptor.

Ejemplificando lo anterior, supóngase 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.⁵ Entonces, el espacio morfológico resultante es un subconjunto de \mathbb{R}^2 tal que todo vector será de la forma $\vec{m}(t) = (m_0(t), m_1(t))$ y representará objetos sonoros de acuerdo a los descriptores establecidos. De este modo, para cada objeto sonoro, la primer 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 (la media aritmética y el coeficiente de variación) es claro que el espacio morfológico no sólo es un subconjunto de \mathcal{M} sino más aún de $[0, 1] \times [0, 1]$ ⁶. De este modo, para cada instante de tiempo t , dados dos vectores $\vec{m}^0, \vec{m}^1 \in \mathcal{M}$ la distancia entre ellos representará un nivel de similaridad de los objetos sonoros de acuerdo a los descriptores y las caracterizaciones numéricas elegidas.

Planteado todo lo anterior, se proponen entonces la definición formal de *caracterización numérica*.

Definición 3.3.3 *Sea $\mathcal{X} \subset \mathbb{R}^n$ un espacio morfológico, espectral o espectromorfológico. Entonces, para cada descriptor que conforma dicho espacio, se denomina **caracterización numérica del descriptor** a la unidad de medida con la que es analizado el objeto sonoro. Dicha caracterización puede ser de dos tipos: fija o paramétrica. En el caso de que la caracterización sea paramétrica, se considerará la misma como una serie de tiempo.*

Se finaliza esta sección con la definición formal de distancia entre dos vectores dentro de un espacio morfológico, espectral o espectromorfológico. 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*⁷ que funciona sobre todo para series de tiempo que son de distinta longitud. De nuevo, dada la generalidad del modelo, desde

⁵Véase el apéndice sección..

⁶Por supuesto para fines específicos, los valores de las caracterizaciones numéricas se pueden escalar.

⁷Véase el apéndice sección..

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 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. Que dos objetos sonoros dentro de un espacio dado estén más o menos separados implicará un grado de similaridad o disimilaridad respectivamente y esta referencia permitirá tener un panorama de organización de los objetos sonoros de acuerdo precisamente a dicho grado relativo de similaridad.⁸

Definición 3.3.4 Sea $\mathcal{X} \subset \mathbb{R}^n$ un espacio morfológico, espectral o espectromorfológico junto con una métrica $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$ donde cada uno de los n descriptores tiene definida una caracterización numérica específica. Entonces, dados dos objetos sonoros y sus vectores $\vec{m}^0, \vec{m}^1 \in \mathcal{M}$ asociados, la distancia $d(\vec{m}^0, \vec{m}^1)$ representará el grado de similaridad entre dichos objetos sonoros de acuerdo a los descriptores y a la caracterización numérica establecidos.

3.4. Espacio Timbre: construcción abstracta

A continuación se presentará la construcción del concepto del *Espacio-Timbre* utilizando sistemas dinámicos. Esta forma de definir dichos conceptos, permitirá formular una contribución teórica formal haciendo uso de una herramienta matemática sólida de tal modo que como se ha venido puntualizando a lo largo del presente trabajo, se establezca un modelo lo suficientemente general como para ser capaz de englobar y explicar la mayoría de los casos particulares.

Un sistema dinámico puede entenderse en términos muy generales como la descripción de la evolución de un sistema cualquiera a través del tiempo dadas ciertas condiciones iniciales. Las condiciones iniciales tienen que ver usualmente, con el punto sobre el cual se elige aplicar la *dinámica* del sistema así como 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; pero con una gran generalidad, que la representación no es para un punto único o específico sino para cualquiera. En esta característica radica la utilidad del sistema dinámico como herramienta para representar de manera general un concepto teórico como lo es el la conciencia espacial audible o el perfil espectromorfológico.

⁸El lector familiarizado encontrará 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.

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 2.2.4 dado un espacio aural $\mathfrak{S}(a)$, un objeto sonoro tiene una representación de localización espacial dada por

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

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

Definición 3.4.1 *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 $(\mathfrak{S}(a), I, \pi)$, donde:*

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

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

$$\Upsilon_m^n(t) = \{\pi(\mathcal{O}, t) | t = \overline{1, n}\}$$

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

$$\begin{aligned} \pi_{\mathcal{O}} : I &\longrightarrow \mathfrak{S}(a) \\ t &\longrightarrow \pi_{\mathcal{O}}(t) \end{aligned}$$

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

Proposición 3.4.1 *Sea $(\mathfrak{S}(a), I, \pi)$ 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 $\mathcal{O} = (r_{\mathcal{O}}, \theta_{\mathcal{O}}, \phi_{\mathcal{O}}) \in \mathfrak{S}(a)$. Por definición de punto fijo de un sistema dinámico,

$$\pi_{\mathcal{O}}(t) = \mathcal{O} = (r_{\mathcal{O}}(t), \theta_{\mathcal{O}}(t), \phi_{\mathcal{O}}(t)) \quad \forall t \in \mathbb{I}$$

i.e. la posición espacial del objeto sonoro no cambia a través del tiempo. ■

Definición 3.4.2 Sea $(\mathfrak{S}(a), I, \pi)$ el *sistema dinámico asociado a un espacio audible*. Dado $\mathcal{O} \in \mathfrak{S}(a)$ y $n \in \mathbb{I}\mathbb{R}$, la *trayectoria espacial continua* de cualquier objeto sonoro asociado a dicha representación está dada por el *recorrido finito* de la *órbita* del sistema dinámico correspondiente.

Para el caso del **espacio espectromorfológico** $\mathcal{S} \subset \mathbb{R}^k$, es necesario recordar que la dimensión k dependía de la morfología y del número de descriptores que se utilicen para construir el *perfil espectromorfológico* 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 3.4.3 Sea \mathcal{S} el *espacio espectromorfológico* e $I \subset \mathbb{R}^+$ un conjunto indexador de tiempo. Entonces el *sistema dinámico espectromorfológico* estará dado por la tripleta (\mathcal{S}, I, ψ) , donde:

$$\begin{aligned} \psi : \mathcal{S} \times I &\longrightarrow \mathcal{S} \\ (s, t) &\longrightarrow \pi(s, t) \end{aligned}$$

Sea $n \in I$, y $s \in \mathcal{S}$, entonces, se denominará *recorrido finito* a la sección de la órbita correspondiente al intervalo de tiempo $\delta t = t_f - t_0$:

$$\mathcal{O}_s^n(t) = \{\psi_s(t) | t = \overline{1, n}\}$$

Definición 3.4.4 Dado un *sistema dinámico espectromorfológico* por la tripleta (\mathcal{S}, I, ψ) , los *objetos espectromorfológicamente estáticos* son asociados a los puntos fijos de dicho sistema.

Es necesario recordar en este punto que cada sistema dinámico asociado tanto al espacio audible como al espectromorfológico corresponde a una modelación de comportamiento y no a cada objeto sonoro en sí. En este sentido, un sistema dinámico puede estar asociado a uno o más objetos sonoros y transformarlos de acuerdo a su operador de evolución individual y dependiendo de las condiciones iniciales.

3.4.1. Donde la espacialidad y el timbre se unen.

Una 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 espectromorfológico de cada objeto sonoro dentro del proceso compositivo de la música electroacústica. Esta conjetura ha sido probada 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. Un camino posible para establecer resultados acerca de la relación espacio/timbre con fuentes sonoras más complejas desde la parte psicoacústica es a través del presente trabajo, en el cual se desarrolla todo el corpus teórico formal que permita fundamentar de manera directa dichas comprobaciones prácticas.

Al plantear como problemática la indisolubilidad entre el espacio y el timbre, se propuso como solución el concepto: *espacio-timbre*.

En este apartado se establecerá dicha correspondencia pero a través de la representación realizada mediante sistemas dinámicos. Como se mencionó anteriormente, cada sistema dinámico (en cualquiera de los dos espacios; ya sea 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 recorra el arreglo multicanal en forma cíclica puede ser modelada por un sistema dinámico específico el cual además puede actuar sobre un número cualquiera de objetos sonoros. De igual modo, una transformación espectromorfológica cualquiera 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 3.4.5 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 la relación *espacio-timbre* de manera matemática se hará uso de la *propiedad universal del producto de conjuntos*⁹ 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.

⁹Véase el apéndice.

Proposición 3.4.2 Sean $(\mathfrak{S}(a), I, \pi) \in \Pi$ un sistema dinámico asociado a la conciencia audible espacial y $(\mathcal{S}, I, \psi) \in \Psi$ un sistema dinámico asociado al perfil espectromorfológico. Entonces, existe un sistema dinámico $\mathcal{E} = (\mathfrak{S}(a) \times \mathcal{S}, I, \zeta)$ tal que:

$$\zeta = \langle \pi, \psi \rangle$$

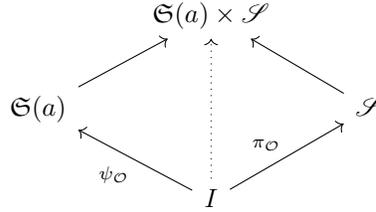
A \mathcal{E} se le denominará **sistema dinámico del espacio timbre**.

Demostración. Sea \mathcal{O} un objeto sonoro con representaciones espacio-espectromorfológicas dadas por los sistemas dinámicos correspondientes:

$$\pi_{\mathcal{O}} : \mathcal{I} \longrightarrow \mathfrak{S}(a)$$

$$\psi_{\mathcal{O}} : \mathcal{I} \longrightarrow \mathcal{S}$$

Por la propiedad universal de productos de conjuntos existe una única función $\zeta_{\mathcal{O}} : \mathcal{I} \longrightarrow \mathfrak{S}(a) \times \mathcal{S}$ tal que el siguiente diagrama conmuta:



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

$$\zeta_{\mathcal{O}} = \langle \pi_{\mathcal{O}}, \psi_{\mathcal{O}} \rangle : \mathcal{I} \longrightarrow \mathfrak{S}(a) \times \mathcal{S}$$

donde:

$$\zeta_{\mathcal{O}}(t) = (\pi_{\mathcal{O}}(t), \psi_{\mathcal{O}}(t))$$

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, es que la definición no dependerá de una visión o un caso particular sino que es expandible a virtualmente cualquier caso.

3.5. 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 espectromorfológico* 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 lo más formal y general posible.

En esta sección, se mostrarán algunos ejemplos concretos en los que se definirán dichos espacios con base en parámetros espectromorfológicos 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 de tal modo que permita dicha generalización, entonces una gran cantidad de casos particulares pueden ser abordados y explicados a través del mismo.

En este punto se reitera la importancia de este aspecto (la generalización) y del hecho de que la aportación central de este trabajo es precisamente dicho objetivo. Enfatizando lo mencionado previamente, una manera alternativa y un tanto informal de decir lo anterior, es que en realidad **se pueden construir tantos espacios espectromorfológicos como combinaciones de descriptores se tengan**.

Por ejemplo, en el apéndice se listaron un total de 15 descriptores morfológicos y 20 descriptores espectrales. Por lo tanto se tendrán el siguiente número de posibles combinaciones y por lo tanto de espacios espectromorfológicos.

La idea anterior muestra que desde el punto de vista conceptual, es en cierto modo irrelevante el conjunto de descriptores que se consideren para conformar los distintos espacios morfológicos y espectrales o espectromorfológicos. 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.

Ahora bien, la pregunta natural que surge como consecuencia de lo anterior es sin duda: ¿cuáles

descriptores conformarán espacios más adecuados para fines de análisis y/o creación? La respuesta a la cuestión anterior radicará definitivamente como ya se mencionó, en los objetivos y metas de cada caso particular. 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 modo que es posible construir cualquier caso particular de los anteriores espacios 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.

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 $\mathcal{O}_1, \mathcal{O}_2$ los objetos sonoros mencionados y que representan dos muestras de audio cuyos espectrogramas y formas de onda se muestran en las figuras... Los dos objetos sonoros son muestras de golpes de madera y cristal respectivamente; ambas procesadas digitalmente mediante resonadores y filtros con transformación de *pitch*. La diferencia principal radica en que la primera muestra tiene contenido tonal más bajo que la segunda (como puede observarse en el espectrograma).

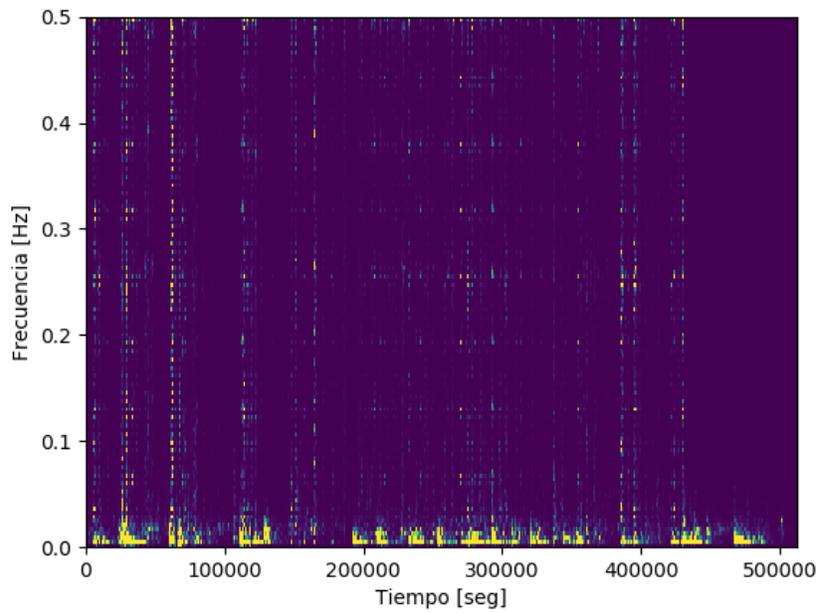


Figura 3-1: Espectrograma para muestra de audio I.

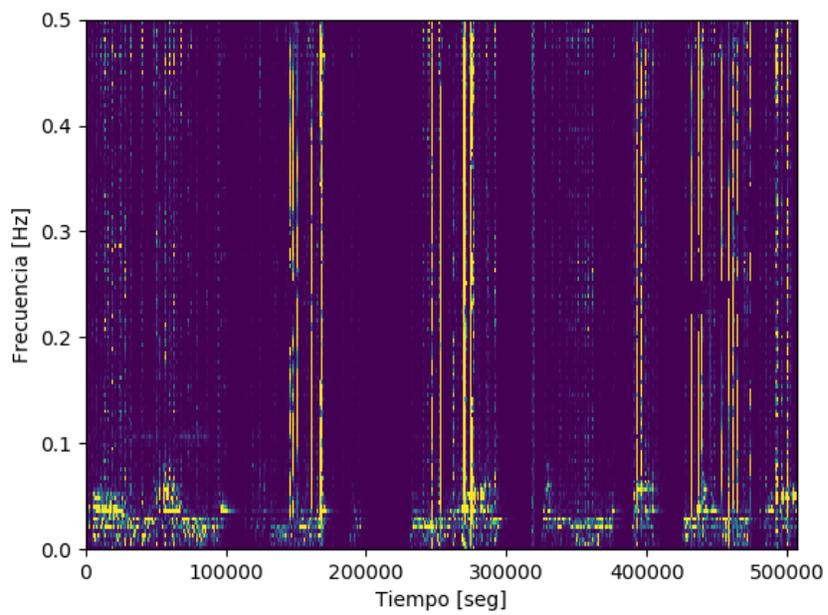


Figura 3-2: Espectrograma para muestra de audio II.

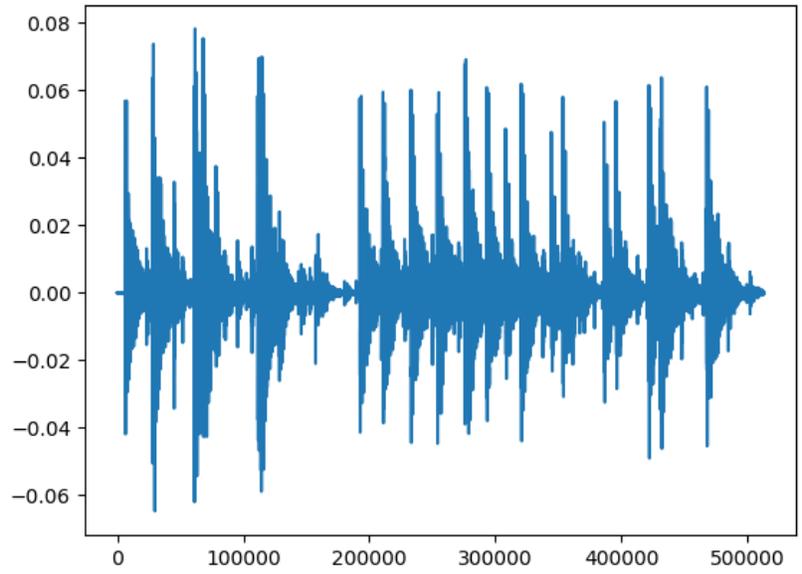


Figura 3-3: Forma de onda para muestra de audio I.

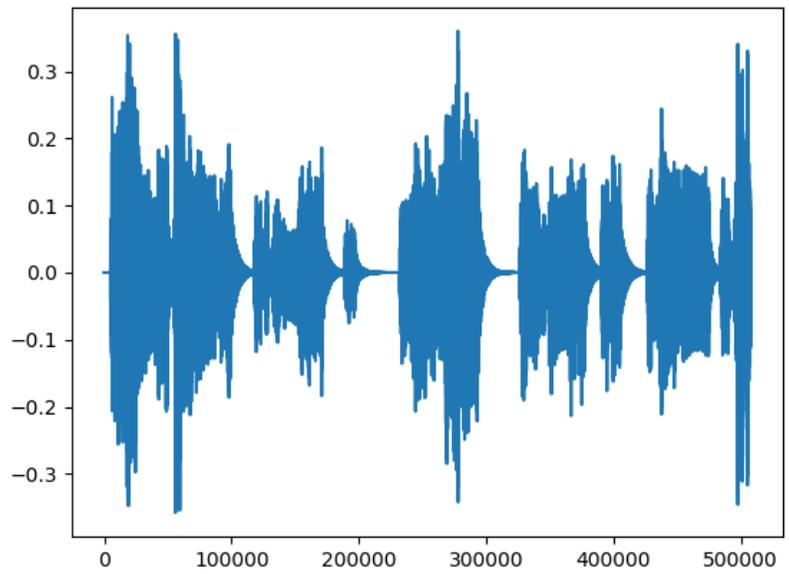


Figura 3-4: Forma de onda para muestra de audio II.

Todos los valores de las caracterizaciones numéricas de los descriptores de dichos objetos sonoros,

fueron obtenidos utilizando la librería *librosa* (McFee et al, 2015) y algunas implementaciones de DSP dentro del lenguaje de programación Python.

3.5.1. Ejemplo I

Considérense como descriptores morfológicos:

- ZCR.
- 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 que se considera un sólo descriptor espectral:

- Coeficiente de tonalidad espectral.

Donde dicho descriptor tiene como caracterización numérica 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, el espacio espectromorfológico estará dado por:

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

Para el espacio morfológico \mathcal{M} los vectores \vec{m}^1 y \vec{m}^2 asociados a los objetos sonoros \mathcal{O}_1 y \mathcal{O}_2 respectivamente, tienen los valores:

$$\vec{m}_0 = [m_{00}, m_{01}] = [2.131538, 0.002888036]$$

$$\vec{m}^1 = [m_{10}, m_{11}] = [0.656021, 0.008939]$$

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 de los objetos sonoros en cuestión.

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

$$\vec{e}^1 = [e_{00}, e_{01}] = [0.02223, 0.1222]$$

$$\vec{e}^2 = [e_{10}, e_{11}] = [0.0143, 0.4522]$$

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

$$\vec{s}_0 = [m_{00}, m_{01}, e_{00}, e_{01}] = [0.02223, 0.1222, 0.33, 0.567]$$

$$\vec{s}_1 = [m_{10}, m_{11}, e_{10}, e_{11}] = [0.0143, 0.4522, 0.1223, 0.13345]$$

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

$$d(\vec{s}^1, \vec{s}^2) = \sqrt{|m_{00} - m_{10}|^2 + |m_{01} - m_{11}|^2 + |e_{00} - e_{01}|^2 + |e_{10} - e_{11}|^2}$$

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, supónganse caracterizaciones numéricas distintas para cada uno de los descriptores. Esto afectará naturalmente 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 el espacio espectromorfológico depende en su construcción, de la elección de los descriptores y de las caracterizaciones numéricas para cada descriptor.

3.5.2. Ejemplo II

Considérense para este segundo ejemplo como descriptores morfológicos:

- ZCR.
- Energía.
- Envoltente 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.
- Coeficientes de contraste espectral.

Supóngase que para los dos primeros descriptores espectrales se considera como caracterización numérica el *coeficiente de variación*, mientras que para el conjunto de coeficientes de contraste espectral se considera la entropía como caracterización numérica.

Entonces, el espacio espectromorfológico estará dado por:

$$\mathcal{S} = \mathcal{M} \times \mathcal{E} \subset [0, 1]^3 \times [0, 1]^9 \subset \mathbb{R}^{12}$$

Para el espacio morfológico \mathcal{M} los vectores \vec{m}_0 y \vec{m}_1 asociados a los objetos sonoros \mathcal{O}_0 y \mathcal{O}_1 respectivamente, tienen los valores:

$$\vec{m}_0 = [m_{00}, m_{01}, m_{02}] = [0.02223, 0.1222]$$

$$\vec{m}_1 = [m_{10}, m_{11}] = [0.0143, 0.4522]$$

donde para cada vector, la primera 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:

$$\vec{e}_0 = [e_{00}, e_{01}, e_{02}, \dots, e_{08}] = [0.02223, 0.1222]$$

$$\vec{e}_1 = [e_{10}, e_{11}, e_{12}, \dots, e_{18}] = [0.0143, 0.4522]$$

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

$$d(\vec{s}_1, \vec{s}_2) = \sqrt{|m_{00} - m_{10}|^2 + |m_{01} - m_{11}|^2 + |e_{00} - e_{01}|^2 + |e_{10} - e_{11}|^2}$$

3.5.3. Ejemplo III

Para este último ejemplo, considérense 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..

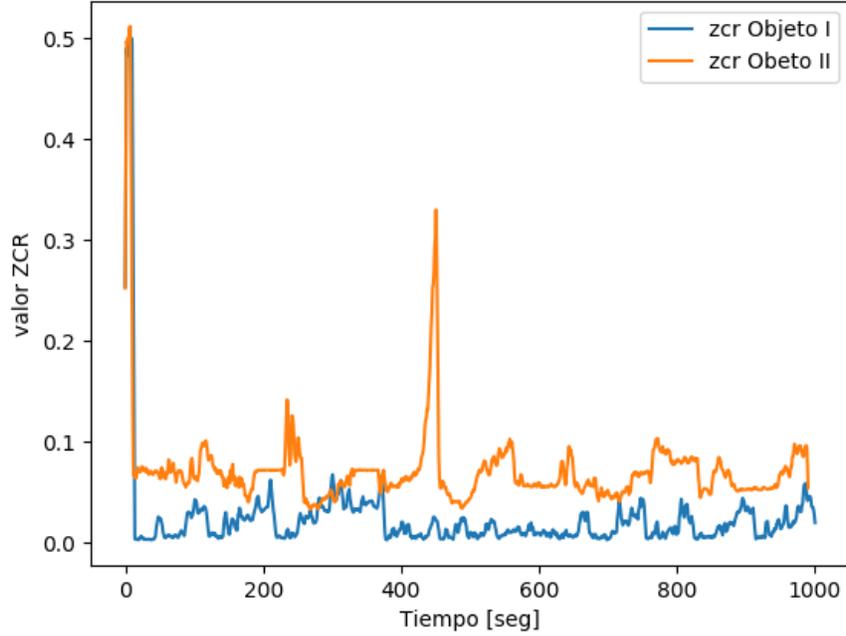


Figura 3-5: Series de tiempo del zcr para los objetos sonoros I y II.

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:

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

Sean entonces $\vec{m}_0(t)$, $\vec{m}_1(t)$ los vectores del espacio morfológico asociados a los objetos sonoros \mathcal{O}_0 y \mathcal{O}_1 respectivamente:

$$\vec{m}_0(t) = [m_0(0), m_0(1), m_0(2), \dots, m_0(k_0)] = [0.2345, 0.123, 0.321, \dots, 0.8596]$$

$$\vec{m}_1(t) = [m_1(0), m_1(1), m_1(2), \dots, m_1(k_1)] = [0.2345, 0.123, 0.321, \dots, 0.8596]$$

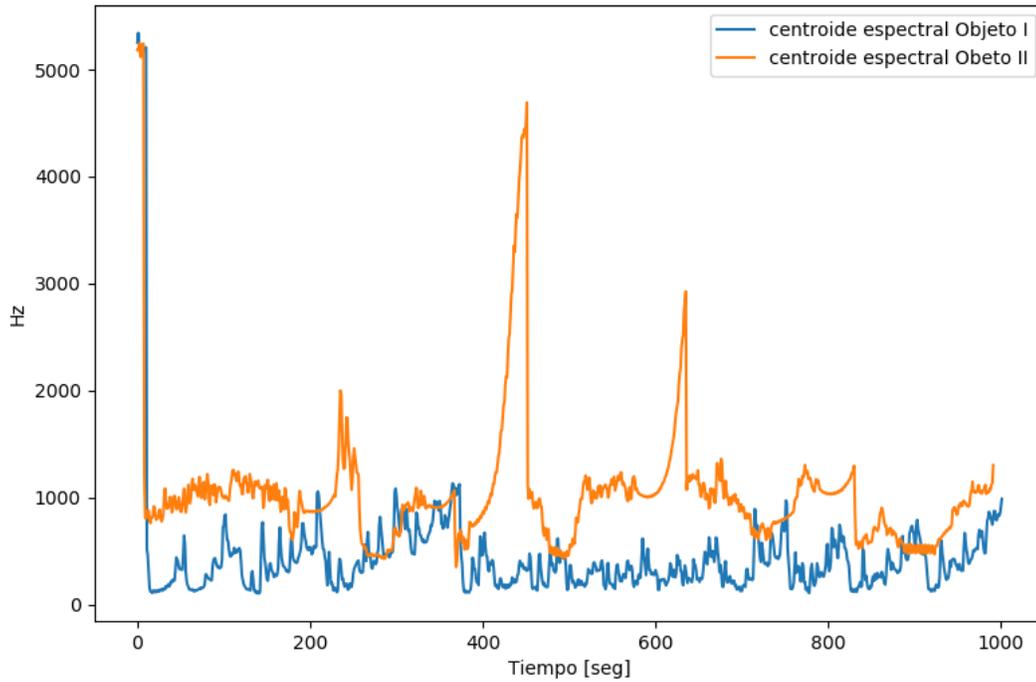


Figura 3-6: Series de tiempo del centroide espectral para los objetos sonoros I y II.

