

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Programa de Maestría y Doctorado en Música

Facultad de Música
Instituto de Ciencias Aplicadas y Tecnología
Instituto de Investigaciones Antropológicas

Estrategias de composición a través de la similitud tímbrica mediante la sinergia entre la escucha del compositor y la escucha de máquina

QUE PARA OPTAR POR EL GRADO DE
DOCTOR EN MÚSICA
(Tecnología Musical)

PRESENTA
OTTO ALEXANDER CASTRO SOLANO

Tutor principal
Dr. Jorge Rodrigo Sigal Sefchovich (ENES, UNAM)

Miembros del comité tutor
Dra. Rossana Lara Velázquez (UNAM)
Dr. Hugo Solís García (UAM, LERMA)

Ciudad de México, México, 18 abril de 2022

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

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

Agradezco a la Universidad de Costa Rica por el apoyo otorgado a través de su programa de becas para docentes. Esta ayuda me permitió llevar a cabo todo el programa de Doctorado en Tecnología en la Universidad Nacional Autónoma de México.

Esta tesis doctoral está dedicada a la memoria del clarinetista, compositor y director de banda, Efraín Chaves Chaves (1912-2011), mi primer maestro de música.

Por envolver con sueños e ilusiones, a través de los sonidos de sus composiciones y enseñanza instrumental, a tantos jóvenes en Curridabat, Costa Rica.

Coyoacán, México, 15 de abril de 2022

Agradecimientos

A mi esposa Adela Marín, quien siempre tuvo una escucha formidable con diálogos que fortalecieron mis ideas en esta tesis, en donde en cada sesión aprendí cada vez más a escuchar mi propia voz.

A mi tutor principal, el Dr. Rodrigo Sigal por el seguimiento y esmerada entrega a mi proceso de investigación. Le agradezco la libertad que me brindó, con el objetivo de que explorara mi propia forma de componer y reflexionar acerca de mis ideas. Además, Rodrigo fue un guía formidable para mi esposa y para mí, en momentos difíciles viviendo en México, siempre estuvo allí con su consejo preciso.

A la Mtra. María Clara Vargas por las continuas pláticas a distancia, sus consejos, experiencia, guía y apoyo incondicional. A la Dra. Laura Otero por aparecer en México en el momento justo y preciso, acompañándome en el transcurso de mis últimos años del doctorado y dándome mayor claridad sobre los procesos. A doña Yamileth Damazzio y Vivian Madrigal, así como a las dos directoras de la Oficina de Becas de la Universidad de Costa Rica: MS.c. Leonora de Lemos y a la Dra. Diana Senior, por el acompañamiento siempre oportuno, y un magnífico seguimiento a mi beca. A los Doctores Fernando Zúñiga, Orquídea Guandique y Gabriel Venegas que confiaron en esta odisea, haciendo recomendaciones oportunas y apoyándonos de manera sincera y espléndida en cada momento.

A mi familia que siempre me ha acompañado en mis emprendimientos. A mi papá por legarme su amor a la música y al conocimiento, antes de partir de este mundo. A mi madre por su acertada decisión de matricularme a los 8 años en la banda de Curridabat, Costa Rica. A mi hermana Margarita por enseñarme otras perspectivas de la vida. Debo de dar un agradecimiento especial a mi padre político que ya no nos acompaña, Don Mario Cordero, una presencia fundamental en mi familia, siempre con un consejo adecuado y prudente. A mi tía Grethel Solano, por estar tan cercana en momentos importantes de mi vida. Sus largas pláticas por “Whatsapp”, siempre las consideré como parte de mi proceso de investigación. A doña Lidiette Villegas, por acompañarnos en tantas conversaciones y recomendaciones a la distancia, así como abrirnos su casa y atendernos con tanto cariño en nuestros acostumbrados viajes a Costa Rica. Así mismo, a Silvia y a Carlos Marín por su apoyo incondicional, esmerada, espléndida ayuda y apoyo en el transcurso de mis estudios.

Me siento muy satisfecho con el trabajo logrado por los colaboradores de esta tesis. Gracias a Alejandro Cabezas por su detallada y precisa ayuda, en cuanto a la transcripción de las piezas musicales. A mi brillante prima Valeria Méndez por su consejo, apoyo y asistencia en cuanto aspectos de formato y a Priscilla Vargas por su colaboración.

Marusita, Santi, Maby, Andrea, Cristian, Elisa, Gabriela, Axel, Aketzalli, Marie-Christine, Tere, Chana, Carole Chargueron, Cristian Bernal, Mario Duarte y Emma, que fueron nuestros amigos cercanos en este largo camino, siempre

permitiéndonos dialogar, dándonos guías para comprender trozos de la cultura mexicana y ayudándonos en cada momento. A Iracema De Andrade y Edgardo Espinosa que compartieron con nosotros camaradería e inquietudes intelectuales y creativas, acompañándonos en todo momento. Quiero agradecer al destino por conocer en un concierto de Steel Drum al percusionista Juan Carlos Cirujeda, quien además de mostrarnos el Mercado de Sonora, El Chopo, Los Sonideros, fue un muy buen amigo a lo largo del periodo de estudios de Maestría y Doctorado. A doña Celina, Don Max, Jesús, Don Pablo, Maggi y doña Ángeles, vecinos y amigos con los cuales aprendimos más de cerca de este hermoso y magnífico país, gracias por todos los momentos cotidianos compartidos.

A mi comité académico, a la Dra. Rossana Lara por proponer un diálogo provocador, lleno de ideas, que me ayudó a escribir cada vez mejor. Sin lugar a dudas, ella fue una magnífica co-tutora que con mucha paciencia me ayudó a reestructurar mi investigación, a fin de lograr un trabajo de mayor profundidad. Al Dr. Hugo Solís, por dedicar tantas horas, con esmero y cariño, propiciando que descubriera los bordes de la tecnología y su implementación en el ámbito creativo. Al Dr. Cristian Bañuelos por su apoyo incondicional, gracias por sus largas sesiones de diálogo, recomendaciones y sobre todo, porque siempre me alentó a insistir en las profundas raíces de la programación.

A la Dra. Iracema De Andrade por llevar a cabo de forma crítica, comprometida y exhaustiva sus lecturas, que me enriquecieron con una serie de herramientas precisas para la investigación.

A mis profesores que dieron tanta luz y guía para lograr este proceso: Dr. Jorge David Castilla que siempre recordaré su frase: “Las cosas sólo se dan si uno las impulsa”, un maestro que siempre fue un detonador que hacía que todo avanzará dentro y fuera de su alumnado. Al Dr. Roberto Kolb, por sus profundos comentarios en el Seminario Doctoral y por ser tan eficiente en los procesos de Maestría que me llevaron a estar en el programa de Doctorado. A la Dra. Lizette Alegre por brindarnos su preciado tiempo con las bellas sesiones de etnografía, las cuales nutrieron mi panorama musical y entendimiento de lo sonoro. Al Dr. Felipe Orduña por enseñarme con su ejemplo a sistematizar mi pensamiento y sobretodo, por insistir en que fuera yo el que se escuchara a través de la escritura, al Mtro. Pablo Silva por su diálogo de empatía hacia su alumnado y sus recomendaciones tan acertadas.

Al proyecto: Conciertos Expandidos, que me permitió implementar el conocimiento tecnológico en un espacio Inter y Transdisciplinario. A nuestros amigos en Morelia, Mara, Jose Luis, Rosario, Luis Fer, Bernardo. A todo el personal del CMMAS, en especial a Silvana, Alma, Tonalli, Viridiana, Luis, Daniel, Doña Abigail, Lupita y Romina que siempre en mis estancias artísticas nos atendieron con mucho calor humano y escucha.

A mis queridos compañeros que en cada sesión a través de sus investigaciones, comentarios y diálogos aprendí siempre algo de ellos: Aaron, Abner, Aldo, Alessio, Byron, Colasanto, Emmanuel, Guillermo, Hernani, Juan Carlos, María Clara,

Mariana, Nicolás, Nonis, Pablo, Santiago, Sarmen y Tomás. A Mónica y Jasmín por su calor humano y atención esmerada con el alumnado del posgrado UNAM, siempre atentas a escuchar y solventar cualquier duda o situación. A nuestro coordinador del Posgrado, el Dr. Fernando Nava por su eficiencia en la resolución de los asuntos administrativos.

A mis maestros de composición que sembraron en mí la motivación a seguir en estos territorios tan desérticos pero apasionantes. Al Dr. Bernal Flores, por sembrar con su ejemplo, el interés de entender la música más allá de la tradición, al Mtro. Benjamín Gutiérrez que fortaleció con sus sesiones mi interés por el color tímbrico en sus enseñanzas de orquestación instrumental. Al Mtro. Luis Diego Herra, por mostrarme por primera vez lo que era la música electroacústica, la pasión por las computadoras y descubrir que los recursos gráficos y la improvisación podían ser parte de la composición musical. Al Mtro. Marvín Camacho por ser mi inspiración de cómo ser un compositor. Al Mtro. Rogelio Lopéz por motivar desde la obra de ballet: “Buenas noches tristeza”, mi pasión por el sonido urbano y los montajes sonoros. Al Dr. Gabrio Zapelli por su siempre luminosas pláticas y sesiones de convivio.

A nuestros amigos costarricenses residentes en México por dejarnos escuchar sus palabras ticas: Lau, Mario, Daniel, DAngelo, Arabela. Y los que dejamos en nuestro terruño: Sussy, Roberto, Ana, Edguitar, Marianita, Pili, Marianela, Don Carlos y Andrés. A Edin Solís por aquellas lecciones de armonía en 1989, y su frase: “dejemos que pase el tiempo, para ver qué de nuestras vidas”. A Alonso Torres, por las largas conversaciones de intercambio de ideas musicales, que me hicieron crecer y revisar con más detenimiento mis certezas.

Un inmenso agradecimiento a la Asamblea por darme la oportunidad de estudiar becado en el extranjero y a la Cátedra de Composición de la EAM-UCR: Maestros Carlos Castro, Eddie Mora y Carlos Escalante, por darme la oportunidad de incorporar mis nuevos conocimientos en la institución. A la Dra. Susan Campos por los diferentes proyectos, diálogos y conocimientos compartidos. A la Dra. Patricia Fumero y a la Dra. Marjorie Ávila por apoyarme y estar al tanto de mi desarrollo como investigador.

Índice de Contenidos

INTRODUCCIÓN 1

CAPÍTULO 1: CRITERIOS DE SEGMENTACIÓN DESDE LA ESCUCHA: HUMANO / MÁQUINA 10

PROCESO DE ESCUCHA Y SELECCIÓN DE LOS MATERIALES SONOROS EN LA CIUDAD 11
LA CIUDAD GENERADORA DE SONIDOS 11
LOS SONIDOS URBANOS EN MIS OBRAS SONORAS ANTERIORES 16
REPOSITORIOS SONOROS: ¿COLECTIVOS O PERSONALES? 21
METODOLOGÍA ARTÍSTICA EN LA EXPLORACIÓN DEL ESPACIO URBANO 27
Preparación para la exploración 28
Grabando mientras se camina 30
Implementación de la escucha 31
Algunos consejos y estrategias para la grabación en el lugar 35
Intencionalidad en la grabación de las fuentes sonoras 38
Errores técnicos aprovechados como recursos estéticos 47
SEGMENTACIÓN DE LOS MATERIALES SONOROS A TRAVÉS DE LA MÁQUINA 50
REFLEXIÓN ACERCA DE LA ESCUCHA HUMANA Y LA ESCUCHA DE MÁQUINA 63
PROGRAMAS INFORMÁTICOS DE REFERENCIA CERCANOS AL *CSPECTRA2021* 66
Contexto general 66
Aspectos generales de los programas seleccionados: AudioGuide y Sononym en relación con CSpectra2021 70
Comparación de algunos aspectos requeridos en las dependencias de los programas 77
Datos de entrada para los sistemas informáticos 79
Datos de salida de los sistemas informáticos 80
Datos de control en los sistemas informáticos 82
Datos almacenados por el sistema informático 84
Datos almacenados para el usuario 85
Otros datos generados por los programas 86

CAPÍTULO 2: IDENTIFICACIÓN DE SONIDOS MEDIANTE LA MÁQUINA A TRAVÉS DE SU COMPARACIÓN ESPECTRAL 88

EL TIMBRE Y EL ESPECTRO COMO ÁMBITOS DE PARTIDA 89
LA IMPORTANCIA DEL TIMBRE EN MI PRODUCCIÓN SONORA 93
ACERCAMIENTO AL TIMBRE Y EL ESPECTRO 98
CARACTERÍSTICAS PERCEPTUALES Y ACÚSTICAS QUE SE BUSCAN 99
Categoría: brillante / opaco 102
Categorías: rugoso / liso 108
ACERCAMIENTO TECNOLÓGICO PARA ENCONTRAR SIMILITUD ENTRE LOS SONIDOS 117
PRIMERA EXPLORACIÓN: CÓDIGO DETONANTE 119
MAT2021A Y MAT2021B COMO ANTESALA A LA HERRAMIENTA FINAL: *CSPECTRA2021* 122
EL CSPECTRA2021 134
Aspectos generales 135
Estructura del CSpectra2021 136
Módulo dedicado a encontrar similitud 140
Preprocesamiento de los datos 141

Creación de la base de datos *143*
Medición de los puntos en un espacio multidimensional *144*
Texturizador *148*
Script de Reaper *148*

CAPÍTULO 3: ESTRATEGIAS DE COMPOSICIÓN MUSICAL 154

CUATRO POSIBILIDADES DE COLOR TÍMBRICO EN LAS PIEZAS MUSICALES *MOSAICO SONORO, ANOMALÍA URBANA Y TEXTURAS URBANAS* 156
SIMILITUD Y CONTRASTE TÍMBRICO EN LAS OBRAS: *MOSAICO SONORO, ANOMALÍA URBANA Y TEXTURAS URBANAS* 164
EL ENMASCARAMIENTO COMO RECURSO COMPOSITIVO EN LA PIEZA *TEXTURAS URBANAS* 168
ATENUACIÓN Y REFORZAMIENTO DEL COLOR TÍMBRICO EN LAS OBRAS *ANOMALÍA URBANA Y TEXTURAS URBANAS* 170
LA DENSIDAD ESPECTRAL COMO PRINCIPIO DE TENSIÓN MUSICAL EN LA PIEZA MUSICAL *MOSAICO SONORO* 172
POSIBILIDADES DE COMPOSICIÓN A TRAVÉS DE LAS CAPAS TÍMBRICAS Y LOS PLANOS SONOROS EN LAS TRES PIEZAS MUSICALES 176
Fusión tímbrica de capas *176*
Independencia de los planos tímbricos *178*
REFLEXIÓN ACERCA DE LOS RESULTADOS ESTÉTICOS LOGRADOS CON LA VERSIÓN FINAL DEL *CSPECTRA2021,* EN LA PIEZA *EJES URBANOS* 183
Algunos aspectos técnicos generales *184*
Estructura general de la pieza *184*
Tres tipos de búsquedas para generar texturas sonoras con el CSpectra2021 ..... *187*
Proceso de composición *189*
Algunas estrategias como parte del proceso de composición *194*
Otras estrategias utilizadas *198*

CONSIDERACIONES FINALES 200

BIBLIOGRAFÍA 208

ANEXO 277

EJEMPLOS SONOROS DE LAS PIEZAS 278
INDICACIONES DE NOMENCLATURA EN LOS EJEMPLOS 279
Rugoso *279*
Brillo *280*
Oscuro *280*
Liso *281*
Diferentes perspectivas sonoras de Barry Truax *281*
Perspectivas sonoras suelo *282*
Atenuación *283*
Sonidos armónicos *283*
Densidad tímbrica *283*
Enmascaramiento *284*
Planos tímbricos *285*
Contraste tímbrico de transición *287*
Contraste tímbrico abrupto *288*
Otros recursos musicales para reforzamiento de dirección musical *289*
REPRESENTACIÓN VISUAL EN FORMATO DE NOTACIÓN 291
GLOSARIO SONORO DE LA NOTACIÓN DE *THORESEN* 292
Mosaico Sonoro *298*
Anomalía Urbana *304*
Texturas Urbanas *311*

Índice de Tablas

TABLA 1 OBRAS SONORAS BASADAS EN SONIDOS URBANOS 17

TABLA 2 MÉTODOS DE DETECCIÓN DE EVENTOS A TRAVÉS DE TRANSITORIOS 51

TABLA 3 COMPARACIÓN ENTRE LA ESCUCHA DE MÁQUINA Y ESCUCHA HUMANA 63

TABLA 4 INFORMACIÓN DE ARCHIVO GENERADO POR AUDIOGUIDE DE SEGMENTO UN FICHERO DE AUDIO CON UNA DURACIÓN DE 2'15" SEGUNDOS 85

TABLA 5 VALORES DE LOS DESCRIPTORES ACÚSTICOS NORMALIZADOS 128

TABLA 6 DESCRIPCIÓN DE LAS SECCIONES DE LA PIEZA EJES URBANOS 186

Índice de Figuras

FIGURA 1 ALGUNOS PUNTOS DEL TERRITORIO EXPLORADO EN EL SUR DE LA CIUDAD DE MÉXICO, COYOACÁN 29

FIGURA 2 DISPOSICIÓN DEL EQUIPO DE GRABACIÓN PARA LOS REGISTROS URBANOS 36

FIGURA 3 TAPAS METÁLICAS Y SUELOS CON ONDULACIONES 43

FIGURA 4 TRES DIFERENTES COLOCACIONES DEL MICRÓFONO PARA LA CAPTACIÓN DEL SONIDO URBANO APROVECHANDO LA ARQUITECTURA 49

FIGURA 5 EJEMPLO DEL CAMBIO EN LA ENERGÍA ESPECTRAL A PARTIR DEL TIEMPO 1'03". EL CAMBIO TÍMBRICO SE INDICA VISUALMENTE EN LA FIGURA POR EL RECTÁNGULO AZUL EN LA FIGURA 53

FIGURA 6 MODULACIÓN DE LAS FRECUENCIAS EN EL SONIDO DE UNA AMBULANCIA 58

FIGURA 7 ANÁLISIS DE UNA SEÑAL A TRAVÉS DEL DESCRIPTOR ACÚSTICO SPECTRAL FLUX 59

FIGURA 8 PRESENTACIÓN GRÁFICA PARA EL FILTRO DE BÚSQUEDA OFRECIDO POR EL PROGRAMA SONONYM 83

FIGURA 9 CONTINUIDAD DE SOBRETONOS 95

FIGURA 10 PRIMERA PARTE DE LA TEXTURA TÍMBRICA IRREGULAR QUE VA DE PARTES BRILLANTES A OPACAS 97

FIGURA 11 MEDICIÓN DEL SPECTRAL CENTROID. CASO TÍMBRICO BRILLANTE 104

FIGURA 12 SPECTRAL CENTROID ENTRE LA BANDA DE 300 HZ Y 1200 HZ. CASO TIMBRE OPACO 106

FIGURA 13 MEDICIÓN DEL SPECTRAL CENTROID. CASO TÍMBRICO OSCURO 107

FIGURA 14 NIVEL DE RUGOSIDAD EN UN SONIDO DE MOTOR DE AUTOBÚS 111

FIGURA 15 FRAGMENTO DE SONIDO RUGOSO "ANOMALÍA URBANA" (4'07" A 4'20") 112

FIGURA 16 FRAGMENTO DE SONIDO LISO DE LA PIEZA: ANOMALÍA URBANA (6'34" A 6'36") 113

FIGURA 17 FRAGMENTO DE SONIDO LISO ANOMALÍA URBANA (2'08" A 2'10") 114
FIGURA 18 MÓDULOS DE CÓDIGO DE LAS VERSIONES MAT2021A Y MAT2021B 123
FIGURA 19 DIAGRAMA DE FLUJO DEL CSPECTRA2021 135
FIGURA 20 INTERFACE DE USUARIO DEL CSPECTRA2021 138
FIGURA 21 REPRESENTACIÓN EN DOS DIMENSIONES DE LOS PUNTOS EN EL ESPACIO 146
FIGURA 22 DIVISIÓN A TRAVÉS DE UN HIPERPLANO DEL ESPACIO 147
FIGURA 23 CAPTURA DE LA PANTALLA DEL PROGRAMA REAPER QUE MUESTRA CÓMO CARGAR UN SCRIPT 149
FIGURA 24 SELECCIÓN DE UN SEGMENTO SONORO 150
FIGURA 25 SELECCIÓN Y EJECUCIÓN DEL SCRIPT CSP_REAPER.PY 150
FIGURA 26 PISTA ORIGINAL Y TEXTURA SONORA RESULTANTE 151
FIGURA 27 INFORMACIÓN DESDE LA CONSOLA DE LA TERMINAL CON RESPECTO A LOS SEGMENTOS SELECCIONADOS PARA CONFORMAR LA TEXTURA SONORA 152
FIGURA 28 DISPOSICIÓN DE LAS CUATRO CATEGORÍAS PERCEPTUALES UTILIZADAS EN LAS TRES PIEZAS COMPUESTAS 160
FIGURA 29 REPRESENTACIÓN DE CUATRO FRAGMENTOS SONOROS A TRAVÉS DE LAS CUATRO CATEGORÍAS PERCEPTUALES 164
FIGURA 30 REPRESENTACIÓN DE LA DENSIDAD ESPECTRAL 174
FIGURA 31 ESTRUCTURA GENERAL DE LA PIEZA EJES URBANOS 186

Palabras clave
(Español)

Estrategias de composición, escucha de máquina, escucha humana, similitud tímbrica, reconocimiento automatizado.

Keywords

(English)

Compositional Strategies, Machine Listening, Human Listening, Timbre Similarity, Automatic Recognition.

Sitio web dedicado a esta investigación

https://ottocastro.com/doctorado/

Introducción

Esta investigación parte de mi experiencia como compositor de obras sonoras vinculadas al sonido urbano, y de la necesidad de desarrollar materiales sonoros a través de relaciones con el timbre entre los sonidos para componer piezas musicales. Al iniciar mi carrera me interesó el trabajo con empastes sonoros, incorporando yuxtaposiciones de sonidos instrumentales, que provocaban cambios en los colores tímbricos. Esto direccionó un modo de componer y de disponer los materiales sonoros, por otro lado, esta investigación se propuso recolocar mis aproximaciones a la composición al incorporar en mi hacer procesos de aprendizaje automático, concretando otro tipo de posibilidades y herramientas diferentes para la creación.

Por otro lado, el objetivo general planteado para esta investigación es crear obras sonoras a través de la colaboración entre escucha de máquina y la escucha humana. Esta colaboración tecnológica puede posibilitar una percepción diferente y suplir una necesidad de organización y sistematización de una gran cantidad de materiales sonoros, sobretodo provenientes de grabaciones de sonidos urbanos.

Por tanto, esta investigación utilizará sonidos urbanos como material principal y buscará resolver esta necesidad mediante un proceso de identificación de eventos sonoros, elaborado por el procesamiento que hace la computadora, a través de la cual encuentra similitudes tímbricas entre los sonidos. Esta labor podrá generar posibilidades compositivas a través de diferentes líneas estéticas y requerimientos de la pieza.

Esta investigación propone el desarrollo de un proceso artístico que conjuga la reflexión acerca de sus etapas. Así mismo, considera la influencia en el ámbito

estético que puede proveer el uso de herramientas tecnológicas, la escucha de máquina, la del compositor y lo que se genera entre estas dos escuchas. Esta pesquisa mostrará una serie de estrategias de composición, derivadas de mi interés por la similitud tímbrica desde mi práctica artística, que se proponen como un punto de partida para que otros creadores puedan: debatir, dialogar o incluso refutar, enriqueciendo el camino de investigación a través de otras proposiciones, pudiendo repensar la composición desde una posición diferente a lo que conocían.

Ahora bien, para redondear un contexto, con relación a mi modo de trabajo anterior a esta investigación y el posible aporte de las nuevas herramientas desarrolladas en esta pesquisa, podría mencionar que, con respecto al uso de materiales sonoros urbanos, en composiciones anteriores a este trabajo, una de las tareas fundamentales era la búsqueda de sonidos similares y contrastantes, con el fin de generar desarrollos en la construcción de las piezas musicales. Por lo general, esta búsqueda se realizaba sin la mediación de procesos automáticos para la identificación de estos sonidos. En esta labor se escuchaban los materiales y se anotaban sus posibilidades individualmente en cada caso, con el fin de encontrar relaciones entre dichos sonidos para implementar el montaje en la obra sonora. Este proceso demoraba mucho tiempo porque tenía que escuchar y comparar los sonidos repetidamente.

Durante el desarrollo de estas tareas, mi escucha me llevaba a tomar decisiones a través de la identificación, comparación de materiales y mi memoria. Sin embargo, después de unos minutos, mi atención y escucha se cansaban. Esta

comenzaba a ser menos atenta, y a presentar patrones recurrentes para sus búsquedas, volviéndose habitual y monótona a mi parecer.

Ahora bien, en esta investigación, aunque la máquina tiene algunas limitaciones en cuanto a coincidencias con la percepción de mi escucha, no posee el inconveniente de cansarse, además identifica sin ningún prejuicio los sonidos. Al mencionar prejuicio, me refiero a que esta audición, consiste en una identificación de parámetros objetivos, que no parte de relaciones con la memoria o la experiencia vivida, presente en la escucha humana. Por consiguiente, la máquina brindará al compositor, otra manera de escuchar que puede beneficiar el horizonte estético y artístico en cuanto a la composición musical.

En esta investigación, el principal material sonoro en la construcción de piezas será el sonido urbano, debido a su diversidad tímbrica. Con respecto a esta diversidad, me interesa trabajar con estos sonidos porque pueden ser armónicos o inarmónicos y contener una frecuencia fundamental con mayor o menor evidencia. Dicha variedad provendrá de fuentes sonoras como: cláxones, pitidos, sirenas de ambulancia, señales sonoras peatonales, el ruido de las muflas de los autobuses, el roce de los neumáticos sobre materiales rugosos, ondulados y ásperos del pavimento, entre otros. La pluralidad de registros sonoros urbanos aportará riqueza a los timbres con los cuales se pretende trabajar en la identificación y clasificación que hará el sistema tecnológico.

El interés por el sonido urbano será dirigido hacia el análisis de estos materiales, desde el ámbito tímbrico y su implementación en la creación musical, y

se descartará el abordaje de géneros musicales cercanos, como es el caso del paisaje sonoro imaginario[1], debido a que se desea delimitar el foco de estudio.

En este sentido, esta pesquisa tratará de ahondar en el proceso creativo de las piezas musicales, a través de categorías tímbricas. Dichas piezas serán el resultado de una ruta de trabajo aunada a procesos computacionales de escucha, y su acompañamiento con labores no automáticas de creación. Es decir, el uso de la tecnología será un medio de apoyo para la composición, que ayudará a confrontar la escucha del compositor en el acto creativo, y por ende, complementado mediante la escucha de máquina. Propiciando un ensanchamiento y reflexión de una escucha mixta.

Es preciso resaltar que cuando se mencione la escucha humana, en todos los casos sin excepción, se hablará de esa escucha desde la posición de mi experiencia con el entorno sonoro de la ciudad, y la de mi experiencia perceptiva con los materiales sonoros resultantes del proceso maquinal que usaré, manipularé y montaré en la construcción de las piezas musicales.

Por consiguiente, la estructura del capitulado de la tesis, está pensada para que las estrategias de composición sean el centro, poniendo atención a los procesos entre la escucha del compositor y máquina, antes mencionados.

En el capítulo 1: “Exploración de escucha y segmentación desde humano/máquina”, se hace un brevísimo contexto de la manera en que se ha entendido la ciudad, a partir de algunos autores canónicos que han abordado este

[1] El paisaje sonoro imaginario es una creación, en la cual se utilizan diferentes sonidos: reales, procesados o creados electrónicamente, para recrear la sensación de un entorno.

tema desde la creación artística, con el fin de tener un acercamiento a la comprensión del fenómeno sonoro de la ciudad. Este contexto también se extiende a la manera en que he abordado obras sonoras anteriores con este tema.

Ahora bien, este capítulo está estructurado en tres grandes áreas que abarcan: 1. Un acercamiento a las exploraciones y la selección de sonidos urbanos desde la escucha humana, 2. desarrolla los criterios de segmentación que efectuará el sistema informático, y 3. construye una revisión de programas informáticos similares a la herramienta tecnológica desarrollada en esta investigación.

En la primera parte del capítulo se trabajarán reflexiones sobre repositorios sonoros colectivos y personales, metodología de exploración del espacio urbano y aspectos de grabación.

Además, al final del capítulo, se abordarán las dos escuchas: de humano y de máquina, tomando en cuenta los límites de sus rangos. Estas escuchas serán de vital importancia para los procesos de segmentación de los materiales y el desarrollo metodológico que se utilizará con criterios particulares para el tratamiento de los materiales. Se llevará a cabo una reflexión entre las dos escuchas para conocer sus potencialidades y limitaciones.

Es importante anotar que cuando mencionemos escucha de máquina, este aspecto estará relacionado a procesos desarrollados con programas informáticos dedicados al reconocimiento de sonidos con determinadas características acústicas.

El capítulo 2, nombrado: “Identificación de sonidos mediante la máquina a través de su comparación espectral”, hará un breve contexto de la implementación

del timbre como recurso creativo en las obras de algunos compositores, desde Claude Debussy (1862-1918), hasta la escuela de música espectral en la década de los setenta del siglo pasado. Así mismo, este capítulo vinculará el concepto del timbre para dirigir su atención hacia el análisis del espectro. Se elegirá el espectro de la señal, debido a su posibilidad de ser medible a través de datos estadísticos. Esto nos encamina a enfocarnos en el área tecnológica asociada a la escucha de máquina. Se abordarán aspectos operativos, llevados a cabo por la herramienta tecnológica *CSpectra2021* (Herramienta de Análisis y Comparación Espectral), desarrollada en esta investigación. Esta lleva a cabo procesos de análisis de los sonidos, extracción de características acústicas, recolección de datos de esos análisis y comparación de los arreglos numéricos arrojados por dicho sistema informático desarrollado. Todas estas funciones son ejecutadas con el fin de encontrar similitudes espectrales entre los materiales sonoros.

A lo largo de este capítulo, estas similitudes espectrales, serán encontradas y agrupadas por determinados descriptores acústicos e interpretadas y categorizadas desde el ámbito perceptual de mi escucha. Esta labor sentará las bases de los materiales que serán trabajados en el capítulo tres.

En el capítulo 3: “Estrategias de composición”, nos enfocaremos en la organización de los materiales obtenidos de la escucha de máquina, con el fin de ser dispuestos de manera estratégica para la creación sonora. Desde mi escucha, se desarrollarán cuatro categorías perceptuales vinculadas al timbre: brillante, opaco, rugoso y liso.