Bajo esta lógica, para calcular una distancia entre ambas series de tiempo se puede proceder del mismo modo se mencionó en el caso anterior; i.e. utilizando DTW.¹¹

¹¹Este proceso de comparación de distancias será mostrado a detalle en el siguiente capítulo.

Capítulo 4

Descripción espectromorfológica computacional y clasificación mediante aprendizaje máquina.

En este capítulo se describirá una aplicación concreta computacional que se desarrolló con el fin de realizar dos tareas específicas: descripción espectromorfológica y clasificación mediante aprendizaje supervisado dado un conjunto de muestras de audio.

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 precisamente a tres parámetros específicos: etiquetados en las muestras de audio: a) instrumento, b) nota o altura y c) dinámica.)¹ 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 *Alineación Global de Kernel k-Medios*² y *K-Vecindades*.

Toda vez que se han obtenido estos centroides, cualquier conjunto de muestras de audio puede ser comparado y/o clasificado (de acuerdo a un perfil espectromorfológico específico) ya sea por grados de similitud utilizando algoritmos como el *Dynamic Time Warping* (DTW) o mediante aprendizaje supervisado utilizando como etiquetas los centroides anteriores. De este modo, será posible asociar mues-

¹Las muestras de audio fueron obtenidas de la página web philharmonia.co.uk/explore/soundsamples.

²Ver apéndice ...

tras de audio cualquiera, a un *espacio timbre* conocido y de este modo, trabajar composicionalmente, dichas muestras de acuerdo a los principios de orquestación ya conocidos para los instrumentos musicales utilizados en el análisis primario.

La metodología anterior presenta además una diferencia notable con la mayoría de los procesos relacionados a la clasificación de audio mediante aprendizaje máquina; y es el hecho de que el análisis se realiza mediante la comparación de las series de tiempo de los descriptores obtenidos. La razón de lo anterior es que a juicio del autor, existen múltiples casos en los cuales no es posible aplicar medidas de tendencia central como la media o la desviación estándar ya que en dichos casos la información derivada de los descriptores no se presenta como una distribución normal. Lo anterior implica que la aplicación de dichas medidas arrojaría información sesgada, incorrecta o incluso incoherente al momento del análisis respectivo.

Por esta razón se decidió realizar análisis y clasificación mediante la comparación de las series de tiempo y haciendo uso de técnicas como el *Dynamic Time Warping*³. Más aún, los centroides son obtenidos mediante una hibridación de herramientas algorítmicas. Dado un *cluster* específico, se obtiene su respectivo *baricentro* y toda vez que se ha obtenido dicho baricentro se calcula una banda alrededor de él mediante el método de LB 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).

El capítulo expone la metodología basada en rutinas escritas en Python para poder realizar las tareas antes mencionadas de la manera más directa y automática posible.

Con estas rutinas, el usuario puede analizar cualquier número de muestras de audio⁴ eligiendo un conjunto de descriptores espectrales y de envolvente predefinido de acuerdo a sus intereses particulares o tomando como guía los resultados obtenidos con los conjuntos utilizados en las simulaciones aquí mostradas.

Las rutinas aquí 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 de este capítulo es entonces, el de establecer primeramente, un *espacio espectromorfológico de referencia* derivado del análisis de muestras de audio de instrumentos musicales convencionales y

³Ver apéndice

⁴Por supuesto la cantidad de muestras a analizar impacta directamente al gasto y procesamiento computacional de la estación en la que se realicen dichas tareas.

a través de clasificación no supervisada basada en un conjunto de descriptores espectromorfológicos específicos. Los centroides se definen como zonas o *bandas* alrededor de los baricentros calculados para cada cluster y con ello se construye el *espacio espectromorfológico de referencia*. Toda vez que se ha 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 a 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 a los principios de orquestación conocidos para los instrumentos musicales especificados y obtener de este modo, un panorama de información referencial para dichas muestras arbitrarias de audio.

Antes de continuar a las siguientes secciones es necesario establecer ciertas observaciones respecto a 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 de 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 similitud 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, así 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 composicional, ésta NO se presenta como una herramienta 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 composicional tales como la asociación con los principios orquestales de los instrumentos convencionales o la clasificación de muestras de audio de acuerdo a ciertos perfiles espectromorfológicos con fines de organización tímbrica.

La metodología aquí propuesta puede ser dividida en las siguientes etapas:

1. Obtención de los perfiles espectromorfológicos de las muestras de audio mediante los descriptores: zcr, rmse, ancho de banda espectral, flujo espectral y contraste espectral.
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.⁵
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 euclideano.
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 similaridad con muestras de audio arbitrarias.
8. Asociación de las muestras de audio arbitrarias al centroide más cercano y sumarización de las características del cluster correspondiente de acuerdo a los tres parámetros etiquetados en las muestras de audio: a) instrumento, b) nota o altura y c) dinámica.

4.1. Análisis primario de muestras

Para el análisis realizado en este capítulo se utilizó la base de datos alojada en philarmonia.co.uk/explore/sound_samp la cual tiene en total 27100 muestras de audio pertenecientes a los siguientes instrumentos musicales: cello, contrabajo, clarinete, bajo, clarinete bajo, basson, flauta, corno francés, guitarra, mandolina, oboe,

⁵Como se verá más adelante, este proceso es muy importante ya que la clasificación no supervisada de las series de tiempo es bastante demandante computacionalmente hablando.

saxofón, trombón, trompeta, tuba, viola y violín. Estas muestras incluyen notas en diferentes alturas tocadas a 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 melódicos⁶ y conservar únicamente aquellas muestras que presentaran notas únicas, se escribió una rutina que identificara de manera automática, las muestras que presentaran 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 presentara 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 arrays; 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 homogeneizar el inicio de las mismas.

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

```
#Rutinas modulares para cargar audios automáticamente de un directorio raíz. Paso 1.
#Ejemplo:
#data_sound, data_names = loadAll(/Users/Documents/samples)

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]
```

⁶Debido a que las variaciones tímbricas y de envolvente correspondientes a dichas muestras presentaban ambigüedad en la información obtenida de los descriptores.

```

    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)) if name[:2]!='. _']
    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 13533 muestras con las características antes mencionadas y este es 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; información que fue almacenada en formato de series de tiempo.

4.2. 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 lo que el uso de medidas de tendencia central como la media o la desviación estándar podrían fácilmente generar resultados espurios o fuera de contexto. 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 poder eficientizar 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). 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.

```

#Clasificación automática de muestras de un sólo onset.
#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_mtrx = factorSet(data_array, -2)
    #factorsMtrx = [int(len(x)/y) for x,y in zip(data_array, factor_mtrx)]
    #chop_array = [chop(z,y) for z,y in zip(data_array, factorsMtrx)]
    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 4-1, 4-2 y 4-3 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.

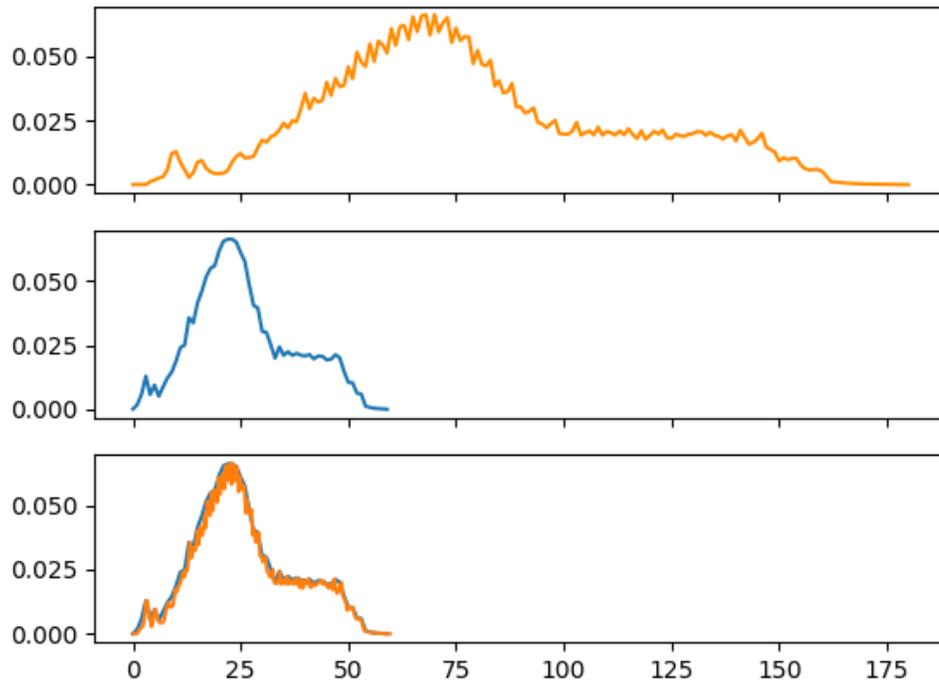


Figura 4-1: Comparación de series de tiempo para RMSE de audio de corno francés. Nota A2 tocada en mezzo forte.

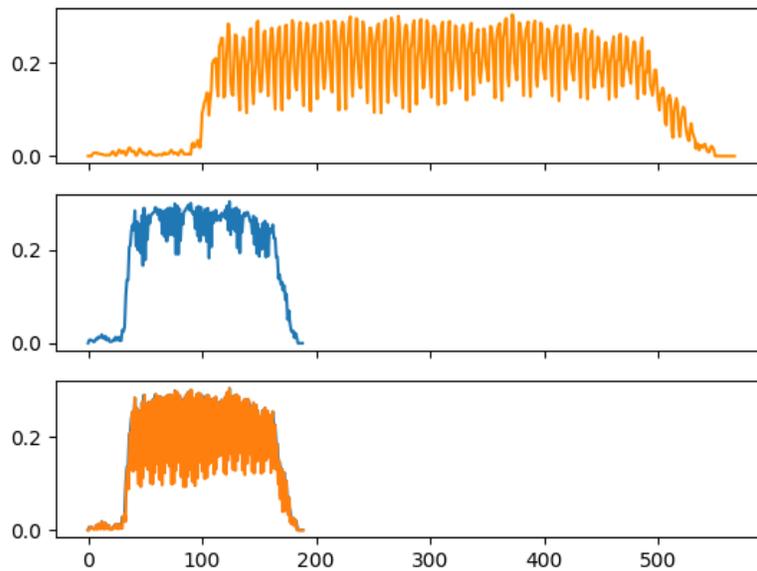


Figura 4-2: Comparación de series de tiempo para RMSE de audio de corno francés. Nota A2 tocada en fortissimo.

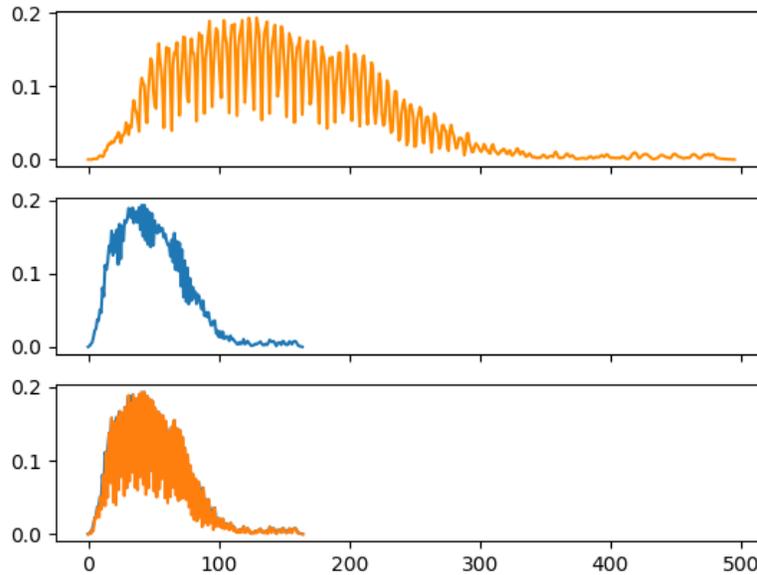


Figura 4-3: Comparación de series de tiempo para RMSE de audio de corno francés. Nota A3 tocada en trino forte.

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 4-4 en la que se pueden observar los onsets identificados en la serie de tiempo original mientras que la figura 4-5 muestra la construcción de la serie de tiempo recortada a partir de la interpolación de dichos valores.

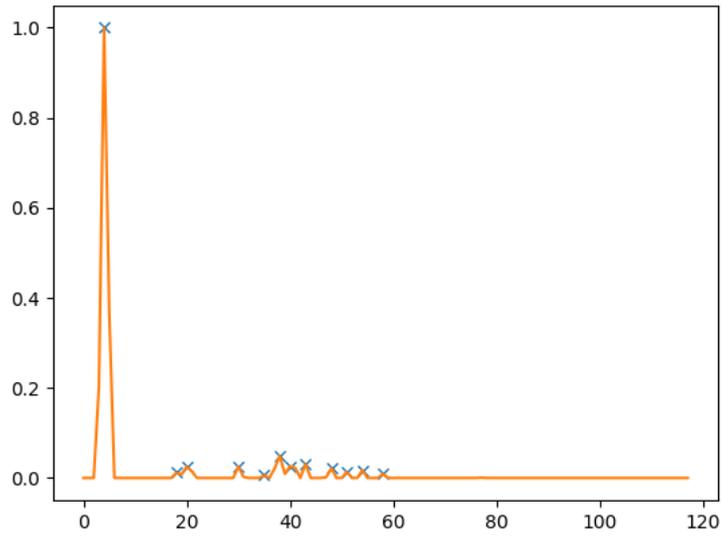


Figura 4-4: Identificación de onsets en la serie de tiempo.

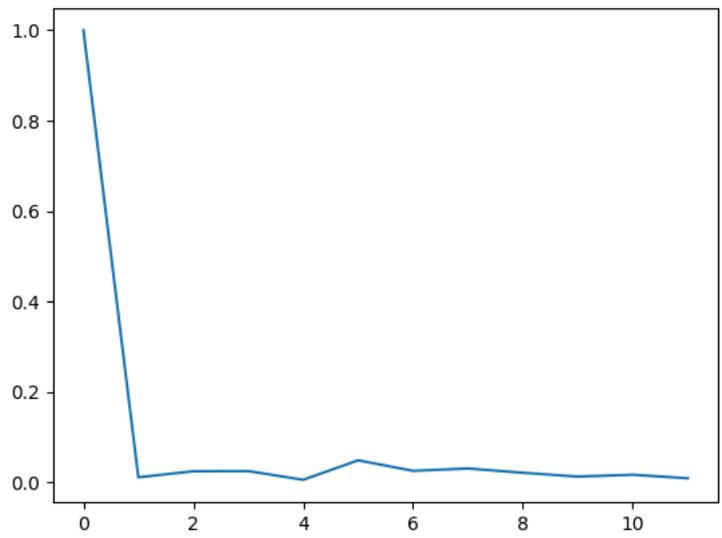


Figura 4-5: Reconstrucción de la serie de tiempo a partir de la interpolación de onsets.

4.3. Clasificación de series de tiempo mediante aprendizaje no supervisado.

Dado que se tiene un conjunto de 13533 de muestras de audio y 10 descriptores, el total de las series de tiempo obtenidas como fuente de información es de 135330. 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)⁷.

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.
2. Aplicación de formato particular de dichas series de acuerdo a 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 a los clusters obtenidos.

Se resalta el hecho de que 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 a su forma (envolvente) y posteriormente realiza la clasificación basada en K-Medios.

#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 r

```
#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))
```

⁷En dicha librería este algoritmo es nombrado *Global Alignment Kernel K-Means* o *GAK*.

```

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 4-6 a la 4-14 muestran ejemplos de clusters obtenidos con el proceso anterior para distintos instrumentos y distintos descriptores.

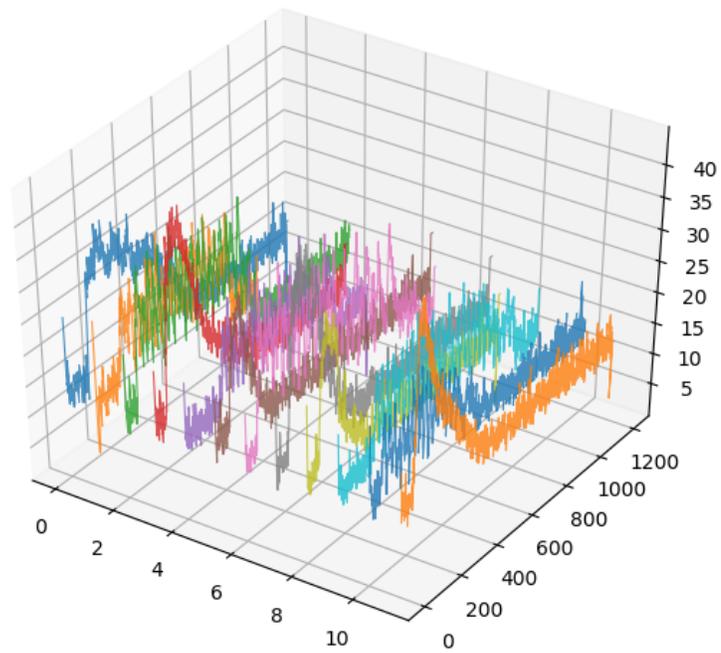


Figura 4-6: Cluster no. 2 de series de tiempo para la subbanda 4 (contraste espectral) de muestras de audio de flauta.

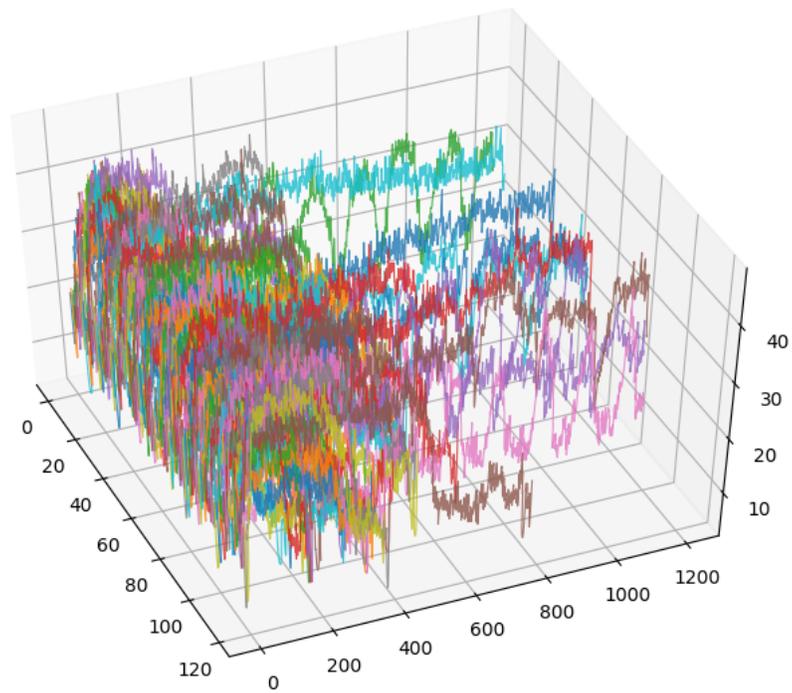


Figura 4-7: Cluster no. 3 de series de tiempo para la subbanda 4 (ancho espectral) de muestras de audio de corno inglés.

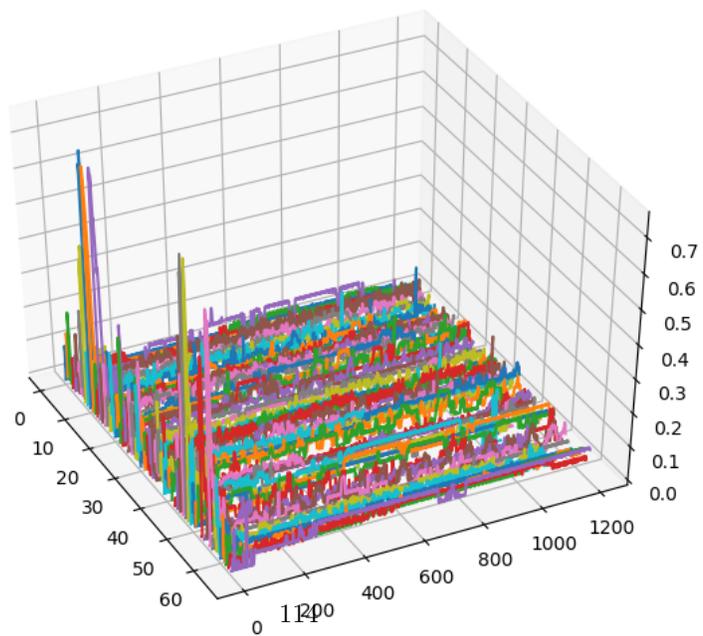


Figura 4-8: Cluster no. 3 de series de tiempo para el zcr de muestras de audio de saxofón.

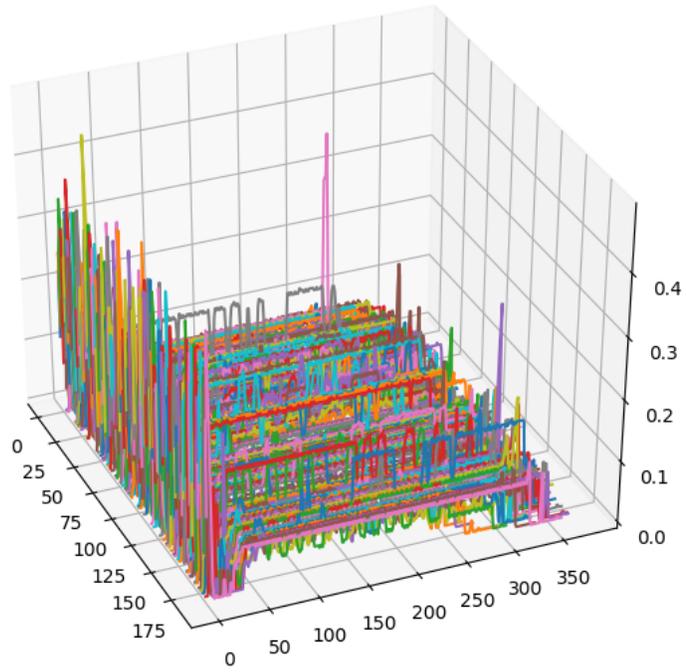


Figura 4-9: Cluster no. 0 de series de tiempo para el zcr de muestras de audio de oboe.

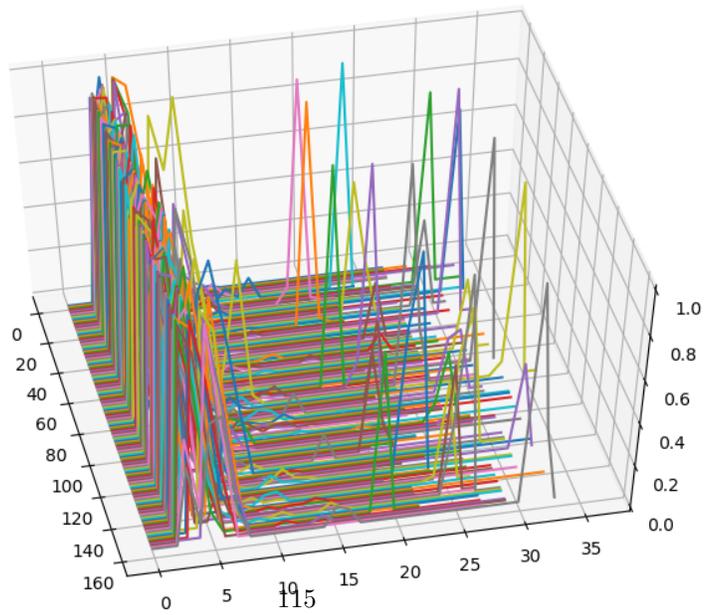


Figura 4-10: Cluster no. 5 de series de tiempo para el flujo espectral de muestras de audio de clarinete.

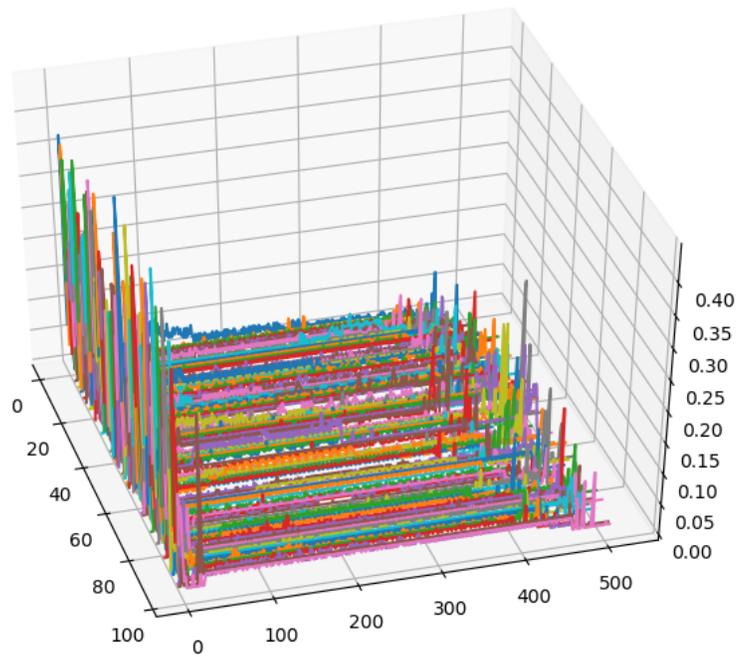


Figura 4-11: Cluster no. 5 de series de tiempo para el zcr de muestras de audio de mandolina.

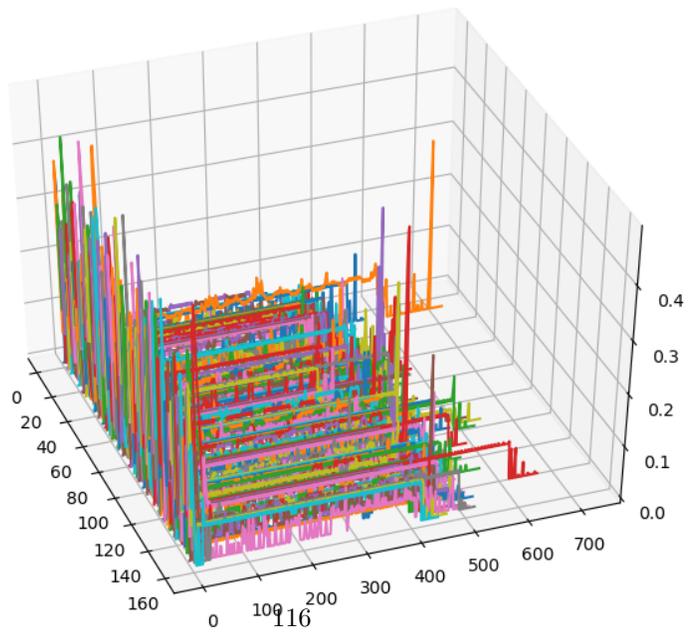


Figura 4-12: Cluster no. 3 de series de tiempo para el flujo espectral de muestras de audio de guitarra.

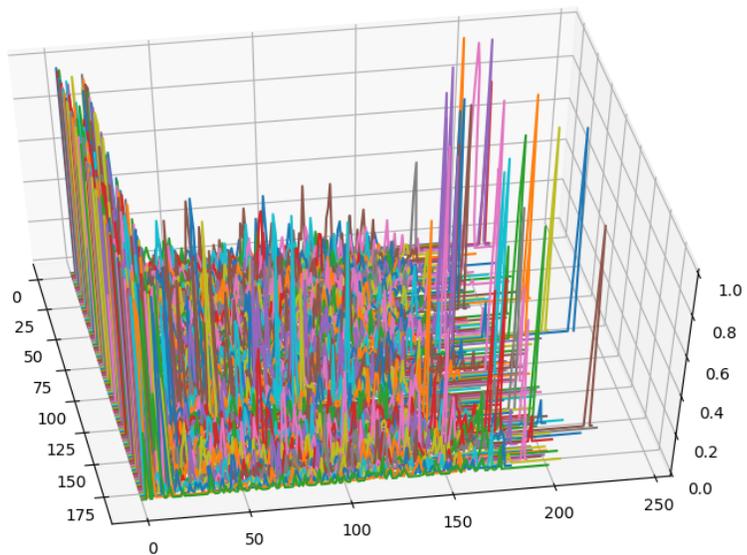


Figura 4-13: Cluster no. 2 de series de tiempo para el flujo espectral de muestras de audio de cello.

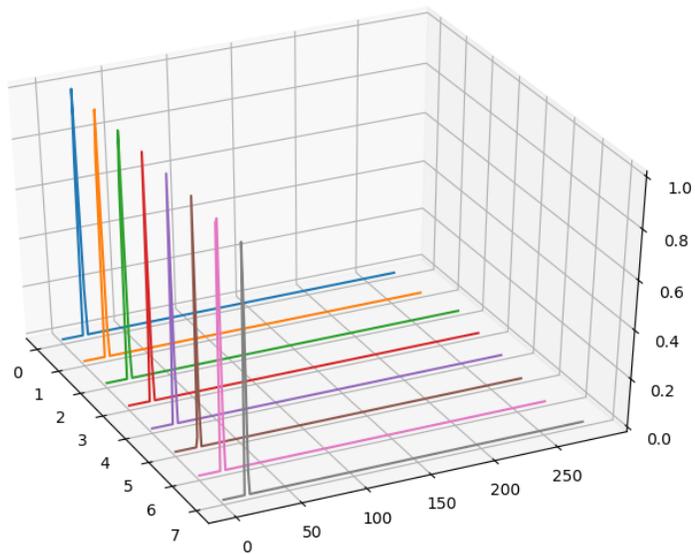


Figura 4-14: Cluster no. 0 de series de tiempo para el flujo espectral de muestras de audio de banjo.

La clasificación posee un alto grado de precisión aunque como puede observarse de 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. 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.

4.4. 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*⁸. Como su nombre lo indica, este método arroja una serie de tiempo que representa el *centro de masa* del cluster en cuestión y la cual puede ser considerada como la serie representativa de dicho cluster.

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 y sus límites inferior y superior están dados por el cálculo de las bandas *LB Keogh*⁹. 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 euclidiano 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:

```
from tslearn.barycenters import euclidean_barycenter, softdtw_barycenter, dtw_barycenter_averaging
```

⁸La 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.

⁹Véase el apéndice

```

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

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 4-15 muestra un ejemplo de cómo se construye el baricentro dado un cluster particular; en este

caso, para el flujo espectral correspondiente a las muestras de audio del violín.

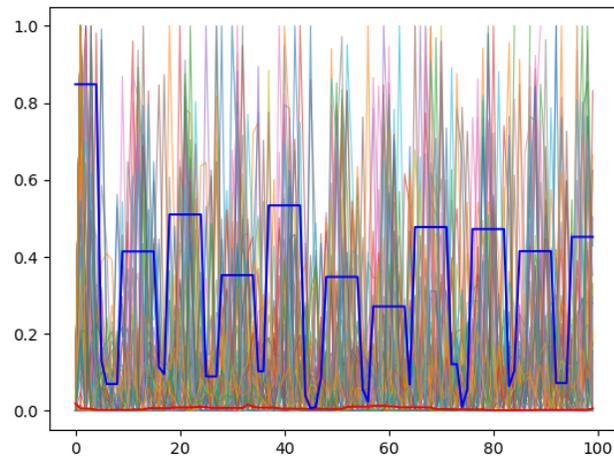


Figura 4-15: Baricentro euclideo para cluster de flujo espectral de muestras de audio de violín.

En las figuras 4-16 a 4-23 se muestran ejemplos específicos de clusters definidos mediante baricentros euclidianos en conjuntos con bandas LB Keogh superior e inferior.

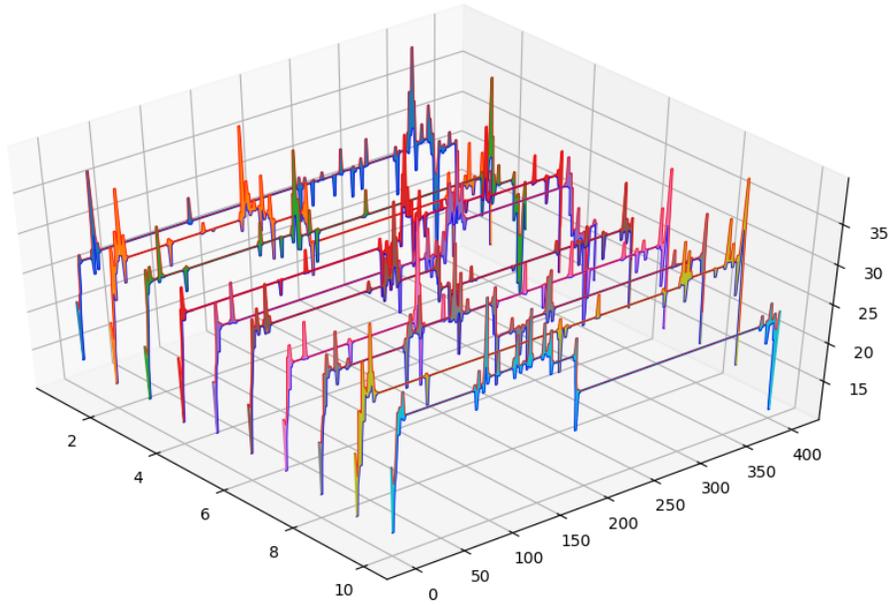


Figura 4-16: Centroide baricentro/bandas LB Keogh para subbanda 2 de ancho espectral de muestras de audio de clarinete

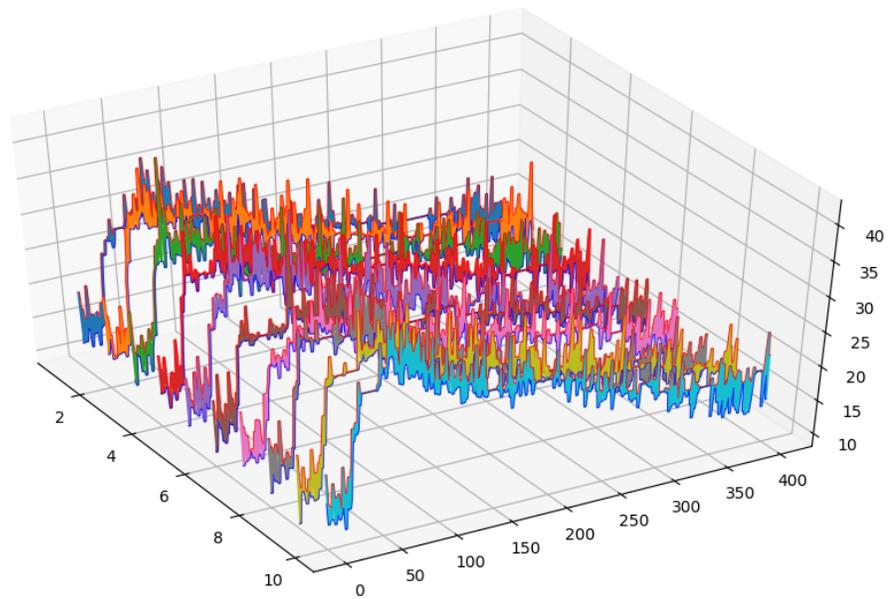


Figura 4-17: Centroide baricentro/bandas LB Keogh para subbanda 2 de ancho espectral de muestras de audio de contrabajo.

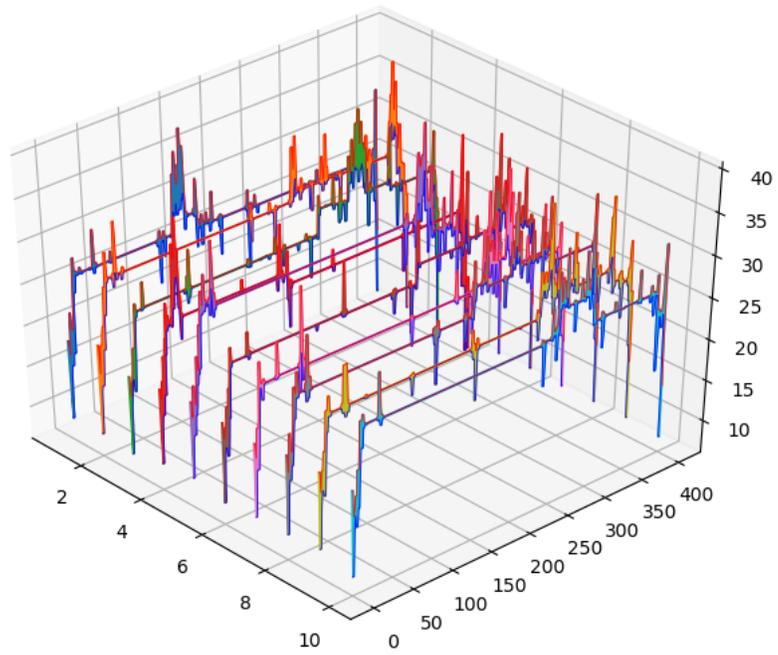


Figura 4-18: Centroide baricentro/bandas LB Keogh para subbanda 2 de ancho espectral de muestras de audio de tuba.

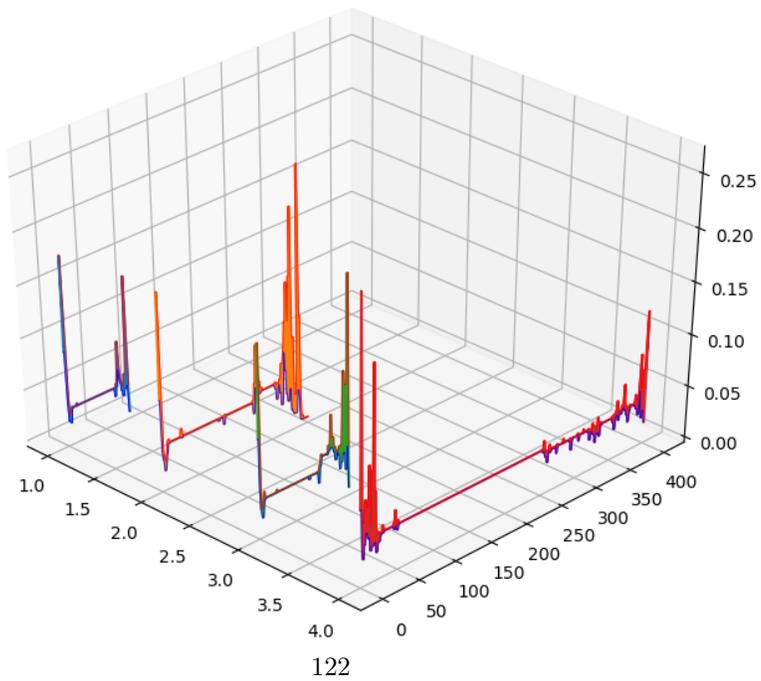


Figura 4-19: Centroide baricentro/bandas LB Keogh para zcr de muestras de audio de trombón.

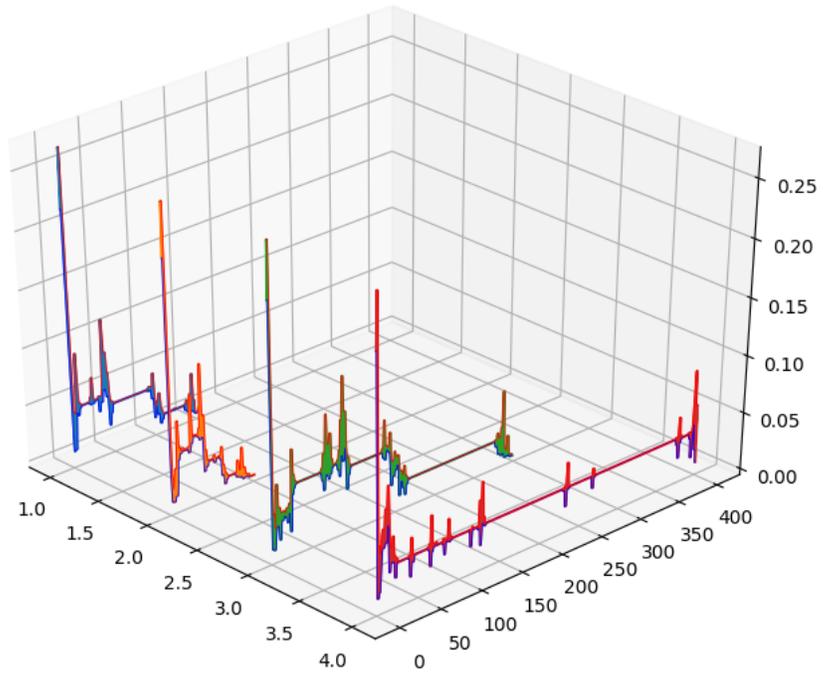


Figura 4-20: Centroide baricentro/bandas LB Keogh para zcr de muestras de audio de tuba.

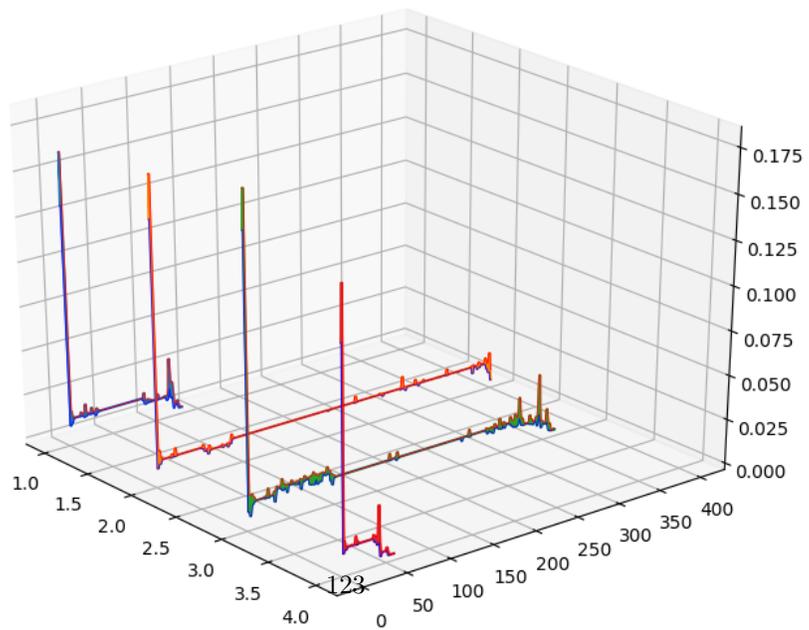


Figura 4-21: Centroide baricentro/bandas LB Keogh para zcr de muestras de audio de trompeta.

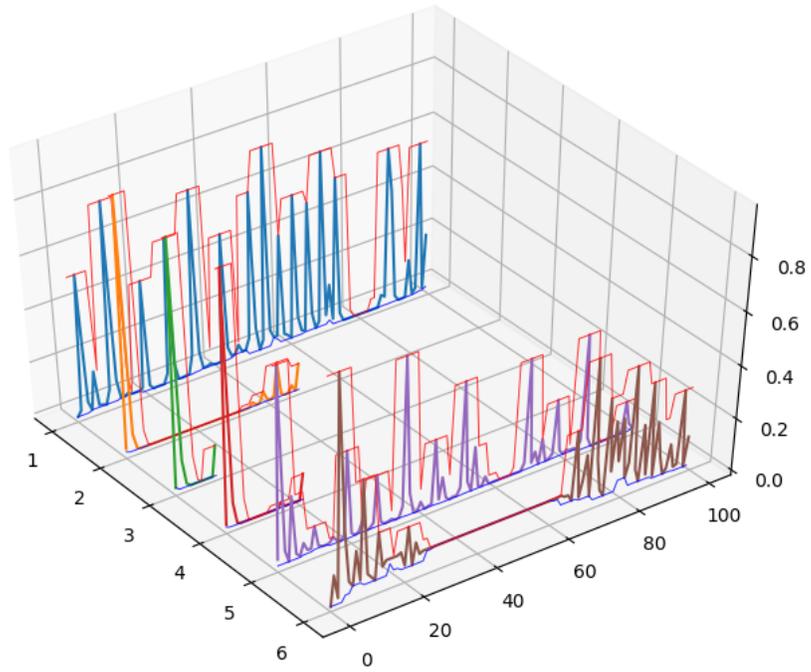


Figura 4-22: Centroide baricentro/bandas LB Keogh para flujo espectral de muestras de audio de clarinete bajo.

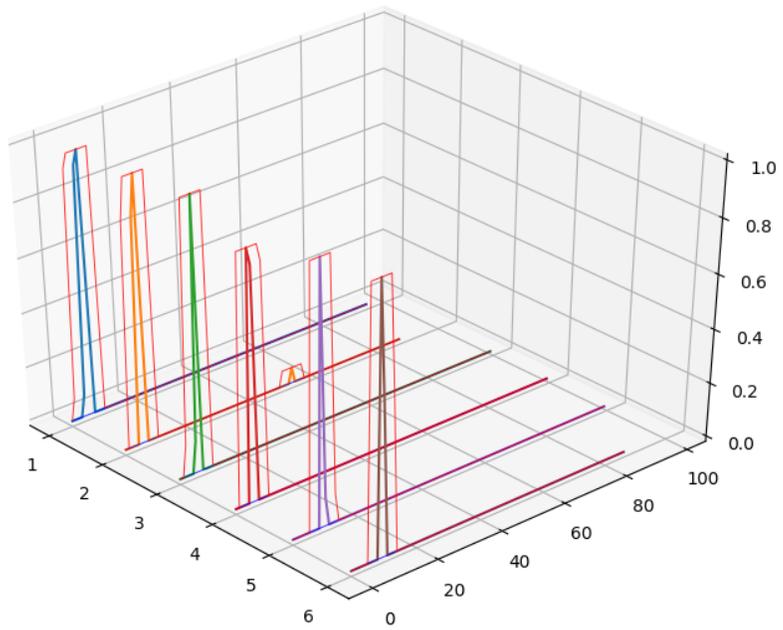


Figura 4-23: Centroide baricentro/bandas LB Keogh para flujo espectral de muestras de audio de clarinete bajo.

La ventaja de haber realizado el *clustering* para descriptores individuales y para muestras correspondientes a cada instrumento musical, es que entonces las muestras arbitrarias pueden ser comparadas de acuerdo precisamente a descriptores particulares asociados a muestras de audio de instrumentos musicales conocidos tocados con una dinámica específica, además del hecho de que dicha comparación no está sujeta o delimitada a un conjunto preestablecido de tales descriptores.

4.5. 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 ha construido entonces el *espacio espectromorfológico de referencia* y con este, 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, realizar la comparación de similaridad mediante DTW y asociar dicho descriptor de esa muestra al cluster correspondiente. Esto permitirá entonces, 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. Sustituir cualquier valor *nan* o *inf* por valores *dummy* para evitar errores.
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 al 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. Las figuras 4-23 a la 4-30 muestran 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*.

```
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})
```

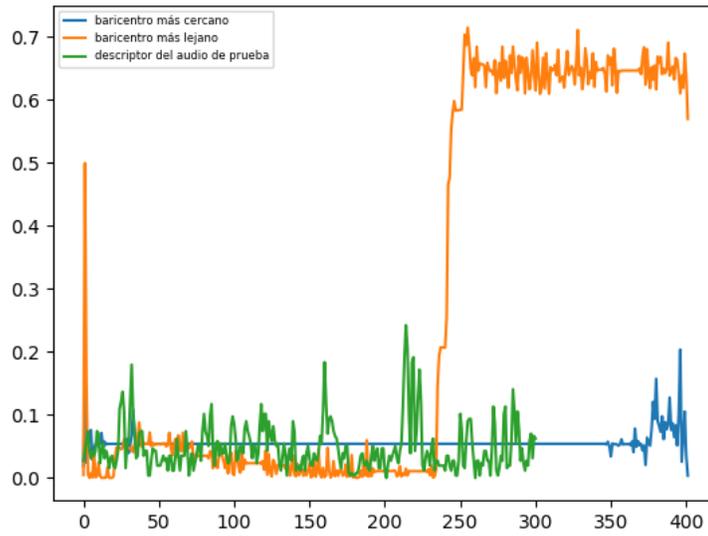


Figura 4-24: Comparación de muestra 1 para baricentros de zcr.

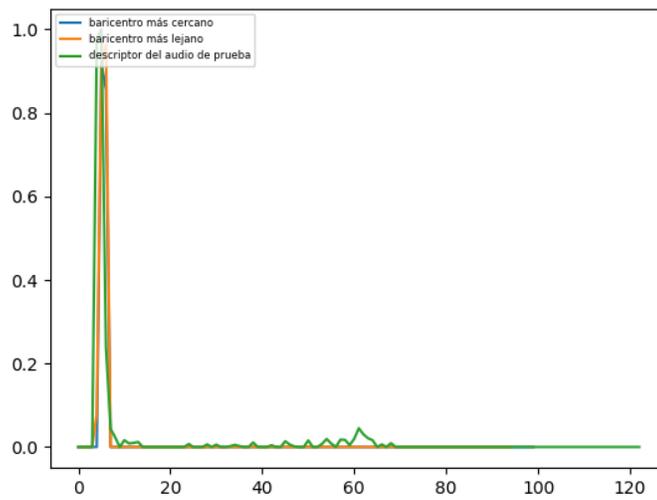


Figura 4-25: Comparación de muestra 1 para baricentros de flujo espectral.

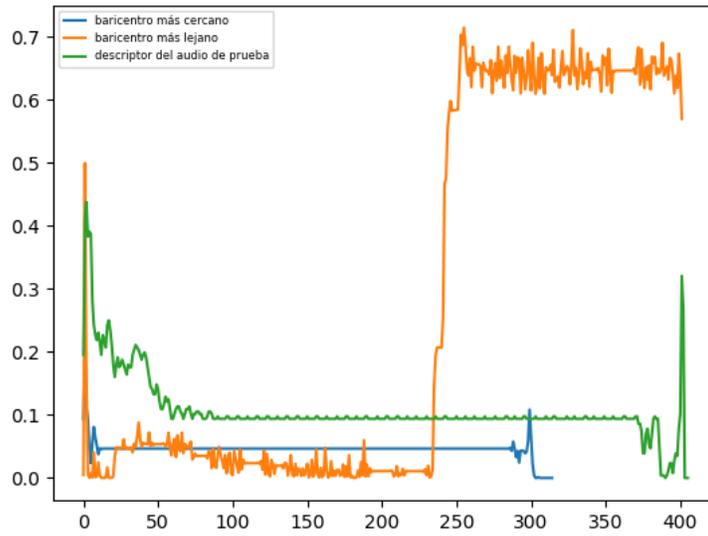


Figura 4-26: Comparación de muestra 2 para baricentros de zcr.

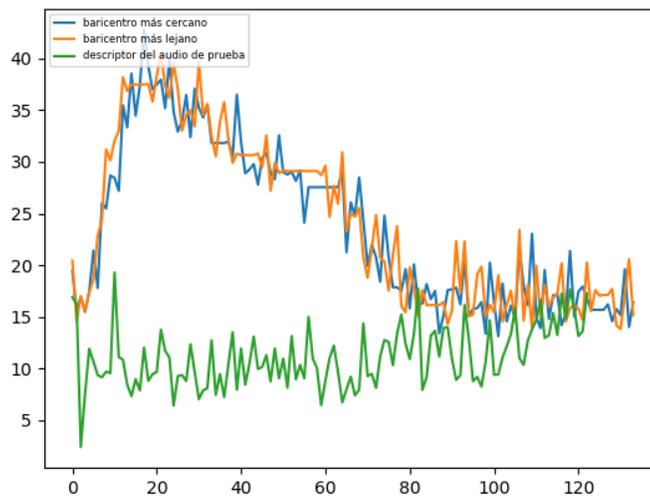


Figura 4-27: Comparación de muestra 2 para baricentros de subbanda 2 de contraste espectral.

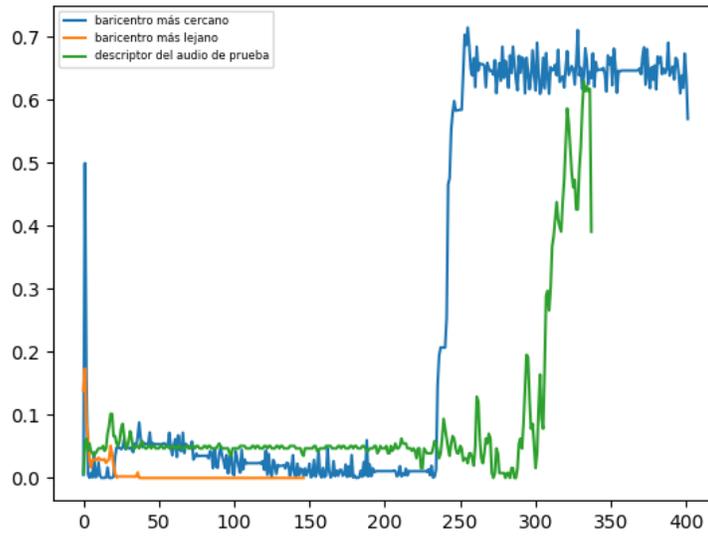


Figura 4-28: Comparación de muestra 3 para baricentros de zcr.

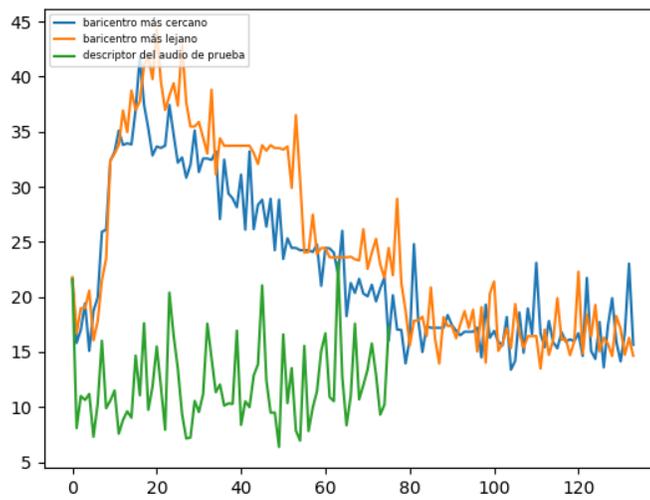


Figura 4-29: Comparación de muestra 3 para baricentros de subbanda 2 de contraste espectral.