Los estudios de caso partirán de las tres piezas musicales desarrolladas a lo largo de esta investigación, las cuales son: *Mosaico Sonoro, Anomalía Urbana y Texturas Urbanas*. A través de estas, se analizarán diversos tópicos estratégicos incorporados en dichas piezas, como: similitud y contraste tímbrico, enmascaramiento como recurso compositivo, atenuación y reforzamiento del color tímbrico, y la densidad espectral como principio de tensión musical. Así mismo, se abordarán posibilidades compositivas mediante las capas tímbricas y los planos sonoros en las tres piezas musicales mencionadas.

Para finalizar, es importante mencionar que, durante este proceso de investigación, con un componente indispensable de interés en el proceso creativo, se desprenderán etapas vinculadas a los procesos de: segmentación, identificación tímbrica y comparación espectral. Estos estarán unidos a aspectos subjetivos de escucha y a estrategias de organización de los materiales, con una dirección estética en progreso, en el acto de componer. De aquí se desencadenan varias preguntas importantes a desarrollar a lo largo de este trabajo.

¿De qué forma es necesario adecuar la segmentación a características determinadas para obtener un material sonoro específico para componer?

¿Qué características acústicas pueden definirse para encontrar niveles de similitud tímbrica?

¿Qué estrategias de composición podrían derivarse de materiales similares tímbricamente?

¿Qué resultados estéticos pueden ser obtenidos a través de un sistema informático de identificación sonora y la creación del compositor, a partir de su experiencia subjetiva?

Estas preguntas muestran un posible camino a seguir dentro de esta ruta de investigación en los materiales, su relación entre la tecnología y aspectos compositivos, una ruta que otros investigadores podrán reelaborar o retomar desde otros enfoques, intereses y estéticas.

Capítulo 1

Criterios de segmentación desde la escucha: humano / máquina

Proceso de escucha y selección de los materiales sonoros en la ciudad

El interés primordial en esta investigación es desarrollar composiciones de obras musicales con sonidos urbanos; para ello, se creó una herramienta tecnológica especializada que encuentra similitudes entre los sonidos. Como el sonido urbano es la materia principal para la creación de estas obras, el primer paso fue explorar el espacio urbano para luego seleccionar los materiales que se introducirían al sistema informático. A partir de las exploraciones, se desarrolló una metodología artística, tanto para la escucha como para la selección de los sonidos que se deseaban recolectar.

La ciudad generadora de sonidos

Cuando se piensa en la ciudad, nos remitimos a un espacio donde habitan personas que realizan actividades industriales y comerciales. La Real Academia Española la define como un lugar: "[…] *cuya población densa y numerosa se dedica por lo común a actividades no agrícolas*" (2021)[2]. Entonces, los sonidos que más se escuchan en la ciudad son los de máquinas, fábricas, medios de transporte (terrestres y aéreos) y los de dispositivos automáticos (como las agujas de paso, alarmas o señales sonoras para el tránsito peatonal en una esquina, entre otros). Por esta razón, hemos conformado un imaginario sonoro de la ciudad compuesto por sonidos en su mayoría "ruidosos", los cuales se contraponen al imaginario sonoro que tenemos de lo rural, la selva o el bosque.

---

2 Ver más en https://www.rae.es/drae2001/ciudad [visitado el 15 de abril de 2021].

Ahora bien, la Real Academia Española define el ruido como un "*sonido inarticulado, por lo general desagradable*"[3] (2021). Aunque es algo que percibimos como molesto, el ruido también es contextual, pues lo que para una persona es un sonido incómodo, para otra podría no serlo. Sin embargo, esto no minimiza los efectos que los sonidos "ruidosos" pueden provocar en la salud, sobre todo los producidos en la urbe. A pesar de ello, podemos acercarnos al ruido desde una perspectiva estética y artística que valore esos materiales desde sus posibilidades acústicas para la creación de piezas musicales.

La estudiosa Ana Lidia Domínguez se refiere al sonido de la ciudad de la siguiente forma:

> El ruido es el referente sonoro de la dinámica urbana, producto de una ciudad que se mueve y cuyo movimiento es indispensable para que ésta funcione. Hoy día, este fenómeno domina el paisaje sonoro de las grandes ciudades, configurando un entorno saturado no sólo de altos volúmenes, sino de una presencia sonora continua que no da tregua, un paisaje invadido por sonoridades producto de una vida altamente tecnologizada. (Domínguez, 2014, p. 90)

Todos estos sonidos conforman lo que el Proyecto de Paisaje Sonoro Mundial, o *World Soundscape Project* (WSP), identificó como "entorno sonoro", vinculado a la formulación del paisaje sonoro como concepto. Al respecto, Raymond Murray Schafer (1933-2021), fundador de la WSP, establece lo siguiente:

---

3 Ver más en https://dle.rae.es/ruido [visitado el 21 de mayo de 2021].

> Denomino *soundscape* (paisaje sonoro) al entorno acústico, y con este término me refiero al campo sonoro total, cualquiera que sea el lugar donde nos encontremos, no está totalmente limitado a espacios exteriores. El entorno que me rodea mientras escribo es un *soundscape*. (Schafer, 2013, p. 50)

Curiosamente, los pioneros del paisaje sonoro se enfocaron en reflexionar y concienciar acerca de los altos niveles de polución sonora que se estaban produciendo en diferentes ciudades del mundo. Ante esto, Murray Schafer afirma: "*La contaminación acústica se da cuando el hombre no escucha con atención. Los ruidos son los sonidos que hemos aprendido a ignorar*" (1977/1994, p. 21).

Así, la WSP impuso una carga negativa a todo sonido "ruidoso" que se produjera en la ciudad, sin considerar que esos mismos sonidos también constituían una representación de la urbe misma, en su condición de espacio comercial e industrial; es decir, los sonidos de las urbes son una proyección cultural, económica y social de un lugar y, por eso, constituyen parte fundamental del paisaje sonoro. Además, generan una vinculación afectiva e identitaria con el lugar por parte de sus habitantes. Por lo anterior, en la presente investigación se entiende la ciudad como un entorno sonoro en donde, como creadores, podemos aprovechar la multiplicidad de sonidos que lo componen: estos se muestran en forma de conglomerados sonoros, los cuales aparecen y desaparecen de manera continua, yuxtapuesta y con diferentes niveles de densidad sonora[4]. Dichas características hacen del

[4] La densidad sonora se refiere a la "*sensación que producen ciertos sonidos de ser más compactos o densos que otros. Esta característica parece depender directamente de la frecuencia, pues*

espacio de la ciudad un lugar ideal para experimentar, recolectar y nutrirse estéticamente.

Ahora bien, a lo largo de los años, se han generado discusiones teóricas con respecto a los entornos sonoros urbanos. Hildergard Westerkamp (1946-), por ejemplo, afirma: "*Más que permanecer marginados produciendo música culta inaccesible y abstracta destinada a audiencias exclusivas y reducidas, concebimos al compositor como un contribuyente valioso en el manejo de asuntos del paisaje sonoro*" (2002). Sin embargo, su posición resulta tajante, pues podría ser más adecuado conectar con la ciudad desde una posición más libre estéticamente, sin tener el peso de resguardar y conservar los sonidos que allí habitan; es decir, nuestra producción sonora puede ser una interpretación de la ciudad, de modo que le devolvamos a esta una parte de nuestra escucha y creemos, a la vez, ciudades paralelas a la real.

Por su parte, en 1913, los artistas Luigi Russolo (1885 - 1947) y Francesco Badilla Pratella (1880-1955) vieron en la ciudad una fuente de sonidos ilimitada, más allá de las consecuencias para la salud. Sus ideas fueron condensadas en la publicación *El arte de los ruidos* (1913) o *L' arte dei Rumori*, en su título original en italiano. En el siguiente extracto de su manifiesto, los rumoristas[5] reflejan la manera en que escucharon la ciudad:

---

aumenta con ella como consecuencia de la mayor velocidad con que se producen las vibraciones en los sonidos agudos" (Calvo-Manzano Ruiz, 2004, p. 136).

[5] *Rumore,* en italiano, significa "ruido"; de allí que "rumorista" se podría traducir también como "ruidista".

> […] el rugido de los motores que bufan y pulsan con una animalidad indiscutible, el palpitar de las válvulas, el vaivén de los pistones, las estridencias de las sierras mecánicas, los saltos del tranvía sobre los raíles, el restallar de las fustas, el tremolar de los toldos y las banderas. (Russolo, 1913/1986, p. 5)

A pesar de las diferentes posturas con respecto al tema, lo cierto es que el espacio urbano contiene una gran diversidad de sonidos de diferentes calidades y facturas. Cuando esos sonidos se yuxtaponen, se convierten en conglomerados sonoros que pueden ser interpretados como texturas sonoras complejas.

A nivel personal, la ciudad es un espacio muy importante, pues he desarrollado toda mi vida en ella; es, por tanto, una extensión de mi propia vida: sus sonidos me remiten no solo a un lugar, sino que constituyen mi propio reflejo. Esto por cuanto la ciudad se conforma por sonidos espontáneos, pero a la vez evoca recuerdos sonoros, vivencias e imaginarios. Su resonancia contiene los sonidos industriales, los cotidianos, los míos, los de los otros, en un espacio de convivencia e intercambio. La ciudad es un espejo de nuestras huellas, tanto en el espacio como en el tiempo, e incluso proyectamos sus sonidos mentalmente en nuestro interior, creando de esta manera otros paisajes sonoros.

Justamente por las características esbozadas, el compartimiento sonoro de la ciudad es tan diverso, dinámico y complejo que esto la convierte en un lugar ideal para recolectar materiales sonoros con características tímbricas heterogéneas, las cuales pueden ser aprovechadas mediante el análisis y la vinculación acústica de un sistema informático. De este modo, se convierte en una herramienta para la

composición musical. A partir del entendimiento de este contexto, surge la presente investigación, sustentada en que los sonidos de la ciudad merecen una relectura dentro de la organización artística del proceso de composición musical.

Los sonidos urbanos en mis obras sonoras anteriores

El interés por los sonidos urbanos en mi producción sonora inicia en el año 2000, cuando el coreógrafo costarricense Rogelio López (1948-) me encargó la composición de la obra para danza *Buenas noches tristeza*, con una duración de 45 minutos, para la Compañía de Danza Universitaria[6] de la Universidad de Costa Rica. Antes de este encargo, mi producción musical se centraba en composiciones creadas para ser tocadas por instrumentos acústicos, con muy poca o nula intervención electrónica; eran piezas con una métrica constante de principio a fin, estructurada según una tonalidad o utilizando recursos de modulación para moverse hacia otras tonalidades.

La obra *Buenas noches tristeza* produjo un giro en mi producción musical debido a que fue la primera pieza que requirió explorar los sonidos de la ciudad, vincularlos y desarrollarlos. Las técnicas de composición para manipular materiales sonoros utilizados en estos primeros acercamientos eran básicamente las técnicas de composición de la música concreta[7]: invertir un sonido, estirarlo, comprimirlo,

---

[6] Ver más en https://danzau.ucr.ac.cr/ [visitado el 25 de noviembre de 2020].

[7] Del francés, *musique concrète.* Esta música nació cerca del año 1929 en los estudios de radiodifusión francesa y fue desarrollada por el compositor Pierre Schaeffer. Utiliza, principalmente, la técnica del montaje, en la cual se colocan los materiales sonoros en el soporte que el compositor considere pertinentes para conformar su obra sonora. Estos sonidos pueden presentarse de manera

leerlo a mayor o menor velocidad, cortarlo, yuxtaponerlo, entre otras. Dos aspectos importantes que caracterizan la estética musical de la producción del año 2000 y 2001 son: su acentuado interés en mantener algún nivel de reconocimiento de las fuentes, así como su enfoque documental cuando se grababa un sonido y se montaba en la obra musical. Así, por ejemplo, se realizaban entrevistas a vendedores ambulantes y estos audios se incluían en secciones o partes de la composición, pero sin perder el contenido semántico de sus entrevistas. A continuación, se muestra un listado de las obras[8] anteriores a las piezas del doctorado que se trabajaron con sonidos de la ciudad.

Tabla 1

Obras sonoras basadas en sonidos urbanos

#	Nombre de la pieza musical	Año de creación	#	Nombre de la pieza musical	Año de creación
1	Buenas noches tristeza	2000	8	TransUrbano	2008
2	Murmuraciones	2000	9	Tsukur	2015
3	Jardín tóxico	2000	10	Zobeide	2014
4	Encuentros televisivos	2001	11	Cuerpo mutable	2015
5	Mala Fe	2002	12	Olinda	2015
6	Código vestimenta	2004	13	Argia	2016
7	Fluctuaciones	2006	14	Ersilia	2016
8	TransUrbano	2008	15	Homenaje a la Ciudades Invisibles	2017

Ahora bien, en todas las piezas producidas por mí antes del año 2014, los criterios de selección de los lugares para grabar eran casuales y los de los sonidos, dependían de cuando detectara un evento sonoro interesante para mí. Entonces,

original, editados o transformados y son fijados en algún soporte, como cinta magnética o tecnologías digitales

[8] Se puede escuchar y conocer más de estas obras en www.ottocastro.com [visitado el 20 setiembre de 2021].

accionaba la grabadora al inicio del evento o en el transcurso de este y la apagaba cuando, a criterio personal, percibía que había acabado. Después los datos se almacenaban en un disco duro a la espera de tener un volumen considerable de audios para empezar a componer.

Posteriormente, con las piezas musicales elaboradas durante el programa de maestría[9], la dinámica cambió: se planificó el registro de las grabaciones, no a partir de un evento sonoro determinado, sino grabando por zonas de la ciudad que tenían ciertas características. Para ello, me basé en las ideas de organización de la ciudad del urbanista estadounidense Kevin Lynch (1918-1984), orientadas a pensar la urbe a través de sendas, hitos, nodos y bordes. De este modo, en el programa de maestría, los sonidos registrados fueron sendas y nodos.

En el caso de las sendas, me interesaron las carreteras, avenidas o caminos; específicamente, algunos callejones o calles con ciertas propiedades acústicas; por ejemplo, los ecos producidos en muros paralelos o los filtros del sonido generado al transitar un vehículo a través de un túnel. Estos últimos provocan variaciones según el diámetro del espacio en cuestión, lo cual atenúa o resalta ciertas frecuencias del sonido de un motor.

Además, el interés en las características acústicas fue aumentando durante la investigación de la maestría, al conocer e implementar la metodología artística de exploración y taxonomía sonora de la Escuela de la *Cresson*[10], específicamente lo

[9] Este programa de maestría en Tecnología Musical se realizó en la Universidad Nacional Autónoma de México, UNAM.
[10] Sus siglas provienen del inglés: *Centre for Research on Sonic Space and the Urban Environment* (CRESSON), que se traduce al idioma español como Centro de Investigación sobre el Espacio Sónico y el Medio Ambiente Urbano. Está ubicado en Grenoble, Francia.

contenido en el libro *Experiencia sónica: una guía a los sonidos cotidianos*[11] (1995), de Jean-François Augoyard (1941-) y Henry Torgue (1950-). Así, en el caso de los nodos, me interesé en el efecto del corte o *cut out*. Según Augoyard y Torgue, "*este efecto es un proceso importante de articulación, entre espacios y ubicaciones; puntúa el movimiento de un ambiente a otro*" (2006, p. 29).

Entonces, el hecho de tomar como punto de partida los nodos y las sendas, junto con el estudio de cómo la arquitectura de la ciudad resalta, atenúa o filtra los sonidos que allí se producen, me llevó a comprender el sonido de la ciudad desde sus características acústicas y a alejarme cada vez más de su contenido semántico ligado a una localidad determinada.

Una vez definida la ruta de estudio, la metodología artística utilizada en la investigación de maestría se organizó en dos etapas: una general y otra específica. La general tenía por objetivo identificar los lugares, las acciones y las estrategias que se utilizarían en la grabación. Incluía también la planificación de los recorridos, la visita a los lugares y la selección de los puntos de registro, así como las grabaciones por los recorridos escogidos y la escucha general de estos registros. Por su parte, la específica se enfocó en seleccionar los materiales sonoros más funcionales para la composición (de manera manual) y determinar los procesos por seguir, desde una escucha puntual de los sonidos y sus posibilidades en la composición musical. De este modo, primero se aplicaba una escucha general y, tras varias escuchas adicionales, se catalogaban los sonidos de acuerdo con sus

[11] Su título original en inglés es *Sonic Experience: A Guide to Everyday Sounds*.

características acústicas. También, en esta etapa se incluyeron los posibles procesamientos o ediciones de los sonidos, con el fin de agruparlos por categorías.

Aunque una parte de los sonidos seleccionados para la investigación de maestría tuvieron altura definida, se prefirieron los sonidos inarmónicos o "ruidosos". Algunas categorías relacionadas con el ruido que se sugirieron en dicha investigación fueron las siguientes: ruido continuo, fluctuante, transitorio[12] o de ataque. Para el presente trabajo, no profundizaremos en cada una de estas categorías; para conocer más detalles se puede consultar el documento de tesis completo[13].

Ahora bien, para el desarrollo de los materiales sonoros que componen *Olinda, Argia* y *Ersilia*, las piezas creadas en la investigación de maestría, se priorizó el gesto que contenía el evento sonoro, pues este último se entiende como una unidad semántica completa y autónoma. Así, por ejemplo, eventos como el inicio y el final del sonido del tránsito de una moto por un callejón de la ciudad, el sonido de tres campanazos de una iglesia o el inicio y final del pregón de un vendedor ambulante, son unidades semánticas completas. En estas obras, siempre se dio prioridad al reconocimiento de la fuente; por ello, el oyente puede identificar las fuentes originales incluso después de haber pasado por las múltiples etapas de procesamiento de los sonidos. Entonces, la construcción de estas obras musicales

[12] Es una parte muy pequeña de la onda en la cual se da una alta amplitud. Esta parte de la onda contiene mucha de la información espectral y tímbrica de un sonido.

[13] Ver más en https://tesiunam.dgb.unam.mx/ [visitado el 10 de enero del 2021].

utiliza las relaciones semánticas como parte de la estrategia de composición en la vinculación de los sonidos.

Además, para favorecer el reconocimiento de las fuentes sonoras originales, los audios se fragmentaban en más de dos segundos durante el proceso de montaje. La duración de los segmentos era variada y fluctuante según las necesidades significativas y el rango siempre se estableció de forma manual.

Posteriormente, en el programa del Doctorado, mi interés sigue estando centrado en la identificación de las fuentes sonoras: el repositorio de sonidos incluye grabaciones con una duración de 20 minutos, enfocadas en diferentes entornos sonoros con el fin de buscar diversidad tímbrica. Sin embargo, la segmentación de los audios, a diferencia de la estrategia seguida en la maestría, utiliza un algoritmo que busca contrastes tímbricos en la señal, por lo cual, la duración de cada corte de audio fluctúa entre 2000 ms[14] y 250 ms, aproximadamente (este algoritmo será abordado en el capítulo dos). Contar con este repositorio de audios permitió pensar la música desde la irregularidad rítmica y aleatoria en su concatenación. Con ello, el nivel semántico pasó a segundo plano y se resaltó, en cambio, su potencial tímbrico, lo que posibilitó el abordaje desde otro enfoque estético.

Repositorios sonoros: ¿colectivos o personales?

Cuando inicié este proyecto, tuve la tentación de acceder a algunos de los repositorios colectivos que existen en la web, para librar la tarea de registrar los

[14] Se utilizará la abreviatura (ms) para indicar las milésimas de segundo.

sonidos yo mismo. Algunos de estos repositorios contienen una gran cantidad de sonidos y la mayoría vienen ordenados siguiendo una taxonomía específica determinada por el autor. Llama la atención que muchos estos repositorios no presentan información acerca de la técnica microfónica utilizada, con excepción *FreeSound*: en este, algunos sonidos muestran los datos, pero no de manera regular, porque depende de la información registrada por el usuario. Algunos de los repositorios existentes son:

- ***Dataset-AOB: urban sounds events classification***[15] [Formato wav, sample rate: 44,100 Khz[16], peso total de los archivos: 1.7 GB[17]]
- ***BCN Dataset: an Annotated Night Urban Sounds dataset***[18] [Formato: *wav*, *sample rate*: 44,100 Khz, peso total de los archivos: 6 GB]
- ***FSDKaggle2019***[19] [Formato: wav, *sample rate*: 44,100 Khz, peso total de los archivos: 26.7 GB]
- ***Freesound***[20] [Formato: *wav*, *sample rate*: 44,100 Khz, peso total de los archivos: 3 TB[21]]

---

[15] Ver más en https://zenodo.org/record/4319802#.YUAJEmZKg_U [visitado el 10 de marzo del 2021].
[16] Se utilizará para indicar los kilohercios en el ámbito de la frecuencia, la abreviatura (Khz).
[17] Gigabyte es la unidad de almacenamiento de información equivalente a 1000 000 000 (mil millones) de bytes. Se utilizará la abreviatura de (GB) para esta unidad.
[18] Ver más en https://zenodo.org/record/3956503#.YUAKjWZKg_U [visitado el 12 de agosto del 2020].
[19] Ver más en https://zenodo.org/record/3612637#.YUAPsGZKg_X [visitado el 3 de marzo del 2019].
[20] Ver más en https://freesound.org/ [visitado el 25 de agosto del 2020].
[21] Terabyte es la unidad de almacenamiento de información equivalente a 1000 000 000 000 (un billón) de *bytes*. Con el fin de indicar esta unidad se utilizará la abreviatura (TB).

Entonces, si se tiene acceso a tantos repositorios colectivos de sonidos, ¿por qué grabar de manera personal? A continuación, se muestran las razones para esta investigación en particular:

1. *Equipo tecnológico*: en todos los repositorios, el uso de uno u otro equipo en el registro de los audios depende de los recursos del investigador. Por tanto, en cada repositorio se utilizan diferentes equipos tecnológicos; es decir, distintos tipos de micrófonos, técnicas microfónicas, modelos de grabadoras y calidades de grabación. Incluso, muchas de estas grabaciones se realizan con celulares, algunas en formatos de baja calidad (por ejemplo, en 22050 KHz). Además, algunos de estos repositorios contienen una gran cantidad de sonidos en formatos comprimidos (por ejemplo, en formato mp3). El proceso de compresión que el MP3[22] utiliza para poder lograr una reducción significativa de peso, hace en muchas ocasiones, que los archivos de audio no sean óptimos de calidad, en contraposición a los formatos sin compresión, como es el caso de los formatos: WAV o AIFF.

   La producción de las tres piezas musicales de esta investigación requería calidad homogénea de los audios, no solo como requisito estético para el

[22] "La compresión con pérdida, por otro lado, implica la destrucción de algunos de los datos para lograr una reducción en el tamaño del archivo y, como resultado, se logra una gran cantidad de compresión. Las estrategias de compresión con pérdida se utilizan en archivos MP3 y JPEG, lo que explica cómo pueden tener un tamaño de archivo tan pequeño y mantener una calidad decente. La cantidad de compresión que se puede lograr al convertir un archivo WAV en un archivo MP3 es de aproximadamente 10 a 1. Por ejemplo, un archivo WAV de 30 MB se puede guardar como un archivo MP3 de 3 MB." (Griffey, 2020, p.50)

nivel de calidad deseada en las piezas, sino para los fines del análisis y comparación efectuados por el sistema informático.

2. *Riqueza de la experiencia en la exploración sonora*: un aspecto importante en el acto de componer es la relación del compositor con los sonidos que va a procesar. Esta va más allá de escoger al azar un conjunto de sonidos: su selección muchas veces implica un interés por algo particular que provoca que esos sonidos le resulten atractivos para su composición; más aún si de sonidos urbanos se trata. Los lugares en que estos sonidos se producen en la ciudad, condicionados por la arquitectura, la cultura y el espacio, hacen que la experiencia de recolección del compositor sea una etapa importante del proceso de selección. Los recorridos de exploración de sonidos en los espacios permiten que el compositor identifique características básicas de sus sonidos e incluso piense en posibles vinculaciones entre estos durante su escucha.
3. *Gestualidad en la grabación*: algunos aspectos importantes que cambian la calidad del timbre son la amplitud, la fase de una señal y la gestualidad de la grabación de un sonido en un periodo determinado. Unido a esto, de acuerdo con Merino de la Fuente, la "*evolución temporal de la intensidad es, además de la composición espectral, otro sello de identidad del timbre*" (2010, p. 29). Por lo anterior, cada grabación es una huella muy específica, no solo de la persona que la realiza, sino del lugar donde

se produce, debido a factores geográficos, atmosféricos o a la utilización de un determinado equipo de grabación y a la posición microfónica.

4. *Intencionalidad en la grabación de un sonido*: uno de los factores claves que influye en el resultado sonoro es la subjetividad del acto de la grabación, en tanto las condiciones del lugar pueden cambiar de manera dinámica y constante: un espacio no contiene los mismos sonidos siempre; su sonoridad es orgánica y cambiante. Por ejemplo, en una calle o avenida, los autos que se movilizan a veces son muchos; en otro momento son pocos; la concentración de ellos y el tipo de transporte cambia a cada instante y, por ende, el sonido se transforma también.

   Entonces, el resultado estético logrado en una grabación estará influenciado siempre por diferentes factores: la posición de los micrófonos, el movimiento o la gestualidad que se realiza con estos, la proximidad a la fuente, el espacio en el cual se está registrando y hasta factores climáticos que hacen que el registro sea único.

5. *Grabación pensando en los materiales para componer*: es relevante anotar que, durante el proceso personal de grabación, se potenciaron posibilidades de selección de sonidos a partir de resultados preliminares obtenidos de la máquina, los cuales marcaron una dirección estética y, por tanto, una orientación en la búsqueda de sonidos para registrar.

   Por ejemplo, una semana experimenté con la introducción de un descriptor acústico que reconocía frecuencias fundamentales en los sonidos que se le brindaban para analizar. Debido a esta particularidad,

durante ese periodo, los sonidos recolectados fueron de cláxones, alarmas sonoras de todo tipo, sonidos de ambulancias, entre otros. Aunque esto no descartó el interés de registrar otros sonidos, lo primordial fue recolectar los que se pretendía que reconociera la máquina. A la vez, los bocetos sonoros de composición giraron en torno a los sonidos tónicos[23] o con una frecuencia fundamental evidente. En el ejemplo sonoro 1.1 se muestra un ejercicio relacionado con el registro sonoro de materiales pensados en función de las posibilidades de selección de la máquina, con una frecuencia fundamental evidente.

Ejemplo sonoro 1.1.

Ejercicio sonoro compositivo de sonidos urbanos con frecuencia fundamental

https://ottocastro.com/o/019

Las razones mencionadas para realizar la grabación constituyen por sí mismas elementos de un proceso de composición diferente al de la música instrumental tradicional no improvisada, pues esta última conlleva una planificación previa en una partitura, mientras que en la composición con registros de sonidos urbanos, el acto de selección e identificación, los recursos de grabación y la tecnología utilizada se generan de forma dinámica durante el momento de la grabación y, por tanto, determinan el resultado de la creación en el mismo instante.

[23] Pierre Schaeffer entiende como un sonido tónico aquel en donde la masa presenta una altura tonal reconocible. Por masa comprende como el "*modo de ocupación del campo de las alturas por parte del sonido*". (Schaeffer, 1988, p.99)

De igual forma, el establecimiento de un guion acerca de lo que se graba suele suceder espontáneamente. En el instante de la grabación, no sabemos qué sonidos vamos a encontrar: será durante los recorridos por el lugar, al enfrentarnos a esos eventos sonoros cambiantes, cuando tomemos diferentes decisiones acerca de su registro.

Metodología artística en la exploración del espacio urbano

Con el fin de proponer una metodología artística que permitiera la exploración y la grabación del espacio urbano, para esta investigación se plantearon las siguientes etapas:

1. Exploración virtual
2. Selección del lugar
3. Exploración presencial del lugar
4. Implementación de la escucha

    a) Escucha general del lugar

    b) Escucha a través de la tecnología. Técnicas microfónicas empleadas y monitoreo de la grabación

    c) Escucha media en la revisión de las grabaciones

    d) Escucha específica de los sonidos

    e) Escucha reducida para el montaje

    f) Escucha general para la revisión de la obra

A continuación, se explica cada una de estas etapas (algunas se muestran de forma agrupada para facilitar la comprensión).

Preparación para la exploración

Para toda exploración es conveniente prepararse previamente, con el fin de determinar qué recursos, posibilidades y condiciones existen en el lugar al que vamos. En la presente investigación, una de las primeras fuentes sonoras establecidas para el análisis fue el sonido del flujo vehicular. Esto por cuanto la ciudad, desde mi escucha, está habitada mayormente por ese tipo de sonidos y los considero emblemáticos de la urbe. No obstante, posteriormente, este interés fue cambiando, conforme se fueron explorando otras zonas urbanas y se requirieron otras posibilidades estéticas para las piezas musicales.

Ahora bien, los sonidos del flujo vehicular siempre me han llamado la atención porque pueden ser inarmónicos o tónicos: algunos eventos, como el escape de gas a través de las muflas de diversos artefactos de transporte o los traqueteos mecánicos de las carrocerías de los autobuses tienen características inarmónicas; mientras que otros, como los cláxones o pitidos de los autos o el sonido de un frenado a gran velocidad, pueden producir tonos muy identificables en altura.

Otro aspecto que llama la atención en este tipo de sonidos son los grupos de automóviles, pues muestran un comportamiento sonoro de enjambre o yuxtaposiciones y aportan densidad. Por ejemplo, en un embotellamiento, la densidad sonora es alta, no solo por el sonido de los cláxones, sino por el de los

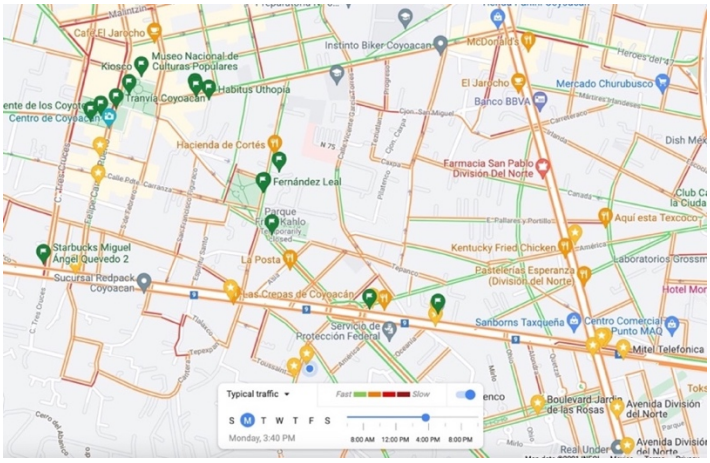
motores o el del escape de los vehículos. Esa gran diversidad siempre me ha resultado fascinante como recurso estético para la composición musical.

En esta investigación, la zona explorada fue Coyoacán, ubicada al sur de la Ciudad de México. Se eligió este lugar debido a que es la zona de residencia del autor y era primordial economizar tiempo en el proceso para alcanzar las metas establecidas.