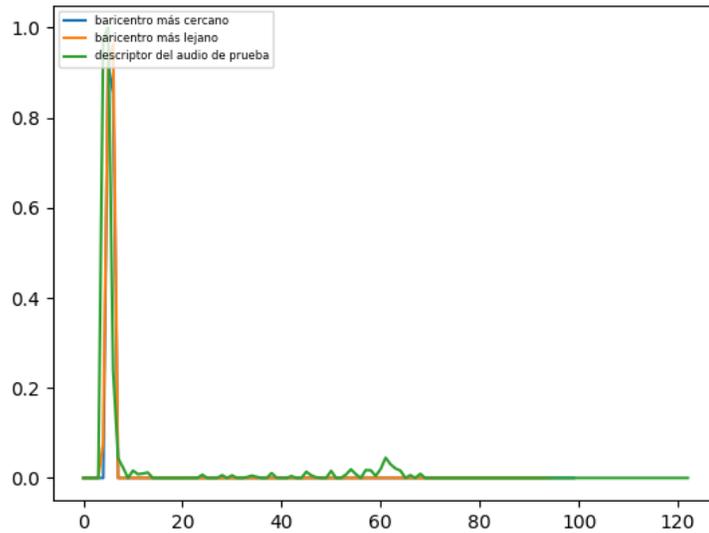


Figura 4-30: Comparación de muestra 4 para baricentros de flujo espectral.

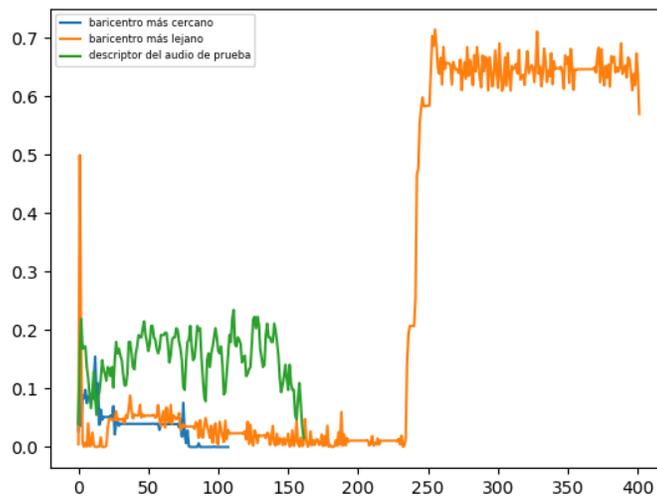


Figura 4-31: Comparación de muestra 4 para baricentros de zcr.

Para concluir el proceso es necesario establecer las relaciones de los baricentros con las muestras comparadas de acuerdo a 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 necesario que se establezca 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 array de tipo *string*.
2. Calcular matriz de combinaciones de pares de comparaciones entre entradas del array anterior para la métrica *Damerau*.¹⁰
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})
```

¹⁰Esta métrica es utilizada para medir grados de similitud entre cadenas de caracteres y la librería utilizada aquí para ese propósito fue *strsimpy* (Zhou Yang Luo, 2019).

Las figuras 4-31 y 4-32 muestran la forma de onda y el espectrograma de la muestra comparada 1 la cual se utilizará como ejemplo único para finalizar esta sección.

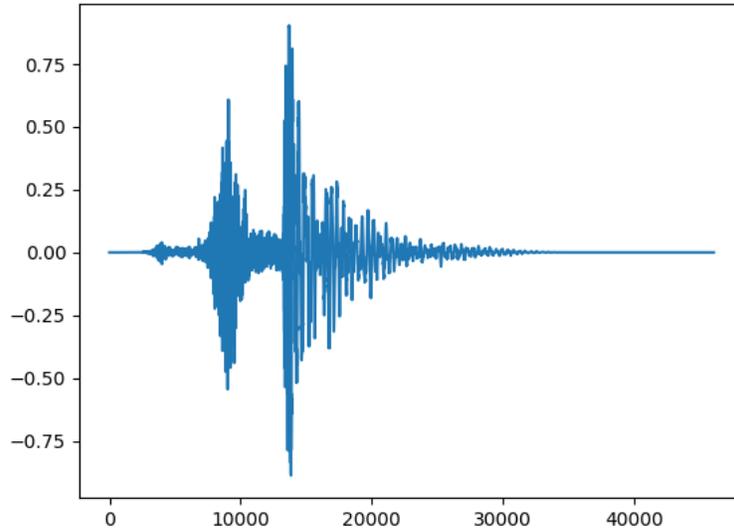


Figura 4-32: Forma de onda de la muestra 1

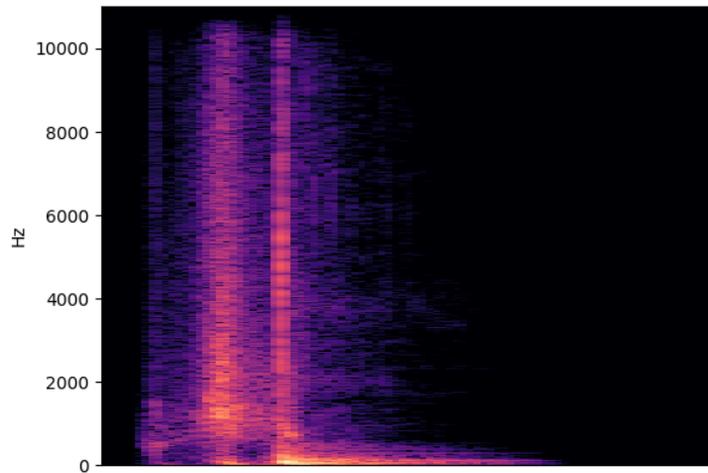


Figura 4-33: Espectrograma de la muestra 1.

La figura 4-33 muestra el cluster correspondiente al baricentro más similar para dicha muestra para el zcr. La figura 4-34 muestra el mismo cluster después de hacer la clasificación mediante la métrica

Damerau y el algoritmo de *Propagación de Afinidad*.

```
['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 4-34: 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 4-35: Cluster clasificado interiormente mediante propagación de afinidad, correspondiente al baricentro más cercano en zcr de la muestra 1

Finalmente, la figura 4-35 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 como que:

- 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*.

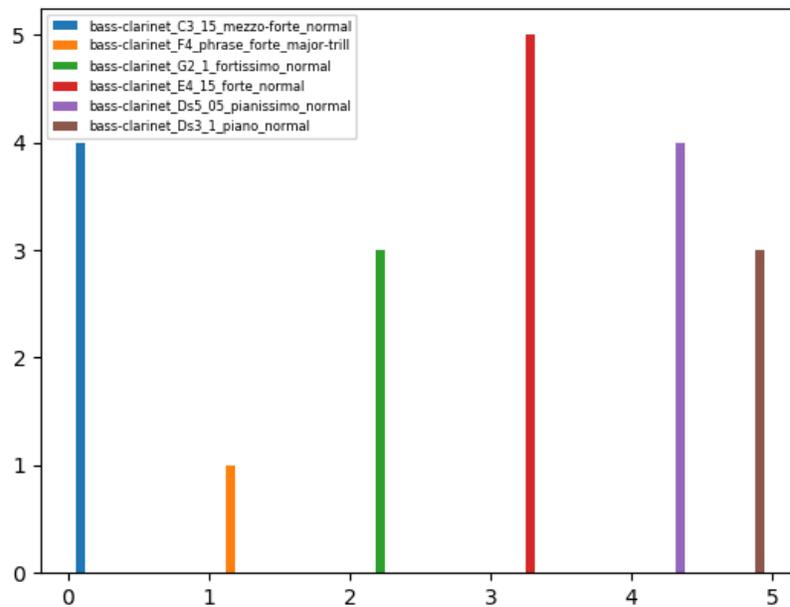


Figura 4-36: Histograma de representación de la muestra 1 para el baricentro más cercano en zcr

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 ciertos símbolos para realizar clasificación no supervisada sobre ese conjunto de descriptores.

La metodología anterior propone establecer un espacio espectromorfológico de referencia para comparar muestras de audio arbitrarias con perfiles espectromorfológicos de instrumentos musicales conocidos, permitiendo de este modo, extrapolar aspectos de orquestación e información acústica conocida de dichos instrumentos a las muestras de audio arbitrarias.

Por supuesto podría establecerse una metodología más usual como el hecho de introducir el conjunto de muestras de audio y realizar clasificación mediante algún algoritmo de aprendizaje no supervisado dado un conjunto de descriptores específicos. En esta sección 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 a su cluster asignado.

4.6. Conclusiones del capítulo

Este capítulo presentó una metodología de clasificación de muestras de audio basada en el análisis de perfiles espectralmorfológicos 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 4.6.1 *El costo computacional de realizar el análisis a través de series de tiempo en vez del de utilizar caracterizaciones numéricas como la media o la desviación estándar, es significativamente más elevado, sin embargo, dicho costo podría justificarse con el hecho de que es posible comparar muestras de audio de distinta longitud (debido a la alineación hecha por DTW) y con un mayor nivel de precisión (derivado del hecho de que la comparación de similitud se hace directamente de la serie de tiempo).*

Observación 4.6.2 *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 4.6.3 *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 4.6.4 *La metodología propuesta basada en establecer centroides de clasificación a partir de las bandas LB Keogh y los baricentros euclídeos, no ha sido abordada de ese modo en la literatura revisada y representa por lo tanto una propuesta del presente trabajo. Las ventajas de establecer los centroides de ese modo, son:*

- *Dado un cluster, el baricentro euclídeo 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 al baricentro o a alguna de las bandas LB Keogh dependiendo del contexto específico, y con la seguridad de que el área establecida por estos 3 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 4.6.5 *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 K-Vecindades o Máquinas de Soporte Vectorial o algunos tipos de redes neuronales. Solamente se debe tener especial cuidado de identificar 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 4.6.6 *Al establecer clasificaciones de muestras de audio tal y como se presentó en el presente capítulo, ya sea en relación a un espacio espectromorfológico de referencia o de manera directa, es posible hablar de **orquespectración virtual** en el sentido de la organización de los objetos sonoros (en un contexto composicional) de acuerdo a su perfil espectromorfológico obtenido computacionalmente y tomando en cuenta los principios y características asociadas a dicho objeto sonoro desde la parte perceptual y acústica.*

Observación 4.6.7

Capítulo 5

Análisis de obra: *Vortex Flow*.

Vortex Flow es una pieza acusmática en formato octafónico que fue compuesta entre mediados del 2017 e inicios del 2018 por el autor del presente trabajo. Dicha obra resultó finalista en el *Electroacoustic Music International Competition SIME 2018* (Lille, Francia). Fue seleccionada para el *New York City Electroacoustic Music Festival 2018*, *MUSLAB 2018* (Buenos Aires, Argentina) y el *Seoul International Computer Music Festival 2018*. De igual modo fue presentada en el Concierto Multimedia en el marco de la exposición *Escuchar con los ojos* del CCEMX (Centro Cultural España de México) 2018 y el *Festival Internacional Visiones Sonoras 2018* (Morelia, México).

La obra fue pensada como una *experiencia narrativa inmersiva* motivada por una perspectiva incidental; a saber, la historia del diluvio tal y como es contada en las tablillas sumerias y en donde se plantea a la pieza como la representación sonoro/inmersiva vivencial del personaje principal: *Ziusudra*.

Esta pieza representa el espacio de laboratorio en donde fueron aplicados, desarrollados, cuestionados y reformulados de manera directa, gran parte de los métodos, técnicas y conjeturas teórico-prácticas acerca del espacio-timbre y desarrollo de la inmersividad que se presentaron en esta tesis. Pero al mismo tiempo esta obra fue el detonante de confrontación creativa para investigar y explorar diversos conceptos y herramientas teórico-prácticas, sobre todo aquellas relacionadas con la acústica y la psicoacústica.

En este sentido, *Vortex Flow* representa el marco referencial tangible dentro del cual se desarrollaron a nivel personal desde la parte técnica, los siguientes puntos:

- Refinamiento de las técnicas de espacialización ya conocidas: VBAP y paneo multicanal por amplitud.

- Exploración de nuevas técnicas de espacialización: Ambisonics y microfónica multiarray.
- Implementación conjunta de las técnicas anteriores.

Desde el punto de vista del uso de material:

- Exploración de organización de objetos sonoros a través de análisis de perfil espectromorfológico.
- Exploración e investigación de los principios acústicos y psicoacústicos que aplicados a los objetos sonoros, permiten en mayor o menor medida una comprensión alternativa a la orquestación espectromorfológica de dichos objetos sonoros.
- Exploración e investigación de los principios acústicos y psicoacústicos que aplicados a los objetos sonoros, permiten en mayor o menor medida una comprensión alternativa a la localización espacial de dichos objetos.

En este sentido, *Vortex Flow* representa el punto de encrucijada el cual por un lado, estableció una suerte de laboratorio individual de constante contraste de principios, observaciones y metodologías respecto a la relación espacio-timbre (tal y como fue presentada en este trabajo) en la creación electroacústica multicanal y por otro, se posiciona (la obra misma) como el detonante para la investigación más profunda de diversos aspectos teórico-prácticos como los principios acústicos y psicoacústicos de la localización espacial, el desarrollo teórico del concepto de espacio y la profundización en el uso de técnicas de aprendizaje supervisado para la clasificación espectromorfológica.

Por lo tanto, la obra no es la culminación de la aplicación de los resultados presentados en este trabajo sino que es un dual teórico-práctico: espacio donde se aplicaron algunos de los resultados aquí mostrados y simultáneamente, detonante para investigar y proponer otros cuantos.

5.1. Herramientas tecnológicas utilizadas para el diseño espacial

La mayor parte de los objetos sonoros que aparecen en esta pieza son transformaciones digitales de procesos complejos sobre grabaciones tipo *sequence jeu*. Existen algunos objetos sonoros que corresponden a grabaciones de diversos objetos físicos y ambientes naturales cuyo procesamiento está limitado a la edición básica (suavizamiento de envolventes, estiramiento-contracción de la muestra, cambio de altura, etc.) y unos pocos más que fueron creados mediante síntesis digital.

Se utilizaron en particular 4 técnicas de espacialización de manera sinérgica a lo largo de la obra, en donde por supuesto, para diversos instantes de la misma, algunas de dichas técnicas eran combinadas simultáneamente o fueron utilizadas de manera individual:

1. VBAP.
2. Ambisonics.
3. Paneo multicanal por amplitud¹
4. Espacialización por microfónica multiarreglos.²

De las técnicas mencionadas con anterioridad las tres primeras resultan bastante familiares, mientras que la última posiblemente no lo sea tanto. A continuación se presentará de manera panorámica, un breve análisis de las herramientas antes mencionadas en relación a su funcionalidad dentro del contexto del diseño espacial multicanal. Se hace hincapie que este análisis, es producto de un proceso tipo laboratorio llevado a cabo durante el periodo de investigación-práctica a través del cual se fueron planteando, formulando, reformulando, reforzando o corrigiendo los distintos resultados teórico-prácticos que se han abordado a lo largo de este trabajo de investigación.

Se plantea desde este punto que las observaciones y análisis desarrollados a continuación respecto a cada una de las herramientas de espacialización, parten del hecho de establecer como premisas, todos los resultados y conjeturas expuestos con anterioridad en este trabajo respecto a los objetos sonoros en sus perfiles espectromorfológicos y espaciales.

5.1.1. VBAP

Existen en la actualidad muchísimas interfaces VST y plugins que permiten hacer uso de esta técnica y de entre ellas son notables distintos casos que permiten hacer diseños de trayectorias espaciales como por ejemplo, OctoGris3 (Groupe de Recherche en Immersion Spatiale at Université de Montréal, 2016), Spatium (Penha 2013), Holophon (Bascou 2010), OMPrisma (Schumacher 2010), los últimos módulos de las *GRM Tools*, así como las últimas integraciones en OpenMusic (Bresson 2010), además claro está, de todos los *patches* y *standalone* desarrollados en MAX MSP por diversos artistas independientes. Algunas

¹Esta técnica fue utilizada mediante el algoritmo de espacialización desarrollado en Supercollider, por el autor de este trabajo y que puede ser consultado en (Soria,).

²Esta técnica está actualmente siendo estudiada y profundizada por el autor de este trabajo utilizando como referencia directa los trabajos de investigación de grabación multicanal de Michael Williams. Por supuesto, uno de los antecedentes más conocidos al respecto es el proceso desarrollado por Stockhausen para *Kontakte*.

herramientas como FLUX Spat (Carpentier 2015) o Spatium (Peña y Oliveira 2013) funcionan como integraciones de distintos métodos de espacialización por lo que pueden ser usadas tanto para VBAP, como Ambisonics y paneo por amplitud.

De acuerdo a los resultados y conclusiones personales del autor de este trabajo, se puede afirmar que VBAP es un método de espacialización que funciona de manera eficaz para localización de objetos sonoros fijos (con el perfil espectromorfológico adecuado, claro está) dentro de un espacio audible. Sin embargo, es bastante complicado obtener una sensación de distancia con un nivel de naturalidad aceptable para dichos objetos sonoros; incluso estableciendo espacios sonoros virtuales mediante reverberación y filtrado de frecuencias altas³ Lo anterior implica entonces, que mediante VBAP, es posible desarrollar escenas espaciales de múltiples objetos fijos con un nivel aceptable de localización y dirección pero con una baja resolución en la percepción de la distancia y que para subsanar dicha limitación se requiere un trabajo digital adicional que no siempre resulta en una naturalidad aceptable. Por supuesto, existen objetos sonoros muy particulares para las cuales la percepción de distancia puede ser lograda de manera más sencilla únicamente con amplitud y filtrado; sobre todo, aquellos con onset cortos y contenido espectral arriba de los 3-4 kHz.

Para el caso de trayectorias continuas, VBAP funciona eficazmente para los objetos sonoros mencionados con anterioridad y dependiendo de la interface propia del plug in, el diseño de dichas trayectorias puede resultar mayor o menormente complejo. Sin embargo, al igual que en el caso anterior, la percepción de la distancia para fuentes en movimiento tiene una resolución baja, tanto para la velocidad angular como para la velocidad radial. Lo anterior además implica que la percepción final de la distancia de dichas trayectorias, está mayormente concentrada en la periferia del espacio audible por lo que trayectorias que en su diseño serial crucen o pasen cerca del *sweet spot*, tendrán una disimilitud serial-perceptual considerable.

Para el caso de superficies sonoras fijas VBAP arroja una resolución de envolvimiento y dispersión a través del espacio audible, mayor a los casos anteriores por lo que la superposición de superficies texturales fijas (tomando en consideración los resultados presentados en los capítulos anteriores) puede desarrollarse exitosamente con este método de espacialización. Sin embargo, dependiendo del plug in utilizado, aplicar un movimiento a las superficies sonoras dentro del espacio audible, puede resultar bastante complicado por lo que este método de espacialización pudiera no ser el más adecuado para ese

³Debe recordarse que en la percepción acústica del sonido, sobre todo en espacios abiertos, un componente importante que permite discriminar la distancia es precisamente el filtrado de frecuencias altas debido tanto a la absorción del medio como a la disminución de la capacidad de propagarse a través de dicho medio debido a la longitud de onda.

propósito.

En conclusión, VBAP funciona bastante bien para localización fija de objetos sonoros pero con una baja resolución en la percepción de distancia. Debido a esta razón, las posiciones fijas y trayectorias continuas tienden a percibirse más hacia la periferia del espacio audible⁴. Lo anterior también implica que dicho método presenta dificultades considerables para lograr un grado de naturalidad en la creación y diseño del espacio, sin embargo, como herramienta base de espacialización y debido a la gran popularidad que tiene en su uso, es una alternativa viable que puede funcionar de manera bastante eficaz sobre todo para objetos sonoros sintéticos u objetos concretos que han sido tan procesados que resulte difícil establecer una analogía relacional perceptual con el material físico de dicho objeto sonoro.

5.1.2. Ambisonics

Dependiendo del uso que se le dé, Ambisonics puede resultar una opción realmente viable para establecer campos sonoros distribuidos dentro de un espacio audible con un nivel de localización bastante aceptable. Por supuesto como es bien sabido, la arquitectura del espacio audible debe ser altamente simétrica y todos los parámetros de distancias y ángulos físicos de dicha arquitectura deben estar correctamente medidos para que el algoritmo pueda ofrecer una resolución espacial adecuada.

En este sentido, Ambisonics resulta muy útil para realizar traslaciones de campos sonoros entre distintas configuraciones de bocinas mediante el proceso codificación-decodificación en *B-Format* siempre y cuando como ya se dijo, la arquitectura del espacio audible sea óptima. A pesar de que la resolución en la localización puntual o de trayectorias continuas de objetos sonoros puede ser baja (sobre todo si la arquitectura del espacio audible no es lo suficientemente simétrica), Ambisonics, siendo un algoritmo de distribución de campos sonoros, ofrece una alternativa muy eficaz para el uso de superficies texturales tanto fijas como dinámicas, posicionándose de este modo, como una herramienta bastante útil para desarrollar la sensación de envolvimiento.

5.1.3. Paneo multicanal por amplitud

Este método es comúnmente utilizado en distintas plataformas digitales y su referente básico análogos la espacialización en vivo mediante la manipulación de *faders* de una mezcladora digital para conciertos multicanal y/o *acusmonium*; lo que es comúnmente conocido como *live diffusion*.

⁴Por supuesto, esto variará de acuerdo al perfil espectromorfológico de cada objeto sonoro.

El autor de este trabajo de investigación, desarrolló un algoritmo que implementa este método en Supercollider y permite realizar diseño espacial dentro de cualquier configuración de bocinas (Soria 2012). Se nombró a dicho algoritmo *PanW* y está escrito a modo de archivo del tipo *clase* dentro de Supercollider. Al haber sido desarrollado como un algoritmo basado en funciones coseno, su portabilidad hacia otros lenguajes de programación no resulta demasiado complicada.

Este algoritmo es el que fue y ha sido utilizado por el autor de este trabajo, tanto para fines de creación como de performance en tiempo real en todas las obras desarrolladas. El algoritmo utiliza un patrón complejo de combinación de paneo por funciones coseno de modo tal que permite tener control sobre el movimiento espacial de fuentes sonoras en tres parámetros principales: diseño arbitrario de trayectorias⁵, velocidad de la trayectoria del objeto sonoro y el *spread* o número de bocinas entre las cuales se distribuye el paneo.

Cuando se utilizan objetos sonoros con perfiles espectromorfológicos adecuados en relación a las velocidades óptimas de los mismos, la disimilitud serial-perceptual de las trayectorias móviles de dichos objetos sonoros disminuye de manera considerable utilizando este algoritmo. Del mismo modo, este algoritmo permite posicionar superficies texturales a través del espacio audible de manera bastante sencilla y con una buena resolución de localización.

Puesto que esta herramienta de espacialización está basada en el método de paneo multicanal por amplitud usando funciones coseno, la continuidad de las trayectorias móviles se percibe en la mayoría de los casos, con un nivel bastante aceptable. Del mismo modo, el método mencionado parece poseer un nivel de resolución mayor a VBAP para el caso de trayectorias que cruzan o pasan cerca del *sweet spot*.

5.1.4. Microfonía multiarray

Este método tiene su antecedente directo como ya se mencionó, en el proceso composicional de la obra *Kontakte* (1960) de Stockhausen quien utilizó un arreglo de microfonía cuadrafónica para captar el movimiento rotatorio de una bocina.

Por otro lado, el referente teórico más importante es sin duda, todo el corpus de investigación realizado por Michael Williams a lo largo de varias décadas acerca de los distintos métodos de grabación multicanal mediante arreglos de microfonía multiarray. Si bien Williams ha centrado su trabajo en formatos surround comerciales, el material teórico y práctico que se desprende de ello, puede ser extrapolable

⁵Es decir, el usuario puede especificar el camino o sucesión arbitraria de bocinas a través de la cual se trasladará o posicionará el sonido.

a configuraciones multicanal orientadas a la experimentación sonora o al campo de la electroacústica.

Tomando como referencia este material, el autor de este trabajo se ha enfocado en realizar diversas prácticas de exploración y experimentación usando grabación multicanal a través de microfónica multiarray con fines de creación electroacústica.

En este sentido, *Vortex Flow* representó para el autor, el punto en el cual surgió la necesidad de explorar formas alternativas de espacialización y la microfónica multiarray fue una de ellas. Cabe señalar que el material espacializado por microfónica multiarray que aparece en la obra, es bastante puntual y se limita a unos cuantos objetos sonoros que aparecen a la mitad de la misma. Sin embargo, como se mencionó anteriormente, la construcción de la pieza y el proceso creativo relacionado a ella permitieron detonar las primeras exploraciones y prácticas empíricas sobre dicho método de espacialización, el cual hoy en día está siendo explorado y estudiado de manera más profunda por el autor de este trabajo. Además, diversos resultados presentados en este trabajo fueron desarrollados con base al estudio y reflexión de este método; sobre todo los relacionados con la *disimilitud serial-perceptual* y la construcción teórica del *espacio virtual asociado*.

Habiendo trabajado con configuraciones de 4,6 y 8 microfófonos de condensador del tipo cardioides, se puede afirmar que este método ofrece una resolución de localización muy por encima de los métodos digitales. Dependiendo de la distancia entre los microfófonos la resolución de localización puede presentar disminuciones naturales debido a la intersección de los patrones polares, como sucede en el caso de una microfónica cuadrafónica. El arreglo anterior mejora considerablemente cuando se introducen dos microfófonos más con el fin de establecer frentes de captación que cubran en su totalidad el espacio delimitado por dicho array. Algunos ejemplos se muestran en las figuras...

Este método puede pensarse como una *reconstrucción digital de la espacialidad natural* y debido a esta misma razón, el nivel de naturalidad percibida es considerablemente elevado. Cabe mencionar que aún a pesar de la disposición de los microfófonos, existe una captación muy eficiente de la acústica del recinto, contribuyendo de este modo a un aumento en la resolución de la *espaciosidad*⁶. Las configuraciones utilizadas hasta ahora, presentan ciertos problemas de discontinuidad para el caso de trayectorias móviles de velocidades superiores a los 3°/seg aproximadamente. Lo anterior está por supuesto, directamente relacionado con el perfil espectromorfológico del objeto sonoro, el cual estará a su vez directamente relacionado con el *sequence-jeu* que se realice con la fuente acústica. Sin embargo, para objetos sonoros con morfología percusiva y perfil espectral arriba de los 2 KHZ aproximadamente, las

⁶Véase sección...

trayectorias móviles son bastante bien captadas si la velocidad de las mismas se encuentra dentro del rango ya mencionado.