En la exploración previa de Coyoacán, uno de los recursos utilizados para comprender el flujo vehicular fue la aplicación *Google Maps Typical Traffic*[24], la cual muestra al usuario la mayor o menor afluencia de tráfico según el día de la semana, además de las franjas horarias. Este recurso ayudó a establecer los recorridos que se realizarían para las grabaciones. El lector puede ver lo anterior en la figura 1.

Figura 1

Algunos puntos del territorio explorado en el sur de la Ciudad de México, Coyoacán



[24] Ver más en https://ottocastro.com/o/020 [visitado el 21 de agosto de 2021].

Nota: En la imagen se muestra el mapa generado por la aplicación web de *Google Maps*. En la parte inferior, se observa la franja horaria; los colores, en una escala de verde a rojo, indican la saturación de tráfico (el color verde significa pocos autos y con el color rojo, muchos vehículos). Con banderillas de color naranja y un asterisco en el centro, se indican los lugares en que se realizaron las grabaciones.

Para esta etapa de la investigación, me interesó considerar los niveles de afluencia del flujo vehicular que transita por las calles, el cual puede ser mayor o menor dependiendo de la franja horaria. Esta característica del flujo vehicular es importante debido a que los sonidos cambian según el rango de afluencia del tráfico. No obstante, es importante aclarar que las planificaciones de horario no siempre son precisas en cuanto a las fuentes sonoras que se desea encontrar. Por ello, la experiencia previa en el lugar, de manera presencial, es fundamental para tener mayor claridad sobre la dinámica de las fuentes sonoras.

Grabando mientras se camina

Los espacios de la ciudad se caracterizan por su arquitectura y el paisaje visual urbano cambia de un sitio a otro. La arquitectura de cada espacio, a su vez, condiciona la proyección del sonido. Al inicio de este proyecto, las grabaciones se realizaban en puntos fijos de la ciudad, pero poco a poco se determinó que los resultados mejoraban al moverse por la ciudad (es decir, grabando mientras se camina), pues se podía recolectar mayor diversidad de sonidos. Por ello, caminar mientras se registra se convirtió en una de las metodologías artísticas utilizadas para recolectar los sonidos en la ciudad.

Por lo general, las caminatas duraban 30 minutos y llegaron incluso a ser de 45 minutos. El criterio de tiempo establecido se definió con el fin de que el autor pudiera escuchar las grabaciones realizadas al finalizar el mismo día de la recolección, de modo que pudiera tener todavía presentes los recuerdos e ideas que le generó el proceso.

Implementación de la escucha

La escucha es un tema ampliamente abordado en diversos textos; sin embargo, acá nos referiremos exclusivamente a las etapas de la escucha humana para repasar las características principales de este tipo de percepción.

La escucha humana contiene diferentes niveles: "*una escucha discriminativa, escucha precisa, escucha estratégica, escucha crítica, escucha apreciativa*" (Opitz y Zbaracki, 2004, p. 3), entre otros; pero, además, está determinada por factores como la edad, la actitud, la religión, las emociones, la experiencia, el bagaje cultural, la educación, la familia o la condición física, entre otros (Ortiz, 2004, p. 34). Aunque todos estos aspectos intervienen en la audición, el órgano que interpreta esa escucha (incluyendo las proyecciones mentales que cada percepción nos genera) es nuestro cerebro. Al respecto, Garman y Carman afirman: "*Nosotros oímos con nuestros oídos, pero escuchamos con nuestra mente*" (1992, p. 2). Unido a esto, Torsten Wissmann establece que "*lo que se escucha siempre debe interpretarse en función del significado individual, el significado social y el contexto cultural*" (2014, p. 45).

Teniendo en cuenta todos los aspectos mencionados acerca de la escucha, así como su complejidad y particularidad en cada persona, se entiende que, para esta investigación, la escucha del compositor parte de un filtro particular y de una intencionalidad que se fue desarrollando y transformando a lo largo de todo el proyecto. A partir de esto, la escucha de los lugares elegidos para grabar fue desarrollada en diferentes etapas, según el objetivo por alcanzar**.** Así, para explorar el lugar, se utilizaron cuatro tipos de escucha, los cuales permitieron la obtención del material sonoro inicial:

1. Escucha de exploración
2. Escucha de grabación y monitoreo
3. Escucha general en la revisión de audios
4. Escucha reiterativa en el mismo lugar de la grabación inicial

De manera breve, se puede decir que el objetivo de la primera escucha es tener una idea general del espacio: una vez que se ha recorrido, se identifican puntos, lugares y áreas de la ciudad en los que se llevarán a cabo los registros sonoros.

En la segunda etapa, existe una diferencia en la percepción del sonido con respecto a la primera, causada por los dispositivos utilizados en las grabaciones. La intermediación tecnológica no solo afecta el registro obtenido, sino que está mediada por un dispositivo de reproducción con características particulares, lo que influye en la manera de recibir el estímulo por parte de quien escucha. A eso se suma un tema corporal y emotivo propio de la escucha presencial, que claramente no existe en las grabaciones. Así, en la escucha realizada en un espacio físico de

la ciudad, la percepción de los sonidos no solo es auditiva, sino que provoca otras sensaciones de vibración en nuestro cuerpo. Por ejemplo, si escuchamos un autobús pasar muy cerca de nosotros, no solo escuchamos su sonido ronco y grave con nuestros oídos, sino que esa fuente sonora resuena en todo nuestro cuerpo, agregando a la escucha información que condiciona la forma en que se percibe.

De igual manera, en la escucha posterior de las grabaciones la experiencia se transforma: por las bocinas o los audífonos el sonido es vacío, hueco; es decir, algo le falta a la experiencia que tuvimos de manera presencial. Quedan huellas fantasmales de un entorno sonoro real. Aunque parece una situación obvia, lo cierto es que la elección final de los puntos de grabación está relacionada directamente con las experiencias visuales, olfativas o corporales que suceden en el sitio. Por ejemplo, un generador eléctrico, la fricción de las llantas en superficies rugosas en una autopista a gran velocidad o el acercamiento del metrobús a la estación no se perciben de igual forma en una grabación sonora.

A partir de lo expuesto, la diferencia entre la primera, segunda y tercera etapa reside en que las dos últimas están mediadas por dispositivos de grabación y de monitorización del sonido.

En este punto, resulta conveniente mencionar algunas reflexiones del teórico Michel Chion (1947-) con respecto al fenómeno de la reducción sonora en la "*fonofijación*" (vinculado con la percepción a través de la intermediación tecnológica). Chion afirma: "*Lo que se fija no es en ningún caso el relato fiel y exhaustivo de lo que caracterizó la onda sonora emitida en el momento de la fijación*" (Chion y Steintrager, 2016, p.133). Entonces, los resultados sonoros

alojados en dispositivos electrónicos y fijados por medio de cualquier soporte (como la cinta magnética o cualquier tipo de memoria digital) sufrirán una especie de compresión: se convertirán en una representación fragmentada y lejana del sonido original y, por consiguiente, se transformarán en una experiencia sonora diferente.

Además, la fonofijación "*descontextualiza el sonido, en particular haciéndolo acusmático*"; por esto, el sonido registrado y fijado tendrá determinadas características acústicas, conservadas a través de la fonofijación. Justamente, ese "*principio de repetibilidad del sonido fijo es lo que nos permite convertirlo en un objeto y estudiarlo*" (Chion y Steintrager, 2016, p.133); es decir, la fijación posibilita conservar las características acústicas de un sonido fijado.

Así pues, si bien es cierto el sonido de la ciudad es complejo y dinámico en su entorno, al utilizar la *fonofijación* estamos tomando un segmento de sus fuentes sonoras, en el tiempo, lo que nos permite tener una muestra delimitada y filtrada de esa complejidad sonora. Precisamente, debido a esta delimitación, la máquina puede encontrar similitudes entre los timbres de los sonidos registrados y realizar sus identificaciones.

Para finalizar, la escucha reiterativa en el mismo lugar de la grabación inicial constituye la etapa número cuatro, la cual consiste en regresar al sitio donde se elaboraron las grabaciones, ahora con la experiencia previa de exploración del lugar, los datos obtenidos del análisis y habiendo reflexionado sobre los registros sonoros efectuados.

En esta etapa se produce una situación especial: debido a toda esa experiencia previa que nos influencia, tenemos más información, la cual nos da

mayor seguridad a la hora de determinar qué sonidos grabaremos en esta ocasión, cuáles serán los puntos en la zona seleccionada, qué nos gustaría registrar, cómo es el terreno en el que registraremos, qué tipo de trayectorias recorreremos con nuestros micrófonos cuando aparezca tal o cual sonido, cómo serán las fuentes sonoras y, posiblemente, a qué hora tendremos más o menos afluencia de eventos sonoros.

Así, la primera y cuarta etapa están vinculadas, por tratarse ambas de exploraciones del lugar, solo que la primera constituye una manera de dotarnos de insumos que se pondrán en acción en la cuarta etapa.

Algunos consejos y estrategias para la grabación en el lugar

Registrar en el espacio urbano requiere de ciertas precauciones previas a la grabación, las cuales se van estableciendo en la práctica mediante prueba y error, durante las sesiones que se realizan en el sitio. En este apartado comparto algunos consejos que pueden ser útiles a la hora de grabar.

Primero, cuando se realiza el registro sonoro caminando, es preciso evitar la introducción de sonidos mecánicos que puedan producirse durante el traslado. Para esto, se recomienda apartar el micrófono lo más posible del equipo de grabación, así como de la ropa, para evitar el roce. Una estrategia que utilicé en esta investigación fue colocar la grabadora al lado izquierdo del cuerpo, en un bolso, y sostener el micrófono con la mano derecha. Es decir, es importante que la grabadora y el micrófono nunca estén del mismo lado del cuerpo, con el fin de evitar los sonidos mecánicos.

Figura 2

Disposición del equipo de grabación para los registros urbanos



Como segundo consejo, siempre es recomendable usar audífonos durante la grabación, pues si se graba sin monitoreo es difícil saber qué se está registrando. Yo utilicé unos audífonos semi-cerrados *K712 PRO25*, de la marca *AKG*, que, por sus características, no solo permiten escuchar lo que se está grabando, sino también parte de lo que está sucediendo en el entorno. Esa posibilidad de doble escucha es positiva, pues se puede direccionar la grabación en caso de detectar otras fuentes que no esté captando el micrófono por su diagrama polar26 o por su posición. Otro punto importante a favor de estos audífonos es que su estructura de audífonos semi-cerrados disminuye notoriamente la fatiga auditiva.

---

[25] Ver más en https://www.akg.com/Headphones/Professional%20Headphones/K712PRO.html [visitado el 10 de mayo del 2020].

[26] Es la radiación que un determinado sistema es capaz de capturar. Cuando se habla de los diagramas polares de los micrófonos, se refiere, al patrón de sensibilidad que es capaz de captar en relación con las fuentes sonoras.

Por otro lado, es importante mencionar que las grabadoras utilizadas fueron la *H127*, de la marca *ZOOM*, y la *DR-44WL28*, de la marca *TASCAM*. La primera, la utilicé por su portabilidad y tamaño pequeño. Esta grabadora se convirtió en un dispositivo que siempre andaba conmigo; incluso, la incorporé de forma fija en la parte trasera exterior de mi mochila, en una bolsa de malla textil. Para eludir el sonido del viento y que no se filtrara directamente en los micrófonos, produciendo batimientos sonoros molestos, envolví la grabadora con una espuma y les puse un paraviento a sus micrófonos. De esta forma, cuando usaba mi mochila, iba registrando sin que las personas lo percibieran. Además, con esta técnica, el registro tenía mayor continuidad y flujo, a diferencia de cuando utilizaba la grabadora *TASCAM*, a la cual haré referencia a continuación.

A diferencia de la grabadora *H1 ZOOM*, que tiene una entrada de micrófono de 1/8" de cable telefónico, la grabadora *DR-44WL / TASCAM* cuenta con dos entradas de conector *XLR*[29], por lo que existe la posibilidad de conectar dos micrófonos monoaurales o un micrófono estéreo. En mi caso, utilicé un micrófono estéreo modelo *AT8022 Audio-Technica,* con patrón X-Y omnidireccional. Este micrófono tiene la ventaja de contar con dos pastillas microfónicas, las cuales generan dos audios monoaurales, con registros diferentes uno del otro. Lo anterior ocurre porque cada pastilla está ubicada de forma opuesta con respecto a la otra. Esto hizo posible captar sonidos y situaciones distintas simultáneamente, a la

[27] Ver más en https://www.zoom.co.jp/products/handy-recorder/h1-handy-recorder [visitado el 20 de setiembre de 2020].
[28] Ver más en https://tascam.com/amer-es/product/dr-44wl/top [visitado el 15 de mayo del 2020].
[29] Este tipo de conector eléctrico cuenta con tres pines y es utilizado tanto en audio como en video.

izquierda y a la derecha. Por ejemplo, cuando grababa caminando por una alameda angosta, sucedían eventos sonoros diferentes a ambos lados; entonces, colocaba el micrófono de manera paralela entre una fuente sonora y la otra. De la misma forma, cuando caminaba por una calle, podía encontrar, de un lado, sonidos de autos, y del otro, un grupo de personas (por ejemplo, en un café con sus sillas afuera), por lo que grababa de manera simultánea tanto los autos como el barullo de las personas y se generaban dos pistas de audio distintas.

Intencionalidad en la grabación de las fuentes sonoras

La explicación de la metodología artística utilizada pretende visibilizar el proceso del que se generan, posteriormente, los resultados. No obstante, es importante aclarar que esta es solo uno de muchos caminos posibles para alcanzar los objetivos. Además, las explicaciones sobre las formas de acercamiento son subjetivas y los resultados están permeados por mi escucha. Habida cuenta de lo anterior, se explica a continuación la manera en que se efectuaron las grabaciones en el lugar, las etapas y el método que se aplicó para realizar esos registros.

Todo entorno sonoro urbano posee una imbricada manera de mostrar sus sonidos. Algunos aparecen claramente mientras otros se presentan de manera fugaz; unos, en cambio, se superponen entre sí y producen, en algunos casos, sonidos híbridos y complejos. La multiplicación de diferentes fuentes sonoras es lo que llamamos en esta investigación *conglomerados sonoros.* Estos, por lo general, son los que más suceden en la ciudad. Desde el ámbito sonoro, entonces, la urbe es un espacio de múltiples capas, orgánico y dinámico, en donde el sonido es fluido.

Acerca de la percepción de esas múltiples capas, mencionamos, seguidamente, un ejemplo de nuestra subjetividad perceptiva: cuántas veces no nos ha pasado que vamos caminando en compañía de otra persona, esta nos indica que le llamó la atención un sonido y nos damos cuenta de que no lo hemos escuchado; o incluso escuchamos otros que esa persona no escuchó. Ese enfoque aural se replica a la hora de registrar un sonido, debido a que buscamos con nuestra subjetividad los sonidos que queremos percibir. Dirigimos, entonces nuestro equipo con el fin de lograr cierta grabación, con cierto acabado. Además, iniciamos el registro en el momento en que para nosotros es importante o particular.

Así, debido a que el acto de grabar no solo implica el proceso mecánico de hacerlo, sino que está permeado por una serie de decisiones personales que afectan lo que grabamos, podemos considerar la grabación como un acto de subjetividad fijado en el tiempo. Nuestras intenciones quedan impregnadas en la grabación como gestos que influyen en lo sonoro; de allí, que la intencionalidad sea una de las marcas importantes de la grabación en el espacio urbano.

Tomando en cuenta que en el espacio de la ciudad se presentan tantos sonidos en múltiples planos y capas, surge la pregunta: ¿qué queremos grabar, tomando en cuenta alguna perspectiva de todos esos planos en la ciudad?

Para responder a esta pregunta, recordé una cita del movimiento Dada[30]:

El 14 de abril de 1921, en París, a las tres de la tarde y bajo un diluvio torrencial, Dada fija una cita frente a la iglesia de Saint-Julien-le-Pauvre. Con

[30] Movimiento artístico que nace en 1916 en el Cabaret Voltaire en Zúrich, Suiza. Fue propuesto por el poeta Hugo Ball (1886-1927) y por el artista rumano Tristán Tzara (1896-1963).

esta acción, los dadaístas pretenden iniciar una serie de incursiones urbanas a los lugares más banales de la ciudad. (Careri, 2009, p. 68)

Esta anécdota del movimiento Dada nos hace pensar que, para los primeros reconocimientos, lo importante es dejarse llevar: no tener sumamente planificada la búsqueda o abrigar una idea preconcebida de que aparecerá una cantidad determinada de sonidos o cierto tipo de ellos, pues, incluso, es posible que tengamos una experiencia totalmente contraria a la que imaginamos.

Por lo anterior, lo primero en toda exploración es llegar al lugar y deambular, tratar de comprender las dinámicas del espacio y del ambiente, experimentar con diferentes lugares para grabar sin mucho afán de que salga perfecto; en resumen, darse la oportunidad de sentir, de percibir ese espacio, de vivirlo. Así realizamos la primera etapa de este proceso.

Una vez que se ejecutó la etapa de exploración, en las visitas posteriores de recolección de sonidos, las primeras muestras que se catalogaron fueron los sonidos registrados a lo lejos y los que pasaban de forma cercana, es decir, sonidos de fondo y sonidos frontales o cercanos.

Los sonidos de fondo se vinculan al ambiente y no tienen una configuración principal. Una de sus características primordiales es que no se puede reconocer en ellos tan claramente las fuentes sonoras; por ende, se presentan con un efecto de difuminado. Es decir, estos sonidos se perciben como si estuvieran combinados y producen una percepción de continuidad y flujo sonoro. En el siguiente ejemplo se pueden apreciar los sonidos de fondo.

Ejemplo sonoro 1.2.

Sonidos urbanos de fondo

https://ottocastro.com/o/024

Ahora bien, para los sonidos frontales o cercanos, se determinaron dos niveles:

a. Sonidos cercanos con reconocimiento de su fuente sonora
b. Sonidos extremadamente cercanos en donde se pierde el reconocimiento de su fuente sonora y se resaltan sus características acústicas.

A continuación, presentamos un ejemplo sonoro del primer grupo de sonidos cercanos.

Ejemplo sonoro 1.3.

Sonido cercano[31]

https://ottocastro.com/o/013

Como se mencionó, al iniciar las exploraciones, estuve interesado en grabar únicamente flujo vehicular, pero este objetivo fue cambiando y decidí incorporar otro tipo de sonidos que habita el espacio urbano. No obstante, para registrar el flujo vehicular, ideé dos planos de grabación para implementar en las calles, banquetas y alamedas. A estos dos planos los llamé: plano suelo y plano cintura.

[31] Para explicar y escuchar el sonido extremadamente cercano (lo que se conoce como efecto de proximidad), se recomienda leer el apartado "*Errores técnicos aprovechados como recursos estéticos*", en este mismo capítulo.

En el plano suelo, la grabadora se coloca a nivel del suelo o montada sobre algún trípode pequeño que permita un máximo de elevación de 30 centímetros. El resultado que se deseaba con este tipo de colocación era el registro de los vehículos o personas en relación con la superficie de la calle o de la banqueta; por ejemplo, el sonido de las llantas de los autos al pasar por alguna superficie rugosa o el sonido de una carretilla metálica de venta de alimentos al moverse por la calle empedrada.

Ejemplo sonoro 1.4.

Sonido de auto pasando por suelo con ondulaciones

https://ottocastro.com/o/001

También utilicé esta técnica para registros sonoros de movimiento a gran velocidad en una avenida central de la ciudad. Con esto, el resultado sonoro es más incisivo que el que se obtendría si el micrófono estuviera más alejado del suelo.

Por último, otros sonidos que deseaba grabar con esta técnica eran los que se producían cerca de superficies onduladas o con textura áspera, como tapas o cintas metálicas que cubren la calle (ver figura 3). Este tipo de superficies me llamaba la atención por el sonido que se producía cuando un vehículo, una bicicleta u otro transporte pasaba encima de ellas: al entrar en fricción, generan calidades tímbricas diferentes, las cuales van desde un sonido explosivo y muy brillante, hasta sonidos roncos, ásperos o chillidos.

Figura 3

Tapas metálicas y suelos con ondulaciones



Ahora bien, para el registro de las fuentes sonoras se utilizaron también algunas técnicas microfónicas propuestas por el compositor Barry Truax (1947-): la captación fija, móvil y de trayectoria. A continuación, se describe cada una:

(a) la perspectiva espacial fija enfatiza el flujo del tiempo o una serie discreta de perspectivas fijas;

(b) la perspectiva espacial en movimiento o viaje enfatiza un flujo de espacio/tiempo suavemente conectado;

(c) la perspectiva espacial variable enfatiza un flujo de espacio/tiempo discontinuo (Truax, 2002, p. 8).

Así, las técnicas de posicionamiento microfónico del compositor Truax se implementaron con el fin de lograr una mayor intencionalidad del material grabado

y un mejor aprovechamiento de las características de la fuente sonora y de su gestualidad en el espacio público. Por ejemplo, en la perspectiva espacial fija, el detalle y la contemplación de la percepción de las fuentes es mayor; por tanto, se utilizó esta técnica para registrar fuentes sonoras que no tenían movimiento: señales sonoras de paso, generadores de electricidad, personas conversando sentadas afuera de un café. Estas fueron registradas a una distancia de entre dos y cinco metros del dispositivo de grabación y se utilizó un micrófono colocado en un soporte.

Es importante aclarar que la fuente sonora puede ser fija, como en el caso de un generador eléctrico estático, o puede ser un sonido en movimiento en un punto de la calle en particular, pero obtenido con el micrófono posicionado sin moverse. A continuación, se muestran dos ejemplos sonoros: en el 1.5 se presenta una perspectiva fija para la grabación del sonido de un generador eléctrico, y en el 1.6 se implementa la grabación fija de una serie de sonidos, que posteriormente fueron yuxtapuestos para una sección de la pieza *Mosaico Sonoro*.

Ejemplo sonoro 1.5.

Perspectiva sonora fija con sonido de generador eléctrico

https://ottocastro.com/o/139

Ejemplo sonoro 1.6.

Utilización del plano suelo en la pieza Mosaico Sonoro (1’38” a 2’09”)

https://ottocastro.com/o/140

No obstante, la perspectiva espacial en movimiento fue la que más utilicé al moverme por la ciudad y grabar mientras caminaba: en cada recorrido, trataba de ir componiendo mientras grababa, de contar una historia y relacionarla con las fuentes sonoras que encontraba a mi paso. A diferencia de la técnica fija, esta es más compleja en lo que respecta a resolver la grabación de los eventos sonoros simultáneos en el momento mismo de su utilización, debido a que en este proceso suele improvisarse constantemente para determinar qué se graba y cómo se enfocan las fuentes sonoras con el movimiento.

Ejemplo sonoro 1.7.

Perspectiva sonora de movimiento con sonido de generador eléctrico

https://ottocastro.com/o/003

Ejemplo sonoro 1.8.

Perspectiva sonora móvil en la obra Anomalía Urbana (6'26" a 6'35")

https://ottocastro.com/o/141

Por último, la perspectiva espacial variable la utilicé poco: únicamente en caso de que deseara darle cierto protagonismo a alguna fuente sonora en movimiento. En esta técnica, el micrófono se mueve junto con la fuente sonora, es decir, la sigue en su recorrido. El resultado de la grabación es que la fuente sonora se registra en primer plano y el entorno sonoro se percibe en segundo plano y en movimiento. Por ejemplo, se puede seguir un carrito pequeño, de los que se utilizan

para llevar verduras, jugando con la relación entre el micrófono y la fuente. Con esta técnica, es posible enfocarse en diferentes puntos de la fuente sonora: se puede registrar las llantas del carrito que pasan por el pavimento o los sonidos que se producen por el movimiento de la estructura metálica y su contenido. También se podría enfocar en las conversaciones de la persona que maneja el carrito de verduras, si es que va acompañada, o en el sonido de su calzado, si lleva algún tipo de zapato con suela dura.

Todos estos sonidos se pueden captar acercándose o alejándose de la fuente, girando el micrófono, realizando movimientos rápidos de arriba-abajo o viceversa, o jugando con las velocidades de esos movimientos. Con esta técnica se puede explorar la fuente sonora de una forma más dinámica y, por tanto, las vinculaciones de los sonidos, con los movimientos del micrófono en el espacio y la fuente pueden lograr grabaciones muy particulares.

Ejemplo sonoro 1.9.

Perspectiva espacial variable con sonido de campana

https://ottocastro.com/o/004

Ahora bien, es importante señalar que el resultado sonoro de las grabaciones que se realizaron utilizando las técnicas citadas no es evidente en las composiciones musicales con las que finaliza esta investigación. Esto se debe a los diferentes tipos de procesamiento y segmentación que los sonidos grabados reciben cuando entran al sistema informático. Sin embargo, la implementación de

estas técnicas previas en el registro del paisaje sonoro produce una exploración particular que favorece la variedad tímbrica. Por lo tanto, el repositorio se vio enriquecido por la multiplicidad de posibilidades tímbricas aportadas desde la grabación, gracias su implementación.

Errores técnicos aprovechados como recursos estéticos

En la mayoría de los registros de audio realizados, se produjeron eventos que podríamos considerar como “errores técnicos” en la práctica común de la grabación de audio. Sin embargo, estos, en lugar de ser tomados como errores, se aprovecharon como un recurso estético importante que pasó a formar parte del repositorio de los sonidos que se analizarían a través del sistema informático.

De igual forma, otros sonidos provocados por este supuesto “error técnico” (como chillidos, pitidos o retroalimentaciones sonoras) fueron aprovechados, por ejemplo, como un movimiento de tensión dramática en la composición musical, cuando se deseaba una coloración sonora densa. En muchas ocasiones, estos sonidos se utilizaron de forma breve, para resaltar alguna pulsación de la pieza musical, pero también se procesaron con un estiramiento en el tiempo, con el fin de usarlos a manera de pedal o como sonidos largos con función armónica de tono grave dentro de la obra.

A continuación, se mencionan algunos de los errores técnicos:

Error de proximidad: este se genera al colocar el micrófono muy cerca de la fuente sonora, pues se producen zumbidos o ataques pronunciados, con frecuencias graves muy resaltadas que provocan que el sonido se perciba

distorsionado y muy alejado del reconocimiento de la fuente sonora original. Este tipo de resultados sonoros se utilizó en secciones de la pieza musical en donde se requería una marcada coloración o un reforzamiento de la tensión narrativa.

En el siguiente ejemplo, las fuentes sonoras están más cercanas al micrófono y producen cierto nivel de distorsión, no solo por su amplitud, sino debido a la pérdida de percepción de su contexto general y a una apreciación muy detallada de su materia sonora; por ejemplo, el paso de un automóvil que, por su cercanía, se percibe como un zumbido.

Ejemplo sonoro 1.10.

Sonido por proximidad

https://ottocastro.com/o/014

Ejemplo sonoro 1.11.

Sonidos por proximidad utilizados en la pieza Anomalía Urbana (5'31" a 5'39")

https://ottocastro.com/o/077

Registro fuera del ámbito de la grabación: este puede producirse cuando se graba en un ambiente muy silencioso, pero existe alguna fuente sonora detrás del eje o ámbito de captación del micrófono según su diagrama polar[32]. El resultado sonoro en esta circunstancia, desde mi percepción, es una especie de retumbo que

[32] Para conocer más acerca de los diagramas polares y los tipos de micrófonos se puede revisar el siguiente enlace: https://www.audio-technica.com/es-ar/support/una-breve-guia-de-microfonos-que-es-el-patron/ [visitado el 10 de setiembre de 2019].

no se precisa con claridad y que, en algunas ocasiones, se convierte en una imagen sonora que pareciera tener un velo encima; es decir, se produce un difuminado en la ubicación de su fuente sonora en el espacio.

Durante mi experimentación con el sonido, utilicé está técnica en el registro del entorno, ayudándome con la arquitectura: en lugar de dirigir el micrófono hacia la fuente sonora directa (por ejemplo, el sonido de un auto en movimiento) lo orienté en dirección contraria a la fuente (en este caso, una pared lisa o cóncava) para aprovechar la resonancia o el filtraje de dicho sonido. Esto con el objetivo de atenuar o resaltar ciertas frecuencias de la fuente sonora original.

Figura 4

Tres diferentes colocaciones del micrófono para la captación del sonido urbano aprovechando la arquitectura



Segmentación de los materiales sonoros a través de la máquina

La segmentación a través de la máquina es un proceso de partición aplicado a un continuo sonoro, cuyos criterios de segmentación son realizados según el método escogido para que la máquina realice el corte. Esta acción de separar un fragmento de otro extrayendo un momento determinado de un archivo de audio está vinculada a la composición, forma parte del acto mismo de componer, debido a que esos fragmentos después se convertirán en unidades que podrán yuxtaponerse, unirse o articularse con otros segmentos para crear texturas sonoras particulares que irán conformando la pieza musical. Por ello, desde el ámbito de la composición musical, la segmentación puede entenderse como los "*elementos individuales para la construcción de ideas musicales, considerando su funcionalidad y utilidad dentro de un contexto [...]"* (Sigal, 2014, p. 82).

Por otro lado, a nivel de la máquina, la segmentación de los audios afectará la identificación de sonidos similares en su proceso final, debido a que los datos extractados pueden cambiar dependiendo de la longitud de los segmentos que se comparen. Sin embargo, existen varios métodos para identificar el inicio y final de un evento sonoro. Al respecto, Juan Pablo Bello et al. afirman:

> La forma habitual de detectar los inicios es buscar regiones “transitorias” en la señal, una noción que conduce a muchas definiciones: una explosión repentina de energía, un cambio en el espectro a corto plazo de la señal o en las propiedades estadísticas, etc. (2005, p. 1)

Ahora bien, en el proceso de segmentación que realiza la máquina, “*existe una gran complejidad para saber con precisión la efectividad de la detección en un inicio*

del evento sonoro" (Dixon, 2006, p. 133). El resultado son aproximaciones de corte, las cuales están mediadas por métodos de análisis de la señal realizados por la máquina. En estas aproximaciones, muchas veces existe un nivel de imprecisión con respecto a lo que, como humanos, desearíamos recibir como resultado o a cómo imaginaríamos que pudiera ser cada segmento resultante. Sin embargo, esto, más que ser una deficiencia, nos enfrenta a una diferencia entre nuestra escucha y la de la máquina.

Así, existen tres criterios de segmentación a través de los cuales la máquina puede llevar a cabo los cortes de audio:

a. por identificación de picos de amplitud, considerando un umbral de sensibilidad;
b. por contraste de la fase de la onda;
c. por cambios en el espectro.