De igual modo, aún para configuraciones de micrófonos 2D y sistemas de reproducción también 2D, existe una limitada pero notablemente perceptible, resolución en plano de elevación, lo que posiciona a este método como una herramienta bastante eficaz para fines de naturalidad en la percepción de localización, de distancia y de espacialidad. Se ha observado además, que este último elemento no afecta en la percepción de la naturalidad cuando se superponen objetos sonoros espacializados por métodos digitales; dicho de otro modo, la sensación de recinto captada por la microfónica multiarray, no entra en conflicto con los objetos sonoros espacializados mediante los algoritmos digitales (los cuales por lo descrito anteriormente, ofrecen una sensación de distancia no significativa).

5.1.5. Comparación de las herramientas.

Las observaciones anteriores, se reitera, son producto de haber experimentado, explorado, estudiado y aplicado de manera individual o conjunta los métodos de espacialización antes descritos, en obras y ejercicios de composición electroacústica multicanal. En este sentido, *Vortex Flow* representa el punto cumbre de análisis-laboratorio, que permitió asentar de manera clara y definida dichas observaciones y por supuesto, gran parte del material expuesto a lo largo de este trabajo.

A continuación se describirá a manera de sumario, las conclusiones pragmáticas a las cuales se han llegado, acerca del uso de las herramientas de espacialización antes descritas, dentro del proceso de creación electroacústica personal. Este sumario servirá como base para establecer un análisis posterior del diseño y construcción del espacio dentro de la obra en cuestión.

Localización puntual fija

- La herramienta que presenta una mayor resolución para localización puntual fija y percepción de distancia es sin duda la microfónica multiarray. Esta herramienta presenta la natural limitación de que sólo puede ser utilizada con fuentes acústicas físicas o mediante la captación de reproducción de fuentes digitales a través de bocinas colocadas en posiciones específicas.
- VBAP presenta una facilidad sobresaliente en su uso debido a la gran cantidad de plug ins existentes y a las distintas interfaces gráficas que varios de ellos ofrecen. A pesar de tener una resolución baja para la percepción de distancia, es una herramienta que representa de manera muy aceptable la

localización direccional de los objetos sonoros.

- Ambisonics no presenta una ventaja sobresaliente para la percepción de la localización direccional o de distancia sobre alguno de los otros métodos de espacialización. Aunado a las restricciones de las condiciones óptimas en la arquitectura del espacio audible, podría resultar una herramienta no tan eficiente para estos fines. Sin embargo, debido a su naturaleza propia, resulta bastante útil para trasladar ubicaciones espaciales entre distintas configuraciones multicanales y/o formatos, como por ejemplo de binaural a octafónico o haciendo uso de grabación mediante micrófonos especializados como el *Ambeo*.⁷
- El paneo multicanal por amplitud, en su versión del algoritmo *PanW*, presenta una resolución similar al VBAP pero para propósitos de ubicación específica de objetos sonoros fijos, su uso podría resultar más complicado que el de VBAP, siendo similar en este aspecto (el de la facilidad del uso) a Ambisonics.

Como conclusión de lo anterior, para el diseño espacial que requiere la colocación objetos sonoros fijos en ubicaciones específicas del espacio audible, la microfónica multiarray es la que arroja la resolución más alta en cuanto a dirección, distancia y espacialidad, sin embargo, al combinarse con VBAP (para fuentes digitales), es posible desarrollar espacios virtuales bastante definidos. Más aún, la interacción de objetos sonoros espacializados con microfónica en relación a aquellos espacializados con VBAP, no generan un conflicto en la percepción de la naturalidad, sino que más aún, y presumiblemente por el principio de contraste, dicha interacción permite establecer una diferenciación notable en la percepción de la distancia espacial lineal entre dichos objetos y en la percepción del espacio virtual que surge como consecuencia de ello.

Trayectorias móviles

- La microfónica multiarray ofrece una alta resolución en la continuidad de las trayectorias móviles si las fuentes acústicas mantienen una velocidad relativamente baja y son desplazadas dentro del espacio de microfónica sin ningún obstáculo que pudiera crear una sombra acústica. Ofrece además una resolución bastante alta para la velocidad angular y radial con las condiciones antes descritas. Para velocidades arriba de los $3^\circ/\text{seg}$ aproximadamente, la continuidad se comienza a ver fragmentada dependiendo por supuesto del perfil espectral de cada objeto sonoro.

⁷© 2019 Sennheiser electronic GmbH & Co. KG.

- VBAP y el algoritmo *PanW* presentan resoluciones bastante similares, sin embargo, *PanW* ofrece la ventaja de poder programar cualquier tipo de trayectoria (por compleja que sea) entre cualquier tipo de configuración multicanal de manera mucho más eficiente y rápida. Dependiendo de las necesidades composicionales, las interfaces gráficas de VBAP pudieran resultar más simples de utilizar para creadores acostumbrados a ello.
- Las trayectorias con Ambisonics pueden resultar difíciles de diseñar pero como se mencionó con anterioridad, esta herramienta resulta útil cuando se extrapolan los componentes espaciales entre distintos formatos o configuraciones.

Superficies sonoras

- Para este componente del diseño espacial, *PanW* funciona bastante bien (aumentando el valor del *spread*) sobre todo si se requiere aplicar un grado de movilidad a dichas texturas. Si se aplican correctamente las conjeturas.. para múltiples texturas sonoras, la resolución en la percepción de la individualidad de las mismas es alta. De nuevo, la facilidad en términos de uso, es mayor que VBAP, aunque la resolución de la percepción de envolvimiento resulta bastante similar entre ambas.
- La mayor resolución para la percepción de envolvimiento está dada por Ambisonics, aún cuando la arquitectura del espacio audible no sea lo suficientemente ideal. En este sentido, la distribución de campos sonoros extendidos responde mucho mejor utilizando este método. Ambisonics por lo tanto funciona bastante bien para crear volúmenes sonoros con un grado de cohesión interna bastante elevado, sobre todo si se utilizan configuraciones 3D.
- La microfónica multiarray debido a su alto grado de resolución en la localización y distancia, puede no ser la herramienta inmediata más adecuada para este fin. Sin embargo, se pueden crear superficies texturales interesantes mediante la superposición (en la edición dentro del DAW) de diversas tomas (o grabaciones) de un mismo objeto sonoro u objetos sonoros con un perfil espectromorfológico muy cercano y con características primariamente resonantes.

5.2. Materiales sonoros

El material sonoro de la obra puede ser agrupado en tres grandes conjuntos: objetos concretos, objetos acusmáticos y objetos por síntesis digital.

Debido a la naturaleza incidental-narrativa de la obra, gran parte de los objetos sonoros presentados son del tipo concretos y fueron trabajados de modo tal que la fuente acústica pudiera ser ocultada pero el material físico⁸ permaneciera reconocible; es decir, se desarrollaron objetos sonoros cuyo perfil espectromorfológico fuera reconocible en gran medida, pero de comportamiento inusual o poco familiar. De este modo, existe material que transita entre dos límites; el primero, la literalidad referencial casi total de la fuente acústica, y el segundo, la disociación de la fuente pero con un grado de reconocimiento del material.

El segundo grupo de objetos, el acusmático, está conformado por materiales concretos sometidos a transformaciones digitales complejas con el objetivo de desviar en la medida de lo posible, la analogía relacional. Si bien para estos objetos el material en sí conserva características referenciales, la asociación no es inmediata y permite al espectador no distraerse en tratar de establecer dicha fuente acústica.

Finalmente, el tercer grupo es el constituido por material desarrollado completamente a partir de síntesis digital.

El hecho de que la cantidad de objetos sonoros de los dos primeros grupos es considerable, y tomando en cuenta que surgió como necesidad en el proceso compositivo, el crear escenas sonoras con una complejidad bastante elevada, impulsó la investigación y aplicación de las primeras pruebas de clasificación espectromorfológica mediante aprendizaje supervisado. Estas exploraciones se enfocaron al uso del algoritmo de propagación de afinidad utilizando únicamente como descriptores, los MFCC. Sin embargo, la conclusión de la obra detonó la necesidad de investigar más descriptores y un método alternativo de clasificación (k-medios) derivando en el modelo expuesto en el capítulo IV.

Los materiales anteriores se dividieron en 10 clusters, los cuales sirvieron para establecer puntos de referencia para poder combinar diversos objetos sonoros en escenas complejas y densas de modo tal que cada uno de los objetos fuera claramente perceptible.

Del mismo modo, al enfrentarse con el problema común del enmascaramiento entre objetos sonoros para escenas con gran densidad de elementos, se comenzaron a aplicar de manera meramente pragmática y explorativa, diversos principios acústicos y psicoacústicos que después derivaron en la investigación de los resultados mostrados en la sección 2.3.3.

⁸Es pertinente hacer aquí la diferenciación entre el término *material físico* y *material sonoro*. El primero hace referencia a la analogía relacional en la que el escucha es capaz de asociar el perfil espectromorfológico del objeto sonoro con algún material relacionado a una fuente acústica; por ejemplo, madera, metal, cartón, etc. El material sonoro, término que se usará con frecuencia más adelante, se refiere al conjunto de objetos sonoros dentro del contexto del proceso compositivo.

5.3. Diseño y construcción de la espacialidad.

Al haber sido planteada como una experiencia o narrativa sonora inmersiva, la obra requería el diseño del espacio virtual con un alto grado de naturalidad y de involucramiento pero teniendo un cuidado detallado en las localizaciones fijas y las trayectorias móviles. El objetivo principal al cual se quería llegar, era el de poder crear diversos espacios virtuales asociados a escenas narrativas específicas con un alto grado de inmersividad. En este sentido, la definición de las escenas marcó la forma de la obra, la cual está conformada de tres grandes secciones con subsecciones al interior de cada una:

- Sección I.

1. Escena 1. Anuncio de la catástrofe. Ziusudra está en espera dentro de la construcción encomendada por Enki.
2. Escena 2. Agitación de las aguas. Inicio de la noche oscura del alma.
3. Escena 3. Maremoto, destrucción de la embarcación, naufragio y desmayo (Climax I). Punto de quiebre de la voluntad; pérdida total de la fe.

- Sección II.

1. Escena 1. Despertar en tierra desconocida. Mirada retrospectiva y desoladora hacia el naufragio. El vacío inunda el corazón.
2. Escena 2. Exploración de la nueva tierra. Organismos y geomorfología desconocida.
3. Escena 3. Adaptación al nuevo medio. Aceptación de la tragedia (Climax II).
4. Escena 4. Construcción de herramientas en el nuevo hábitat. Reestructuración del espíritu.

- Sección III.

1. Escena 1. Aparición gradual del Vórtice inverso. Deliberación de los Dioses.
2. Escena 2. El Vórtice comienza a girar con mayor energía alrededor de Ziusudra.
3. Escena 3. El Vórtice inunda el espacio.
4. Escena 4. Ziusudra es suavemente transportado hacia la morada de los dioses a través del Vórtice inverso (Climax III).
5. Escena 5. Coda. Cambio de perspectiva de primera persona a espectador. El espectador se queda dentro del hábitat post catástrofe.

Al establecer la forma y estructura de la obra en relación a escenas incidentales, el diseño espacial de cada sección se pudo plantear y desarrollar de manera directa, buscando en todo momento que la experiencia inmersiva fuera lo más natural posible para el espectador. A continuación se describirán las estrategias de dicho diseño espacial paara cada una de las secciones.

5.3.1. Sección I

Los materiales utilizados en esta sección se van acumulando de acuerdo a la similaridad de su perfil espectromorfológico. Por otro lado, para poder presentar material espectromorfológicamente contrastante se realizó mediante resonancias de colisiones en ubicaciones espaciales específicas. El ejemplo que pudiera resultar más claro de esta metodología es el inicio de la obra en donde el aliento da lugar a la primer corriente de agua.

La construcción del espacio estuvo diseñada de acuerdo al perfil espectromorfológico del material (relación espacio-timbre) y en particular para esta sección, conforme a la densidad espectral de cada objeto (de acuerdo a la previa clasificación en clusters). Se plantearon entonces cuatro capas espacio-tímbricas generales: fluido acuoso, fluido granular, colisiones, granulación percusiva.

La primer capa corresponde al material-fondo del agua que está presente a lo largo de toda la sección junto con variaciones de dicho material para representar corrientes y flujos más violentos. El primer tipo (fondo) corresponde a grabaciones estéreo que fueron espacializadas con ambisonics de manera superpuesta con el propósito de establecer un campo sonoro lo más extensamente distribuido a lo largo del espacio audible. Estas grabaciones fueron previamente procesadas con el fin de establecer distintas centros espectrales para cada una, y de este modo poder plantear una percepción lo más envolvente posible tanto en el plano horizontal como en el de elevación.

El material del segundo caso (las corrientes) fue espacializado mediante *PanW* a lo largo de trayectorias que reiterativamente fluyen de adelante hacia atrás, con distintas velocidades y ondulaciones. La recurrencia en la dirección de dichas corrientes se planeó con el fin de establecer un referente constante perceptual tal que permitiera dar una idea de que el espectador se encontraba inmerso dentro de una escenario que era abatido desde el frente por el fluido acuoso. Para reforzar el efecto de movimiento de dichas corrientes, se agregaron de manera muy sutil, en cada una de dichas trayectorias, objetos sonoros cercanos al fluido granular y de granulación percusiva, para que de acuerdo al principio gestalt de similaridad junto con el efecto del momentum representacional, aumentara la resolución perceptual de las trayectorias.

El material del fluido granular (cuyo material puede ser perceptualmente asociado a madera, y cierto tipo de metal) también fue espacializado mediante *PanW* y funciona como elemento de flujo pero en un rango espectral distinto al del fluido acuoso. Esto permite reforzar la sensación de movimiento externo y de conformación de superficies texturales delimitadas espacialmente pero con un alto grado de cohesión interna.

Estos dos materiales establecen la dimensionalidad virtual del espacio audible de la obra y aunque a través del transcurso de la sección se van acumulando más materiales, debido a las características de los perfiles espectromorfológicos de estos dos primeros y a su extensión como superficies texturales, la percepción de dicha dimensionalidad se mantiene aceptablemente constante, ayudando de este modo, a que la acumulación de los materiales subsecuentes se perciban como elementos que aparecen y desaparecen dentro de esta delimitación virtual.

Las colisiones que aparecen tienen como primer función, la de establecer focos de atención con ubicación espacial fija que permitan tener un contraste perceptual sobre el torrente auditivo de movimiento constante debido a las dos primeras capas. Establecen también inicios y/o finales de distintas trayectorias, reforzando con ello tanto la delimitación de la dimensionalidad como la percepción de la trayectoria en sí. Finalmente, sirven a su vez, como eventos detonadores de resonancias granulares asociadas casi por completo a los objetos sonoros desarrollados por síntesis digital. Casi todos estos objetos sonoros fueron espacializados a través de VBAP.

Finalmente, la capa de granulación percusiva está compuesta casi en su totalidad por objetos de síntesis digital y algunos otros que fueron procesados reiterativamente. Debido al principio de separación espectral en conjunto con el efecto de las bandas direccionales de Blauert, es posible percibir esta capa con una separación aceptable en el plano elevación, ya que existe un contraste espectromorfológico considerable con las capas anteriores. Estos objetos fueron espacializados en su mayoría mediante *PanW*.

La sección tiene por lo tanto, cuatro zonas bien definidas que debido a la diferencia tanto espectromorfológica, como de expansión espacial, localización y comportamiento, permiten crear una sensación de inmersividad con un alto nivel de percepción individual de todos los elementos presentes; es decir con un mínimo efecto de enmascaramiento. Esta separación perceptual de los elementos permite además, crear una acumulación masiva de dinámica previa al climax de modo tal que el espectador se sienta completamente envuelto dentro de un volumen sonoro.

5.3.2. Sección II

Esta sección comienza con objetos sonoros espacializados en VBAP que tienen una ubicación muy puntual. Lo anterior ofrece un contraste abrupto en el diseño espacial al pasar de un volumen masivo a una escena de eventos aislados que si bien ofrecen sensaciones marcadas de distancia espacial, no definen por completo una dimensionalidad virtual.

En esta sección aparecen por primera vez, objetos sonoros específicos espacializados por microfónica multiarray tanto fijos como de trayectoria móvil. Lo anterior en conjunción con objetos sonoros fijos espacializados con VBAP y algunos otros de trayectorias móviles lentas espacializados con *PanW*, permite establecer perímetros perceptuales bien definidos respecto a la distancia, dirección y elevación (incluso aún cuando varios de dichos objetos poseen perfiles espectromorfológicos cercanos); lo anterior por supuesto coadyuvado por la baja densidad de elementos presentes.

Conforme se va desarrollando la sección, la delimitación de la dimensionalidad se construye a través de la reiteración de objetos sonoros fijos en posiciones específicas (objetos sonoros que hacen alusión a insectos o grillos) y trayectorias móviles que transitan entre las posiciones de dichos objetos con el fin de establecer relaciones perceptuales entre lo fijo y lo dinámico, espacialmente hablando.

Después de la presentación intencional de este espacio virtual amorfo y sin dimensionalidad claramente delimitada, se prepara el terreno para volver a construir un espacio virtual con delimitación específica. Esta preparación ocurre mediante el material que empieza a aparecer de manera sutil con trayectoria circundante al espacio audible. Debido al perfil espectromorfológico de este material, además de la percepción de la espacialidad en el plano horizontal, se establece una percepción de elevación debido a la separación de los perfiles correspondientes del material anterior y este es el punto en el que se comienza de nuevo a construir un volumen sonoro.

5.3.3. Sección III

La construcción de la espacialidad en esta sección está basada primordialmente, a través de la superposición de distintos objetos sonoros que viajan de manera circundante al espacio audible. Existe un objeto sonoro principal que debido a su perfil espectromorfológico y por el principio de ... es el más propenso a ser percibido en primera instancia. Derivado de este objeto sonoro, se van añadiendo más objetos sonoros con perfiles espectromorfológicos cercanos que realizan la misma trayectoria pero a velocidades distintas (y a algunos de los cuales se les aplicó además el efecto Doppler).

De este modo, ahora la sensación de envolvimiento es creada principalmente, como producto de la superposición de dichas trayectorias. Una vez que se ha establecido el ritmo del flujo circular, comienzan a aparecer objetos sonoros percusivos que igualmente se incorporan a dicha dinámica, estableciendo con ello una sensación de inmersión que rodea al espectador de forma parecida a una superficie cilíndrica.

Este flujo o vórtice, es contrastado con la reiteración de los objetos sonoros fijos presentados anteriormente, lo que permite establecer una delimitación más clara tanto de dimensionalidad como de la inmersividad.

La pieza concluye con la desaceleración y el desvanecimiento gradual del flujo circular acompañado de los objetos sonoros fijos, permitiendo de este modo, crear una retrospectiva perceptual del espacio virtual construido.

5.4. Comentarios finales

Como se mencionó a lo largo de este capítulo, *Vortex Flow* representa un punto de bifurcación. Por un lado, fue el laboratorio en el cual se aplicaron diversos principios acústico-psicoacústicos acerca de la localización espacial del sonido: sobre todo los relacionados a los indicios espectrales e indicios monoaurales, así como a los relacionados con los procesos de segregación y fusión del ASA. De igual modo, se refinó el uso de los distintos métodos digitales de espacialización antes descritos y se experimentó además por primera vez, con la organización orquestal de material sonoro mediante clasificación de aprendizaje no supervisado y de espacialización con microfonía multiarray. La obra sirvió además para aplicar de manera directa, los conceptos teóricos relacionados a los componentes de la espacialidad tal y como fueron descritos en la sección ... sobre los cuales el autor de este trabajo había venido trabajando con antelación.

Por otro lado, esta pieza funcionó como detonante para explorar más a fondo las técnicas de clasificación mediante aprendizaje supervisado con el fin de tener mejores resultados al momento de agrupar objetos sonoros por su similitud espectromorfológica y por lo tanto tener una herramienta alternativa para la orquestación sonora. Del mismo modo, planteó la necesidad de abordar de manera más profunda, una comprensión acústico-psicoacústica sobre la localización espacial del sonido, tal y como fue planteada en el capítulo II.

Capítulo 6

Conclusiones

Al establecer que el *espacio-timbre* como relación inherente entre la espectromorfología y la localización espacial de los objetos sonoros puede ser considerado como un parámetro por sí mismo, se propuso un modelo teórico original que construyera y definiera el concepto anterior desde tres perspectivas específicas: la perceptual, la acústica y la teórica.

El *espacio-timbre* está entonces constituido de dos elementos: la *espacialidad* (tal y como se planteó en la definición 2.1.2) y el *timbre* (de acuerdo a la definición 3.0.1). Tanto la espacialidad como el timbre 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 manteniendo siempre la visión de que el timbre y el espacio son un sólo parámetro.

Esta construcción dual (perceptual-acústica) quedó claramente establecida en las construcciones de los perfiles y espacios morfológicos y espectrales, los cuales a su vez formaron el perfil y espacio espectromorfológico de acuerdo a lo desarrollado en las secciones 3.1, 3.2 y 3.3.

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 los siguientes resultados obtenidos:

- Realizar una investigación puntual sobre antecedentes y principios generales sobre percepción, psi-

coacústica y acústica, tales que permitan desarrollar y fundamentar la construcción de los conceptos de *espectromorfología* y *espacialidad*.

- Proponer una definición original de *espacialidad* a través de la idea de virtualidad y experiencia estética.
- Establecer el concepto de *disimilitud serial-perceptual*.
- Definir desde una mirada particular, los *componentes primarios de la espacialidad*.
- Construir teóricamente con base en lo anterior, los conceptos de perfil y espacio morfológico, espectral y espectromorfológico.
- Definir la forma de construcción métrica del espacio espectromorfológico.
- Fundamentar el concepto de espacio-timbre en 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 todo lo anterior, se estableció la propuesta teórica de que de manera inexorable el timbre y el espacio son un sólo atributo; el **espacio timbre**.
- Desarrollar una metodología computacional capaz de realizar análisis del perfil espectromorfológico de muestras de audio en series de tiempo y utilizando algoritmos de aprendizaje no supervisado para fines de clasificación y comparación.
- Establecer el concepto de *orquespectración virtual*.

El modelo presentado significa una mirada alternativa al proceso de comprensión y análisis de la música electroacústica en el sentido que plantea emancipar de forma estructurada y sistematizada al espacio como un parámetro primordial e indisoluble del timbre, pero que 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 clave 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. Este planteamiento presupone una postura estética específica que absorbe categóricamente, las construcciones compositivas

dirigidas únicamente por exploraciones espectromorfológicas o conceptuales. Absorber en este sentido significa que el desarrollo de la inmersividad es considerada más 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. Si bien los resultados mostrados en el capítulo I son evidencia suficiente para comprobar la hipótesis de 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, etc.

Sin embargo, este modelo plantea de manera natural una potencial metodología para desarrollar pruebas psicoacústicas enfocadas al análisis de fuentes sonoras asociadas con el ámbito de la música electroacústica. Lo anterior utilizando precisamente la herramienta de clasificación aquí presentada.

6.1. Trabajo a futuro

Derivado de la investigación aquí presentada, existen tres puntos específicos que podrían ser desarrollados de manera inmediata 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 a su perfil espectromorfológico. Lo anterior realizando 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 a su perfil espectromorfológico. Este punto en particular resultaría de gran impacto para el proceso compositivo y sobre todo para aquel asistido por computadora.

Capítulo 7

Apéndice

Este apéndice contiene un desarrollo breve pero formal de la herramienta matemático-computacional utilizada en el presente trabajo de investigación tanto a nivel teórico como práctico. Está dividido en tres secciones: Teoría de Grafos, Teoría de Categorías y Cómputo. En el caso de la parte matemática, las secciones correspondientes contienen en su mayoría las definiciones de los objetos matemáticos 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 redes neuronales y aprendizaje máquina y sobre todo las consideraciones de eficiencia computacional sobre las técnicas utilizadas en este trabajo.

7.1. 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 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 mientras que los segundos, a través de ecuaciones diferenciales y por supuesto tiene que ver con el tipo de parametrización del tiempo. En esta sección se presentarán los conceptos más básicos de los sistemas dinámicos vistos desde su construcción abstracta.

Definición 7.1.1 *Sea M un conjunto cualquiera, $\mathbb{I} \subset \mathbb{R}^+$ para el caso continuo, o $\mathbb{I} \subset \mathbb{Z}^+$ para el caso*

discreto, entonces, un sistema dinámico está definido por la siguiente función:

$$\pi : \mathbb{I} \times \mathbb{M} \longrightarrow \mathbb{M}$$

tal que para cualquier $x \in \mathbb{M}$:

- $\pi(0, x) = x$
- $\pi(t + s, x) = \pi(t, \pi(s, x))$ 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*.

$\pi(t, x) = x_t$ 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:

$$\pi_x(t) : \mathbb{I} \longrightarrow \mathbb{M}$$

Una parte de gran interés para el análisis de los sistemas dinámicos es clasificar el comportamiento de un punto respecto al paso del tiempo. Se tienen tres formas generales respecto a este tópico: 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 a los valores iniciales con los que se evalúa el sistema también conocidos como condiciones iniciales.

Definición 7.1.2 *Un punto $x \in M$ es un **punto periódico** de periodo n si se cumple que $\pi(n, x) = x$ y $\pi(j, x) \neq x$ para todo $0 < j < n$. 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 7.1.3 Dado un sistema dinámico, la *órbita hacia adelante* o simplemente *órbita*, es el conjunto:

$$\mathcal{O}^+ = \{\pi(t, x) = x_t \in M | t \geq 0\}$$

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.

7.2. Algoritmos computacionales.

En el presente trabajo se utilizaron algoritmos computacionales específicos para desarrollar la implementación computacional mostrada en el capítulo IV. En esta sección se describen brevemente dichos algoritmos con el fin de contextualizar su aplicación y referenciar las fuentes correspondientes para el lector interesado.

7.2.1. Dynamic Time Warping

El DTW es un algoritmo bastante popular que se usa en diversas áreas como un método de medida de similitud 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 a Senin (2009), está dado por:

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 7-1: Pseudocódigo para el algoritmo de DTW (Senin, 2009)

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 similitud a partir de ello.

Existen diversas librerías en Python que calculan este algoritmo, entre ellas destacan: *PyPi dtw* (Rouanet, 2014), *dtaidistance* (Meert, 2018), *dtw-python* (Giorgino 2009) y la propia *tslearn* ya mencionada. Para esta última, la figura 7-2 muestra la documentación correspondiente.

tslearn.metrics.dtw

`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 7-2: Documentación para el algoritmo DTW en la librería tslearn.

7.2.2. 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-Medios con la particularidad que es capaz de identificar clusters separables de manera no lineal.* (Tzortzis, 2008).