Al respecto, el investigador Hui Li Tan, en su artículo *Audio Onset Detection using Energy-based and Pitch-based Processing,* del año 2010, señala cuatro métodos de reconocimiento de inicio de un evento:

Tabla 2
Métodos de detección de eventos a través de transitorios

#	Método
1	Detección de cambios en la energía temporal y/o espectral
2	Detección de cambios en la fase espectral
3	Detección de cambios en el tono detectado
4	Técnicas de apredizaje automático

(Tan, L. H. et alt. ISCAS 2010, 2010 IEEE International Symposium on Circuits and Systems Bio Circuit Fabrics andSystems, May 30th-June 2nd, 2010, Paris, France, 2010)

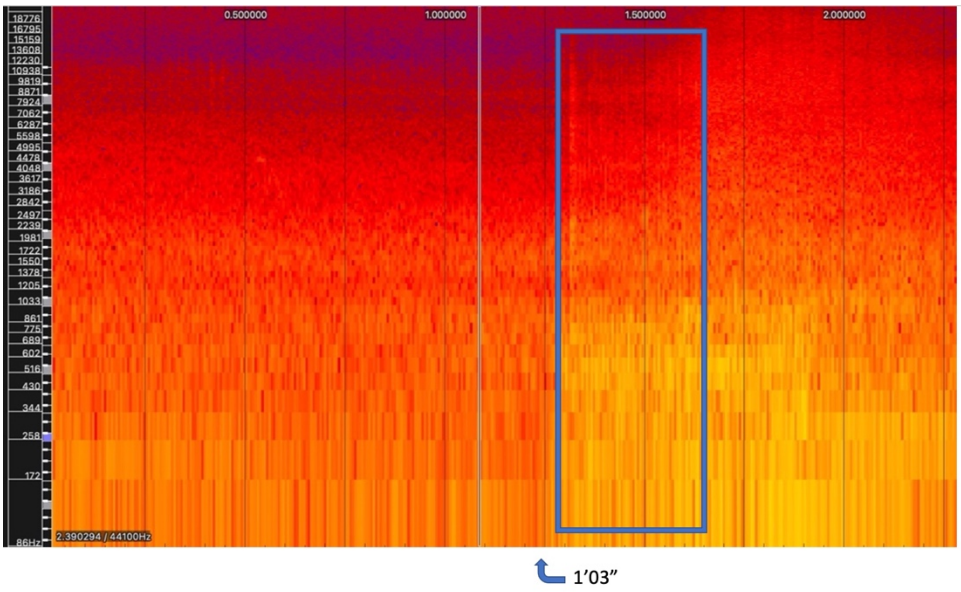
Para efectos de este trabajo, se decidió utilizar un algoritmo que fuera sensible a los cambios tímbricos más relevantes. Esto con el fin de obtener una gran cantidad de fragmentos sonoros, con los cuales el sistema informático pudiera seleccionar grupos de sonidos con características tímbricas similares.

Una de las motivaciones que me llevó a escoger este tipo de segmentación se vincula a mi experiencia en la composición de obras electroacústicas mediante el uso del sonograma. Cuando se compone, es usual que se busquen timbres adaptables al desarrollo sonoro de la pieza que se está construyendo, sobre todo en el momento del montaje. En esta tarea, la escucha ayuda a encontrar sonidos apropiados. Posteriormente, puede usarse el sonograma como un complemento al análisis, que a través de la representación de la energía espectral facilita encontrar patrones en el sonido, pues por lo general, los cambios tímbricos importantes se ven a través del cambio del color: colores intensos o brillantes indican mayor intensidad de la energía en determinada frecuencia. Esas características crean patrones que pueden cambiar abruptamente, lo que indica un cambio tímbrico contrastante.

En la figura 5, el sonograma muestra un cambio en la energía espectral después del segundo 1’03” en adelante. Este cambio se refleja en el color amarillo, que se vuelve más intenso en las frecuencias que van de los 172 Hz a los 1205 Hz. Podemos comprobar auditivamente este cambio en el ejemplo sonoro 1.12.

Figura 5

Ejemplo del cambio en la energía espectral a partir del tiempo 1’03”.



Nota: El cambio tímbrico se indica visualmente en la figura por el rectángulo azul en la figura. El eje X corresponde al rango de frecuencias (Hz) en escala logarítmica[33], y el eje Y representa el tiempo en segundos (s)[34].

Ejemplo sonoro 1.12.

https://ottocastro.com/o/012

Los cambios tímbricos pueden visualizarse en el sonograma de muy diversas maneras, algunas muy evidentes y otras no tanto. Esto se debe a que la distribución de la energía espectral en un sonograma tiene un límite en las dimensiones que puede

[33] "Un logarítmo es una función matemática que reduce los valores numéricos grandes en números menores. Los números logarítmicos aumentan exponencialmente de maneja semejante a como percibimos el volumen de un sonido, por ejemplo, 1,2,4,16, 128, 256, etc. (Huber et al., 2011, p.59)

[34] En adelante la indicación de la frecuencia se realizará en Hertz (Hz) y el tiempo en segundos (s).

representar, lo que imposibilita la visualización del comportamiento de la energía espectral de un sonido.

Ahora bien, además de esta experiencia empírica sobre el uso del sonograma, uno de los descriptores acústicos que permite extractar información espectral sensible a los cambios tímbricos es el *Spectral Flux*[35]. Para esta investigación, se buscó una librería de código que permitiera la segmentación desde los contrastes tímbricos y no solo por contraste de amplitud o de fase. Una de las librerías que mejor cumplía con este objetivo fue *Aubio*[36], la cual contiene el comando *Specflux,* que está compuesto por el descriptor acústico *Spectral Flux*. En la página oficial *Aubio,* se establece la utilidad del descriptor *Spectral Flux*:

> *El Spectral Flux se utiliza para medir la rapidez con que se realiza un cambio en el espectro de potencia en una señal. Es ampliamente utilizado para detectar el cambio de un timbre a otro.* (*Aubio*, 2021)

Como vemos, la función de este descriptor es identificar cambios en el espectro: marca un inicio y un fin de un evento cuando percibe un cambio espectral relevante. Además, utiliza un nivel de umbral con el fin de tener mayor o menor sensibilidad a los cambios espectrales analizados en un audio.

---

[35] Se conoce en español como flujo espectral.
[36] Más información en https://aubio.org/ [visitado el 15 de abril de 2019].

Por otro lado, en la librería de código *Aubio,* también se encuentra la opción de *Aubiocut,*[37] la cual incluye varios algoritmos:[38] *HFC, Complex, Phase, Specdiff, Specflux*.

Estos sirven para realizar diferentes cortes de un audio. En esta investigación se utilizó principalmente *Specflux*, el cual está basado en el descriptor acústico *Spectral Flux.*

En el capítulo segundo se explicará con detalle el funcionamiento de la herramienta *CSpectra2021*. Sin embargo, es conveniente mencionar desde ahora, que todas las librerías y líneas de comando son llamadas para ejecutarse desde el entorno de *JavaScript*[39].

Un ejemplo de una línea de comando, para ejecutar el módulo de *Aubiocut,* es el que se muestra seguidamente, este se ejecuta con el fin de llevar a cabo las segmentaciones de audio.

---

[37] https://aubio.org/manpages/latest/aubiocut.1.html [visitado el 20 de abril de 2019].

[38] Algunos de los algoritmos de *Aubio* son: 1. HFC: son las abreviaturas de *High-Frecuency Content*. El método calcula el contenido de alta frecuencia (HFC) de la entrada de la señal. La función resultante es eficiente para detectar inicios de pulsos de instrumentos de percusión. 2. *Complex*: conocida como función de detección del dominio complejo. Esta utiliza información tanto en frecuencia como en la fase de la señal para determinar cambios en el contenido espectral y así identificar posibles inicios. Es apropiada para música polifónica. 3. *Phase*: traducido al español como fase. Su función es detectar los cambios de fase. 4.*Specdiff*: detección de inicio a través de diferencia espectral.

[39] *JavaScript* es un lenguaje de programación interpretado a través del compilador de una página web, además es imperativo, lo que quiere decir que ejecuta una línea después de otra en forma consecutiva. Es un lenguaje sensible a la letras mayúsculas y minúsculas, de tipado débil, es decir, que el valor de las variables puede cambiar a través del tiempo. Por último, *JavaScript* es un lenguaje orientado a objetos.

A continuación, se muestra un fragmento de su implementación dentro del código[40].

```
function cortar_audio(filename) {
  console.log("entra a aubiocut" + filename);
  var baseroot = __basedir + "/resources/static/assets/uploads/.";
  var fileinput = baseroot + filename + ".wav. ";
  var diroutput = baseroot + filename + "/";
  fs.mkdirSync(diroutput);
  var comando = "aubiocut" + fileinput + " -c -O specflux -t 0.6 -o" + diroutput
```

Acerca de la línea de comando "aubiocut" + fileinput + " -c -O specflux -t 0.6 -o" + diroutput se puede decir acerca del significado de cada comando que:

aubiocut = nombre del programa a ejecutar

archivo01.wav = nombre del archivo

-c = cortar el archivo de entrada

-o = método de detección de inicio que se utilizará

-t = umbral de inicio (Threshold)[41]

El último comando, -t, corresponde a *Threshold* o umbral, el cual influye en la variable que se le asigne cuando la máquina realice la segmentación. Esta

---

[40] El código completo del *CSpectra2021* puede verse en el apéndice digital del repositorio TESIUNAM. Más información en: https://tesiunam.dgb.unam.mx/ [Visitado el 6 de noviembre del 2020]

[41] El valor del umbral suele estar en el rango [0,001 a 0,900]. Los valores de umbral más bajos implican más apariciones detectadas. El aumento de este umbral debería reducir el número de detecciones incorrectas. El valor predeterminado por defecto es 0.3. (Aubio, 2021)

variable establece la sensibilidad en el umbral, con un tiempo que va entre 0,1” y 0,9”.

Una vez establecida la línea de comando, se realizaron algunas pruebas (ver figura 6), pero al final se optó por el umbral de tiempo de 0,6”, pues este mostraba los eventos sonoros de la manera en que se quería que estuvieran dispuestos en las texturas sonoras. La decisión se tomó también debido a que, con segmentos de entre 10 ms y 100 ms, que determinaban eventos sonoros muy cortos, se perdía completamente el origen de la fuente; por el contrario, con segmentos de entre 2000 ms y 5000 ms, que constituían extractos muy largos, el nivel semántico de las fuentes era tan evidente que diluía la percepción de las vinculaciones tímbricas entre los sonidos.

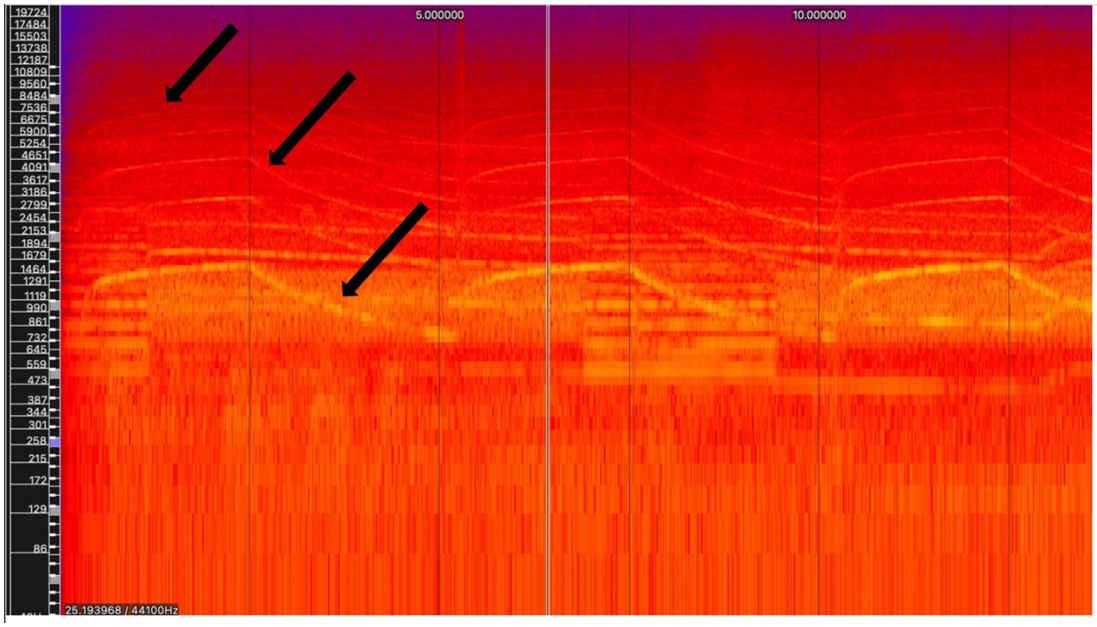
Así, se decidió mantener un umbral de tiempo de 0.6”, como el que vemos en en donde la segmentación va entre 250 ms y 1500 ms. Empíricamente, desde mi escucha, noté que este era un buen valor de corte y funcionó bien para realizar las yuxtaposiciones, uniones y montajes necesarios para que las texturas sonoras se escucharan fluidas, así como para lograr que, a lo largo de varios audios, se pudieran escuchar con claridad las relaciones tímbricas buscadas por la máquina.

Ahora bien, es importante resaltar que los cortes realizados a través del algoritmo por contraste tímbrico no son del todo precisos. En el ejemplo de la figura 6, podemos observar en el sonograma la representación del sonido de una ambulancia. Además, algunos sonidos pueden presentar dificultades de reconocimiento del inicio de un evento; por ejemplo, sonidos como los *glissandos* producidos por el roce abrupto de los neumáticos en el asfalto o los zumbidos de

frecuencias graves de motores de autos presentan esta dificultad. Esta condición se da debido a los cambios en la modulación de frecuencia que afectan el timbre del sonido original.

Figura 6

Modulación de las frecuencias en el sonido de una ambulancia.



Nota: El eje X corresponde al rango de frecuencias (Hz) en escala logarítmica, y el eje Y representa el tiempo en segundos. Las flechas indican las frecuencias que modulan en el tiempo.

Ejemplo sonoro 1.13.

https://ottocastro.com/o/126

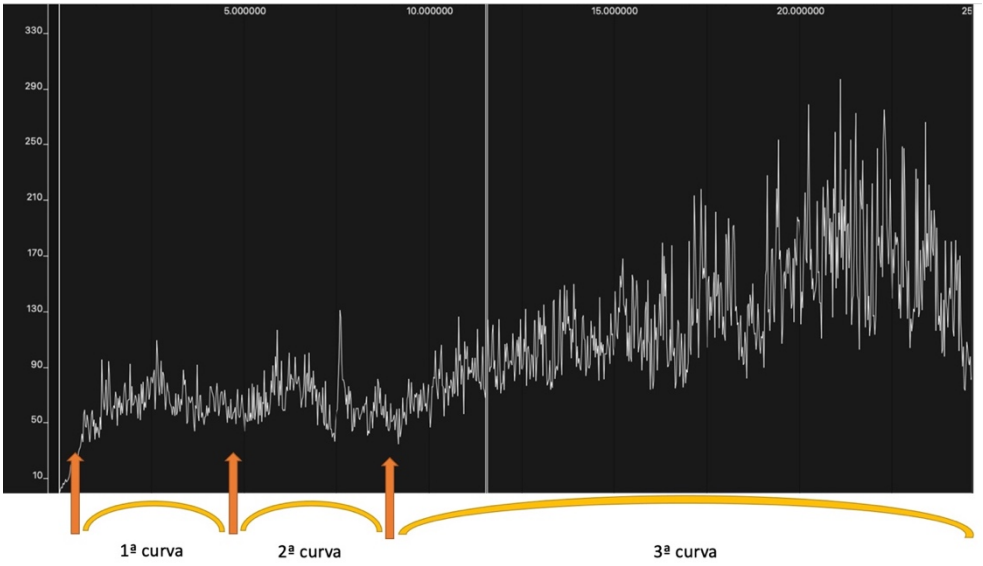
La figura 7 muestra la representación del análisis del descriptor acústico *Spectral Flux*. En este ejemplo, existe un mismo sonido repetido que establece un patrón en los primeros diez segundos del fragmento, el cual está representado por

dos curvas que coinciden con la modulación del sonido de la ambulancia. Sin embargo, después del segundo diez, se observa una gran curva que engloba las otras dos modulaciones. La representación de este análisis es extraña, porque el patrón de sonido es el mismo a lo largo de todo el evento; es decir, debería repetirse sucesivamente el comportamiento de la representación que se da en el tracto del segundo 0 al 5, pero no ocurre así. Esto demuestra que algunos eventos sonoros, como las modulaciones de frecuencias y otros sonidos con características similares, no son identificados por el descriptor acústico.

En la figura 7, el eje Y indica la magnitud de la energía espectral representada por *Spectral Flux*. Es usual encontrar en las representaciones gráficas de los análisis una línea vertical que divide la figura por la mitad, sin embargo, esta no tiene una significación particular.

Figura 7

Análisis de una señal a través del descriptor acústico Spectral Flux



Nota: El eje X corresponde a los valores de 0 a 350 en la escala de *Spectral Flux* que brinda el descriptor para esta muestra, y donde el Y representa el tiempo en segundos.

En esta investigación, se realizaron varios intentos de segmentación con otras librerías de código y algoritmos, antes de utilizar *Aubiocut*; entre ellos *SBIC*[42], un algoritmo de la librería Essentia. En la página oficial de esta librería se afirma acerca del algoritmo *SBIC*:

> Este algoritmo segmenta el audio utilizando el *Criterio de Información Bayesiano,* dada una matriz de características de la trama. El algoritmo busca segmentos homogéneos para los cuales los vectores de características tienen la misma distribución de probabilidad, según la implementación en estos. (Essentia, 2021)

El algoritmo *SBIC* busca el mejor inicio de un transitorio a través de probabilidades estadísticas. Sus segmentos resultantes tienen la particularidad de iniciar en silencio o atacante. Sin embargo, cuando estos fragmentos se colocan en el montaje para conformar las texturas sonoras, se evidencia la heterogeneidad entre los sonidos. Eso no sucede con la segmentación del algoritmo *aubiocut,* con la opción de *specflux*. Por consiguiente, estos resultados inciden en la estética de la composición musical obtenida a través de los diferentes procesos algorítmicos. En el ejemplo sonoro 1.14, se muestra un fragmento de la pieza *Anomalía Urbana*

[42] Ver más información en https://essentia.upf.edu/reference/streaming_SBic.html [Visitado el 10 de noviembre del 2019]

(1’21” a 1’24”), donde se escucha el tipo de segmentación que produce el algoritmo *SBIC*.

Ejemplo sonoro 1.14.

https://ottocastro.com/o/142

En *Anomalía Urbana*, muchos de los segmentos se pegaron uno tras otro. En pocas ocasiones, se utilizó el *crossfade* o fundido cruzado entre los sonidos que conformaron la textura sonora. Debido a lo anterior, en esta pieza el acabado es más tosco y resulta más evidente el recurso de la herramienta que presenta esta característica de corte.

Por su parte, en la pieza *Texturas Urbanas*, además de *aubiocut*, se utiliza otra librería de código llamada *SoX* [43]. Esta provee una utilidad de línea de comandos con el fin de fusionar audios entre sí cuando se pegan en el *multitrack* de *Reaper*. Gracias a este recurso tecnológico, las texturas sonoras se perciben más suaves y continuas que en la pieza: *Anomalía Urbana*. Esto puede apreciarse en el siguiente ejemplo sonoro.

Ejemplo sonoro 1.15.

https://ottocastro.com/o/014

Ahora bien, ninguna de las dos opciones presentada por los ejemplos sonoros anteriores, es mejor que la otra. Desde mi experiencia en el ámbito compositivo, en la opción utilizada en *Anomalía Urbana* (ejemplo 1.14), se produce

[43] Ver más información en http://sox.sourceforge.net/ [Visitado el 10 de agosto de 2019]

una rítmica compleja entre los segmentos. Esta rítmica hace más evidente el límite de la herramienta. Mientras que en la segunda opción (ejemplo 1.15), se logra homogeneidad en las uniones de los segmentos, que aunque estos son de fuentes diversas, el resultado estético constituye un continuo sonoro. Este puede resultar útil para un enfoque armónico, más que rítmico.

A partir de lo anterior, se concluye que la diferencia entre la segmentación humana y la de la máquina brinda otras posibilidades creativas y estéticas para el compositor, que pueden aprovecharse de varias maneras:

a. Podría ser una posibilidad de confrontación de la escucha del compositor, que le permite una revisión crítica de su escucha, al no coincidir esta con lo que la máquina descubre. Esto le permite al compositor preguntarse ¿cuál es el límite de su escucha con respecto a la escucha de la máquina?, y ¿qué escucha este que la máquina no puede oír y viceversa?, preguntas relevantes para el proceso de composición musical.

b. Los diferentes momentos en que la máquina presenta algún factor de imprecisión para identificar un inicio y fin de un evento sonoro pueden potenciar otras posibilidades estéticas no contempladas por el compositor, más que convertirse en un obstáculo.

El uso de diferentes tipos de algoritmos de identificación de eventos sonoros como parte de la segmentación, a partir de sus diferencias, puede transformarse en una manera de provocar diversos resultados estéticos en la construcción de las texturas sonoras.

Reflexión acerca de la escucha humana y la escucha de máquina

La escucha humana y la de máquina son muy contrastantes entre sí. La virtud de su unión radicará en la complementación que puede hacer una con la otra para crear una especie de híbrido que beneficia al compositor en su toma de decisiones. Por ejemplo, en la escucha de máquina[44], la velocidad de cómputo y la meticulosidad, precisión y repetición de tareas para las que se le programó es muy eficiente. En cambio, la escucha humana tiene la capacidad de integrar, flexibilizar y adaptar los resultados en una escucha más global y de imprimir su perspectiva subjetiva en sus decisiones de selección.

Tabla 3

Comparación entre la escucha de máquina y escucha humana

Escucha Humana	Escucha de Máquina
Muy variable	Variable hacia tareas específicas
Subjetiva	Realiza las tareas para las que se les programó
A partir del contexto cultural, social, psicológico y fisiológico	Contexto algorítmico, matemático
Encuentra patrones por una habilidad desarrollada desde la sobrevivencia de la especie	Encuentra patrones por procesos estadísticos
Jerarquiza por percepción a partir de reconocimiento de patrones de múltiples orígenes	Jerarquiza por patrones estadísticos

Para este proyecto, la naturaleza complementaria de las dos escuchas es muy importante: el compositor escucha subjetivamente en el momento en que graba sonido, le da sentido a un contexto caótico y filtra lo que más le interesa. A su vez, esta audición humana es tamizada por un instrumento de grabación, el cual, de

[44] Este tipo de escucha está vinculada al reconocimiento de patrones y su clasificación, utilizando para ello métodos automáticos computacionales. La *escucha de máquina o listening machine* se ha incrementado después de los años noventa gracias al poder de las computadoras.

alguna manera, constituye otra escucha derivada de la máquina, que le aporta un material sonoro que deberá volver a escuchar y dar sentido. Así, el compositor planifica los elementos que necesita para que la máquina (el sistema informático) escuche y sea una herramienta óptima para trabajar datos. Posteriormente a esta escucha, es muy importante nuevamente la escucha del compositor, quien es el que va a dar sentido, pensando en el oyente y en su propia experiencia; un sentido que pasa por su formación profesional y su aprendizaje social.

Ahora bien, con respecto a la escucha humana, la alta variabilidad, la poca atención durante largos periodos y el apoyo en experiencias previas atravesadas por aspectos psicológicos, sociales o culturales la hacen una recepción altamente subjetiva. Además, este tipo de escucha cambia de persona en persona, pues se modifica según su formación, el campo de implementación y la tecnología que utilice, porque no es lo mismo componer pensando en que los sonidos serán montados en una estación de trabajo de audio digital, que pensar en esos mismos sonidos desde el código sin proceso de fijación digital.

En cambio, la escucha de máquina dispone de búsquedas muy específicas, (para eso fue programada). Puede realizar una tarea repetitiva por horas sin agotarse, recurre a su contexto matemático y algorítmico para realizar sus predicciones y modelos y jerarquiza mediante sus análisis estadísticos. No obstante, la escucha de máquina puede considerarse una extensión a través de la cual el compositor puede reflexionar y descolocar su propia escucha al contrastarla con lo que le provee la escucha de máquina. Por tanto, retomar los resultados dejados por la máquina y configurarlos nuevamente por medio de una escucha humana intuitiva

es una manera de acercar al oyente a una forma de escucha a la cual está habituado.

En mi trabajo, la escucha de máquina se ha convertido en un par que me permite oír de otra manera, confrontar mi propia escucha y ratificar o criticar las certezas o debilidades de mi sistema auditivo. Coincido con las afirmaciones de Jeremy Malasky, en su texto "*Toma de decisiones colaborativas entre hombre-máquina en un sistema de optimización complejo*":

> El objetivo de este enfoque es combinar la intuición y la experiencia de un ser humano con la velocidad computacional de un sistema informático para generar soluciones de manera oportuna que mejoren las generadas únicamente por un humano o una computadora. (2005, p. 15)

Así, los métodos utilizados por las máquinas nos ayudan a optimizar y realizar varias tareas en un mismo instante, situación que es difícil de ejecutar desde la escucha humana. Sin embargo, esta última aventaja a las máquinas en su capacidad de adaptarse a las situaciones que escucha, incorporando diferentes niveles de intuición y tomando decisiones a partir de sus experiencias previas.

Toda esta reflexión acerca de la alternancia de dos escuchas se refleja en el proceso de composición de este proyecto de doctorado, en donde media el *CSpectral2021*, el planeamiento del discurso musical y la experiencia creativa del compositor.

Programas informáticos de referencia cercanos al *CSpectra2021*

Contexto general

CSpectra2021 es una herramienta tecnológica creada para este proyecto. Se enfoca en encontrar similitudes espectrales mediante el análisis, la identificación y la comparación de los sonidos que se introducen al sistema, para generar texturas sonoras que se convierten en posibles materiales de la composición musical.

Con el fin de lograr que esta herramienta tecnológica cumpliera con las necesidades estéticas de la investigación, se analizaron varios programas informáticos similares a lo que se deseaba conseguir con el *CSpectra2021*; algunos de estos fueron: *Corpus-based Concatenative Synthesis (CataRT System)*[45], *Musical Mosaicing*[46], *Orchids*[47], *Flucoma*[48], *EarGram*[49], *AudioGuide*[50] y *Sononym*[51]. Se determinó que las propuestas tecnológicas existentes pueden agruparse en dos tipos:

a. Los programas que analizan la señal extractando sus características acústicas y que muestran al usuario la información encontrada.

---

[45] *Corpus-based concatenative Synthesis* (CataRT System) fue creado en Francia, en el año 2006, por Diego Schwarz.
[46] *Musical Mosaicing* fue creado en Francia, en el año 2011, por Aymeric Zils y François Pachet.
[47] *Orchids* fue creado en Francia, en el año 2009, por Carmine-Emanuele Cella y Yan Maresz. Se puede encontrar más información en http://www.orch-idea.org/ [visitado el 12 de setiembre del 2019].
[48] *FluCoMA* fue creado en Inglaterra, en el año 2001, por Pierre Alexandre Tremblay, Owen Green, Gerar Roma, Alex Harker, Jacob Hart y Frédéric Dufeu. Se puede encontrar más información en https://www.flucoma.org/ [visitado el 21 de abril del 2020].
[49] *EarGram* fue creado en Portugal, en el año 2014, por Gilberto Bernardes. Se puede encontrar más información en https://sites.google.com/site/eargram/ [visitado el 1 de mayo del 2019].
[50] *AudioGuide* fue creado en Inglaterra en el año 2010 por Ben HackGarth, Norbert Schnell, Diemo Schwarz, Philippe Esling, Arshia Cont y Gilbert Nouno. Se puede encontrar más información en http://www.benhackbarth.com/audioGuide/docs_v1.41.html [visitado el 3 de junio del 2018].
[51] *Sononym* fue creado en Alemania en el año 2019 por Eduard Müller, Bjørn Næsby Nielsen e Ingolf Wagner. Se puede encontrar más información en https://www.sononym.net/ [visitado el 9 de marzo del 2018].

b. Los programas que ejecutan yuxtaposiciones o combinaciones de audios aprovechando la información acústica que se obtuvo del análisis.

Como ejemplo del primer grupo, puede mencionarse el programa informático *Sonic Visualizer,*[52] el cual muestra al usuario el análisis de las características acústicas de un sonido mediante gráficos, colores o anotaciones en su interfaz. Inclusive, algunos programas más recientes, como *Sononym*, presentan los grados de cercanía en aspectos muy puntuales de las características acústicas, tales como niveles de intensidad, brillo o armonicidad, entre otros.

Por otro lado, en el segundo grupo destacan los siguientes programas: *CataRT System*, *Musical Mosaicing*, *Orchids, Flucoma*, AudioGuide, *EarGram* y *CSpectra2021*. Estos, además de lo mencionado, extraen características acústicas determinadas por el primer grupo.

De acuerdo con Owen Green *et al*. los programas informáticos del segundo grupo presentan las siguientes características:

1. Posibilidades de descomponer y recomponer sonidos en nuevos híbridos. Estos procesos están vinculados con la elaboración del corpus[53].
2. Técnicas y tácticas para encontrar el camino a través de colecciones de sonidos.

---

[52] *Sonic Visualizer* fue creado en Inglaterra por Chris Cannam en el año 2005. Se puede encontrar más información en https://www.sonicvisualiser.org/ [visitado el 11 de agosto del 2017].
[53] El corpus se entiende desde la teoría de la síntesis concatenativa, es decir, el repositorio de sonidos desde donde el sistema analiza sus sonidos.

3. Producto de la combinación de las dos anteriores, tienen el potencial para desarrollar montajes sónicos, trayectorias o interpolaciones (2019, p.1).

Todos los programas informáticos mencionados no podrían ser comparados entre ellos utilizando pruebas de resultados que visibilicen relaciones o patrones de procesamiento, aún utilizando un conjunto determinado de sonidos de entrada para todos. Estas comparaciones serían complejas de realizar pues cada uno de los programas ejecuta tareas diferentes. Por ejemplo, *Orchids* analiza un segmento de sonido de entrada y, mediante la extracción de sus características acústicas, propone una transcripción con sonidos instrumentales y técnicas extendidas, con el fin de que el resultado instrumental sea similar al sonido de entrada. En el caso del programa *EarGram*, este, a partir de un repositorio de sonidos, analiza sus posibles vinculaciones, manteniendo el tempo y el ritmo del segmento original, además de generar posibilidades de mezclas o *mashups*[54] finales. Por último, *Flucoma* analiza la señal de un conjunto de sonidos con el fin de concatenarlos mediante la sustitución de segmentos de audio; para ello, aplica diferentes técnicas de resíntesis, es decir, reconstruye una síntesis de sonido a partir de un análisis inicial.