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 $(K, k, C_1^*, \dots, C_{k-1}^*, C_k = \{\mathbf{x}_n\})$ 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 \mathbf{x}_n and every cluster C_i compute $\|\phi(\mathbf{x}_n) - \mathbf{m}_i\|^2$ using (3)
 2. Find $c^*(\mathbf{x}_n) = \operatorname{argmin}_i (\|\phi(\mathbf{x}_n) - \mathbf{m}_i\|^2)$
 3. Update clusters as $C_i = \{\mathbf{x}_n | c^*(\mathbf{x}_n) = i\}$
 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 7-3: Pseudocódigo para el algoritmo de Kernel Global de K Medios (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 método *tslearn.clustering.GlobalAlignmentKernelKMeans* el cual se muestra en la figura 7-4.

tslearn.clustering.GlobalAlignmentKernelKMeans

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-learns' [Glossary](#) for more details.
- verbose** : *bool (default: False)*
Whether or not to print information about the inertia while learning the model.

Figura 7-4: Documentación para el algoritmo GAK de tslearn.

7.2.3. 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 distorsiones, puede ser usada para eficientizar el procesamiento de la transmisión de series de tiempo* (Ratanamahatana & 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. Sin embargo, las bandas LB Keogh son particularmente eficientes cuando se comparan series de tiempo a través de DTW.

En la figura 7-5 se muestra la documentación de dicho algoritmo tal y como está descrito en la librería *tslearn*.

tslearn.metrics.lb_envelope

`tslearn.metrics.lb_envelope` (*ts*, *radius=1*)

[\[source\]](#)

Compute time-series envelope as required by LB_Keogh.

LB_Keogh was originally presented in [1].

Parameters: **ts** : *array-like*

Time-series for which the envelope should be computed.

radius : *int* (default: 1)

Radius to be used for the envelope generation (the envelope at time index *i* will be generated based on all observations from the time series at indices comprised between *i*-radius and *i*+radius).

Returns: **array-like**

Lower-side of the envelope.

array-like

Upper-side of the envelope.

Figura 7-5: Documentación para el algoritmo GAK de tslearn.

7.2.4. Métricas de similaridad 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*. Esta métrica es un tipo particular de métrica 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. & 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.

$$\text{lev}_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1, j) + 1 \\ \text{lev}_{a,b}(i, j-1) + 1 \\ \text{lev}_{a,b}(i-1, j-1) + \mathbf{1}_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

Figura 7-6: Métrica Levinshtein

7.3. 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 en el capítulo III, 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 dichas 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 7-6 muestra el algoritmo correspondiente para calcular la *ApEn*.

Step 1: Form a time series of data $u(1), u(2), \dots, u(N)$. These are N raw data values from measurement equally spaced in time.

Step 2: Fix m , an **integer**, and r , a **positive real number**. The value of m represents the length of compared run of data, and r specifies a filtering level.

Step 3: Form a sequence of vectors $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N - m + 1)$, in \mathbf{R}^m , real m -dimensional space defined by $\mathbf{x}(i) = [u(i), u(i + 1), \dots, u(i + m - 1)]$.

Step 4: Use the sequence $\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N - m + 1)$ to construct, for each $i, 1 \leq i \leq N - m + 1$

$$C_i^m(r) = (\text{number of } x(j) \text{ such that } d[x(i), x(j)] \leq r) / (N - m + 1)$$

in which $d[x, x^*]$ is defined as

$$d[x, x^*] = \max_a |u(a) - u^*(a)|$$

The $u(a)$ are the m **scalar** components of \mathbf{x} . d represents the distance between the **vectors** $\mathbf{x}(i)$ and $\mathbf{x}(j)$, given by the maximum difference in their respective scalar components. Note that j takes on all values, so the match provided when $i = j$ will be counted (the subsequence is matched against itself).

Step 5: Define

$$\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log(C_i^m(r)),$$

Step 6: Define approximate entropy (ApEn) as

$$\text{ApEn} = \Phi^m(r) - \Phi^{m+1}(r).$$

where \log is the natural logarithm, for m and r fixed as in Step 2.

Parameter selection: typically choose $m = 2$ or $m = 3$, and r depends greatly on the application.

Figura 7-7: Algoritmo para calcular la entropía aproximada. Fuente (Wikipedia)

7.4. Clasificación no supervisada utilizando medidas de tendencia central.

Como contraste a la metodología propuesta en el capítulo IV se muestran aquí los resultados de clasificación no supervisada utilizando descriptores numéricos fijos.

Se recolectaron 650 muestras de audio de diverso contenido espectralmorfológico que forman parte de la biblioteca personal del autor de este trabajo para fines composicionales y que incluyen básicamente: grabaciones de objetos físicos en tipo *sequence jeu*, grabaciones procesadas digitalmente y sonidos creados a partir de síntesis digital.

Se calculó a continuación, la matriz de coeficientes de correlación de Pearson a pares, para todos los descriptores mencionados en la sección anterior y de acuerdo a dicho procedimiento se obtuvieron los pares que mostraron un índice menor a 0.5.

```
soundsArray, names_array = load_sounds('C:/Users/Edmar/Documents/AllSounds')
features_array = ['mfcc', 'spectral_contrast', 'zero_crossing_rate',
'zero_crossings', 'energy', 'spectral_flatness',
'spectral_centroid', 'rmse']
spktrlArray = np.array([spectralMorph_profile(i, features_array, 'median_raw') for i in soundsArray])
correlation_matrix = correlation(spktrlArray, ['mfcc', 'spectral_contrast', 'zero_crossing_rate',
'zero_crossings', 'energy', 'spectral_flatness',
'spectral_centroid', 'rmse'])
c_threshold = features_correlation(correlation_matrix, 0.5)

/////post/////

[['mfcc4', 'mfcc5'), 0.34344876638350824],
[['mfcc4', 'mfcc6'), 0.38514001739784121],
[['mfcc4', 'mfcc7'), 0.3643904682132692],
[['mfcc4', 'mfcc9'), 0.38375158459884112],
[['mfcc4', 'mfcc10'), 0.35652978158235404],
[['mfcc4', 'mfcc11'), 0.42461149380431729],
[['mfcc10', 'zero_crossings'), 0.4832090249489594],
[['mfcc11', 'mfcc13'), 0.43621928332628801],
[['mfcc11', 'mfcc14'), 0.46586590354446611],
[['mfcc14', 'spktrl_contrast3'), 0.33072056200650612],
[['mfcc14', 'spktrl_contrast4'), 0.40284895967041467],
[['mfcc14', 'spktrl_contrast6'), 0.33569486675834442],
[['mfcc14', 'zero_crossing_rate'), 0.30273664027855479],
[['mfcc17', 'zero_crossings'), 0.12065104007861815],
[['mfcc18', 'spktrl_contrast6'), 0.1347801611458079],
[['spktrl_contrast1', 'spktrl_contrast5'), 0.28083470821044093],
[['spktrl_contrast1', 'energy'), 0.35270716284983977],
[['spktrl_contrast1', 'spectral_centroid'), 0.46115366886694287]]
```

Tomando en cuenta que se tenía además, la herramienta de reducción dimensional mediante PCA, después de realizar un total de 50 simulaciones, se plantearon 4 simulaciones óptimas específicas usando el algoritmo de k-medios.

Por otro lado, después de realizar distintas simulaciones con el algoritmo de propagación de afinidad se observó que existía una relación entre el número de descriptores-número de muestras y el número de valores atípicos. Esta relación parece ser no lineal como se puede observar en las figuras correspondientes y si bien las razones matemático-computacionales que expliquen este efecto están fuera del alcance de este trabajo, de dicha visualización se pudieron seleccionar 4 conjuntos de parámetros específicos para formular 4 simulaciones utilizando dicho algoritmo.

Consecuentemente, el presente capítulo ofrece los resultados de estas 8 simulaciones las cuales representan los modelos más representativos para los fines de clasificación espectromorfológica de esta tesis, de acuerdo a cada conjunto de descriptores y muestras de audio utilizados.

A continuación se muestran todas las simulaciones que se realizaron utilizando el algoritmo k-medios.

7.4.1. Simulación Base

Esta simulación representa un caso base en el cual se tomaron en cuenta todos los descriptores disponibles y se obtuvo una aproximación del número óptimo de componentes a usar en el algoritmo de reducción dimensional PCA mediante la proporción de la varianza explicada de todos los descriptores. El objetivo de realizar esta simulación fue el de tener un punto de comparación explícito a partir del cual analizar los subsiguientes casos.

De la figura 7-8 se observa que el intervalo de relación *número de componentes vs porcentaje de la varianza total* es: $[(10, 0.9), (20, 0.96)]$. De este resultado, se decidió utilizar el valor medio de dicho intervalo por lo que el número de componentes a utilizar fue de 15.

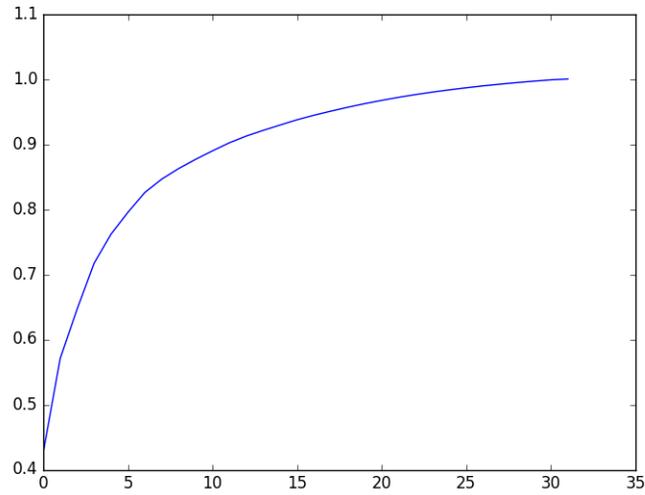


Figura 7-8: Número de componentes óptimo para PCA vs varianza explicada

Para decidir el número óptimo de clusters para el algoritmo k-medios, se utilizó el método del codo, el cual arrojó el resultado mostrado en la figura 7-9. Como puede observarse, para este conjunto de descriptores, el número óptimo no es completamente contundente y parece encontrarse dentro del intervalo de 8 y 12, por lo que se decidió utilizar 10 como dicho valor.

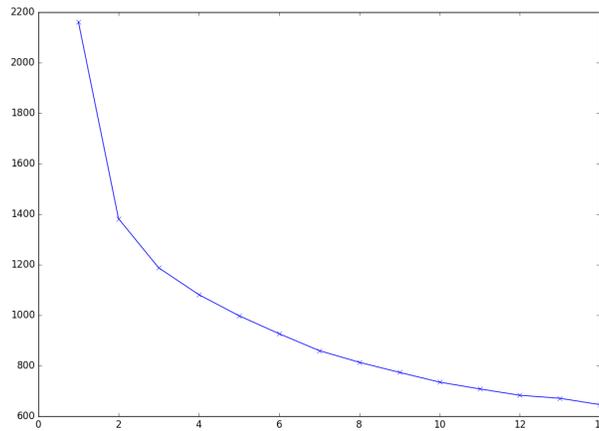


Figura 7-9: Método del codo, k-medios.

A continuación, se ejecutó la simulación con esos valores y se obtuvieron los clusters mostrados en las figuras 7-10, 7-11 y 7-12.

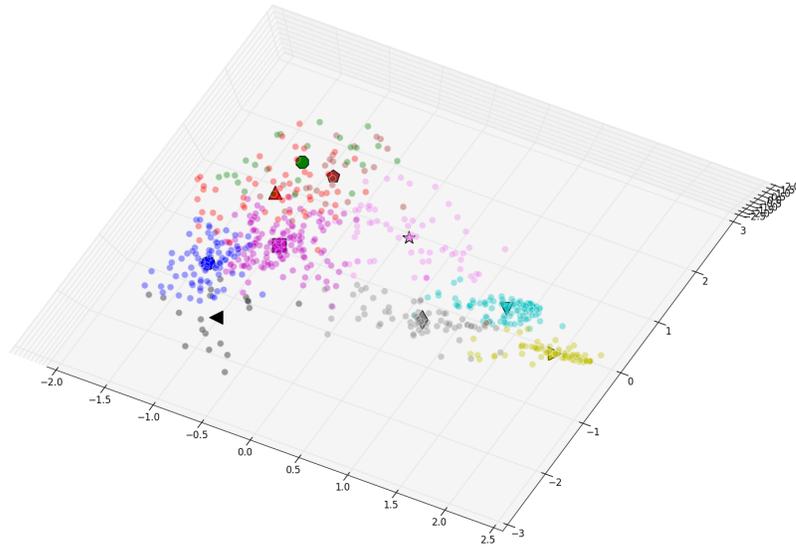


Figura 7-10: Clustering de 10 etiquetas.

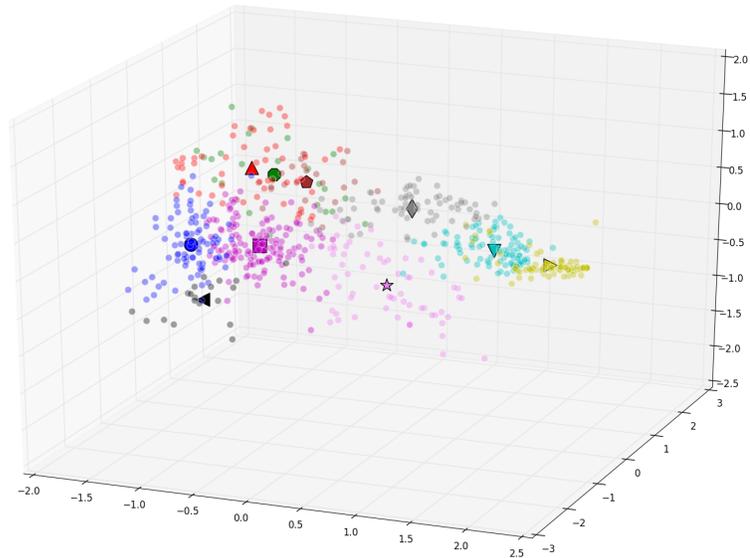


Figura 7-11: Clustering de 10 etiquetas

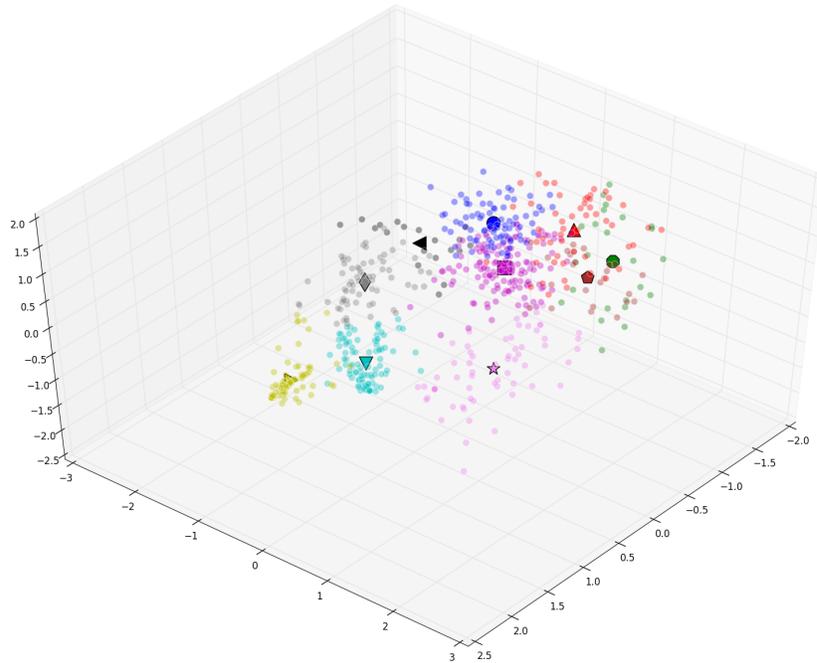


Figura 7-12: Clustering de 10 etiquetas

Los resultados del proceso anterior se muestran a continuación:

```
pca_k_selection(soundsArray,features_names)
data_frame, features, labels, centroids = cluster_model(soundsArray, features_names,['k-means',10,15])
/////////
      pc13    pc14  cluster_id    sample_name
0 -0.090383 -0.006176         8      aaaH1.aif
1 -0.039161 -0.036058         8      aahs2.aif
2 -0.114289  0.286006         6      ADNA1.aif
3 -0.084658 -0.030617         8      ADNA2.aif
4 -0.115120  0.077077         9      ADNA3.aif
5  0.002299  0.044317         1      ADN1.aif
6 -0.124968  0.067106         8      ADN2.aif
7 -0.046423  0.037094         1      ADN3.aif
8 -0.009517  0.040908         8  aerealBassLoop.aif
9 -0.009217 -0.043438         8      aerealLoop.aif
10 -0.032432  0.159961         3      afterProl.aif
11 -0.032209  0.140978         3      afterProlF.aif
12 -0.094068  0.000298         3      aguaTri.wav
13 -0.021744 -0.072459         6      aguaTri2.wav
14  0.002944 -0.026882         4      air1.aif
15  0.073646  0.000260         4      air2.aif
16  0.036901  0.015191         4      air3.aif
17  0.010844 -0.044043         4      air4.aif
18 -0.136459 -0.014188         8      air5.aif
```

19	-0.024539	0.005196	4	air6.aif
20	-0.059814	-0.018022	4	air7.aif
21	0.012585	-0.001426	4	airRes.aif
22	0.004669	-0.146894	8	arrastreA.aif
23	0.149233	-0.150850	1	arrastreB.aif
24	0.099409	-0.131210	8	arrastreC.aif
25	0.112516	-0.183351	1	arrastreD.aif
26	0.075644	-0.140344	8	arrastreE.aif
27	-0.006611	0.011940	1	arrastreF.aif
28	-0.014664	-0.061780	4	arrastreFill.aif
29	-0.027686	0.021321	1	arrastreG.aif

Si bien se tiene una buena separación para los primeros clusters, en esta simulación se observa que los últimos agrupamientos presentan un nivel considerable de dispersión lo cual se debe en gran medida a la presencia del alto grado de correlación entre algunos de ellos. Esto es un indicador de que un número elevado de descriptores no está asociado del todo a la precisión de la clasificación del algoritmo sobre todo si existe un nivel considerable de correlación. También hay que tomar en cuenta que los MFCC si bien son herramientas bastante útiles para aplicaciones de voz, su uso en fuentes sonoras de otra naturaleza parece no estar del todo optimizada.

7.4.2. Simulación 2

Básandose en la matriz de correlación obtenida al inicio y a partir del hecho de que existe una fuerte correlación entre los MFCC y los restantes descriptores espectrales, se seleccionó un conjunto de descriptores que incluyera un representante espectral y dos representantes de envolventes tal y como se muestra a continuación.

```
features_name = ['mfcc', 'energy', 'spectral_centroid']
spectralMorph_profile(i, features_array, 'median_raw')
data_frame, features, labels, centroids = cluster_model(soundsArray, features_name, ['k-means', 11])
```

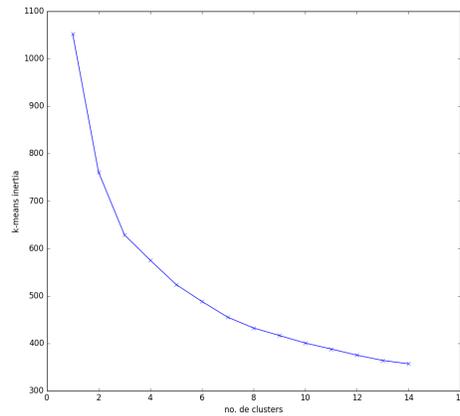


Figura 7-13: Número de clusters óptimos para k-medios

Usando como referencia lo obtenido por el método del codo, se procedió a ejecutar el algoritmo de k-medios sin reducción dimensional con un número de 8 clusters y se obtuvieron los resultados mostrados en las figuras 7-13 a 7-15. Vale la pena mencionar que en esta simulación, parece ser que los MFCC contribuyen a un incremento considerable en la dispersión del agrupamiento final cuando es usado en conjunto con otros descriptores espectrales o de envolvente dentro del contexto del algoritmo k-medios.

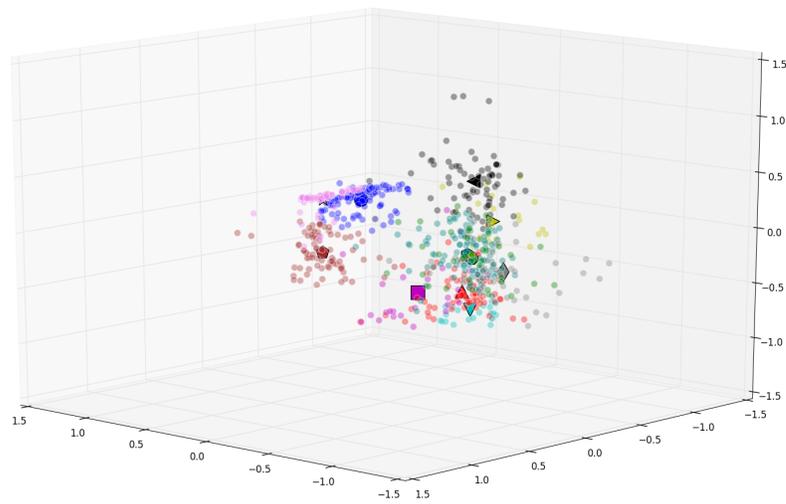


Figura 7-14: Clustering para 8 etiquetas.

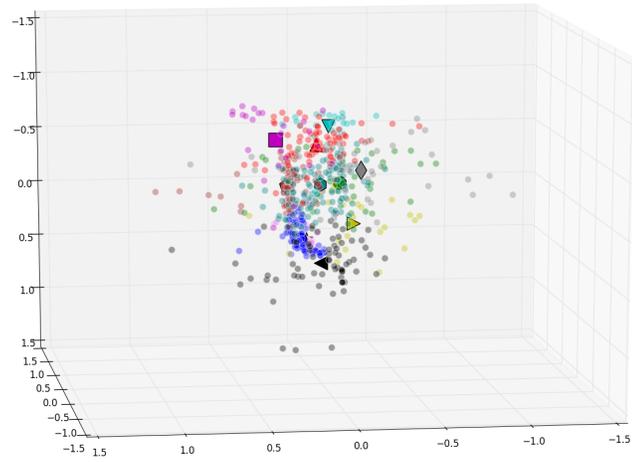


Figura 7-15: Clustering para 8 etiquetas.

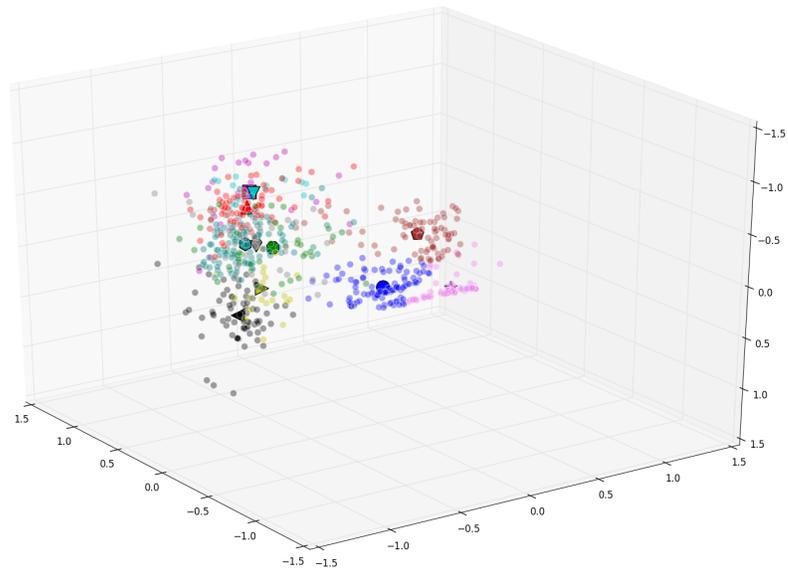


Figura 7-16: Clustering para 8 etiquetas.

7.4.3. simulación 3

Para la siguiente simulación, se eleigieron los descriptores de: rmse, coeficiente de tonalidad, centroide espectral y tasa de cruce cero. Para este conjunto de descriptores, el método del codo arrojó un valor óptimo de clusters entre 6 y 8 como se muestra en la figura 7-16, por lo que se decidió utilizar el primer número para realizar la simulación

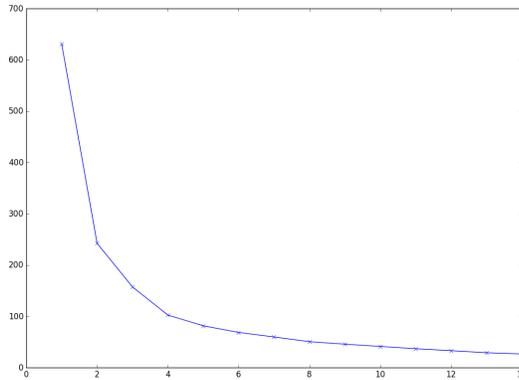


Figura 7-17: Número de clusters óptimos para k-medios.

Esta simulación arrojó un grupo de clusters bastante bien diferenciados y separados tal y como puede observarse en las figuras 7-17 a 7-19.

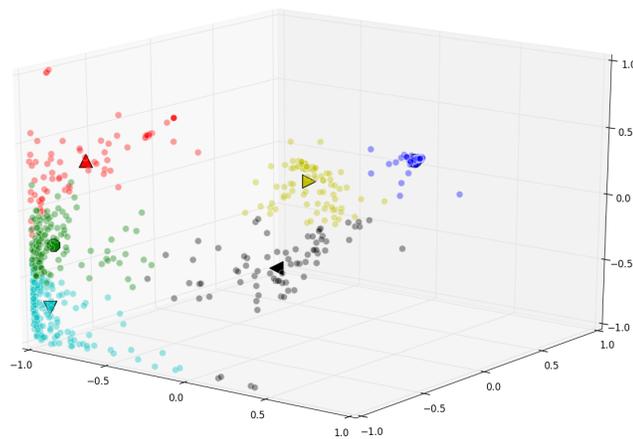


Figura 7-18: Clustering para 6 etiquetas.

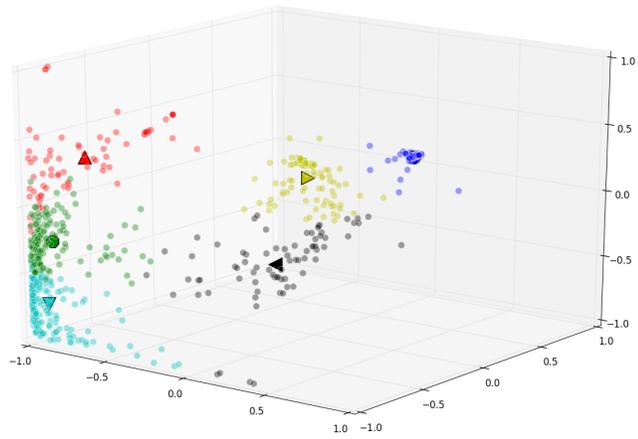


Figura 7-19: Clustering para 6 etiquetas.