Debido a las diferencias entre los programas, para analizarlos, describirlos y compararlos se tomó como criterio el acercamiento a la ruta de usuario y se tomaron

[54] El *mashup* es una técnica compositiva que mezcla dos o más canciones, superponiendo la melodía de una canción sobre la otra para crear una nueva canción. Se puede considerar un subgénero de la remezcla o *remix*. También se le conoce como *blend, mesh o bootleg*. La manera más tradicional de aplicar un *mashup* es colocar la línea vocal de una canción encima de otra con su acompañamiento instrumental. Este tipo de técnica nació en los años noventa y pone en jaque los conceptos de autor, copia y obra original, así como también abre un debate sobre los derechos de autor.

como referencia los programas con las tareas más cercanas al desempeño deseado para el *CSpectra2021*.

Así, mediante la comparación de los programas, fue posible imaginar las necesidades fundamentales de ejecución que debería tener la herramienta tecnológica *CSpectra2021*:

1. Configurar un conjunto de descriptores acústicos[55] existentes en una librería de código para el análisis acústico.
2. Analizar con facilidad repositorio sonidos de 10 a 30 GB[56].
3. Realizar tareas como recolección de datos mediante el análisis acústico de sonidos, almacenamiento de esa información, gestión de los datos recogidos y comparación de estos entre sí.
4. Encontrar similitudes entre los sonidos en el ámbito espectral.
5. Crear conjuntos de sonidos similares, como resultado del análisis acústico.
6. Generar texturas sonoras para disponer de posibilidades compositivas.

[55] Es un dispositivo creado para extractar información acústica de una señal. Existe en la actualidad una serie de librerías que ya se pueden instalar como *frameworks*, tales como: *Librosa, Essentia, Aubio, LibXtract, Marsyas, Meyda, Mir Toolbos, Timbre Toolbox, YAAFE,* entre otras. Los descriptores acústicos ayudan a brindar información con el fin de establecer medidas de similitud, síntesis estadística, síntesis basada en funciones, minería de datos, clasificación de datos y procesamiento de señal.

[56] El gigabyte es una unidad de almacenamiento de información cuyo símbolo es el GB.

Gracias a esta exploración, se determinó que los dos programas más cercanos a las características deseadas para la herramienta tecnológica diseñada en esta investigación son *AudioGuide* y *Sononym*.

Aspectos generales de los programas seleccionados: AudioGuide y Sononym en relación con CSpectra2021

A continuación, se analizan los programas *AudioGuide* y *Sononym* con el objetivo de visibilizar las características generales relacionadas con la manera en que se despliega la información, sus funciones y su accesibilidad. Esta revisión permitirá determinar de qué forma se podría desarrollar un prototipo con características similares, pero con funciones tecnológicas específicas, ausentes en los otros dos programas.[57]

A continuación, se describe el uso de dichos programas:

Sononym posee una interfaz gráfica que le permite al usuario interactuar con las posibilidades del programa y accionar desde ahí diferentes procesamientos de análisis de la señal, a través de niveles como intensidad, sonoridad o *loudness*, frecuencia, tempo en pulsos por minuto, brillo espectral, armonicidad espectral y niveles de inarmonicidad. Además, muestra otros valores básicos de información general del audio analizado, como longitud del segmento, rango de sampleo o *sample rate*, profundidad de *bit* o *bit depth* y los canales de cada segmento de audio analizado.

---

[57] Algunas de estas particularidades requeridas en el *CSpectra2021* se citaron en el apartado de contexto general.

Ahora bien, los parámetros intensidad, frecuencia y *tempo* en pulsos por minuto se muestran como valores numéricos, mientras que los relacionados con el timbre (brillo, armonicidad e inarmonicidad), se presentan en forma de gráficos deslizantes o *faders*, que indican de forma muy general los niveles de dichos parámetros. El control que el usuario tiene de todas estas posibilidades en una sola pantalla es posible debido a que los parámetros son muy básicos.

El objetivo general del programa *Sononym* es mostrar al usuario niveles de similaridad entre los sonidos. Estos niveles se indican en números romanos, que van del I al IV, donde I es el más similar y IV, el nivel más lejano de similaridad. El usuario no dispone de la información sobre cuáles o cuántos descriptores acústicos realizarán el trabajo de similitud o de qué manera podría enfocar ese nivel de similitud; esto por cuanto *Sononym* cuenta con un conjunto específico de descriptores acústicos que ejecutan su trabajo tal y como se le programó de fábrica y solo arroja un resultado general en el análisis que realiza al usuario. Los niveles de similaridad los obtiene, muy posiblemente, a través de los tres valores predominantes en el análisis sonoro: brillo, armonicidad e inarmonicidad.

En contraposición, *AudioGuide* envía sus acciones o procesos desde la consola de la terminal[58] o interfaz no gráfica. Esto es funcional, debido a que el programa ofrece múltiples posibilidades de ejecutar comandos, además de que cada algoritmo puede modificarse con el fin de disponer de diferentes parámetros

[58] La terminal es la interfaz que permite la comunicación con la computadora, desde una pantalla que muestra solo caracteres alfanuméricos. Se trata de un contenedor de comandos del intérprete. Existen varios tipos de lenguajes de comandos que son utilizados por la terminal, tales como: *Bash*, *CSH*, *ZSH*. La terminal contiene un arreglo de comandos de línea para invocar funciones, que se ejecutan en la computadora.

para el análisis de la señal. Este programa también especifica qué tipo de descriptores acústicos pueden implementarse en el análisis, así como la estadística computacional que se desarrolle. Si todas estas posibilidades se presentaran en una pantalla, como lo hace *Sononym,* la interfaz gráfica se vería saturada, a menos que pudiera disponerse de múltiples pantallas; pero esta opción también complicaría su uso.

Así, como puede extraerse de la descripción anterior, estos dos programas realizan actividades cercanas pero distintas, más allá de la manera en que presentan la información al usuario. Por un lado, *Sononym* dispone de información para el usuario acerca del análisis espectral y la posible cercanía tímbrica de los sonidos analizados. *AudioGuide,* en cambio, no solo analiza las características acústicas de cada grupo de sonidos e identifica los niveles de similitud tímbrica, sino que sustituye cada segmento del sonido referente según su análisis tímbrico, utilizando su repositorio de sonidos, lo que permite crear una textura sonora.

En resumen, se puede concluir que el *CSpectra2021* presenta la información de forma similar a la de *AudioGuide* y tiene similitudes en la forma en que ofrece múltiples posibilidades para modificar los parámetros de análisis por parte del usuario.

Además, *CSpectra2021* sugiere posibilidades de color tímbrico, a diferencia de *Sononym,* que únicamente identifica los sonidos según ciertas características acústicas. También difiere de *AudioGuide* en que sustituye segmentos de sonido por otros, según su análisis tímbrico. Por tanto, *CSpectra2021* favorece el trabajo estratégico de composición musical gracias a las posibilidades que brinda para

hacer yuxtaposiciones vinculando categorías tímbricas como brillo, opaco, rugoso y liso, lo que contribuye al proceso de creación.

Otro punto importante de *CSpectra2021* es que aprovecha el poder de una variedad de librerías de código que son ejecutadas en la herramienta tecnológica desde el entorno de programación de *JavaScript*. Así, *Essentia*[59] le ayuda a realizar los análisis acústicos, *Annoy*[60] mide las distancias entre los resultados de las características acústicas y *Aubiocut*[61] se utiliza para tareas de procesamiento: de amplitud, de frecuencia o de identificación de inicios y finales de un evento sonoro, entre otros.

Además, en lo referente a la búsqueda de sonidos similares, este programa brinda al usuario un resultado en dos presentaciones: como conjunto de sonidos separados o como yuxtaposiciones, a manera de textura sonora, a diferencia de los otros programas, que solamente ofrecen un tipo de resultado.

Ahora bien, con respecto a las texturas sonoras que obtiene a través de yuxtaposiciones, *CSpectra2021* tiene la particularidad de aplicar rampas de la amplitud de inicio y final[62] de cada evento sonoro, con lo cual sugiere posibles trayectorias sonoras en cada iteración

---

[59] Essentia es una librería de código abierto, la cual se implementa con *JavaScript* o lenguaje *Python*. El paquete de algoritmos que contiene y la gran versatilidad de estos la colocan como una librería apta para desarrollar el área de Recuperación de Información Musical o (*MIR, Music Information Retrieval*). Ver más en https://essentia.upf.edu/ [visitado el 12 de marzo del 2018].

[60] *Annoy*: *Approximate Nearest Neighbors* es una librería de código construida en el lenguaje informático C++ para implementarse en el lenguaje *Python*. Es utilizada para implementar medidas de similitud. Esta librería es utilizada por la plataforma *Spotify* para encontrar géneros musicales similares.

[61] Ver más en https://aubio.org/manpages/latest/aubiocut.1.html [visitado el 3 de agosto del 2018].

[62] Conocido en el idioma inglés como *fade in* y *fade out*.

Unido a esto, el proceso automático de yuxtaposición de texturas sonoras es una buena herramienta para desarrollar diferentes estrategias compositivas vinculadas con factores de densidad. Además, provee factores de color, en el ámbito de orquestación acusmática, cuyo concepto se explicará en el capítulo segundo.

Posibles utilidades de los programas Sononym y AudioGuide como asistentes a la composición musical

Sononym es un programa dedicado a identificar y sugerir posibles taxonomías a partir de los sonidos analizados. En sus análisis, este programa no genera resultados sonoros que puedan incorporarse directamente como posibles nuevos sonidos en el montaje. Por tanto, es funcional para brindar resultados analíticos generales, dejando de lado posibles trayectorias, recorridos o texturas sonoras. En el caso de AudioGuide, el proceso de análisis lo utiliza para saber cuáles son los sonidos similares de un repositorio a partir de un sonido de referencia. Este programa constituye una herramienta de asistencia a la composición musical. Debido a esa característica, es utilizado por diferentes creadores sonoros en varios enlaces y publicaciones. A continuación, se muestran diferentes obras en las que se ha utilizado AudioGuide como una herramienta digital la asistencia a la composición:

1. ***Six synthèses*** *(2020)* de Yan Maresz (1966-). Esta obra utiliza *AudioGuide* para conformar sonidos nuevos mediante la unión de fragmentos de materiales sonoros. La estética resultante es cercana a la generada por la síntesis granular.

Ejemplo sonoro 1.16.

https://ottocastro.com/o/042

2. ***Sinking Air*** (2015) de Fred Szymanski (1956-). En esta pieza el autor crea gestos de forma libre a través de la electrónica. Mediante la aplicación del programa *AudioGuide*, los datos son procesados y entregados como un material resultante que responde a los gestos anteriores. Así se obtiene una especie de contrapunto de timbres que se suceden unos a otros, de modo que en ese diálogo se olvida cuáles son los gestos originales y cuáles los creados con el programa.

   Ejemplo sonoro 1.17.

   https://ottocastro.com/o/041

3. ***Tensus*** (2014), de Fred Szymanski. Esta pieza utilizó *AudioGuide* para reordenar una serie de sonidos de cuerda, tales como arco, raspados, golpeados y frotados. Con ello se obtuvieron conglomerados sonoros diversos.

   Ejemplo sonoro 1.18.

   https://ottocastro.com/o/040

4. ***Fanfare*** (2019) de Davor Branimir Vincze (1983-). Obra comisionada para la Bienal de Música de Zagreb, Croacia, del 2019. Es una pieza para 6 instrumentos de bronce y 6 bocinas. Utiliza la síntesis concatenativa para reorganizar diferentes segmentos de canciones de música tradicional popular de Zagreb, en diálogo con instrumentos de metal. De este modo, se establece un mosaico sonoro entre los gestos que propone el compositor y los generados por el programa.

Ejemplo sonoro 1.19.

https://ottocastro.com/o/039

5. ***L'entaille*** (2019) de Grégoire Lorieux (1976-). Esta pieza toma como referente un repositorio de sonidos de cantos de pájaros, utilizando la síntesis concatenativa que provee *AudioGuide,* y los sustituye por sonidos instrumentales diversos, con lo cual crea una textura musical que sugiere de manera lejana los referentes originales.

Ejemplo sonoro 1.20.

https://ottocastro.com/o/038

6. ***Au-dessus du carrelage de givre*** (2019) de Louis Goldford (1983-). Un trabajo sonoro para voz tenor, electrónica en tiempo real y video. En la pieza, la voz del cantante aparece sola por muchos momentos,

pero en algunos lugares es transformada por *AudioGuide*, generando sonidos nuevos a través de la síntesis concatenativa, que, al comparar su tímbrica, sustituye los sonidos originales por instrumentales.

Ejemplo sonoro 1.21.

https://ottocastro.com/o/037

En conclusión, si bien los programas *Sononym* y *AudioGuide* constituyen una herramienta para la asistencia de la composición musical y cuentan con diversas funciones para obtener resultados distintos, sus alcances son específicos y limitados, a diferencia del programa *CSpectra2021,* que tiene una gama más amplia de posibilidades y mayor capacidad de procesamiento de sonidos.

Comparación de algunos aspectos requeridos en las dependencias de los programas

De los tres programas analizados, incluyendo el *CSpectra2021*, el más sencillo de instalar es *Sononym,* debido a que proporciona un archivo ejecutable correspondiente para cada plataforma en la que se instale, ya sea *OSX*, *Windows* o *Ubuntu*. Este ejecutable instala todas las librerías y dependencias que se requieren de manera automática.

En el caso de *AudioGuide* y *CSpectra2021,* se debe instalar primero una serie de dependencias y librerías, además del lenguaje *Python,* para su funcionamiento.

Para *AudioGuide,* debe instalarse una librería llamada *NumPy*[63], la cual se implementa con el lenguaje *Python*. *NumPy* está dedicada al soporte de funciones matemáticas de alto nivel, tales como matrices multidimensionales y otros procesos matemáticos. *AudioGuide* utiliza el programa *Csound*[64] para llevar a cabo sus procesamientos de audio. *Este* es un programa escrito en el lenguaje *C++* y está construido para ejecutar procesamientos de señal y síntesis de audio.

Por otro lado, en el caso del *CSpectra2021*, se requiere instalar dentro del entorno de programación de *JavaScript*, las siguientes librerías:

Essentia Freesound Extractor:[65] "*Es un extractor de características de línea de comandos configurable para análisis de sonido. Calcula un gran conjunto de descriptores de sonido, que incluyen varias características espectrales, de dominio del tiempo, rítmicas y tonales*". (Essentia, 2021)

Annoy[66] es una librería construida en el lenguaje informático C++ para ejecutarse en el lenguaje *Python*. Se usa para implementar medidas de similitud y crear modelos de clasificación. Además, tanto *CSpectra2021* como *AudioGuide* requieren instalar un archivo binario[67] que extrae características acústicas de los sonidos analizados.

---

[63] Librería creada por Travis Oliphant en el año de 2006. Puede obtenerse más información en https://numpy.org/doc/stable/user/whatisnumpy.html [visitado el 10 de noviembre del 2019].
[64] Es un programa libre enfocado en el procesamiento de audio. Está escrito en el lenguaje C. Para más información ver https://csound.com/ [visitado el 6 de noviembre del 2018].
[65] Puede consultarse más información en https://essentia.upf.edu/freesound_extractor.html [visitado el 11 de abril del 2019].
[66] Sus siglas provienen del idioma inglés: Approximate Nearest Neighbors (ANNOY), que se traduce como: Aproximadamente vecinos más cercanos. Más información en https://github.com/spotify/annoy [visitado el 3 de marzo del 2019].
[67] Este se entiende como un archivo informático que contiene información de cualquier tipo, codificada en formato binario. En este texto se refiere a archivos ejecutables que le indican al sistema que realice una acción.

Datos de entrada para los sistemas informáticos

Existen tres intenciones con las que pueden introducirse los datos en estos programas analizados anteriormente:

1. Introducir un conjunto de sonidos a un programa que realiza un análisis y devuelve información correspondiente a ese análisis.
2. Incorporar un sonido que puede ser referente de otros sonidos. Tanto el sonido referente como el conjunto de sonidos son introducidos al sistema para llevar a cabo su análisis y la comparación entre ellos.
3. Incluir un sonido que es analizado y sustituido por otros sonidos incorporados con anterioridad al sistema informático.

En el caso de los programas *Sononym* y *CSpectra2021,* los audios de entrada son analizados por medio de datos numéricos. El sistema informático entrega un resultado que cataloga los sonidos y busca relaciones entre ellos.

De forma básica, en *Sononym,* los sonidos pasan por un proceso de análisis, que se muestra al usuario como resultado final, de manera numérica y gráfica. El programa permite relacionar los datos obtenidos con el fin de presentar niveles de similitud entre los sonidos a partir del análisis arrojado por el sistema. Sin embargo, los niveles de similitud son tan generales que dan una idea difusa de la cercanía tímbrica de estos.

Tanto en *Sononym* como en *CSpectra2021*, la extracción de características acústicas puede entenderse como un proceso común. No obstante, *AudioGuide* agrega un paso más: sustituye diferentes segmentos, que toma del repositorio de

sonidos analizados, para emular el sonido de referencia; es decir, crea un mosaico sonoro de todos los segmentos similares al sonido de referencia.

Finalmente, en lo que respecta a la entrada de datos a los programas, *AudioGuide* y *CSpectra2021* ejecutan sus acciones a través de una serie de instrucciones de comando, que se ejecutan desde la consola de la terminal, con el fin de generar diferentes procesos de análisis. Por su parte, en *Sononym*, la entrada de datos es recibida por una interfaz visual que brinda acciones básicas de procesamiento de los sonidos por parte del usuario.

Datos de salida de los sistemas informáticos

En la etapa de salida, existen dos tipos de datos que son devueltos por los programas informáticos:

1. Como resultado final del procesamiento de análisis, datos útiles para comprender los niveles de similitud e información acerca de los sonidos que introdujo el usuario.
2. Datos que le sirven al sistema informático para llevar a cabo procesos a corto plazo o para analizar eventos futuros.

Ahora bien, existen varios formatos disponibles para almacenar los datos resultantes. Algunos de los más populares en la actualidad son:

json[68] (por sus siglas en inglés: *JavaScript Object Notation*). Es un formato de archivo para almacenar e intercambiar datos. El formato *JSON* consta de

[68] Ver más en https://www.json.org/json-en.html [visitado el 12 de mayo del 2020].

dos tipos de elementos estructurales: valores tipo diccionario y lista ordenada de valores. Los tipos de datos que acepta son número, cadena o *string*, valores *booleanos*, arreglos numéricos y objetos.

yamls[69]*.* Creado por Clark Evans en el año 2001. Los datos en este formato pueden representarse como listas o matrices. Este formato organiza jerárquicamente los datos de la siguiente manera: primero, un conjunto de nombres o valores y, segundo, un conjunto ordenado de valores.

csv[70]*.* Sus siglas provienen del idioma inglés: *Comma-separated values,* que puede traducirse como "valores separados por comas". Es un archivo de texto que utiliza la coma como separador de campo. Se usa como un formato de intercambio de datos.

Los archivos con formatos: json, yamls y cvs, son utilizados por *Sononym*, *AudioGuide* y *CSpectra2021* para sus procesos internos. En el caso de *Sononym,* el programa genera, tras su análisis, un archivo en formato .db. Por lo general, almacena información de un modo estructurado en una tabla. Los archivos .db en este programa, se almacenan en un formato de datos manejados y leídos por el programa de software libre, *SQLite*. Este tipo de información en formato .db no es accesible para el usuario desde *Sononym*, sino que es una base de datos utiliza a lo interno del programa informático.

---

[69] Ver más en https://yaml.org/spec/history/2001-12-10.html [visitado el 21 de agosto del 2019].
[70] Ver más en https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml [visitado el 06 de febrero del 2019].

AudioGuide genera archivos en formato .json para sus análisis; no utiliza archivos .db para sus bases de datos. Por su parte, *CSpectra2021* genera archivos en formato .json como resultado de su análisis de las características espectrales de los sonidos y a través del sistema de base de datos *NoSQL*[71] *como es MongoDB,* que es un; es decir, que en vez de generar datos de manera estructurada utiliza bases de datos no relacionales y con un esquema dinámico, con el objetivo de que la búsqueda sea más rápida. Al utilizar este recurso tecnológico, genera archivos en formato .db.[72]

Así, de la comparación de los tres programas se puede concluir que el factor de la eficiencia en la manipulación de los datos es un aspecto importante al implementar las funciones para obtener los datos resultantes. También, la generación de este tipo de datos estará relacionada con el procesamiento específico que cada programa haga de los datos de entrada, así como con el objetivo general de este.

Datos de control en los sistemas informáticos

Los tres programas ofrecen un ámbito de control de funciones con respecto a las búsquedas[73] que se desean realizar. En el programa *Sononym,* las búsquedas por filtro se establecen en cada parámetro analizado y se presentan al usuario en

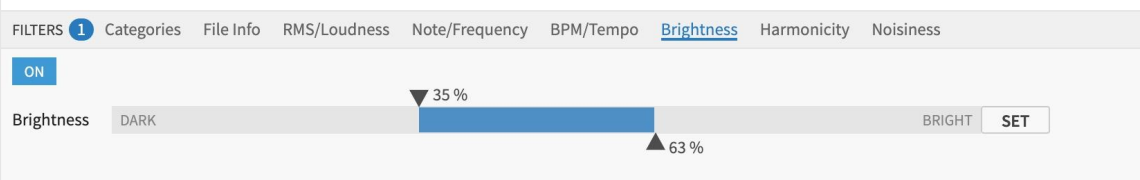
---

[71] NoSQL se refiere a que los sistemas de administración y gestión de bases de datos no utilizan el lenguaje SQL.
[72] Ver más en https://docs.mongodb.com/manual/tutorial/getting-started/ [visitado el 12 de agosto del 2019].
[73] En la literatura de tecnología, este término se encuentra como *queries,* en idioma inglés.

una escala de 0 a 100%. El usuario define la franja general que desea buscar en los sonidos. No obstante, el rango que ofrece es muy general.

Figura 8

Presentación gráfica para el filtro de búsqueda ofrecido por el programa Sononym



En *AudioGuide,* las búsquedas se realizan desde la interfaz no gráfica o terminal a la base de datos*,* con lo cual ejecutan diferentes funciones de acceso a la información. En *CSpectra2021* se puede especificar el tipo de descriptor o el conjunto de descriptores acústicos por utilizar y la clase de estadística computacional; en cambio, en *AudioGuide* se puede solicitar el porcentaje de uso del descriptor acústico o método estadístico.

Además, el nivel de precisión en el filtro de búsqueda de sonidos similares varía dependiendo del objetivo final que tiene el programa informático. Por ejemplo, en *Sononym*, uno de los objetivos principales es visibilizar, en la pantalla de usuario, una posible taxonomía de los sonidos analizados. En contraposición, en *AudioGuide,* el objetivo es encontrar niveles de sustitución tímbrica en los sonidos de su repositorio con respecto al sonido referente. Debido a esto, el usuario debe saber anticipadamente lo que desea obtener en esas sustituciones y qué parámetros acústicos desea comparar, con el fin de que la máquina pueda llevarlos a cabo. Por su parte, en el programa *CSpectra2021,* la similitud de los sonidos no

solo se logra a través del tipo de descriptores acústicos, sino que depende de las características espectro-morfológicas del sonido de referencia.

Datos almacenados por el sistema informático

Existen dos tipos de datos que generan estos programas:

a. Datos que sirven al sistema informático para realizar procesos de análisis interno y que son mostrados en diferentes formatos, como base de datos .db o intercambio de datos en formatos .json o .yaml.
b. Datos que se presentan como resultado final de un proceso de análisis y comparación producido por la máquina, tales como conjuntos de sonidos agrupados para el montaje de texturas sonoras, sonidos separados o colecciones de sonidos identificados como similares, en formato .wav o .aiff.

Con respecto a los datos almacenados, *Sononym* arroja únicamente un archivo de datos permanentes para la creación, en la base de datos, en formato .db. Mientras, el *CSpectra2021* incorpora diferentes archivos con extensiones .json y .yamls*,* que convierten la información contenida en formato .sig y este, a su vez, puede ser convertido a una base de datos en formato .db.

Por otro lado, el programa *AudioGuide* utiliza archivos en formato .json*. Este* almacena los diferentes datos del análisis de la señal. Genera también un archivo en formato .txt que contiene los valores de segmentación del sonido referente y de los sonidos que seleccionó como similares. En la tabla 4, puede apreciarse un fragmento de la información contenida en el archivo de texto generado por el

programa *AudioGuide*. En este se muestran cuatro columnas: la primera y segunda, en sentido de izquierda a derecha, son el inicio y el final de la segmentación del audio. La tercera columna es el ataque de la envolvente que se aplica a cada sonido y se activa con el umbral o *threshold,* según los decibeles (*dB*) que indique el sistema informático.

Tabla 4

Información de archivo generado por AudioGuide de segmento un fichero de audio con una duración de 2'15" segundos

Inicio	Final	Umbral en decibeles
0.14336	0.256	-5.02
0.27648	0.43008	-31.83
0.4608	0.5632	-32.87
0.57344	0.6656	-36.72
0.67584	0.89088	-4.28
1.05472	1.15712	-11.68
1.16736	1.30048	-12.02
1.31072	1.5872	-34.84

Datos almacenados para el usuario

Los datos brindados al usuario para su utilización inmediata completan la etapa final de la obtención de datos generados por el sistema informático. Los

programas *AudioGuide* y *CSpectra2021* generan archivos de audio únicamente en formato .wav[74] o .aiff[75]. Existen tres maneras de presentar estos archivos al usuario:

1. Conjunto de sonidos comparados por el sistema informático en relación con las características acústicas de un sonido referente que fue introducido por el usuario a la máquina.
2. Lista de texto o información gráfica que visibiliza los audios similares tímbricamente, relacionados con el sonido referente (como sucede en el programa *Sononym)*.
3. Yuxtaposición de los sonidos a manera de una textura sonora, presentados en los programas *CSpectra2021* y *AudioGuide*.

Ahora bien, la interpretación que el usuario le pueda dar a esos audios estará vinculada a la forma en que el programa los exhiba. En este sentido, *Sononym* muestra algunas características acústicas y nivel de similitud, a través de su interfaz visual, utilizando gráficos que marcan diferentes niveles de magnitud y colores. Este tipo de exhibición de los datos es apropiada para un usuario con intereses y necesidades en la información más general.

Otros datos generados por los programas

Algunos programas no solo generan datos necesarios para el sistema informático o para el usuario; también es posible generar otro tipo de datos que

[74] La extensión .wav es un tipo de formato con o sin compresión creado por Microsoft e IBM para almacenar datos digitales de audio. Sus siglas provienen del inglés *Waveform Audio Format* o formato de onda de audio.
[75] La extensión .aiff fue creada por Apple en 1988. Al igual que el formato .wav, este tiene la finalidad de almacenar datos digitales en formato de audio sin compresión.

podrían cargarse en otros programas que utilicen las mismas extensiones de los archivos obtenidos. Estos pueden ser analizados o tomados como archivos de control para generar datos en otros programas. Es el caso del programa *AudioGuide,* el cual genera archivos tanto para el programa MAX/MSP, como para *Csound*.

En el programa Max/MSP, *AudioGuide* genera archivos en formato .json, que sirven para analizar y graficar la secuencia de los eventos. El archivo que genera el programa Csound en formato (.csd) funciona para producir texturas sonoras, a través de segmentaciones generadas por el control de un archivo de texto en formato .txt.

A modo de conclusión, podríamos decir que existe una gran oferta de programas que realizan análisis de señal con el fin de brindar al usuario información acerca de las características acústicas del sonido. Sin embargo, muy pocos de estos programas disponen de información pertinente sobre los niveles de similitud tímbrica. Además, estos programas no utilizan su análisis para generar relaciones entre los materiales con similitudes y muchos no realizan una concatenación de los materiales sonoros para el uso inmediato en la composición musical. Por estas razones, el *CSpectra2021* se convierte en una herramienta útil en la asistencia del proceso de creación, pues brinda posibilidades para entender la composición de otra manera desde el ámbito del análisis espectral: búsqueda de similitud, resultados sonoros individuales, agrupamiento de sonidos separados y las texturas sonoras que genera.

Capítulo 2:

Identificación de sonidos mediante la máquina a través de su comparación espectral

El objetivo principal de este trabajo de investigación es generar estrategias de composición musical partiendo del análisis de comparación espectral entre diversos sonidos. Los procesos de análisis de los sonidos, extracción de características acústicas, recolección de datos de esos análisis y comparación de los arreglos numéricos arrojados por el sistema informático, son llevados a cabo por la herramienta tecnológica *CSpectra2021* (Herramienta de Análisis y Comparación Espectral). Esta herramienta digital es creada en colaboración y supervisión

con el Dr. Hugo Solís y tiene como objetivo facilitar el proceso de selección, identificación y creación de las yuxtaposiciones sonoras con los sonidos resultantes del análisis.

Por otro lado, es importante aclarar que este capítulo solamente abordará el proceso tecnológico. Los aspectos estéticos y artísticos implementados en la composición se desarrollarán posteriormente en el capítulo tercero.

El timbre y el espectro como ámbitos de partida

El sonido tiene cuatro características: altura, intensidad, duración y timbre[76]; de ellas, esta última resulta la más difícil de definir. Hasta el momento, no se ha logrado establecer una convención de lo que es el timbre, debido a sus múltiples variables propias, que además son modificadas continuamente por la influencia de las tres cualidades restantes del sonido.

Sirvan como ejemplo estas dos definiciones: para la *American National*

[76] El timbre "*depende del grado de complejidad del movimiento vibratorio que origina dicho sonido. Esta cualidad no es mesurable, ni existe por tanto una unidad que permita comparar timbres de distintos sonidos*" (Calvo-Manzano Ruiz, 2004, p.122).

Standard: "*el timbre es ese atributo de la sensación auditiva en términos del cual un sujeto puede juzgar entre dos sonidos similares que tengan diferente sonoridad y altura*" (1973, p. 56). Por su parte, Pratt y Doak definen el timbre como "*el atributo de la sensación auditiva mediante el cual un oyente puede juzgar que dos sonidos son diferentes a partir de cualquier criterio que no sea el tono, el volumen o la duración*". (1976, p.36).

La problemática de evaluar el timbre llevó a Wei Jiang et al. a afirmar: "*La naturaleza subjetiva del timbre complica el proceso de evaluación, que generalmente se basa en evaluaciones subjetivas, procesamiento de señales y análisis estadístico*" (2020, p.1). Sin embargo, como contraparte puede haber dos formas importantes con las que se puede aproximar al estudio del timbre, las cuales son: por un lado, la psicoacústica, que estudia la respuesta psicológica de un sujeto ante los "*estímulos acústicos con sensaciones auditivas*". (Roederer, 2008, p.9), y por otro lado, la acústica que "estudia el sonido, incluyendo la producción, transmisión y sus efectos" (Pierce, 2019, p.1).

Sin embargo, aunque no es posible establecer una definición unívoca que abarque o delimite todos los aspectos del timbre, debido a su complejidad es muy difícil implementar un método que logre mesurar y comparar de manera precisa distintos timbres, como sucede con ámbitos como la altura o la intensidad, lo que si es posible es: "obtener el llamado espectro acústico, consistente en la obtención de la distribución de sus armónicos o sobretonos en frecuencias e intensidades" (Calvo-Manzano Ruiz, 2004, p.131).