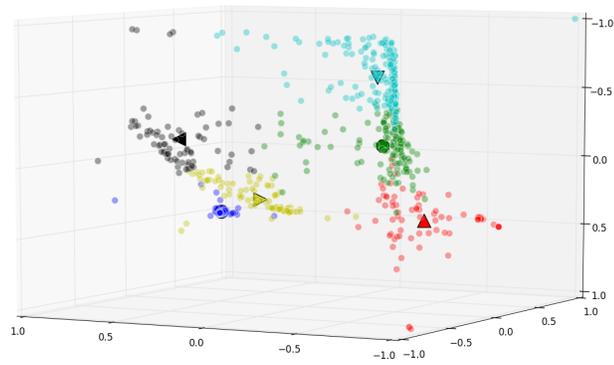


Figura 7-20: Clustering para 6 etiquetas.

De la matriz de correlación mostrada en la figura 7-20 puede observarse que únicamente la tasa de cruce cero y el centroide espectral muestran un grado de correlación significativo.

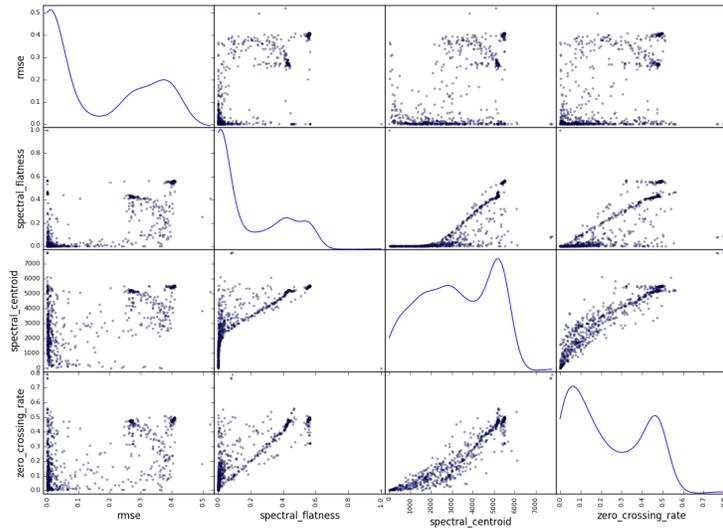


Figura 7-21: Matriz de correlación para los descriptores utilizados.

simulación 4

En esta simulación se utilizaron como descriptores los siguientes: rmse, coeficiente de tonalidad, centroide espectral y mfcc. Tras realizar el cálculo del número de componentes óptimos PCA, se obtuvo el resultado mostrado en la figura 7-21 mediante la cual se eligió que dicho número fuera 14, además, aplicando el método del codo mostrado en la figura 7-22, se utilizó 10, como número óptimo de clusters.

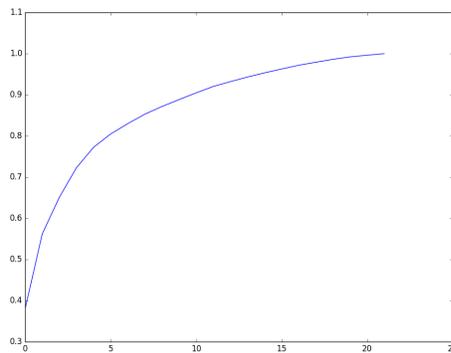


Figura 7-22: Número de componentes óptimos para PCA.

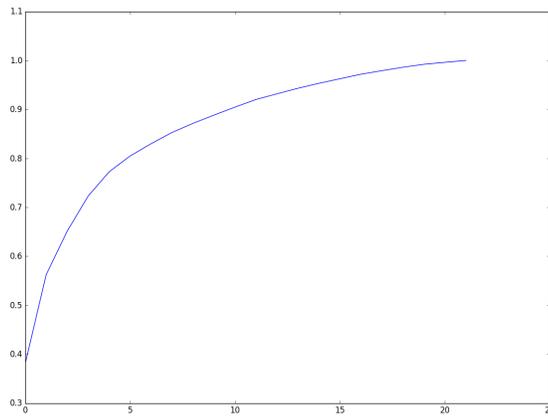


Figura 7-23: Número de clusters óptimos para k-medios.

Para este conjunto de descriptores y con los parámetros antes mencionados se obtuvieron los resultados obtenidos en las figuras 7-23 a 7-25 de donde puede observarse que si bien existe una clasificación bastante óptima entre los elementos de cada cluster, de igual modo existen algunas pocas muestras asignadas a clusters que se encuentran a una distancia considerable del mismo, lo que podría probablemente implicar una asignación errónea para estas muestras específicas.

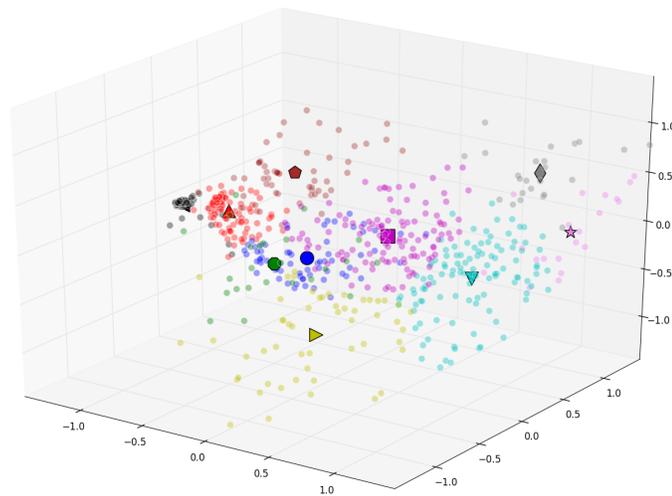


Figura 7-24: Clustering para 10 etiquetas.

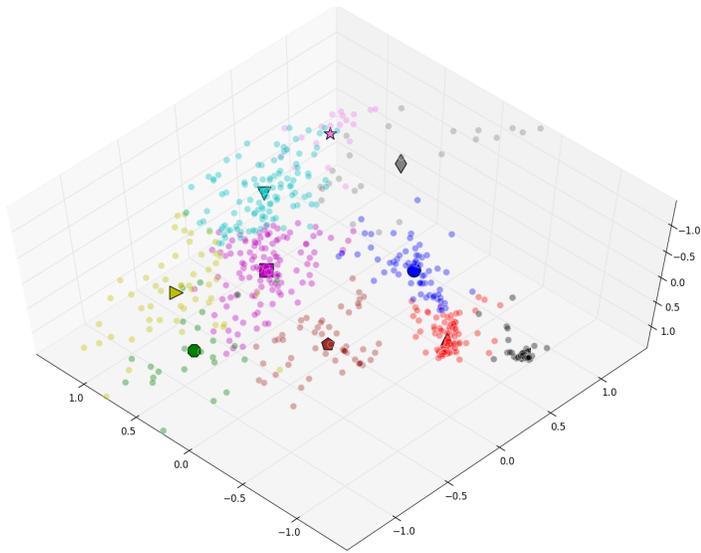


Figura 7-25: Clustering para 10 etiquetas.

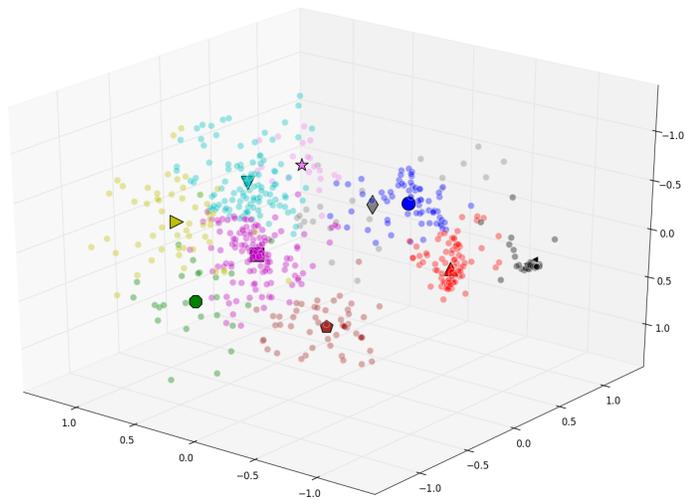


Figura 7-26: Clustering para 10 etiquetas.

7.4.4. simulación 5

Para esta simulación los descriptores utilizados fueron: rmse, coeficiente de tonalidad, contraste espectral, mfcc y energía, con 14 componentes de PCA y 10 clusters. Las figuras 7-26 y 7-27 muestran el cálculo de los componentes óptimos de PCA y el número óptimo de clusters para este caso respectivamente.

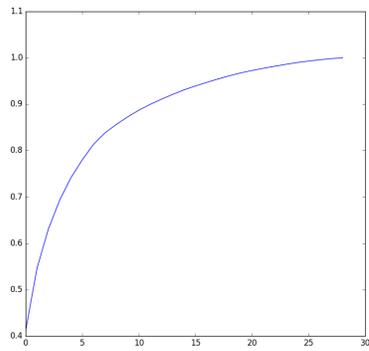


Figura 7-27: Número de componentes óptimos para PCA.

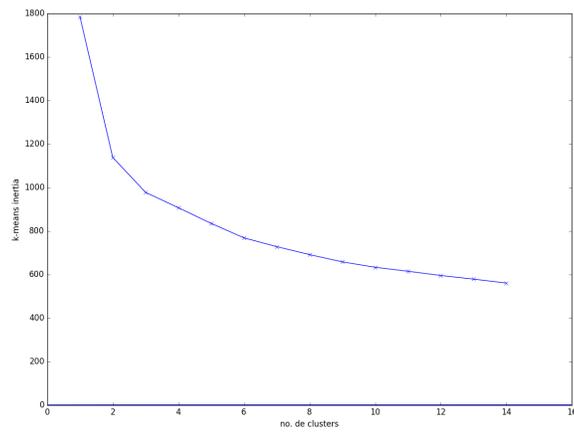


Figura 7-28: Número de clusters óptimos para k-medios

Los resultados de la clasificación son mostrados en la figura 7-28 la cual muestra un comportamiento bastante similar al de la simulación anterior respecto a la dispersión de ciertas muestras en relación a sus centroides asignados.

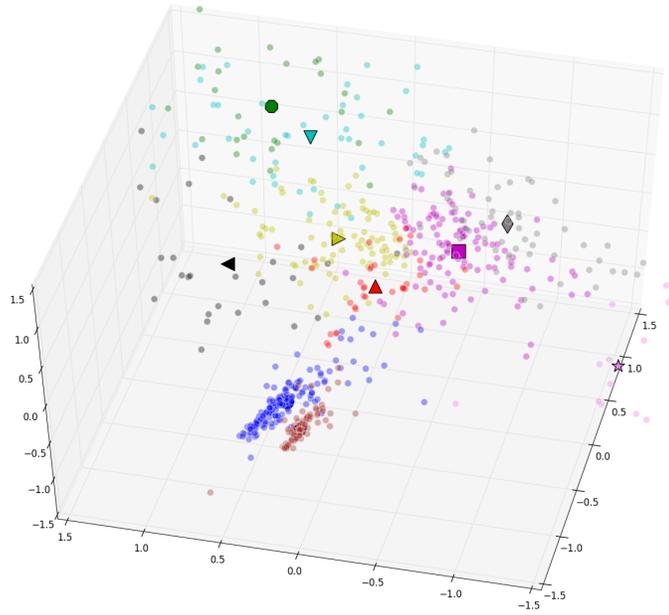


Figura 7-29: Clustering para 10 etiquetas.

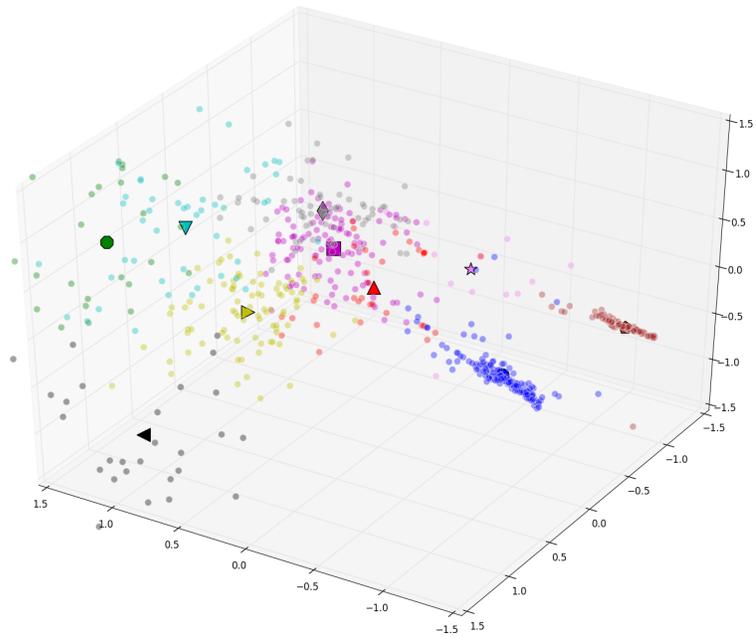


Figura 7-30: Clustering para 10 etiquetas.

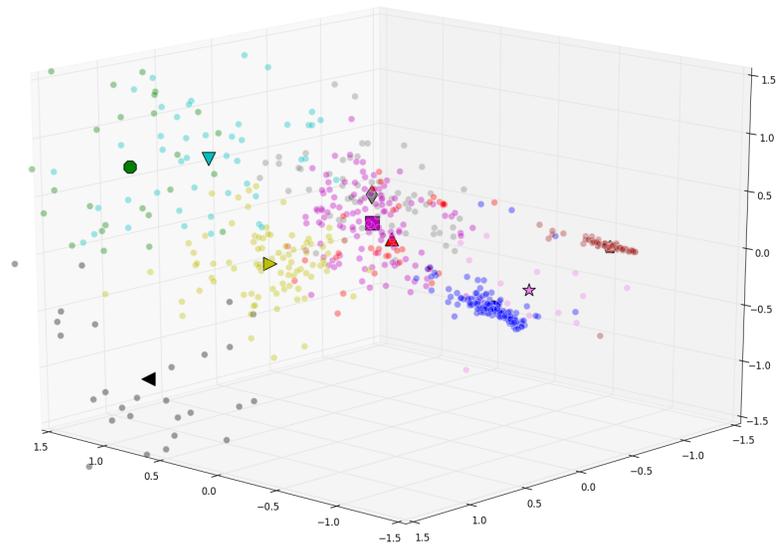


Figura 7-31: Clustering para 10 etiquetas.

7.4.5. simulación 6

En esta simulación se eligieron como descriptores: el rmse, el conjunto de contraste espectral y la energía. Usando 5 componentes de PCA y 11 clusters se obtuvieron los resultados mostrados en las figuras 7-29 a 7-31:

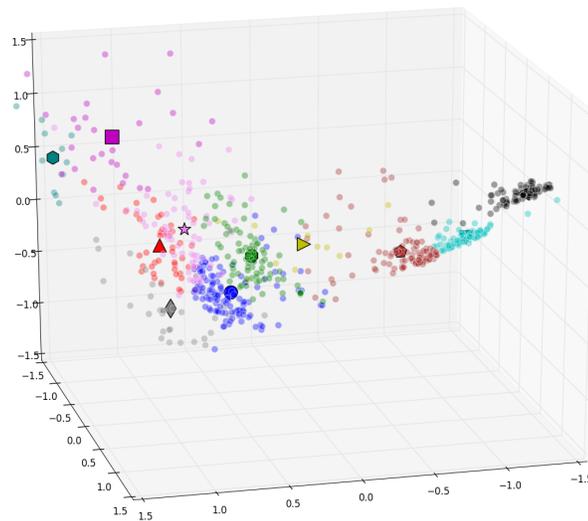


Figura 7-32: Clustering para 11 etiquetas.

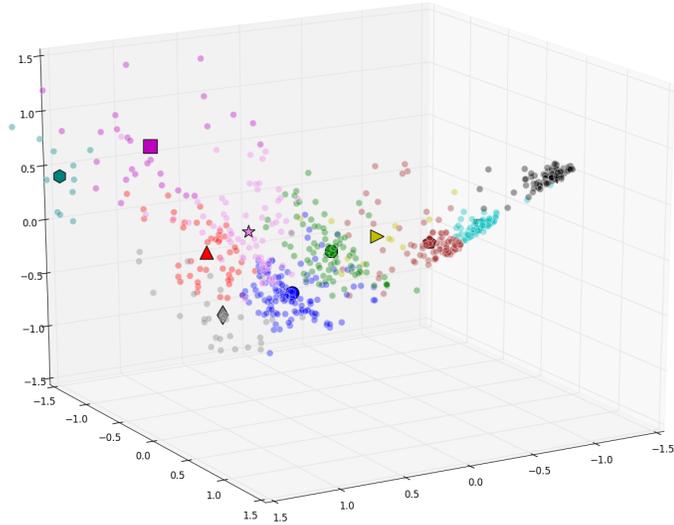


Figura 7-33: Clustering para 11 etiquetas.

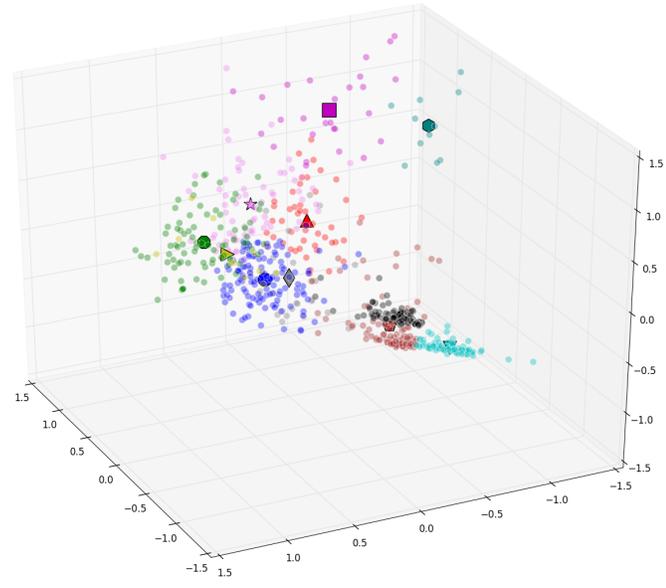


Figura 7-34: Clustering para 11 etiquetas.

Esta simulación presenta una mejora considerable en el agrupamiento de las muestras con respecto a sus centroides, sin embargo, continúan apareciendo algunas muestras con un grado considerable de

dispersión.

7.4.6. Simulaciones con el algoritmo propagación de afinidad

Este algoritmo presentó un comportamiento no estable con las muestras de audio utilizadas. Se detectaron variaciones importantes en el número de etiquetas obtenidas de acuerdo no sólo a los descriptores utilizados sino principalmente en función del número de muestras analizadas. Para obtener un análisis más formal de la observación anterior se escribió la rutina descrita al inicio de la sección mediante la cual se ejecutaba dicho algoritmo de manera agrupada para distintos descriptores y distintos números de muestras, donde además estas últimas eran seleccionadas al azar a partir del conjunto original de ellas.

Para todos los distintos conjuntos de descriptores se utilizaron como números representativos de muestras: 100, 150, 300, 450, 500 y 600. Dado que se analizaron 7 conjuntos distintos de descriptores, se ejecutaron en total 42 simulaciones para realizar las comparaciones y los análisis correspondientes, todos ellos con parámetros de iteraciones y convergencia de 400 y 20 respectivamente.

En la figura 7-32 se muestran los resultados obtenidos de dichas ejecuciones para el número de muestras respecto al número de etiquetas obtenidas y de acuerdo al conjunto de descriptores siguientes:

1. Coeficiente de Tonalidad y rmse.
2. Energía, centroide espectral, coeficiente de tonalidad y rmse.
3. Coeficiente de tonalidad, energía, centroide espectral, tasa de cruce cero y rmse.
4. MFCC y rmse.
5. Coeficientes de contraste espectral y rmse.
6. Coeficientes de contraste espectral, tasa de cruce cero y rmse.
7. Coeficientes de contraste espectral, MFCC, tasa de cruce cero y rmse.

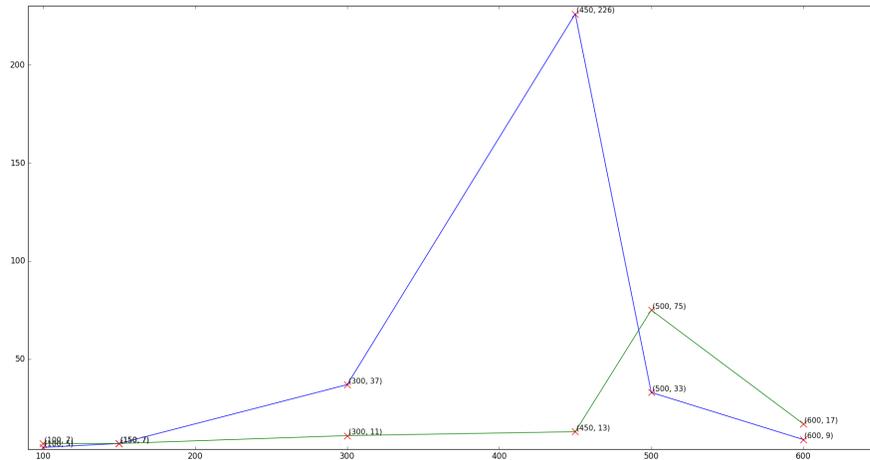


Figura 7-35: Comparación de ejecuciones del algoritmo: número de muestras vs clusters obtenidos.

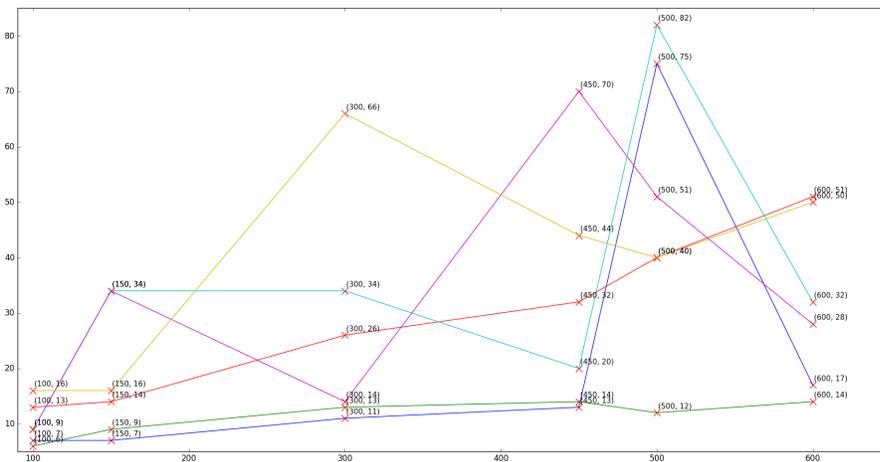


Figura 7-36: Comparación de ejecuciones del algoritmo: número de muestras vs clusters obtenidos

De los resultado anteriores se observa claramente que el conjunto de descriptores que responde de manera más estable al algoritmo en cuestión de acuerdo al número de muestras es el numeral 3. Una posible interpretación de lo anterior es que para ese conjunto de descriptores en particular y para las muestras seleccionadas, el algoritmo en cuestión no está altamente correlacionado con el número de muestras que le son asignadas para la clasificación.

En base a lo anterior se decidió ejecutar el modelo con el conjunto de descriptores 3 y para el número

total de muestras, obteniendo los resultados mostrados en las figuras siguientes:

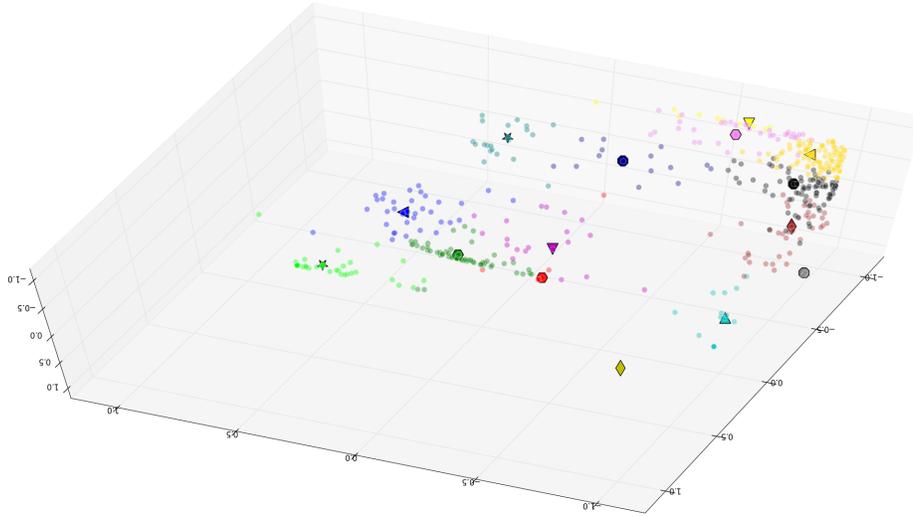


Figura 7-37: Clustering para 15 etiquetas.

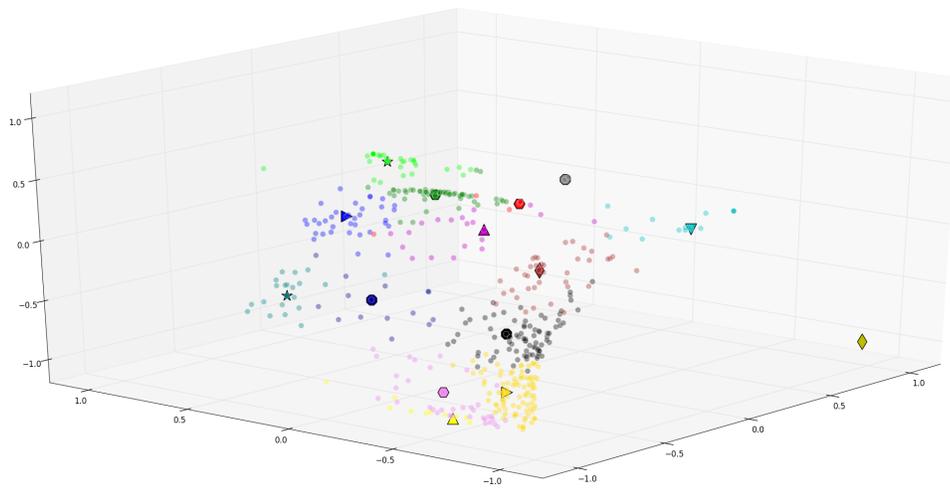


Figura 7-38: Clustering para 15 etiquetas.