Un acercamiento a comprender el amplio y profundo terreno del timbre es

mediante una representación cuantificable. Es por esto por lo que algunos métodos de composición contemporánea, como la música espectral, han optado por la extracción de datos a través del análisis del espectro de la señal, cuyo proceso da información exacta de parámetros muy determinados y acotados del comportamiento de la energía espectral de un sonido. Más adelante, se comentará algunos exponentes de este género musical como muestra de la implementación de la extracción de información espectral como estrategia de composición musical.

A pesar de la dificultad de definir este fenómeno, la exploración del timbre en la composición musical ha suscitado interés en los creadores musicales y sonoros a lo largo de la historia. Esto ha sido un aspecto cada vez más presente en la música del siglo XX y XXI. Algunos ejemplos, podemos verlos en compositores como: Claude Debussy (1862-1918) en su pieza musical: *Prélude à l'après-midi d'un faune* (Preludio a la siesta de un fauno) (1894) en donde está presente la intención de yuxtaponer diferentes agrupamientos tímbricos, como parte de la estrategia de composición de esta obra. Otro ejemplo es el concepto: *KlangFarbenmelodie* (melodía de timbres), expuesto en el Tratado de Armonía (1911) del compositor Arnold Schönberg (1874-1951), con el que propone un componente de una metodología de composición. De igual forma Edgar Varèse (1883-1965), a través de su obra musical: *Ionisation* (Ionisación) (1929-31), establece una distribución de los materiales sonoros a través de características de vinculación tímbrica más que melódica. Por último, la compositora Kaija Saariaho[77] (1952-) en su obra: "*Vers Le*

[77] Muestra de la importancia de la exploración del timbre en la obra de Saariaho se puede ver en su página web oficial de la compositora, en donde se afirma: "Antes de incorporarse al IRCAM, Saariaho aprendió a conocer a los compositores espectralistas franceses, cuyas técnicas se basan en el

blanc" (Hacia el blanco) del año 1982, abordará el timbre estableciendo estrategias compositivas que utilizan información acústica extractada por la herramienta tecnológica[78] implementada para esta pieza musical.

Cabe destacar que la escuela de música espectral que vio la luz a inicios de 1970 marcó una dirección específica en cuanto a comprender las características acústicas del sonido como estrategia en la organización del sonido en una obra musical. Algunos compositores representativos de esta escuela son: *Hughes Dufourt* (1943-), *Gérard Grisey* (1946-1998), *Tristan Murail* (1947-), *Michael Levinas* (1949-) o *Georg Friedrich Haas* (1953-). Algunas de sus obras más representativas son: *Dérives*, para pequeño ensemble y gran orquesta (1973-74), *Partiels* para 16 ejecutantes (1975) de Grisey, *Désintegrations* o *Territoires de* l´*Oubli*, de Murail.

Otros compositores que han utilizados la técnica espectral pero no así tan abanderados del movimiento son: *Magnus Lindberg* (1958-), George Benjamín (1960-), *Jonathan Harvey* (1939-2012) y *Julián Anderson* (1967-). A manera de resumen de lo que significa la música espectral el compositor *Hughes Dufourt (1943-)* en su libro: Música espectral. Una revolución epistemológica, afirma: "Podemos definirlo como el arte de la modulación del color." (2014, p.16)

---

análisis informático del espectro sonoro. Este enfoque analítico la inspiró a desarrollar su propio método para crear estructuras armónicas, así como la notación detallada utilizando armónicos, microtonalidades". Disponible en https://saariaho.org/biography/ [visitado el 15 de marzo del 2021].

[78] *Saariaho* realizó residencias de creación e investigación en el IRCAM (Instituto de investigación y coordinación acústica/musical) ubicado en Paris, Francia, en donde trabajo con el programa informático *CHANT*. Este programa está diseñado para la generación de síntesis de voz. Ver más en: http://anasynth.ircam.fr/home/english/category/logiciel-associ%C3%A9/chant [visitado el 10 de octubre del 2021].

La importancia del timbre en mi producción sonora

A continuación, se aborda la importancia del timbre en obras de mi autoría, anteriores al programa de doctorado, su concepción desde un ámbito estético y artístico en la composición musical. La explicación que se ofrece en este apartado embarga una percepción intuitiva, que no tiene que ver con el análisis espectral del sonido para fines compositivos, a diferencia del ámbito abarcado en mi investigación doctoral.

El timbre en mi obra sonora ha jugado un papel importante desde la implementación de técnicas de orquestación acusmática[79]. En piezas como *Argia, Olinda, Ersilia*[80] se incorporaron estas técnicas, con el objetivo de implementar estrategias de composición musical partiendo de los recursos y posibilidades tímbricas. La música acusmática[81] implementa una serie de recursos taxonómicos y de técnicas de desarrollo de materiales sonoros, que están vinculados desde una perspectiva del timbre, sin embargo, en este género musical usualmente no se explora la posibilidad del timbre como recurso de "color". La palabra "color" es entendida acá desde el uso que hacen los diferentes tratados de orquestación

[79] Puede obtenerse más información acerca de este tema en el artículo bajo el título: "De la idea a la obra: los caminos de la creatividad. Orquestación electroacústica y polifonía del espacio" del compositor argentino Mario Mary (1961-).

[80] Este conjunto de obras acusmática fue compuesto para el programa de Maestría en Tecnología Musical de la UNAM en los años 2014 al 2016.

[81] Término que proviene del término griego Akousma y que Pierre Schaeffer y Jérôme Peignot lo usaron para referirse a la escucha de música concreta. El término hace referencia a la leyenda de la escuela de Pitágoras en donde el maestro enseñaba a sus alumnos separado por una cortina. Lo que se escucha y no se ve. Esta idea se vincula con la obra artística producida por el arte radiofónico y la música concreta en donde las fuentes sonoras no se reconocen por lo cual el escucha solo percibe los sonidos.

instrumental, los cuales designan un sonido o conjunto yuxtapuesto de sonidos como: oscuro, brillante, suave, rugoso, mórbido, entre otros. Estas son designaciones semánticas construidas con el fin de poder identificar los procesos, no sólo delimitan la calidad del sonido, sino las posibilidades de yuxtaposición de estos sonidos en una composición instrumental, que para efectos de la música acusmática son similares.

Es así como el timbre, dentro de la orquestación acusmática, se convierte en una nueva manera de lograr una articulación de los sonidos, además de obtener nuevos recursos tímbricos con un mayor control en el ámbito del montaje sonoro.

Como mencioné anteriormente, el timbre ha sido un recurso que en mis obras antecesoras a esta investigación, ha cambiado la manera de concebir las estrategias de composición. Esto debido a que ha hecho evidente los diferentes niveles de densidad, una mayor dirección y control en la yuxtaposición de los materiales sonoros, en el proceso de la composición, contribuyendo a lograr determinados resultados de color tímbrico, en diferentes combinaciones de este en capas de sonidos que se convierten en texturas sonoras.

En este sentido el empleo del timbre como recurso para resaltar o atenuar secciones, subsecciones o partes es vital en crear nuevas formas de articulación de los sonidos, así como proveer en el acto de la composición diferentes posibilidades de vinculación entre los materiales.

Retomando el tema de la orquestación acusmática, la valoración del timbre me ha permitido llevar a cabo un trabajo, a nivel micro[82], de cohesión y vinculación de los sonidos desde su interrelación interna. Un ejemplo de la cohesión tímbrica se puede escuchar en el inicio de la obra sonora: *Olinda* en el tiempo 0’00” a 0’55”.

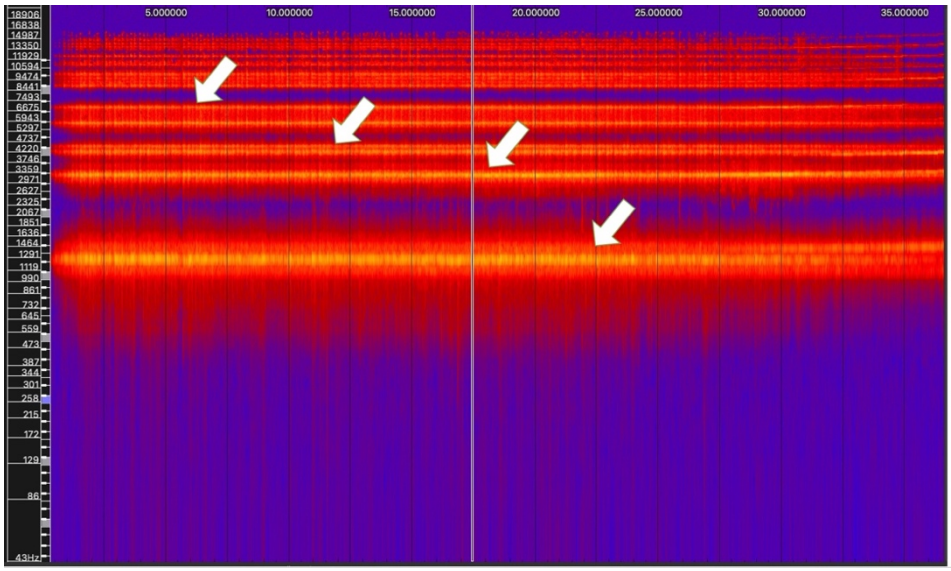
Ejemplo sonoro 2.1.

https://ottocastro.com/o/115

En la figura 9 se puede observar en el sonograma, unas franjas de color rojo y amarillo que indican los sobretonos[83], estas se muestran continuas a pesar de que este segmento graficado en el sonograma está conformado por diferentes muestras de sonido. Pieza musical *Olinda* (0’00” a 38’00”).

Figura 9

Continuidad de sobretonos



[82] Macro y micro se utilizan en este texto como términos que se refieren a la parte de la estructura musical de la obra. Macro se entenderá acá como las secciones, subsecciones, frases o gestos largos, y micro, en el sentido de la constitución del sonido y sus elementos internos de constitución tímbrica.

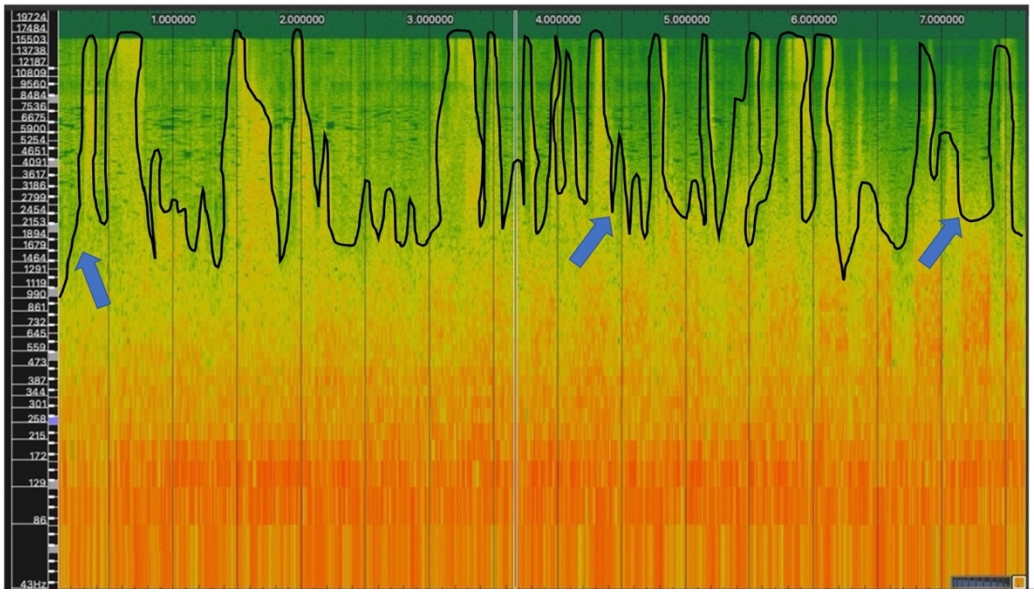
[83] Los sobretonos son conocidos con la palabra: armónicos de un sonido.

Nota: *El eje Y corresponde al rango de frecuencias (Hz) en escala logarítmica, y el eje X representa el tiempo (s). Las fechas blancas indican los sobretonos continuos.*

Otro aspecto importante para resaltar es la utilización del timbre para trabajar en una composición sonora a nivel macro, llevando el entendimiento de la vinculación de los materiales a un ámbito mayor y más general. En esta estrategia el control del ritmo es entendido como la continuidad de los eventos sonoros, la distribución de las secciones y subsecciones favorece la construcción del discurso musical. Un ejemplo de lo anterior se puede escuchar en la obra: *Argia* (ejemplo sonoro 2.2) en dos momentos que se comparan en esta pieza musical. El primer segmento en el tiempo 8’20” a 8’28 y el segundo segmento en el tiempo 12’12” a 12’26. En este ejemplo se combinan diferentes sonidos, para producir una textura sonora con una morfología compleja y niveles de densidad que varían en el tiempo de manera irregular. La estrategia de composición utilizada presenta de forma similar la disposición orquestal acusmática, con el fin de establecer nodos en la estructura musical, que dan dirección al desarrollo sonoro narrativo. En la primera sección el carácter de textura irregular y brillante es más evidente que en la segunda sección, como se ve en el sonograma (figura 10).

Figura 10

Primera parte de la textura tímbrica irregular que van de partes brillantes a opacas



Nota: El eje Y corresponde al rango de frecuencias (Hz) en escala logarítmica, y el eje X representa el tiempo en segundos. Para resaltar la irregularidad entre partes brillantes y más opacas, se le dibujó un contorno en el espectrograma, con flechas indicando este contorno. Este contorno no pertenece al sonograma original.

Ejemplo sonoro 2.2.

https://ottocastro.com/o/006

Ejemplo sonoro 2.3.

https://ottocastro.com/o/005

A modo de conclusión de este apartado**,** inicialmente en mi trabajo el timbre tuvo una aproximación intuitiva, relacionada con el color y la orquestación acusmática, sin embargo, para esta investigación actualmente me interesa este fenómeno desde la posibilidad de utilizar la comparación del análisis espectral como parte del desarrollo de estrategias de composición.

Acercamiento al timbre y el espectro

Con el fin de brindar una explicación sencilla del espectro, se puede decir que el sonido está constituido por ondas de diferente frecuencia, amplitud y fase. El análisis espectral por medio de un descriptor acústico permite extractar información de la frecuencia, amplitud y fase de cada onda de un sonido. Un ejemplo similar, pero en el ámbito óptico es cuando pasa un haz de luz a través de un prisma. El resultado que obtenemos son los componentes de frecuencia, amplitud y fase de las diferentes ondas de luz que componen ese haz. De esta manera podemos tener un acercamiento analítico a la constitución de la luz, y en nuestro caso al del sonido.

Antes de seguir adelante es necesario establecer la diferencia entre timbre y espectro. Podemos describir y estudiar un sonido desde dos acercamientos, en un ámbito perceptual-cualitativo y en otro cuantitativo. El primero vinculado a la escucha humana y el segundo, que por el contrario, entendiendo al sonido desde la constitución de sus frecuencias como fenómeno acústico. La primera posibilidad nos acerca al fenómeno que conocemos como: el timbre, y la segunda a lo que se conoce como: espectro sonoro.

Acerca de lo anterior el teórico Antonio Calvo-Manzano, afirma:

> “La cualidad del timbre de un sonido no es mensurable como las de altura o intensidad, ni existe por tanto una unidad que permita comparar timbres de distintos sonidos. [...] Por lo que el análisis del timbre de un sonido se refiere, lo que sí se puede obtener es el llamado espectro acústico” (Calvo-Manzano Ruiz, 2004, p.129).

Habida cuenta de lo anterior, a continuación, se explica la manera en que se

extrajeron los datos de los sonidos urbanos que alimentaron el sistema computacional en este trabajo investigativo, así como el proceso y los criterios para comparar dicha información. Este material permitió mostrar los resultados sonoros obtenidos por la máquina y, a partir de ellos, se derivaron estrategias para la composición musical en una etapa posterior.

Características perceptuales y acústicas que se buscan

Para la presente investigación, se desarrolló una herramienta tecnológica que utilizara descriptores[84] acústicos con sensibilidad a cuatro características del comportamiento del espectro acústico. Las características tímbricas que se deseaban identificar fueron:

- Brillante
- Opaco
- Rugoso
- Liso

Más adelante, en este mismo capítulo se explicará cada una de ellas y se pondrán ejemplos sonoros que ilustran las categorías. Estas características fueron propuestas a través de la experiencia lograda con mi escucha en composiciones anteriores. Aunque pueden existir varias maneras de nombrar estas categorías tímbricas, yo quise sugerir estas, desde cuatro adjetivos que me podrían ser útiles, no sólo como una etiqueta de categorización, sino porque para mí escucha los

[84] Un descriptor "*es una representación de una característica definida sintáctica y semánticamente*." (Sheu, p.183, 2018).

sonidos podrían ordenarse a través de cuatro características perceptuales contrastantes entre sí, lo que me permitiría tener cuatro tipos diferentes de color tímbrico en las obras a través de las identificaciones e incluso, que el sistema realice identificaciones a través de las combinaciones de estas cuatro categorías tímbricas.

En el capítulo tres, se explicará con mayor profundidad las razones estéticas que se propusieron a través de este tipo de clasificación sonora dejando para este, la explicación más técnica en el proceso de identificación por la herramienta tecnológica.

Con el fin de extractar información del sonido que pueda ser útil a la hora del análisis y a fin de comparar el comportamiento tímbrico se usarán por lo general descriptores acústicos lo cuales extractan información de la señal, brindando al usuario un análisis muy específico del comportamiento acústico de dicho sonido.

En la herramienta *CSpectra* se utilizó la librería de código *Essentia Freesound*[85], la cual dispone de una gran batería de descriptores acústicos con características espectrales, de dominio del tiempo, ritmo y tonal. Para esta investigación se seleccionaron únicamente los descriptores de bajo nivel, conocido en inglés como *low-level*.

Los descriptores de bajo nivel son calculados directamente de la señal de audio y analizan la amplitud, frecuencia o energía de la señal. Estos procesos, también

[85]Lista completa de descriptores acústicos de Freesound https://freesound.org/docs/api/analysis_docs.html#analysis-settings [visitado el 6 de febrero del 2018].

conocidos como análisis de series de tiempo[86], por lo general son resultados numéricos, uno de ellos es el que se cálculo a través de la Transformada Rápida de Fourier (FFT)[87], a través de una serie de procesos de muestras de ventana, en donde los productos finales son arreglos numéricos.

Al inicio de esta investigación se utilizó la librería de código: *Librosa*[88], pero poco después se implementó la librería: *Essentia*. El cambio se llevó a cabo debido a que *Essentia* cuenta con más recursos de análisis que *Librosa*. Además, *Essentia*[89], es uno de los recursos tecnológicos más utilizados en la actualidad con la implementación para los lenguajes de *Python*, *JavaScript*, entre otros. Esto facilita en gran medida su uso en diferentes proyectos tecnológicos.

Se presenta a continuación, de manera representativa, algunos descriptores acústicos que son sensibles al brillo/opacidad, rugosidad/lisura que fueron utilizados en esta investigación. Debido a la naturaleza compleja del timbre, cada uno de estos descriptores, dan información parcial de cada uno de los aspectos de las categorías que se buscan en esta investigación. Con el fin de tener una búsqueda más precisa para cada categoría se sumaron varios descriptores acústicos en el ámbito

---

[86] Los análisis de series de tiempo se realizan con grandes listas de valores numéricos, obtenidas a través de intervalos regulares de tiempo. Su representación se realiza a través de largas cadenas de arreglos numéricos.

[87] La Transformada Rápida de Fourier (TRF) o, en inglés, *Fast Fourier Transform* (FFT), se utiliza ampliamente en el análisis y síntesis de señales de audio. El algoritmo de la FFT fue desarrollado por los matemáticos estadounidenses J.W. Cooley y J.W. Tukey en el año 1965.

[88] Es una librería dedicada al análisis de audio y de música que utiliza la potencialidad y versatilidad del lenguaje *Python*. Dispone de diferentes módulos para llevar a cabo tareas de recuperación de información musical o *Music information retrieval* (MIR). Ver más en: https://librosa.org/doc/latest/index.html [visitado el 9 de febrero del 2018].

[89] "Essentia está diseñado con un enfoque robusto de los descriptores musicales proporcionados, y está optimizado en términos del costo computacional en la implementación y uso de los algoritmos." (Souza Britto, 2013)

espectral, para extractar más información acerca del espectro que diera un perfil más preciso a la hora de su identificación.

Categoría: brillante / opaco

A manera de ejemplo ilustrativo y como metáfora, consideremos la experiencia de percibir un objeto metálico iluminado por la luz del sol: si se ajusta la posición y el ángulo del observador, puede notarse más o menos energía lumínica y eso es lo que percibimos como brillo a nivel visual. De la misma manera, en un sonido puede haber una mayor concentración de energía espectral en una banda de frecuencias de sonido. No obstante, en el sonido, la percepción de brillo está determinada por la concentración de energía, dependiendo de la zona alta o baja de frecuencia que ocupe. También, el brillo se vincula con la intensidad; es decir, a una mayor concentración de energía espectral, en bandas de frecuencias altas y con una intensidad fuerte, un sonido dará la impresión de ser más brillante que otro sonido con una concentración de energía espectral en bandas de frecuencias graves, pero con una intensidad de moderada o leve.

Existen muchos experimentos que confirman lo anterior. Uno de estos fue desarrollado por el teórico Lawrence E. Marks, de la Universidad de Illinois, quien afirma:

> Como veremos, la mayoría de los sujetos efectivamente equipararon tanto el tono y volumen al brillo de la superficie: tono y volumen bajos al negro, tono alto y alto volumen al blanco. (Marks, 1974, p. 175)

Uno de los descriptores acústicos que puede identificar brillo y opacidad

tímbrica es el *Centroide Espectral*[90]. En este caso, "*los valores altos corresponden a más energía de la señal concentrada en las altas frecuencias*" (Kulkarni, N. y Bairagi, V., 2018, p. 44). A nivel de la percepción humana el brillo es evidente en una mayor concentración de energía en las altas frecuencias y varia según la intensidad, es decir "*los sonidos con más energía en bajas frecuencias se perciben como apagados, los sonidos con más energía en las frecuencias altas se perciben como brillantes o nítidos*" (Lemaitre & Susini, 2019, p.7).

Algunos descriptores acústicos utilizados para identificar el brillo y la opacidad entran en lo que se conoce como: características de la energía espectral, tales como: *Spectral Centroid, Kurtosis, Skewness o Spread*, brindan valores estadísticos en la forma que la distribución de la energía espectral se da en un momento determinado del análisis de un sonido. Esta información ayuda a conocer de qué manera se concentra la energía espectral con relación a las frecuencias y así se determina que tan brillante u opaco es un sonido. A continuación, se dan las definiciones de estos descriptores acústicos:

Spectral Centroid

> "El centroide espectral es una medida que indica dónde está el "centro de masa" del espectro. Perceptualmente, tiene una fuerte conexión con la impresión de "brillo" de un sonido y, por lo tanto, se utiliza para caracterizar el timbre musical. Se calcula como la media ponderada de las frecuencias presentes en la señal, con sus magnitudes como ponderaciones." (Essentia,

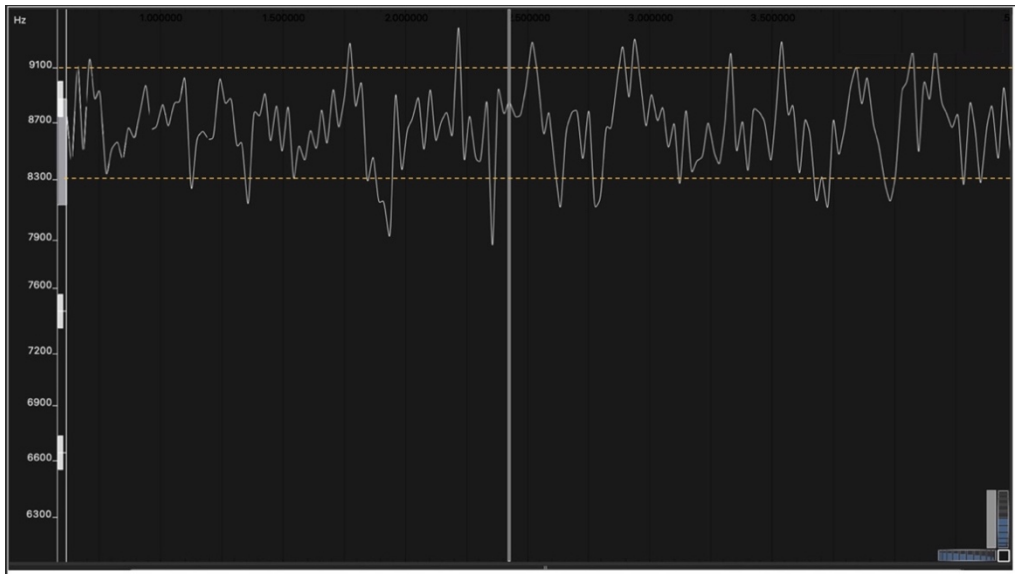
[90] Debido a la dificultad de traducción al idioma español para el caso del nombre de cada uno de los descriptores acústicos, estos se conservan en el idioma inglés.

2021).

Este descriptor es uno de los más utilizados y sensibles para identificación del brillo tímbrico. Para todas las visualizaciones en este capítulo se utilizó el programa: *Sonic Visualizer*[91] que permite el análisis de la señal con diferentes tipos de descriptores acústicos. A continuación, con el fin de ilustrar la categoría brillante, se puede apreciar en la figura 11, con el ejemplo sonoro 2.4, que la magnitud del *Spectral Centroid* se concentra entre 8300 Hz y 9100 Hz, un rango de frecuencias altas, debido a que los valores del descriptor se concentran en ese rango de frecuencias altas, esto nos indica brillo tímbrico en el ejemplo aquí mostrado.

Figura 11

Medición del Spectral Centroid. Caso tímbrico brillante



[91] Uno de los primeros programas gratuitos en permitir a cualquier usuario, la visualización del análisis de señal a través de determinados descriptores acústicos, que el usuario puede establecer.

Nota: El eje Y corresponde al rango de frecuencias en escala logarítmica, y el eje X representa el tiempo en segundos. La línea punteada de color naranja indica la actividad del *spectral centroid* concentrada entre 8300 Hz y 9100 Hz.

Ejemplo sonoro 2.4.

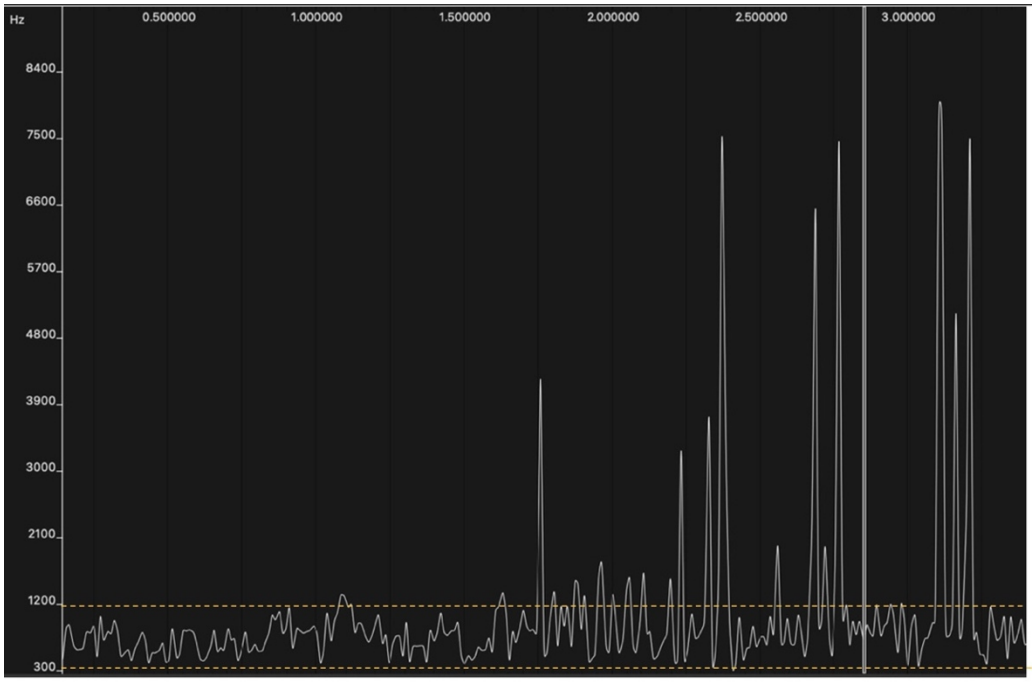
Mosaico Urbano en el tiempo 9’20” a 9’24”

https://ottocastro.com/o/122

Por el contrario, los sonidos que contienen una mayor concentración espectral en las frecuencias bajas son percibidos por el oído humano como oscuros u opacos en su timbre. En la figura 12 vemos a través de la línea blanca, como el *spectral centroid* se concentra en una banda de frecuencia entre 300 Hz y 1200 Hz. Debido a que en este este ejemplo, el descriptor concentra su actividad entre este rango de frecuencia grave por lo cual es que se puede considerar un caso de timbre opaco. Este ámbito dentro del rango frecuencial se percibe como grave. Existen picos que ocasionalmente llegan a los 7536 Hz, pero son casos esporádicos. El ejemplo sonoro de la figura 12 se puede apreciar en el 2.5.

Figura 12

Spectral Centroid entre la banda de 300 Hz y 1200 Hz. Caso de timbre opaco



Nota: El eje Y corresponde a los valores de frecuencia, y el eje X representa el tiempo en segundos.

Ejemplo sonoro 2.5.