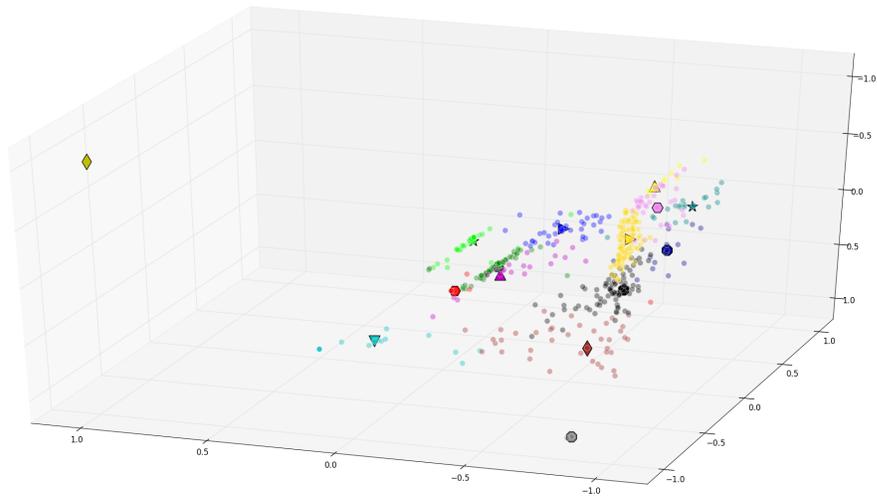


Figura 7-39: Clustering para 15 etiquetas.

7.4.7. Interpretación de los resultados

Para el algoritmo k-medios los resultados más precisos respecto a la clasificación fueron los obtenidos en la simulación 3, la cual utiliza 2 descriptores espectrales y 2 descriptores de morfología y con la que se obtuvieron 6 etiquetas o clusters. Con base en lo anterior, se ejecutó la misma simulación pero variando el número de etiquetas a 8,12,15 y 20 respectivamente, con lo que se obtuvieron los resultados mostrados en las figuras siguientes. De dichas figuras puede observarse que si bien 6, sigue siendo el número de etiquetas óptimo para la clasificación usando dicho conjunto de descriptores, para los otros casos se presenta un nivel tolerable de dispersión de ciertas muestras.

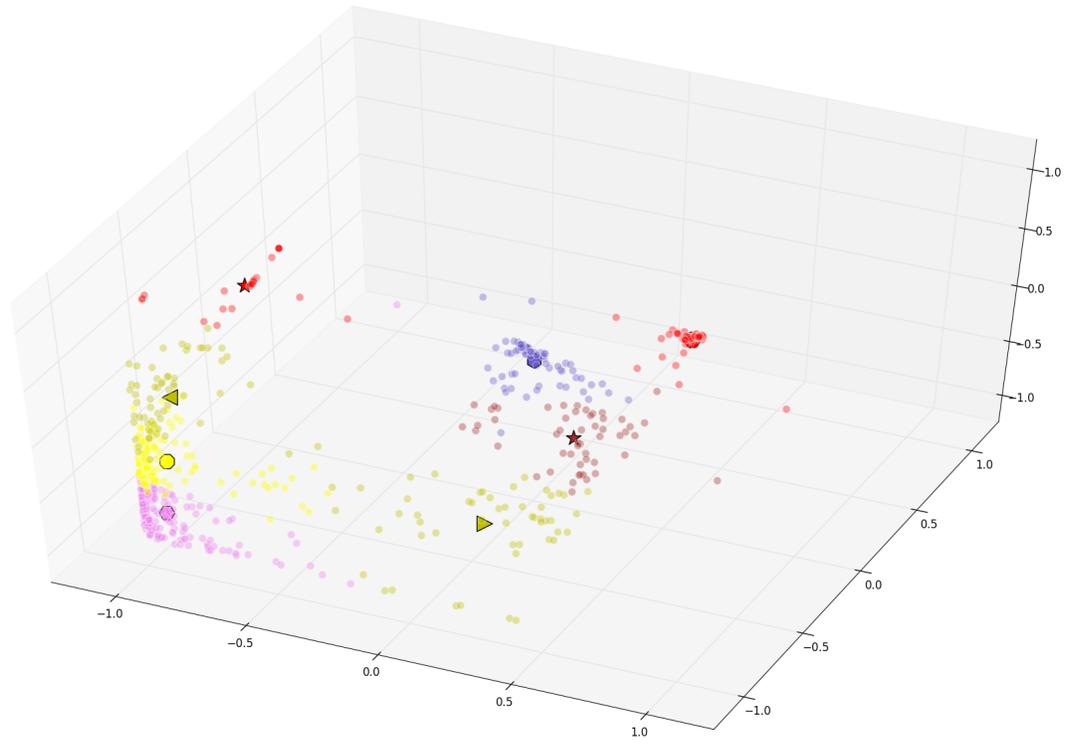


Figura 7-40: Clustering para 8 etiquetas.

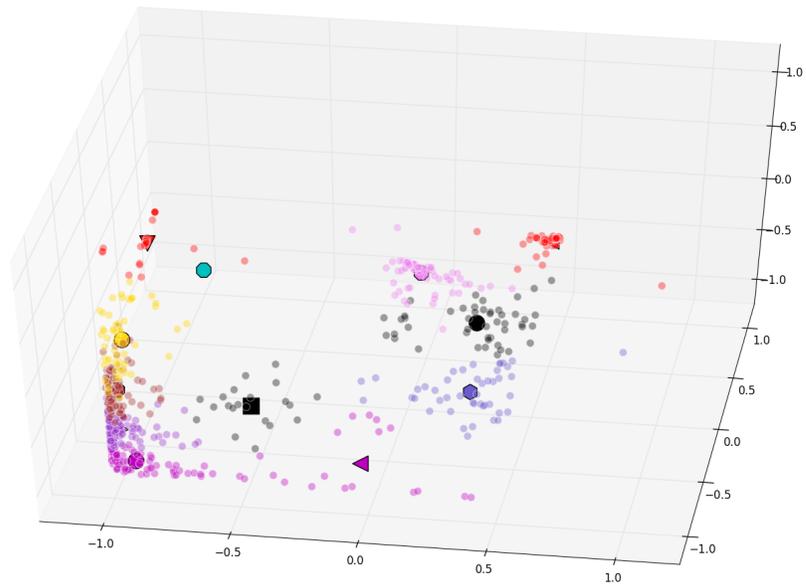


Figura 7-41: Clustering para 12 etiquetas.

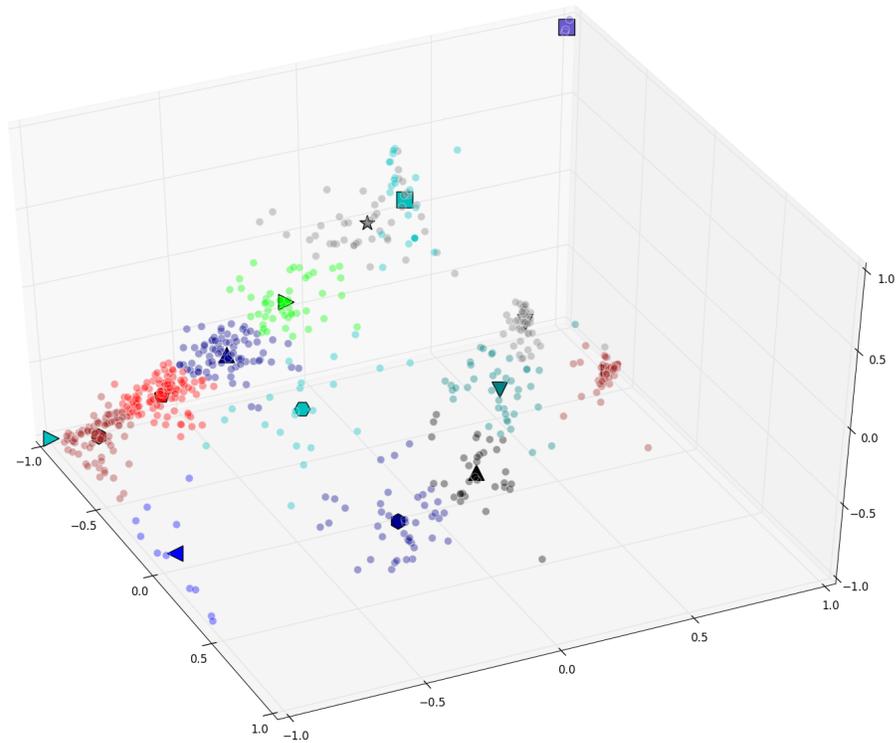


Figura 7-42: Clustering para 15 etiquetas.

En las simulaciones que incluían los MFCC se observó un grado considerable de dispersión de varias muestras de audio respecto de sus centroides asignados; este grado de dispersión de hecho parece ser mayor que cuando se utilizaron los coeficientes de contraste espectral en sustitución de los mfcc. Las razones de los resultados anteriores pueden ser variadas, sin embargo debe notarse que en todos esos casos se realizó reducción dimensional con PCA, con lo que en cierto grado se pudo asegurar que los componentes utilizados para cada simulación eran los más significativos.

De lo anterior se puede concluir que para las muestras de audio utilizadas y utilizando el algoritmo k-medios la mayor precisión se obtuvo utilizando un número reducido de descriptores (4 en este caso) y que de esos descriptores la mitad eran espectrales y la otra mitad morfológicos. De este modo, podría afirmarse que si el usuario desea precisión al momento de la clasificación espectromorfológica con un número bajo de etiquetas, puede usar los parámetros mencionados. Por supuesto, un número bajo de etiquetas (sobre todo para una cantidad tan grande de muestras) podría implicar una sobreajuste en el agrupamiento, considerando como similares, muestras que quizás no lo parezcan tanto al momento de la composición.

Para el caso del algoritmo propagación de afinidad se observó que el conjunto de descriptores con el cual dicho algoritmo responde de manera más estable fue el compuesto por el coeficiente de tonalidad, la energía, el centroide espectral, la tasa de cruce cero y el rmse, obteniendo para este caso, un número total de 15 etiquetas que podría considerarse bastante funcional para fines compositivos de manera muy general.

Todo lo anterior puede resumirse en los siguientes puntos:

1. Para el algoritmo k-medios el conjunto óptimo de parámetros fue el descrito al inicio de la sección.
2. Para el algoritmo k-medios, la introducción de los MFCC implica una elevación considerable en el grado de dispersión de varias muestras de audio respecto de sus centroides asignados, incluso aplicando reducción dimensional mediante PCA. Una opción alternativa sería calcular la correlación a pares (como en la primera simulación) y elegir aquellos coeficientes MFCC con menor correlación de manera manual.
3. Para el algoritmo k-medios, el uso de los coeficientes de tonalidad espectral supone un grado de dispersión respecto a los clusters obtenidos, menor que cuando se utilizan los mfcc.
4. Para el algoritmo k-medios parece ser que existe una relación no lineal entre el número de etiquetas y el grado de dispersión de las muestras de audio respecto a sus centroides, para el conjunto de descriptores óptimo y en general para los conjuntos de descriptores utilizados en todas las simulaciones.
5. Para el algoritmo propagación de afinidad la inclusión tanto de los mfcc como de los coeficientes de tonalidad espectral implican una variación no monótona respecto al número de etiquetas obtenidas. Tal y como se mencionó, el conjunto de descriptores con el cual dicho algoritmo se comportó de manera más estable fue el compuesto por el coeficiente de tonalidad, la energía, el centroide espectral, la tasa de cruce cero y el rmse.
6. Derivado de lo anterior, la evidencia parece indicar que para sonidos muy generales (esto es, sonidos no restringidos a un sólo perfil espectromorfológico como el de la voz), los coeficientes de tonalidad espectral son más eficientes que los MFCC.
7. Dependiendo de los intereses y requerimientos compositivos específicos del usuario, éste puede elegir entre obtener una clasificación con un número más elevado de clusters, sabiendo que existirán

algunas muestras con dispersión respecto a los centroides o decidir obtener un número bajo de etiquetas con la consciencia de que podría existir un sobreajuste de similaridad entre algunas muestras.

Bibliografía

- Zelli, Bijan (2010). *Musique Acousmatique and Imaginary Spaces*. eContact! 13.3, Canadian Electroacoustic Community.
- Schaeffer, Pierre, and G. Reibel (1967). *Solfege De L'Objet Sonore*, Reedition 1998, preface by Daniel Teruggi, Guy Reibel, and Beatriz Ferreyra. Paris: Coedition Ina-Publications.
- Smalley Dennis (1997). *Spectromorphology: explaining sound-shapes*. Department of Music, City University, Northampton Square, London EC1V 0HB, UK.
- Henriksen Frank (2002). *Space in Electroacoustic Music: Composition, Performance and Perception of Musical Space*. City, University of London, Department of Music.
- Macedo Fredserico (2015). *Space as Metaphor: The Use of Spatial Metaphors in Music and Music Writing*. Signata, annals of semiotics.
- Macedo Fredserico (2015). *Investigating Sound in Space: Five meanings of space in music and sound art*. Cambridge University Press.
- Zelli Bijan (2010). *Musique Acousmatique and Imaginary Spaces*. Canadian Electroacoustic Community.
- Van De Gorne Annette (1998). *Space, sound and acousmatic music*. Academia.edu.
- Schumacher Federico, Fuentes Claudio (2017). *Space-Emotion in Acousmatic Music*. Organised Sound December 2017.
- Soria Edmar (2014). *Espacialización Multicanal Algorítmica: Un modelo teórico con implementaciones en Supercollider*. Tesis de Maestría, Posgrado de Música, UNAM.
- Smalley Dennis (2007). *Space-form and the acousmatic image*. Cambridge University Press.

- Harley Maria Anna (1994). *Space and spatialization in contemporary music*. McGill University, Montreal.
- Rumsey Francis (2001). *Spatial Audio*. Focal Press.
- Chowning John M. (2004). *The simulation of moving sound sources*. The Center for Computer Research in Music and Acoustics (CCRMA).
- Gunther Theile, Helmut Wittek (2004). *Waveeld synthesis: A promising spatial audio rendering concept*. Institut für Rundfunktechnik (IRT), Munich, Germany.
- Kendall Gary (2010). *Spatial Perception and Cognition in Multichannel Audio for Electroacoustic Music*. Cambridge University Press.
- Kendall Gary, Ardila Mauricio (2009). *The Artistic Play of Spatial Organization: Spatial Attributes, Scene Analysis and Auditory Spatial Schemata*. Springer Verlag. Volume 4969/2008.
- Bresson Jean, Agon Carlos (2007) *Représentation des données de contrôle pour la spatilisation dans openmusic*. IRCAM - CNRS UMR STMS.
- Pulkki Ville (2001). *Spatial sound generation and perception by amplitude panning techniques*. IRCAM - CNRS UMR STMS.
- Alexander Muller and Rudolf Rabenstein (2009). *Physical Modeling for Spatial Sound Synthesis*. Telecommunications Laboratory University Erlangen-Nuremberg Erlangen, Germany.
- Marcus Lauks (2010). *Computer-Aided Algorithmic Composition*.
- Dan Hosken (2011). *An Introduction to Music Technology*. Ed. Routledge.
- Eduard Rec Miranda (2002). *Computer Sound Design*. Focal Press.
- Alpern Adam (1995). *Techniques for Algorithmic Composition*. Hampshire College.
- Boon Jean Pierre, Decroly Olivier (1994). *Dynamical systems theory for music dynamics*. Université libre de Bruxelles.
- Nguyen Dinh Cong (1997). *Topological Dynamics of Random Dynamical Systems*.
- J. De Vries (2000). *Elements Of Topological Dynamics*.

- Scheinerman R. Edward (1996). *Dynamical Systems*. Department of Mathematical Sciences the Johns Hopkins University.
- Sternberg Shlomo (2000). *Invitation to Dynamical Systems*. Instituto de Investigaciones Fisicas, Carrera de Fisica Universidad Mayor de San Andres.
- Finn Jacobsen (2011). *The sound field in a reverberation room*. Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark,
- Spickler Don, Bergner Jennifer (2011). *The mathematics behind anamorphosis*. Salisbury University.
- Wang DeLiang, Brown Guy J (2005). *Computational Auditory Scene Analysis*. (eds.) John Wiley Sons, Inc.
- Marlon Schumacher, Jean Bressons (2010). *Spatial Sound Synthesis in Computer Aided Composition*. Schulich School of Music of McGill University, Montreal, QC, Canada.
- Nils Petersa, Trond Lossiusb, Jan Schacherc, Pascal Baltazard, Charles Bascoue, Timothy Place (2009). *A stratified approach for sound spatialization*. SMC 2009, July 23-25, Porto, Portugal.
- Juha Merimaa (2006). *Analysis, synthesis and perception of spatial sound binaural localization*. Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing.
- Lambert M. Surhone (Editor), Mariam T. Tennoe (Editor), Susan F. Henssonow (Editor). *Ambisonics*. Beta Script Publishing.
- Büchler, Michael (2002). *Algorithms for Sound Classification in Hearing Instruments*, Tesis doctoral, Swiss Federal Institute Of Technology Zurich.
- Rumsey, Francis (2001). *Spatial Audio*, Focal Press.
- Pressnitzer, Daniel. De Cheveigne, Alain. McAdams, Stephen and Collet, Lionel (2005). *Auditory Signal Processing. Physiology, psychoacoustics, and models*, Springer.
- Rumsey, Francis and McCormick Tim (2009). *Sound and Recording*, Focal Press, sixth edition.
- Bregman, A. S. (1993). *Auditory scene analysis: hearing in complex environments*, Oxford University Press.

- Bregman, A.S., Woszczyk, W. (2004). *Controlling the perceptual organization of sound: Guidelines derived from principles of auditory scene analysis (ASA)*, In K. Greenebaum and R. Barzel (Eds.) *Audio Anecdotes: Tools, tips and techniques for digital audio*. Vol.1. Natick, MA.: A. K. Peters.
- Bregman, A.S., Woszczyk, W. (2005). *Creating mixtures: The application of auditory scene analysis (ASA) to audio recording.*, In K. Greenebaum and R. Barzel (Eds.) *Audio Anecdotes: Tools, tips and techniques for digital audio*. Vol.1. Natick, MA.: A. K. Peters.
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.*, Cambridge, Massachusetts: The MIT Press, 1990 (hardcover)/1994 (paperback).
- Brochard, R., Drake, C., Botte, M.-C., & McAdams, S. (1999). *Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation.*, *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1742-1759.
- Bregman, A.S., A. Ahad A.P. (1990). *Demonstrations to accompany Bregman's Auditory Scene Analysis The perceptual organization of sound.*, Department of Psychology Auditory Research Laboratory McGill University.
- Payri Blas, Miralles Bono José Luis (2007). *Auditory Scene Analysis And Sound Source Coherence As a Frame For The Perceptual Study Of Electroacoustic Music Language.*, Universidad Politécnica de Valencia, Campus de Gandía, 46730 Grao de Gandía, Spain.
- Brown J. G., and Cooke Martin (1994). *Computational auditory scene analysis.*, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, U.K.
- Brown J. G., Wang DeLiang. (2005). *Computational auditory scene analysis. Principles, algorithms, and applications*, The Institute of Electrical and Electronics Engineers, Inc.
- Mountain, Rosemary (2007). *Auditory Scene Analysis and Electroacoustics.*, Music Department, Concordia University.
- Treisman M. A., and Gelade Garry(1980). *A Feature-Integration Theory of Attention.*, Academic Press, Inc.
- Attneave, Fred.(1954). *Some Informational Aspects Of Visual Perception*, *Psychological Review* Vol 61, No. 3.

- Kollmeier, B., Klump, G., Hohmann, V., Langemann U., Mauermann, M., Uppenkamp S., Verhey, J.(2007). *Hearing - From Sensory Processing to Perception*, Springer.
- Bajcsy, Ruzena.(1988).*Active Perception*, Proceedings Of The IEEE, Vol. 76, No. 8.
- Grey John (1976). *Multidimensional perceptual scaling of musical timbres*. Center for Computer Research in Music and Acoustics, Stanford.
- Barlow Horace (1990). *Conditions for versatile learning, Helmholtz unconscious inference and the task of perception*. Physiological Laboratory, Cambridge.
- Windsor Luke (1997). *Frequency structure in electroacoustic music: ideology, function and perception*. Organised Sound 2(2): 77–82.
- Fodor Jerry (1983). *The modularity of mind*.Cambridge, MA: MIT Press
- Reuter Christoph, Siddiq Saleh (2017).*The colourful life of timbre spaces - Timbre concepts from early ideas to meta-timbre space and beyond*. <https://www.researchgate.net/publication/318467904>.
- Stefani Ewan, Lauke Karen (2010). *Music, Space and Theatre: Site-specific approaches to multichannel spatialisation*. Organised Sound / Volume 15 / Issue 03 / December 2010.
- Lindsay Peter, Norman Donald (1972). *Human Information Processing*. Academic Press.
- Rogers Elliot (1968). *Simple visual and simple auditory reaction time: A comparison*. Psychon. Sci., 1968, Vol. 10.
- Barth Friedrich G.,Giampieri-Deutsch Patrizia ,Klein Hans-Dieter (eds.) (2012). *Sensory Perception*. Springer-Verlag/Wien.
- French Robert M. (2002) *The computational modeling of analogy-making*. TRENDS in Cognitive Sciences Vol.6 No.5.
- Kootstra Gert, Bergstrom Niklas, and Kragic Danica (2012) *Gestalt Principles for Attention and Segmentation in Natural and Artificial Vision Systems*. Swedish Foundation for Strategic Research.
- Kofka Kurt (1922) *Perception: An introduction to the Gestalt-theorie*. Etext Conversion Project - Nalanda Digital Library.
- Kofka Kurt (1935) *Principles of Gestalt Psychology*. Lund Humphries, London.

- Shin-ichi Sato, Yoichi Ando (2002). *Apparent Source Width (ASW) of Complex Noises in Relation to the Interaural Cross-correlation Function*. Graduate School of Science and Technology, Kobe University.
- Motokuni ITOH, Kazuhiro IIDA (2006) *Individual differences in directional bands*. WESPAC IX.
- Hafta Ervin, Carrier Samuel (1971). *Binaural Interaction in Low-Frequency Stimuli: The Inability to Trade Time and Intensity Completely*. <https://acousticalsociety.org/>
- Gregory Richard L (1997). *Knowledge in perception and illusion*. Phil. Trans. R. Soc. Lond. B.
- Getzmann Stephan, Lewald Jorg (2007). *Localization of moving sound*. Psychonomic Society, Inc.
- Letowski Tomas, Letowski Szymon (2012). *Auditory Spatial Perception: Auditory Localization*. U.S. Army Research Laboratory.
- Theile Günther (1980). *On the localisation in the superimposed soundfield*. Technische Universität Berlin.
- Disbergen Niels, Valente Giancarlo, Formisano Elia & Zatorre Robert (2018) *Assessing Top-Down and Bottom-Up Contributions to Auditory Stream Segregation and Integration With Polyphonic Music*. Front Neurosci/12.
- Lambert M. Surhone (Editor), Mariam T. Tennoe (Editor), Susan F. Henssonow (Editor). *Ambisonics*. Beta Script Publishing.
- Büchler, Michael (2002). *Algorithms for Sound Classification in Hearing Instruments*, Tesis doctoral, Swiss Federal Institute Of Technology Zurich.
- Rumsey, Francis (2001). *Spatial Audio*, Focal Press.
- Pressnitzer, Daniel. De Cheveigne, Alain. McAdams, Stephen and Collet, Lionel (2005). *Auditory Signal Processing. Physiology, psychoacoustics, and models*, Springer.
- Rumsey, Francis and McCormick Tim (2009). *Sound and Recording*, Focal Press, sixth edition.
- Bregman, A. S. (1993). *Auditory scene analysis: hearing in complex environments*, Oxford University Press.
- Bregman, A.S., Woszczyk, W. (2004). *Controlling the perceptual organization of sound: Guidelines derived from principles of auditory scene analysis (ASA)*, In K. Greenebaum and R. Barzel (Eds.) *Audio Anecdotes: Tools, tips and techniques for digital audio*. Vol.1. Natick, MA.: A. K. Peters.

- Bregman, A.S., Woszczyk, W. (2005). *Creating mixtures: The application of auditory scene analysis (ASA) to audio recording.*, In K. Greenebaum and R. Barzel (Eds.) *Audio Anecdotes: Tools, tips and techniques for digital audio*. Vol.1. Natick, MA.: A. K. Peters.
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.*, Cambridge, Massachusetts: The MIT Press, 1990 (hardcover)/1994 (paperback).
- Brochard, R., Drake, C., Botte, M.-C., & McAdams, S. (1999). *Perceptual organization of complex auditory sequences: Effect of number of simultaneous subsequences and frequency separation.*, *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1742-1759.
- Bregman, A.S., A. Ahad A.P. (1990). *Demonstrations to accompany Bregman's Auditory Scene Analysis The perceptual organization of sound.*, Department of Psychology Auditory Research Laboratory McGill University.
- Payri Blas, Miralles Bono José Luis (2007). *Auditory Scene Analysis And Sound Source Coherence As a Frame For The Perceptual Study Of Electroacoustic Music Language.*, Universidad Politécnica de Valencia, Campus de Gandía, 46730 Grao de Gandía, Spain.
- Brown J. G., and Cooke Martin (1994). *Computational auditory scene analysis.*, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, U.K.
- Brown J. G., Wang DeLiang. (2005). *Computational auditory scene analysis. Principles, algorithms, and applications*, The Institute of Electrical and Electronics Engineers, Inc.
- Mountain, Rosemary (2007). *Auditory Scene Analysis and Electroacoustics.*, Music Department, Concordia University.
- Treisman M. A., and Gelade Garry(1980). *A Feature-Integration Theory of Attention.*, Academic Press, Inc.
- Attneave, Fred.(1954). *Some Informational Aspects Of Visual Perception*, *Psychological Review* Vol 61, No. 3.
- Kollmeier, B., Klump, G., Hohmann, V., Langemann U., Mauermann, M., Uppenkamp S., Verhey, J.(2007). *Hearing - From Sensory Processing to Perception*, Springer.
- Bajcsy, Ruzena.(1988). *Active Perception*, *Proceedings Of The IEEE*, Vol. 76, No. 8.

- F. Gouyon, S. Dixon, E. Pampalk, y G. Widmer (2004). *Evaluating rhythmic descriptors for musical genre classification*. AES 25th International Conference, London.
- Gomez Emilia, Herrera Perfecto (2004). *Automatic extraction of tonal metadata from polyphonic audio recordings*. AES 25th International Conference, London.
- Termens Gaus (2009). *Audio content processing for automatic music genre classification: descriptors, databases and classifiers*. Department of Information and Communication Technologies, Universitat Pompeu Fabra.
- Senin Pavel (2009). *Dynamic Time Warping Algorithm Review*. Information and Computer Science Department, University of Hawaii at ManoaHonolulu.
- Tzortzis Grigorios (2008). *The Global Kernel k-Means Clustering Algorithm*. IEEE World Congress on Computational Intelligence.
- Giorgino Toni (2009). *Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package*. Journal of Statistical Software.
- Keogh E. & Ratanamahatana A. (2005). *Exact indexing of dynamic time warping*. Knowledge and Information Systems.
- Zhao C. & Sahni S. (2019). *String correction using the Damerau-Levenshtein distance*. BMC Bioinformatics.
- Pincus S. (1995). *Approximate entropy (ApEn) as a complexity measure*. Chaos (Woodbury, N.Y.).