Mosaico Urbano en el tiempo 8’58” a 9’02”

https://ottocastro.com/o/121

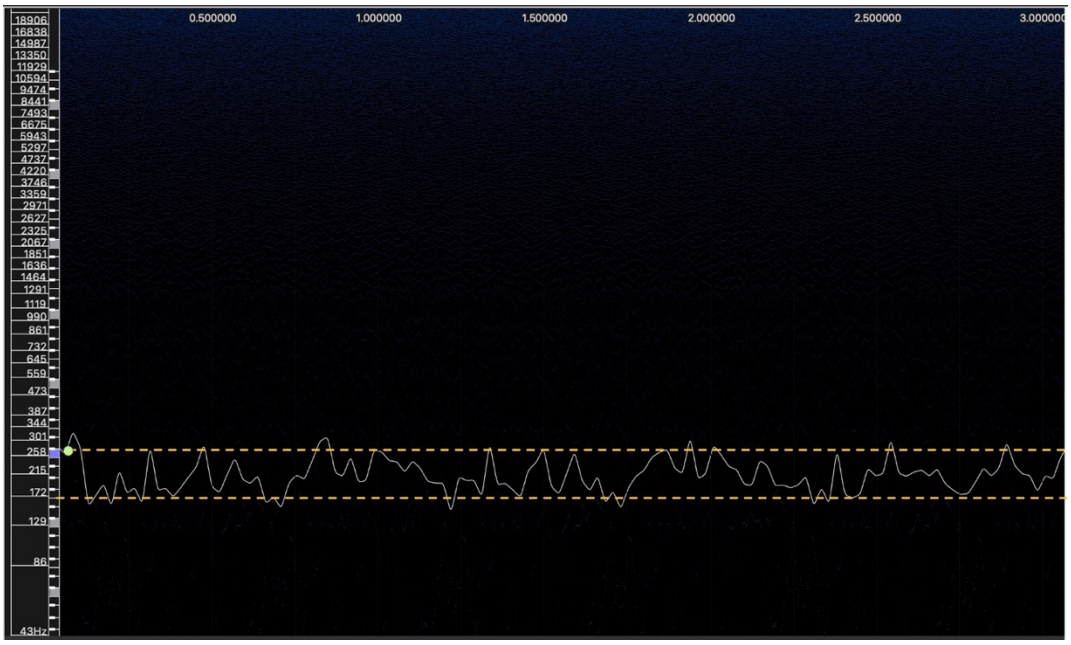
En la figura 13, se muestra un nivel de opacidad mayor al de los ejemplos presentados anteriormente. En dicha figura existen dos capas de lectura, en una se ve un sonograma con colores morado, naranja y amarillo. Los colores más intensos como el amarillo indican mayor concentración de energía espectral. La otra

capa de lectura es la curva de color blanco que indica los niveles de *Spectral Centroid*. El programa *Sonic Vizualizer*, con el cual se realizaron las mediciones, muestra en el eje Y: la relación de la intensidad del *Spectral Centroid* con las frecuencias donde existe mayor concentración de energía espectral. Vemos entonces como hay una concentración de energía, que va de las frecuencias entre 172 Hz a 258 Hz, teniendo algunos eventos esporádicos en donde se dispara la medición del descriptor hasta 7536 Hz. Pero esto se debe al sonido del ambiente, es decir al sonido de fondo y no a la fuente sonora que aparece en primer plano, que es la que interesa analizar.

El ejemplo sonoro de la figura 13 se puede apreciar en el 2.6.

Figura 13

Medición del Spectral Centroid. Caso tímbrico oscuro



Nota: El eje Y corresponde al rango de frecuencias en escala logarítmica, y el eje X representa el tiempo. Las líneas punteadas de color naranja, muestran la mayor

actividad del *spectral centroid* con respecto al rango de las frecuencias que abarca. El ejemplo sonoro de la figura 13 se puede apreciar en el 2.6.

Ejemplo sonoro 2.6.

Mosaico Sonoro en el tiempo 0'00" a 7'05"

https://ottocastro.com/o/143

Categorías: rugoso / liso

La categoría de rugosidad fue estudiada por primera vez por el físico Harvey Fletcher (1884-1981) en 1933. Este término se refiere a un fenómeno de percepción que se lleva a cabo en el oído interno humano, específicamente en la cóclea, con relación a una banda de frecuencia. En la cual hay una interferencia de un tono sobre otro, produciendo una apreciación diferente a lo que realmente está sonando, debido al fenómeno acústico de enmascaramiento.

Sushen *et al*. (2018, p. 109) definen lo rugoso[92] como *"…una medida para estimar la cantidad de disonancia sensorial en cada momento sucesivo a lo largo de una señal*". Sottek & Genuit definen la rugosidad también como:

> "La sensación de rugosidad ocurre siempre que existe una envolvente variable en el tiempo dentro de una banda crítica, por ejemplo, cuando los

[92] Se recomienda leer y escuchar el ejemplo de rugosidad entendido desde la banda crítica que brinda el investigador y compositor Barry Truax desde su página web: https://www.sfu.ca/sonic-studio-webdav/handbook/Critical_Band.html [visitado el 12 de marzo del 2020].

tonos exhiben una estructura temporal debido a una variación de su amplitud o frecuencia. Con incremento rugosidad, las emisiones de ruido se perciben como cada vez más perceptibles y, por lo general, como cada vez más agresivas y molesto" (2013, p.2)

Con el fin de ilustrar esto de una manera más clara auditivamente, a continuación, se presentan dos ejemplos sonoros. En el ejemplo 2.7 se escucha un tono sinusoide de 1000 Khz que para efectos perceptivos del oído no constituye un sonido áspero. Mientras que en el ejemplo sonoro 2.8 se presenta un sonido de 1000 Khz con una modulación de otro tono a 70 Hz. Los batimentos producidos perceptualmente conllevan un fenómeno de aspereza o rugosidad.

Ejemplo sonoro 2.7

https://ottocastro.com/o/tonosimple

Ejemplo sonoro 2.8

https://ottocastro.com/o/tono1000zod70

Es conveniente hacer una aclaración entre los términos que yo propongo desde mi escucha y los resultados que puede identificar la máquina en cuanto a aspectos de energía espectral del sonido. Para efectos estéticos la rugosidad la percibo como un sonido con cierta aspereza, disonancia y con poca armonicidad, en contraposición con los sonidos lisos, los cuales, dentro de mi percepción son sonidos que se escuchan como continuos, armónicos y suaves.

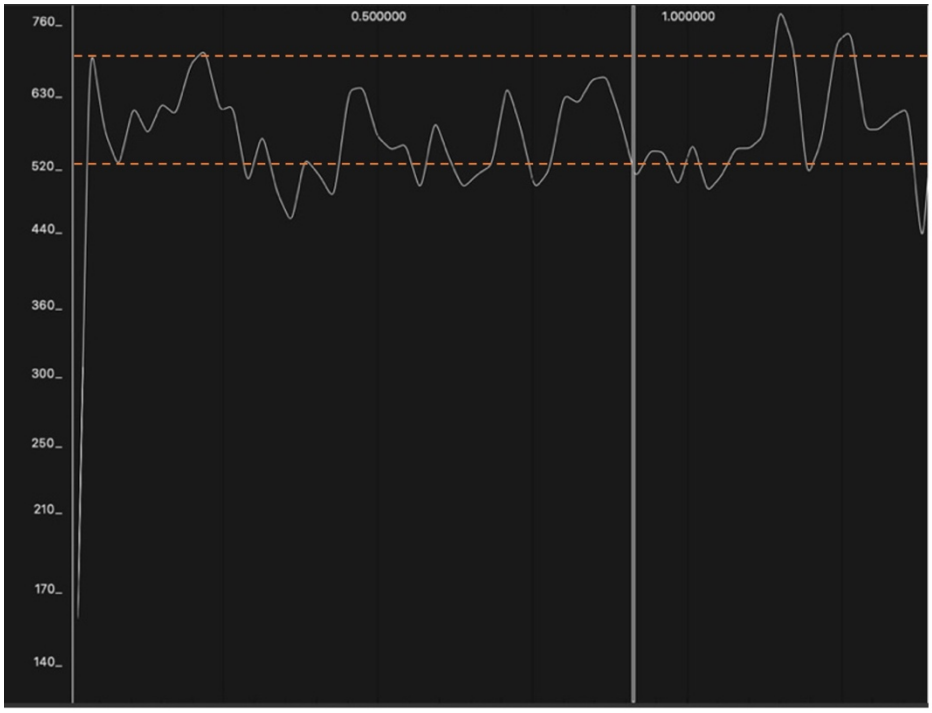
Uno de los descriptores acústicos que se utilizaron en esta investigación para la identificación de los niveles de aspereza es el *Spectral Flux*. Este descriptor se define como: "*Una medida del cambio del espectro de una señal. Se calcula la diferencia entre el espectro actual y el del cuadro anterior. A menudo corresponde a la aspereza o rugosidad perceptiva de un sonido*" (Meyda, 2021) También se define como "*la medida de la dispersión media del espectro en relación con su centroide. Para sonidos ruidosos los niveles son altos, mientras que para los sonidos tonales, con menos banda ancha, mostraría niveles bajos*." (Essentia, 2021)

A continuación, se muestran dos ejemplos de rugosidad, uno con mucho nivel de rugosidad (ejemplo sonoro 2.7), y otro más moderado, ejemplo 2.8. Los dos pertenecen a sonidos de motores de autobús. En el 2.7 se escucha un sonido solo de motor de autobús, en el 2.8 un montaje con diferentes sonidos de motores de autobús, e incluso algunos de estos procesados.

En la figura 14 observamos dos ejes X y Y, donde el eje X es el tiempo en segundos, y el eje Y, la escala de rugosidad que va de 0 a 760. Siendo 0 menor rugosidad y 760 mayor rugosidad. Podemos entender el nivel de rugosidad por la escala propuesta por el programa, que lleva a cabo el análisis que demarca una magnitud alta, que va de 520 a 760.

Figura 14

Nivel de rugosidad en un sonido de motor de autobús



Nota: El eje Y corresponde a los niveles de rugosidad según el programa, y el eje X representa el tiempo en segundos. Las líneas punteadas de color naranja muestran la banda desde donde demarca el nivel de rugosidad del sonido analizado. El ejemplo sonoro de la figura 14 se puede apreciar en el 2.7.

Ejemplo sonoro 2.7.

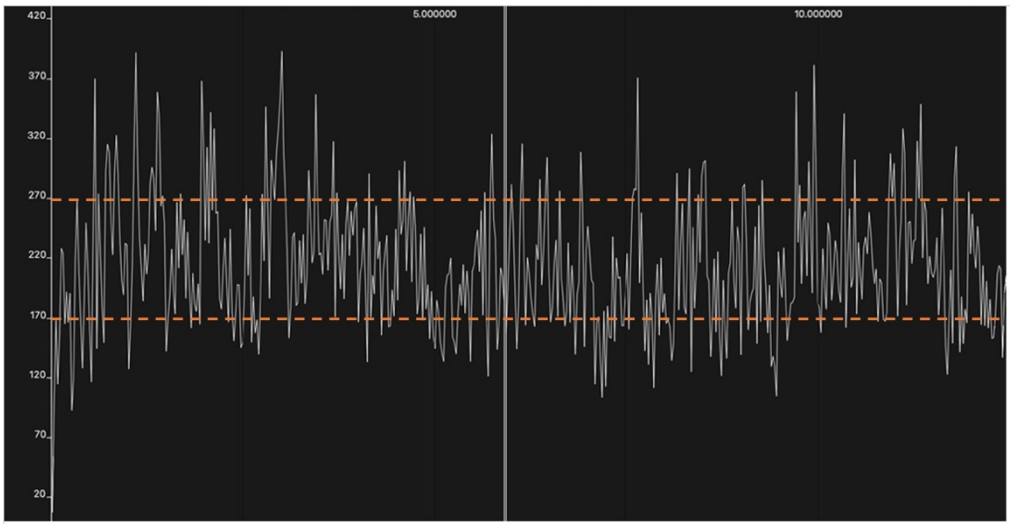
https://ottocastro.com/o/008

En la figura 15 se puede apreciar otro ejemplo de rugosidad, pero esta vez de menor magnitud. En esta, tenemos en el eje X el tiempo en segundos, y en el eje Y, la medida de los niveles de rugosidad va de 0 a 420. Siendo 0 menor rugosidad y 420

mayor rugosidad. El ejemplo sonoro de la figura 15 se puede apreciar en el 2.8.

Figura 15

Fragmento de sonido rugoso "Anomalía Urbana" (4'07" a 4'20")



Nota: El eje Y corresponde a los niveles de rugosidad según el programa, y el eje X representa el tiempo en segundos. Las líneas punteadas de color naranja muestran la banda desde donde demarca el nivel de rugosidad del sonido analizado que va de 170 a 220, los cuales indican menos rugosidad que en ejemplo de la figura 14.

Ejemplo sonoro 2.8.

https://ottocastro.com/o/009

Por otro lado, los sonidos con categoría de lisura desde mi percepción son suaves, continuos con pocos batimentos y armónicos. Dos de los descriptores que

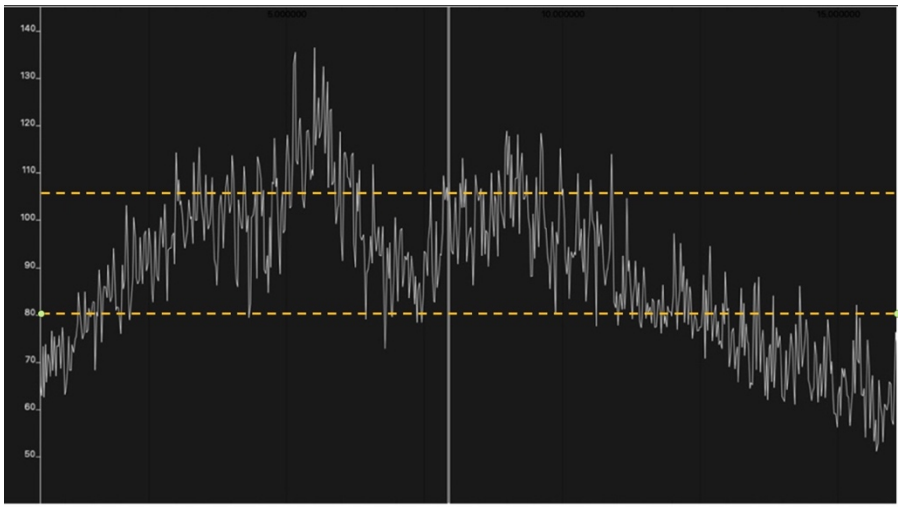
se utilizan en la identificación de la lisura son el *Spectral Flux* y *Spectral Roll-Off.* El *Spectral Flux* se explicó en el apartado anterior, del *Spectral Roll-Off* se define como:

> “Este algoritmo calcula la frecuencia de caída de un espectro. La frecuencia de caída se define como la frecuencia por debajo de la cual está contenido algún porcentaje (corte) de la energía total del espectro. La frecuencia de atenuación se puede utilizar para distinguir entre sonidos armónicos (por debajo de la atenuación) y ruidosos (por encima de la atenuación).” (Essentia, 2021).

A continuación, dos ejemplos de lisura tímbrica. En la figura 16 un caso de lisura con un rango amplio que van desde un poco a mayor lisura, en la figura 17, un caso de lisura con una banda de rango más estrecha en donde los valores son más constantes.

Figura 16

Fragmento de sonido liso de la pieza: Anomalía Urbana (6’34” a 6’36”)



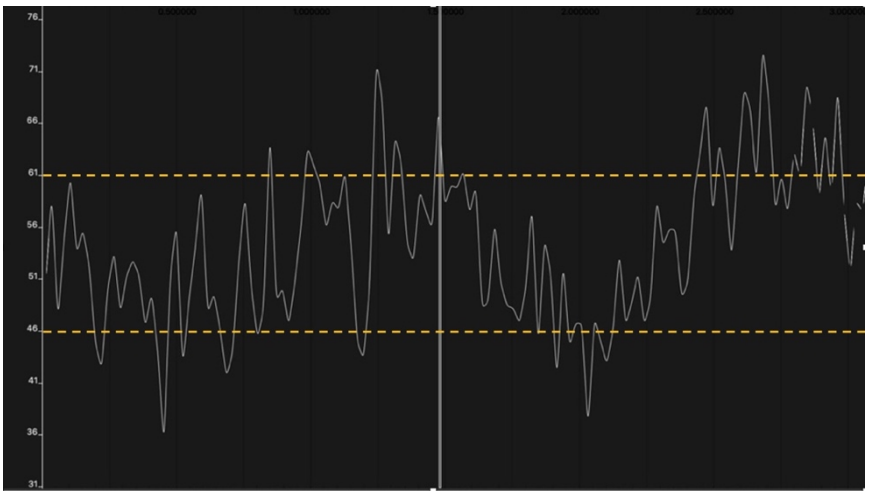
Nota: El eje Y corresponde al nivel de lisura según el programa, y el eje X representa el tiempo en segundos. Las dos líneas punteadas denotan el rango desde donde la lisura se presenta, es decir, niveles de 62 a 120 lisura, estos representan niveles altos de lisura. El ejemplo sonoro de la figura 16 se puede apreciar en el 2.9.

Ejemplo sonoro 2.9.

https://ottocastro.com/o/144

Figura 17

Fragmento de sonido liso Anomalía Urbana (2'08" a 2'10")



Nota: El eje Y corresponde al nivel de lisura según el programa, y el eje X representa el tiempo en segundos. Las dos líneas punteadas denotan el rango de lisura entre los valores 46 a 61 del programa. Estos representan niveles más bajos de lisura que en el ejemplo 16.

Ejemplo sonoro 2.10.

https://ottocastro.com/o/145

Existe una serie de descriptores acústicos que extractan información de la distribución y forma de la energía espectral, arrojando datos estadísticos de una abstracción mayor, en donde la escucha a nivel perceptual, tímbrico a veces no es tan clara, aunque son resultados importantes con el fin obtener un perfil acústico más amplio del comportamiento espectral. Cabe resaltar en este punto que una cosa es lo que la máquina analiza y extracta a nivel espectral, y otra distinta es lo que el oído humano en el ámbito del timbre puede reconocer. Por ejemplo, para varios de ellos como el *Spectral Skewness*, *Kurtosis* miden cómo la energía se distribuye, por ejemplo, para efectos del brillo tímbrico, el conocer qué tipo de distribución se da en cada análisis, refuerza la información acerca de la energía en relación con las frecuencias, y cuales atenúa o resalta, brindando información valiosa en cuanto al brillo u opacidad del sonido.

Se explica a continuación varios de estos descriptores utilizados:

Spectral Complexity: "*Este algoritmo calcula la complejidad espectral de un espectro. La complejidad espectral se basa en el número de picos en el espectro de entrada.*" (Essentia, 2021). Se utiliza para saber qué tanta densidad espectral tiene un sonido. (Meyda, 2021)

Spectral Decrease: "*representado por la cantidad de la caída de la amplitud espectral*" (Peeters, 2003, p.26).

Spectral Energyband High, Middle, Low: se refiere a cómo se distribuye la energía espectral en las bandas altas, medias y bajas. Se utiliza como: "indicador del *loudness* de la señal" (Meyda, 2021).

Spectral Flatness: "*Brinda información acerca de la planitud en la distribución de la energía espectral. Es la medida del nivel de ruidosidad o sinusoidad de un espectro o parte de esta."* (Peeters, 2003, p.26). Se utiliza para determinar "que tan ruidoso es un sonido" (Meyda, 2021).

Spectral Crest: "*Calcula el radio de los valores máximos dentro de la banda a la media aritmética del valor del espectro de energía.*" *(Peeters, 2003, p.26)*
Pitch: "Este algoritmo calcula la prominencia del tono de un espectro". (Essentia, 2021).

Spectral Roll-Off: "*Este algoritmo calcula la frecuencia de caída de un espectro. La frecuencia de caída se define como la frecuencia por debajo de la cual está contenido algún porcentaje (corte) de la energía total del espectro. La frecuencia de atenuación se puede utilizar para distinguir entre sonidos armónicos (por debajo de la atenuación) y ruidosos (por encima de la atenuación).*" (Essentia, 2021). Se utiliza para saber qué frecuencia en un momento determinado contiene el máximo de

energía espectral. (Meyda, 2021)

Spectral Skewness: La librería de código Essentia lo define como: "*Es el grado de asimetría observado en una distribución normal de probabilidad que se desvía de la distribución normal simétrica (curva de campana) en un conjunto de datos dado.*" (Essentia, 2021).

Spectral Kurtosis: "*es una medida estadística, ya sea que los datos sean de cola gruesa o de cola ligera en una distribución normal.*" (Essentia, 2021). Indica que tan "*tonal*" es un sonido. (Meyda, 2021).

Acercamiento tecnológico para encontrar similitud entre los sonidos

En la actualidad, la velocidad de cómputo ha permitido acortar procesos que hace tres décadas tomaban un tiempo considerable o que eran imposibles de ejecutar. En ese momento, quizás hubiéramos tenido que construir desde cero un algoritmo que nos permitiera resolver una situación en particular y repetir de nuevo esta dinámica en cada problema de nuestra investigación. Eso ha cambiado: actualmente, un estudio como el que se presenta en este documento se nutre de recursos digitales disponibles a través de la internet. Tal es el caso de los diferentes módulos y librerías de código, los cuales posibilitan la implementación de algoritmos sofisticados, que permiten extractar información de los audios, procesar datos o realizar otros procesos complejos.

Un área que se ha puesto en boga es el aprendizaje automático, un

subcampo de la *Inteligencia Artificial*, el cual se utiliza, por ejemplo, en la detección de fraude de tarjetas de crédito, en diagnósticos médicos, para generar recomendaciones de piezas musicales similares en *Spotify*, así como en juegos y análisis de valores de mercado, entre otros. Se entiende por aprendizaje automático "*la ciencia de hacer que las computadoras actúen sin estar programadas explícitamente*" (Andrew Ng, 2021). En la presente investigación, el aprendizaje automático se utilizó como el método con el cual la máquina compara los datos obtenidos del análisis acústico y predice posibles sonidos similares.

Ahora bien, existen diferentes algoritmos que se utilizan en el aprendizaje automático, los cuales se dividen en dos tipos: aprendizaje supervisado y no supervisado. El aprendizaje supervisado "*consiste en mostrar al sistema una serie de ejemplos de datos con las etiquetas adecuadas*" (Bañuelos, 2019, p. 31); es decir: al sistema informático se le da una serie de etiquetas que se usan para predecir algún valor de los elementos analizados. En este tipo de modelo, el sistema aprende de los resultados obtenidos e incorpora ajustes mediante la predicción con nuevos ejemplos. Por su parte, el aprendizaje no supervisado "*consiste en entrenar un sistema para aprender a distinguir agrupaciones de interés en una base de datos, donde no se tiene información estructural de etiquetas*" (Ibid., p.32). Por tanto, se utiliza cuando se desea saber qué tipo de clases pueden existir en un conjunto de datos, sin tener etiquetas previas para ellos. Este modelo permite agrupar los datos y asignarles una clase.

Una vez aclarados de forma básica estos conceptos, explicaremos el proceso de construcción de la herramienta tecnológica que nos ayudó a resolver el problema

planteado para este trabajo.

Primera exploración: código detonante

La herramienta construida para esta investigación tiene el objetivo de: identificar y agrupar sonidos acústicamente similares y, además, de ejecutar un proceso automático en la creación de una textura sonora[93]. Dicha herramienta brinda al usuario asistencia en la composición desde el *multitrack* de Reaper.

Como primer punto, es conveniente explicar de qué forma se configuraron las fases de la herramienta tecnológica[94]. Un texto fundamental para esta investigación es "*Machine Learning with TensorFlow",* del teórico y programador Nishant Shukla[95], el cual permitió determinar la ruta de exploración que se tomaría para conseguir el objetivo. En el capítulo tercero de dicha publicación, se presenta un ejercicio para diferenciar sonidos ruidosos y sonidos de altura definida, a través del descriptor acústico *chromogram*[96]. Según las características de los sonidos con alturas definidas, el código compuesto por dicho descriptor acústico identifica estos sonidos como tónicos y, por tanto, los clasifica como gritos. Mientras que los sonidos inarmónicos o ruidosos los identifica como sonidos de tos.

---

[93] En esta investigación, se entiende por textura sonora la concatenación de varios sonidos similares que pueden percibirse a la escucha como una "textura", ya sea rugosa, lisa, brillante u opaca, o bien, una textura sonora que sea híbrida en la combinación de estos cuatro tipos de características perceptuales. *Emmerson y Landy* (2016, p. 102) definen textura como "*la organización de la simultaneidad*". Teniendo en cuenta esta definición, se entienden las texturas como una intencionalidad en la distribución de los sonidos simultáneos que incorpora el compositor en su obra.
[94] Las herramientas tecnológicas elaboradas se concretaron gracias a la contribución, el seguimiento y la supervisión del Dr. Hugo Solís. También es importante resaltar el apoyo de asesoría que hiciera el Dr. Cristian Bañuelos a este proyecto de investigación.
[95] Más información en https://shukla.io/ [visitado el 15 de setiembre de 2019].
[96] Traducido al español como cromagramas.

A continuación, se muestran algunos segmentos del código (este se puede ver completo en Shukla, N. y Fricklas, K., 2018, p. 103).

```
//A
import tensorflow as tf
import numpy as np
from bregman.suite import *

//B
filenames = tf.train.match_filenames_once('./audio_dataset/*.wav')

//C
def get_next_chromogram(sess):
        audio_file = sess.run(filename)
        F = Chromagram(audio_file, nfft=16384, wfft = 8192, nhop=2205)
        return F.X
//D
def extract_feature_vector(sess, chromo_data):
        num_features, num_samples = np.shape(chromo_data)
        freq_vals = sess.run(max_freqs, feed_dict = {chromo: chromo_data})
        hist, bins = np.histogram(freq_vals, bins = range(num_features + 1))
        return hist.astype(float) / num_samples
```

Este código fue fundamental en la investigación, pues permitió desarrollar herramientas tecnológicas como el *MAT2021 (Herramienta para el Montaje y el Análisis Tímbrico)* y el *CSpectra2021 (Herramienta para la Comparación Espectral)*. El análisis del código develó una forma básica para programar las herramientas tecnológicas y lograr el objetivo.

A continuación, se presentan, de manera general, las partes de dicho código:

A. Librerías de código, enfocadas en la extracción de características acústicas

B. Carga un audio de entrada

C. Utilización de un descriptor acústico

D. Se extraen las características acústicas del audio que se introdujo al código, a través del descriptor implementado.

Además, en el código existen dos componentes fundamentales para la identificación de los sonidos: el primero consiste en utilizar el descriptor acústico *Chromagram*, que ayuda a identificar sonidos con altura definida o, en caso contrario, sonidos inarmónicos. El segundo es el uso del algoritmo *K-means,* que procesa la información obtenida por el descriptor para obtener grupos de sonidos similares.

```
//E

def recompute_centroids(X, Y):
        sums = tf.unsorted_segment_sum(X, Y, k)
        counts = tf.unsorted_segment_sum(tf.ones_like(X), Y, k)
        return sums / counts

with tf.Session() as sess:
        sess.run(init)
        X, names = get_dataset(sess)
        centroids = initial_cluster_centroids(X, k)
        i, converged = 0, False
        while not converged and i < max_iterations:
        i += 1
        Y = assign_cluster(X, centroids)
        centroids = sess.run(recompute_centroids(X, Y))
```

No obstante, un inconveniente del código es su nivel de imprecisión en cuanto a la identificación tímbrica, pues esta es muy general en casos de sonidos con características tímbricas evidentemente extremas. Por ejemplo, si se ingresan a este código dos sonidos de características muy definidas, uno claramente "ruidoso" y otro con altura de frecuencia muy definida, se generará un resultado preciso. Sin embargo, si se ingresan sonidos con características híbridas, el código

no los identifica o lo hace de forma incorrecta; es decir, en el caso de sonidos que no sean totalmente tónicos y tengan algún porcentaje de ruido, el resultado obtenido no será tan preciso.

Como mencioné, este código fue el detonante para desarrollar la investigación. Sin embargo, aunque se partió de la estructura original, nuestra herramienta se amplió con una batería mayor de descriptores acústicos y se utilizaron otros métodos sofisticados de clasificación, que explicaremos más adelante.

*MAT2021*a y MAT2021b como antesala a la herramienta final: *CSpectra2021*

Después de analizar el código propuesto por *Nishant Shukla*, se realizaron otros ejercicios, con perfil formativo y pedagógico, para comprender mejor la dinámica de los métodos de aprendizaje automático existentes e implementarlos en nuestro trabajo. Así, se realizaron varios cursos en línea y de este aprendizaje surgieron algunos códigos enfocados en la clasificación a través de algoritmos de agrupamiento por clase.

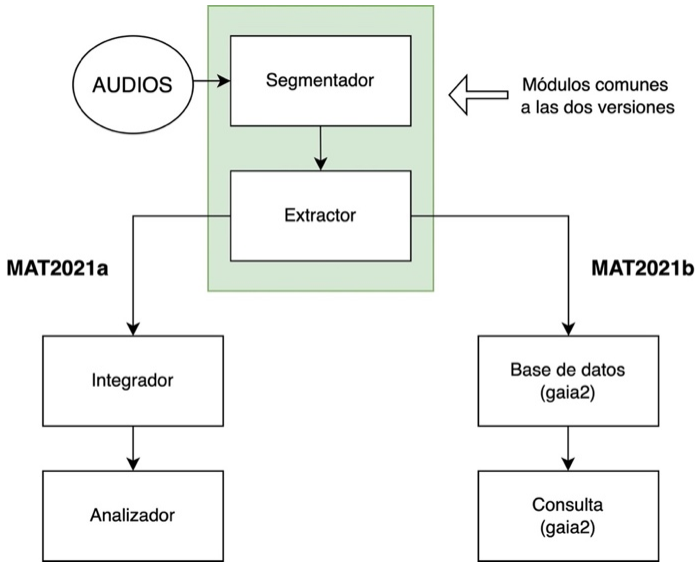
Los códigos iniciales que se generaron tuvieron dos objetivos: primero, probar con diferentes librerías de código como *Librosa* y *Essentia,* para conocer acerca de la extracción de datos a partir del análisis de diferentes sonidos; segundo, realizar ejercicios con algoritmos de *clustering*[97] por clases, con el fin de lograr los primeros resultados de agrupamiento por similitud.

[97] Se puede traducir como “conglomerados”. Este análisis consiste en agrupar objetos por similitud, con el fin de que las clases identificadas tengan las mismas características entre ellas. Es una de las técnicas de análisis dentro del campo del aprendizaje automático.

Habiendo tenido esta experiencia de entrenamiento, se procedió a construir dos versiones de la herramienta *MAT2021: el MAT2021a y el MAT2021b.* Ambas se construyeron utilizando una computadora MacBook Pro, con un procesador 2.2 Ghz Intel Core i7 de cuatro núcleos y memoria RAM de 16 GB 1600 Mhz DDr3, con una tarjeta de video interna Intel Iris Pro 1536 MB, a la cual se le instaló el sistema operativo de Ubuntu Studio versión 20.04 LTS[98]. Las dos versiones están construidas en cuatro módulos y comparten los dos primeros, como se puede ver en la figura 18. Cada módulo tiene un código que realiza una función determinada.

Figura 18

Módulos de código de las versiones MAT2021a y MAT2021b



[98] LTS es la abreviatura en inglés de *Long Term Support*, que se traduce al español como soporte de largo plazo. El LTS es una distribución del sistema operativo LINUX que se caracteriza porque da más soporte que una versión normal. Más información en https://ubuntustudio.org/2020/04/ubuntu-studio-20-04-lts-released/ [visitado el 3 de marzo del 2020].

Para el funcionamiento de dichos módulos se requieren datos de entrada, los cuales se alojan en carpetas ya existentes o creadas por el sistema. Los archivos se almacenan en formato Yamls[99], JSON[100], CSV[101] o .txt y audio. Las dos versiones entregan, como resultado final, una serie de audios en una o varias carpetas.

Además, todos los módulos se ejecutan con comandos desde la terminal[102]. Esto se realiza de la siguiente manera: primero se invoca el lenguaje *Python* junto con su respectiva versión (en este caso, la tercera); se agrega un espacio y luego el nombre del código que se desea ejecutar; se coloca un espacio más y la ruta de origen (en este caso, el nombre del audio que se desea introducir al código) y se añade otro espacio y la ruta de destino o ruta de la carpeta a donde se desea que el código devuelva los audios segmentados. Siguiendo esta ruta, el usuario puede ejecutar cada módulo del sistema.

A continuación, se muestra un ejemplo del comando anterior:

python3 codigo.py Ruta origen Ruta destino

A continuación, explicaremos brevemente el funcionamiento de cada módulo:

Segmentador: el código contenido en este módulo toma como entrada los audios en formato estéreo (2 canales) y los remuestrea para convertirlos en una señal

[99] *Yet Another Markup Language* se traduce como "otro lenguaje de marcado", para resaltar su propósito de trabajar con datos más que con documentos.
[100] *JavaScript Object Notation* se traduce como notación de objetos de *JavaScript*.
[101] Sus siglas provienen del inglés *Comma Separated Values,* que se traduce al español como "valores separados por comas".
[102] La terminal es un programa a través del cual se le dan órdenes a la computadora por medio de líneas de comando.

mono (1 canal) en formato de 44100 *kHz* a 16 *bits*. Este procedimiento permite procesar todos los audios de forma homogénea.

Luego, el código corta los segmentos, utilizando la librería de código *Essentia,* con el algoritmo *SBIC*[103]. Esto permite identificar segmentos con contraste tímbrico. Pero, antes de este proceso, el *SBIC* recibe la información obtenida por el algoritmo *MFCC*[104]*,* la cual se requiere para establecer el criterio de corte al audio.

Finalmente, luego de estos procesos, el código devuelve segmentos de audio de diferentes duraciones. Para los audios urbanos que se introdujeron en este módulo hubo cortes entre 100 ms y 3000 ms, pero este tipo de segmentaciones puede cambiar dependiendo de los materiales sonoros que se introduzcan al segmentador.

A continuación, se muestra un ejemplo del código utilizado para la implementación del algoritmo *SBIC*:

```
sbic=SBic(size1=size1,inc1=inc1,size2=size2,inc2=inc2,cpw=cpw,
minLength=minimumSegmentsLength)
segments=sbic(np.array(features))
grabadorDeSegmentos(audio,segments)
```

Extractor: el código contenido en este módulo toma los audios segmentados por el módulo anterior, los analiza uno por uno y extracta la información acústica por medio de los descriptores. Durante esta acción, ejecuta también un subproceso, en el cual

---

[103] *SBIC* son siglas que provienen del inglés *Segments Bayesian Information Criterion* ,que se traduce al español como "segmentación por criterios bayesianos". Puede verse más información en https://essentia.upf.edu/reference/streaming_SBic.html [visitado el 1 de marzo de 2019].

[104] Este algoritmo calcula los coeficientes de *cepstrum* de frecuencia de *Mel* de un espectro. Las siglas provienen del inglés *Mel Frequency Cepstral Coefficients,* que se traduce al español como "coeficientes cepstrales en las frecuencias de Mel". Más información en https://essentia.upf.edu/reference/streaming_MFCC.html [visitado el 10 de marzo de 2019].

invoca un archivo binario (*Extractor Freesound*[105]) mediante una línea de comando, desde el entorno del lenguaje *Python*. Es decir, esta librería de código extrae información acústica del audio.

A continuación, se muestra un fragmento del código:

```
def analizador(root,fileName):
    file_name, file_extension = os.path.splitext(fileName)
    subprocess.call(["essentia_streaming_extractor_freesound", root + "/"+
    fileName, rootSalida + "/"+ file_name + ".json"])
```

Una vez que el código tiene la información acústica extractada, la guarda en un archivo de formato *JSON* (es un formato para almacenar datos, que tiene la estructura de una etiqueta y su valor correspondiente). Allí están alojados todos los datos de la extracción de sus características. Este proceso se realiza para cada segmento de audio analizado. En el fragmento de código que se muestra a continuación, se ilustra cómo es un archivo JSON: la categoría general (en este caso, *lowlevel*) indica los descriptores que cubren en ese bloque; después se detalla cada descriptor acústico y sus valores estadísticos correspondientes.


```
{"lowlevel": { "barkbands_crest": {
      "dmean": 3.67974710464,
      "dmean2": 6.50118494034,
      "dvar": 8.09882926941,
      "dvar2": 23.0071048737,
      "max": 24.6706180573,
      "mean": 12.6164541245,
      "median": 12.1984729767,
      "min": 4.17003440857,
      "stdev": 3.79192757607,
      "var": 14.3787155151},
```


[105] Ver más información en https://essentia.upf.edu/freesound_extractor.html [visitado el 3 de marzo de 2019].

Los módulos *Segmentador* y *Extractor* se implementaron de la misma forma en las dos versiones del *MAT2021*.

A continuación, explicaremos la diferencia que caracteriza los otros módulos del *MAT2021a* y el *MAT2021b.* Pero antes, es conveniente indicar que las versiones *a y b* del *MAT2021* son diferentes entre sí, pues la versión *a* clasifica grupos de sonidos por su misma clase, mientras que la versión *b* identifica sonidos similares a partir de un sonido referente. Además, el *MAT2021a* utiliza el algoritmo *K-means* para realizar conglomerados de los sonidos similares (más adelante nos referiremos a este algoritmo), mientras que el *MAT2021b* es la antesala a la versión final del *CSpectra2021*, pues realiza su medición desde sonidos más cercanos a un sonido de referencia. Este procedimiento lo explicaremos cuando describamos el funcionamiento del *CSpectra2021*.

Entremos entonces a la descripción de los módulos del *MAT2021a*: el integrador y el analizador.

Integrador: este código agrupa la información obtenida de la lectura de los archivos JSON. El objetivo principal es filtrar los datos según las necesidades del usuario y generar un nuevo archivo en formato CSV. El archivo CSV tiene el diseño de una tabla y distribuye el valor de cada descriptor espectral por columna. Debe tomarse en cuenta que el valor que toma como referencia es la media y que, al generar el CSV, estos valores están normalizados de 0 a 1. Además, el archivo CSV dispone los datos de forma estructurada y estos sirven de entrada para el código siguiente.

Tabla 5

Valores de los descriptores acústicos normalizados

Descriptor 1	Descriptor 2	Descriptor 3	Descriptor 4	Descriptor 5
0.49479806	0.49454758	0.54217565	0.4719924	0.47624463
0.47624463	0.48764586	0.48775512	0.50317222	0.44814029
0.48676986	0.53117162	0.49907163	0.51620746	0.56205875
0.53613669	0.50089407	0.47860369	0.49173212	0.51215148

A continuación, se presenta un fragmento del código utilizado:

```
def integrador(root, file):
    file_name, file_extension = os.path.splitext(file)
    with open(root + '/' + file) as f:
        data = json.load(f)
        datosDeUno = []
        datosDeUno.append(file_name)
        datosDeUno.append(data.get('lowlevel').get('average_loudness'))
        datosDeUno.append(data.get('lowlevel').get('mfcc').get('mean'))
```

Analizador: este último módulo del *MAT2021a* tiene la tarea de encontrar grupos similares entre sí. Para llevarla a cabo, primero lee línea por línea los datos en el archivo CSV y los guarda en un arreglo vacío dentro del código. Luego toma todos los datos de este arreglo y les aplica un proceso de normalización con un rango de 0 a 1, para después aplicar a esos datos un algoritmo de clasificación no supervisado llamado *K-means*, el cual agrupa los datos en K grupo. El valor de K lo especifica el usuario de acuerdo con sus necesidades y este le sirve al algoritmo para saber el número de grupos en que dividirá los objetos.

Entonces, el algoritmo toma cada dato, convertido en puntos en un espacio, y coloca los primeros centroides[106] aleatoriamente. Luego, mide la distancia más cercana de los centroides a los puntos, con el fin de saber cuáles puntos son más próximos. Posteriormente, coloca otro centroide a media distancia de los puntos más cercanos y conforma otro grupo. Este proceso de asignación de centroides se realiza todas las veces que sea necesario hasta que los grupos asignados por los centroides ya no se muevan. Es cuando finaliza el ciclo de búsqueda y se obtienen los grupos posibles.

Una vez realizado el proceso de agrupamiento por clase, el *MAT2021a* entrega una lista en formato .txt. Como se ve en el siguiente fragmento, esta tiene dos columnas: la primera indica la clase a la que pertenece el sonido, mientras que la segunda indica el número de fragmento sonoro al que pertenece la clase:

1 parte00000
1 parte00007
1 parte00005
1 parte00010
1 parte00012
1 parte00020
2 parte00006
2 parte00008
2 parte00013
2 parte00015

A continuación, se muestra un fragmento del código que sirvió para distribuir los sonidos en el multipistas digital de *Reaper.* Se decidió correrlo dentro de este

[106] "El centroide de un *cluster* se define como el punto equidistante de los objetos pertenecientes a dicho cluster" (Larrañaga, 2022, p. 2).

programa porque tiene la posibilidad de ejecutar scripts[107] escritos en alguno de los lenguajes informáticos, tales como *EEL, Python o Lua*. Los *scripts* que se programaron para todas las herramientas tecnológicas aquí desarrolladas se escribieron en el lenguaje *Python*:

```
import shutil, os
f = open('texto.txt',"r")
lines = f.readlines()
f.close()

result = []
for x in lines:
        splits = x.split(' ')
        if len(splits) == 2:
                clase = int(splits[0])
                nombreArchivo = splits[1].strip()+".wav"
                print(clase)
                print(nombreArchivo)
                result.append(clase)
                if clase == 0:
                        print('Copiar archivo a folder 0')
                        shutil.copy(nombreArchivo, 'clase_0')
                if clase == 1:
                        print('Copiar archivo a folder 1')
                        shutil.copy(nombreArchivo, 'clase_1')
etc…
```

Al disponer los sonidos en el multipistas digital de *Reaper,* se evidenció la necesidad de que la misma herramienta pudiera concatenarlos y brindar al usuario una textura sonora. También era necesario que la herramienta posibilitara mayor

[107] Se recomienda revisar algunos repositorios con *scripts* hechos para ejecutarse en el programa *Reaper,* los cuales distribuyen, editan y procesan los segmentos de audio, tales como: https://github.com/cfillion/reascripts, https://github.com/MichaelPilyavskiy/ReaScripts, https://github.com/actonDev/Reaper-Scripts, https://github.com/jalovatt/Lokasenna_GUI, https://github.com/AlbertoV5/Reaper-Scripts, https://github.com/fuzzblob/chtammik_Reaper_Scripts [visitado el 3 de noviembre de 2019].

fluidez al usuario en su labor de composición, pues, en esta primera versión, se tenía que ejecutar código por código para obtener resultados preliminares, los cuales posteriormente debían ser retomados por el usuario utilizando otro código que realizara la distribución en carpetas, para luego montar los fragmentos de forma semiautomática. Todo esto implicaba una demora y dificultaba que el usuario pudiera enfocarse en la composición musical.

Además, en el *MAT2021a* los resultados de identificación sonora eran muy generales, pues arrojaba la información de una única categoría perceptual. Esta limitación técnica de la herramienta, aunque se convierte en una potencialidad compositiva que orienta una estética particular, también constituía una limitación en cuanto a la variedad de los sonidos similares obtenidos. Por todo lo anterior, se inició la construcción de la versión *MAT2021b.*

Seguidamente, describiremos la estructura del *MAT2021b*. Sus dos últimos módulos implementan la librería de código *gaia2*[108].

Base de datos (gaia2): este código convierte los archivos de formato *JSON* a *SIG*. La base de datos que utiliza *gaia2* usa los archivos *SIG* para asignarle a cada uno un índice, que en la librería *gaia2* se llaman puntos o *points*.

Gaia2 es una librería para tratar puntos en un espacio métrico. La estructura del espacio métrico está determinada por el objeto *PointLayout*. Entonces, cada punto tiene su representación con el objeto *PointLayout*, el cual contiene, a su vez,

[108] Librería Gaia2 está disponible en https://essentia.upf.edu/gaia/ [visitado el 3 de junio de 2019].

atributos de cada punto. Además, este módulo permite la construcción rápida de un conjunto de datos, como un *DataSet*, el cual se comporta como una colección de puntos. Como proceso final, el código devuelve un archivo .db, que contiene la base de datos construida.

Además, *Gaia2* cuenta con algunos métodos adicionales, como el de limpieza o *Clean,* que asegura que cada conjunto de datos tenga el mismo diseño. A continuación, se muestra un fragmento del código:

```
from gaia2 import *
dsRaw = DataSet()
dsRaw.load("base-datos.db")
dsClean = DataSet()
dsClean.load("base-datos-clean.db")
dsClean.setReferenceDataSet(dsRaw)

ds = DataSet()
directory = "/home/ottocastro/Escritorio/SIG."

for file in os.listdir(directory):
    if file.endswith(".sig"):
        p = Point()
        p.load(os.path.join(directory, file))
        p.setName(file)
        ds.addPoint(p)
```

Consulta (gaia2): este último código brinda al usuario la posibilidad de realizar consultas específicas[109] a la base de datos de *gaia2*. La idea es filtrar aquellos aspectos que le interesen al usuario. La manera en que se realizan las consultas es programándolas desde la estructura del código mismo, como se muestra en el siguiente fragmento:

[109] En inglés, a estas llamadas se les conoce como *query* o *querries*.

```
//Línea para realizar la consulta
filter = 'WHERE value.lowlevel.pitch.mean>440'

//Consulta de un sonido de referencia en este caso el 13.sig
similarSongs2 = euclideanView.nnSearch('13.sig', euclideanMetric,filter).get(10)

//Prepara el archivo txt para devolver los sonidos 10 sonidos con las características
//solicitadas
arch = open("filtro.txt","w")
for eachSnd in similarSongs2:
    arch.write(eachSnd[0][:-4]+ "\n")
    print(eachSnd[0][:-4])
arch.close()
```

El código presentado anteriormente devolvía un archivo en formato .txt, que guarda la información acerca de los sonidos y sus respectivas clases. Debido a que esta herramienta fue desarrollada hasta este punto, este archivo.txt se retoma a través de un código creado en el lenguaje *Python,* el cual lee este tipo de archivo y distribuye los sonidos en carpetas creadas por el usuario con cada una de las clases que el usuario requiera.

En el capítulo tercero, se explicarán con mayor detalle las estrategias de composición usadas a partir de los sonidos similares obtenidos. No obstante, es importante mencionar que las versiones de la herramienta tecnológica *MAT2021* fueron utilizadas de la siguiente manera: *MAT2021a* se implementó en la composición de la obra *Anomalía Urbana* (https://soundcloud.com/mat2021/2-anomalia-urbana), mientras que la versión *MAT2021b* fue utilizada en el desarrollo compositivo de la pieza musical *Texturas Urbanas* (https://soundcloud.com/mat2021/texturas-urbanas). Al final de este capítulo se explican las estrategias utilizadas para la composición de la cuarta pieza: *Ejes Urbanos* creada con la versión final del CSpectra2021 (https://soundcloud.com/mat2021/4-ejes-urbanos).

El CSpectra2021

Este apartado se enfocará en cuestiones técnicas de la *Cspectra2021*. Para los resultados estéticos obtenidos con esta herramienta, se recomienda revisar el capítulo tercero, en donde se detallan los alcances en este ámbito.

El *CSpectra2021* fue el resultado de la suma de las experiencias recogidas en las herramientas *MAT2021a* y *MAT2021b*. Para el *CSpectra2021* se realizaron varios cambios con respecto a las versiones anteriores, los cuales detallaremos más adelante. A modo de resumen, estos son los ajustes:

- Librerías de código más actualizadas que las utilizadas en las versiones anteriores para las funciones de segmentación y concatenación de los audios.
- Implementación del sistema MongoDB*,* como base de datos NoSQL*.*
- Sistema de medición de puntos más efectivo y rápido, con la librería de código Annoy*.*
- Conjunto de descriptores acústicos enfocados en el ámbito espectral.
- Sistema de módulos con funciones determinadas, programadas como una API en *Node.js* para su instalación local del servidor.
- Programación en el lenguaje de JavaScript y Node.js para construir el servidor y crear el puente de comunicación entre el programa Reaper y el servidor, a través de un *script* desarrollado en el lenguaje Python.

Antes de explicar la estructura y los modos de uso de la herramienta *CSpectra2021* para la asistencia a la composición musical, es necesario aclarar ciertas particularidades de dicha herramienta: ante todo, la razón de programar el

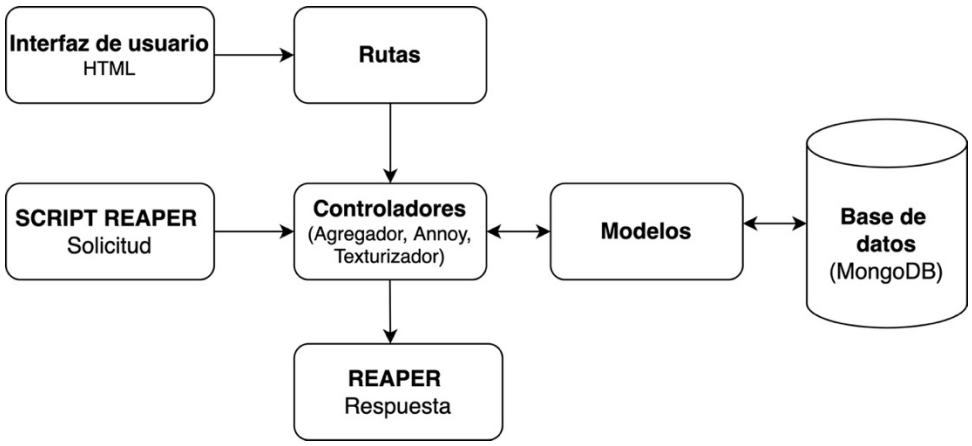
sistema en una arquitectura cliente servidor es disponer de una prevista para, en un futuro, montar la herramienta en un servicio de cómputo en la nube. De esa manera, será posible convertir esta herramienta en una tecnología más accesible para un usuario no especializado en programación que solo la utilice con fines compositivos. Así, la idea para el futuro es convertir esta herramienta en una API[110] web. Sin embargo, debido a los costos económicos que implica este tipo de servicio, por ahora la herramienta es una aplicación para instalarse y usarse en un computador personal, de manera local.

Aspectos generales

A continuación, se presenta la estructura general del *CSpectra2021* y los componentes del sistema (más adelante se abordará la explicación de cada parte de forma más específica):

Figura 19

Diagrama de fujo del CSpectra2021



[110] El término *API* proviene de la abreviatura en inglés de *Application Programming Interfaces*, que en español significa "interfaz de programación de aplicaciones". La *API* permite la comunicación entre aplicaciones a través de un conjunto de reglas.

Antes de referirnos a las partes del sistema modular, es importante mencionar que el primer paso es encender el servidor. Esto se realiza de la siguiente manera: el usuario debe escribir en la *terminal:*

$cd Cspectra2021

Una vez posicionada la ruta desde la *terminal*, se debe escribir:

$node index.js

De esta forma, se inicia el servidor y arranca también el sistema de datos *MongoDB*. Para usuarios no versados en estos aspectos, en el apartado de instalación se explicarán con detalle los requerimientos para instalar el servidor en su computadora. También se recomienda al lector revisar el repositorio de *GitHub* de este proyecto (https://github.com/mat2021/CSPECTRA2021), en donde se detallan los pasos a seguir.

Estructura del CSpectra2021

El *CSpectra2021* está constituido por tres módulos:

- El primero está conformado por las rutas que el usuario debe escribir desde la interfaz. Estas se comunican con el servidor para ejecutar ciertos procesos.
- El segundo lo constituyen los controladores que segmentan, analizan y concatenan los audios, y la comunicación de esta parte con el sistema de base de datos.
- El tercero está constituido por el *script* de Reaper, que posibilita al usuario la selección de un sonido referente. A través de esta solicitud, el servidor

devuelve una textura sonora que monta en el multipistas digital automáticamente.

A continuación, explicaremos en detalle cada una de las partes:

Interfaz de usuario HTML[111]*:*

Este sistema está estructurado en una arquitectura de *cliente servidor*: un *cliente* es una aplicación que permite la comunicación con el servidor. En este caso, el cliente es un navegador de internet que se comunica con el servidor de manera local. Existen varios tipos de navegadores de internet, como *Tor*, *Safari*, *Opera*, *FireFox*, entre otros. Para utilizar esta herramienta se recomienda el uso del navegador *Google Chrome*. El usuario debe escribir en el navegador dos rutas, las cuales se muestran a continuación:

En la primera ruta se debe digitar *localhost:3000/cspectra2021.html,* para que el usuario pueda subir sus audios y que se ejecuten varios procesos que explicaremos más adelante. Es importante recordar que, para la herramienta actual, el servidor está instalado localmente, así que la dirección que el usuario escriba apuntará a la carpeta llamada *public,* en donde está el HTML con el nombre *cspectra2021.html*

En la segunda ruta se debe escribir *localhost:3000/annoybuild,* que tiene el objetivo de crear el archivo *.ann* para brindar información al sistema acerca de la

[111] *HyperText Markup Language,* término en inglés que se traduce al español como "lenguaje de marcado de hipertexto".

medición de los puntos. Más adelante explicaremos con detalle las funciones de esta librería de código.

En la figura 20, se observa la interfaz de usuario del HTML que utiliza el *CSpectra2021*.

Figura 20

Interface de usuario del CSpectra2021



Rutas:

Este módulo se encarga de distribuir las rutas con el fin de direccionar cada una de las partes del sistema para que puedan comunicarse entre ellas y, además, realizar acciones de una determinada tarea en cada ruta. Por tanto, el sistema trabaja de forma modular, flexible y ordenada. Esto es posible gracias al módulo *Express Router,* que ayuda a ejecutar esta distribución en el sistema de la API.

Todas las rutas que se encuentran en este módulo se comunican entre sí de forma interna; el usuario no necesita conocerlas para que se conecten o accionen, sino que el sistema las accesa de manera automática. Algunas de las rutas principales que pone a disposición este módulo

en el sistema son: “/annoybuild”, “/upload” y “/texturabásica”.

Controladores

El módulo de los controladores está compuesto por tres códigos dedicados al análisis y procesamiento de los audios, con el objetivo de encontrar similitudes entre estos y devolver una textura sonora como parte de la concatenación que realiza a los segmentos de los audios seleccionados.

Agregador

Este módulo se encarga, principalmente, de cortar los audios y agregarle a cada uno un número de identificación universal (UUID[112]). Para la segmentación, utiliza la librería de código *Aubio*[113]. [114]Este código llama al módulo *Middleware*[115], con el fin de comprobar el límite de los audios (cada archivo de audio subido al sistema no debe sobrepasar los 200 megas. Si sobrepasan este peso, el sistema registra un error y devuelve al usuario un mensaje en el *script* de *Reaper*, que detallaremos más adelante).

Una vez que cada audio es segmentado, se guarda en la ruta "/resources/static/assets/uploads/". En esta se guardan tanto los audios originales, como los segmentados y los archivos *JSON,* por lo que es un repositorio que puede

---

[112] Conocido en inglés como *Universally Unique Identifier* (UUID), que se traduce al español como "identificador único universal". Etiqueta los archivos a través de 32 dígitos hexadecimales, divididos por guiones de la siguiente manera: 8-4-4-4-12. Un ejemplo de este tipo de numeración es: 442e5511-e50b-23d4-a731-312533440000

[113] Más información en https://aubio.org/ [visitado el 12 de junio de 2019].

114 Al inicio de este capítulo, se explicaron los criterios técnicos en la segmentación de los audios.

[115] El *Middleware* es un pequeño programa o código que se encarga de la gestión de los datos y de la comunicación entre otras partes, según sea el caso, en una API con una estructura de servicio cliente.

alcanzar un peso considerable, debido a la cantidad de archivos y datos que almacena.

Es importante aclarar que, una vez que el usuario clone la versión del *GitHub* para instalarla en su computadora personal, debe escribir esta ruta: "/resources/static/assets/uploads/". Con esto, creará una serie de carpetas vacías para alojar la información operativa que genera para su propio uso. Esta acción debe realizarse para cada usuario que desee instalar el sistema. La ruta está constituida por una serie de carpetas vacías, porque el archivo *.gitignore,* que está alojado en una parte del sistema, avisa al sistema Git cuáles archivos o carpetas de un proyecto puede ignorar. Por esto, cuando se clona la versión del GitHub al escritorio del usuario, no aparecen los archivos de los usuarios anteriores y, por ello, es necesario que el nuevo usuario cree la ruta, con el fin de que el sistema se desempeñe correctamente.

Módulo dedicado a encontrar similitud

En nuestra investigación, uno de los aspectos centrales en el ámbito de la tecnología consiste en que la máquina, mediante el análisis de los audios, pueda encontrar similitudes entre los datos extractados. En este apartado, explicaremos qué tecnología se utilizó para alcanzar este objetivo y cómo funciona.

Para lograrlo, se utilizó Annoy[116], una de las librerías de código más poderosas, flexibles y dinámicas[117] que actualmente está en boga*,* creada por el

[116] Más información en https://github.com/spotify/annoy [visitado el 9 de setiembre de 2021].
[117] Para ver velocidad y eficiencia de la librería *Annoy* con respecto a otras librerías de código, puede consultar: https://github.com/erikbern/ann-benchmarks [visitado el 10 de setiembre de 2021].

programador sueco Erik Bernhardsson[118]. Actualmente, esta librería es utilizada por la plataforma *Spotify*[119] para predecir géneros musicales similares, lo que nos da una idea de su vigencia e impacto en estos momentos.

Annoy ayuda en el procesamiento de datos, para encontrar los puntos más cercanos partiendo de un valor K de referencia. Al ejecutar el procesamiento, esta librería prepara los datos de entrada para darles un formato más accesible que provea un bajo costo computacional. Los datos de entrada corresponden a los vectores extractados por los descriptores acústicos de los sonidos analizados anteriormente.

Ahora bien, para encontrar patrones de similitud en los sonidos, Annoy ejecuta tres etapas; a saber:

1. Preprocesamiento de los datos
2. Cercanía de los puntos
3. Comparación de los puntos para encontrar similitud espectral

Preprocesamiento de los datos

A continuación, explicaremos cómo son las etapas del procesamiento de los datos que lleva a cabo Annoy para encontrar similitud espectral entre los sonidos.

Es importante aclarar que, para llegar a la comparación entre los puntos, los datos obtenidos de los descriptores acústicos deben pasar por la reducción de dimensiones y la normalización de los vectores, para indexarlos. Al mismo tiempo,

[118] Ver su página oficial en https://erikbern.com/ [visitado el 12 de marzo de 2020].
[119] Plataforma de música vía *streaming*.

la longitud de todos los vectores debe ser la misma para facilitar la comparación entre ellos.

La primera etapa de preprocesamiento de los vectores es conocida como *normalización*: aunque sean de la misma longitud, para compararlos entre sí deben tener un rango de magnitud de valores numéricos similar. La normalización es un proceso matemático con el cual se promedia en rangos similares el conjunto de elementos por comparar. Existen varios tipos de normalización; la que se muestra en este trabajo es solo una de ellas. En este caso, el proceso de normalización consiste en que la máquina toma la magnitud máxima y la transporta a un rango o nivel específico, y procede con los otros valores de manera proporcional al valor máximo que definió. Este proceso matemático se ejecuta con todos los conjuntos de arreglos numéricos. Por lo general, el rango al que se proporcionan las diferentes magnitudes es de -1 a 1 o de 0 a 1.

La siguiente etapa de preprocesamiento del vector consiste en un procedimiento conocido como reducción de dimensiones: reduce la dimensionalidad en la representación de los datos contenidos en el vector, con el fin de simplificar gran cantidad de datos y obtener una representación del sonido analizado. De este modo, disminuye el costo computacional de la máquina.

Finalmente, una vez que los datos están preprocesados, la librería Annoy implementa el algoritmo *K-medias*, el cual agrupa objetos por similitud, con el fin de que las clases identificadas tengan las mismas características entre ellas. Es una de las técnicas de análisis dentro del campo del aprendizaje automático.

Creación de la base de datos

En esta investigación, utilizamos el sistema de base de datos[120] *MongoDB* NoSQL. En una base de datos NoSQL, los datos se almacenan en estructuras dinámicas de archivo, como los formatos JSON y Yamls, a diferencia de las bases de datos estructuradas[121], en las cuales la información se guarda en tablas. En MongoDB, los vectores pasan por un proceso de indexación, que servirá posteriormente para identificarlos y ubicarlos como puntos en un espacio multidimensional[122].

A continuación, se muestran algunos comandos de los que dispone el sistema de base de datos MongoDB que pueden ser útiles al lector para obtener información acerca de la base de datos, pero que no son imprescindibles para el funcionamiento de la herramienta[123].

Algo importante de señalar es que MongoDB creará una base datos de forma automática si no se le da alguna otra instrucción. Todas las instrucciones que se mencionan a continuación se deben escribir desde la *terminal*.

El usuario entra al sistema de base de datos MongoDB con la instrucción *$mongo.* Si desea crear una base de datos diferente, puede utilizar *$use*

---

[120] "Una base de datos es una recopilación organizada de información o datos estructurados, que normalmente se almacena de forma electrónica en un sistema informático." (Oracle, 2021)

[121] "Las bases de datos se hicieron predominantes en la década de 1980. Los elementos de una base de datos relacional se organizan como un conjunto de tablas con columnas y filas. La tecnología de bases de datos relacionales proporciona la forma más eficiente y flexible de acceder a información estructurada." (Oracle, 2021)

[122] Para conocer más acerca de cómo se implementa *MongoDB* al sistema, el lector puede referirse al apartado de instalación que está más adelante en este mismo capítulo.

[123] Se puede ver la lista completa de comandos para *MongoDB* en https://www.mongodb.com/es/basics/create-database [visitado el 13 de febrero de 2019].

DATABASE_nombreasignado; pero si quiere ver la base de datos con el nombre asignado, debe escribir *$use nombreasignado*. Para verificar la base de datos, escribirá $db. Por último, para confirmar la lista de la base de datos, la instrucción es $show dbs. Estos comandos pueden ayudar al usuario a conocer el estado de los datos que se han subido al sistema.

Ahora bien, dentro del sistema de MongoDB existen otros programas que facilitan ciertos procesos en la estructuración de los datos. Uno de estos se llama Mongoose[124], que se instala como parte de los paquetes de npm[125] (como se explicará en el apartado de instalación más adelante). *Mongoose* modela los datos y establece una forma particular de información, al filtrar y descartar aquella que no esté declarada en el Schema *o* esquema. A continuación, ilustramos con parte del código el Schema escogido para el *CSpectra2021*:

```
const mongoose = require('mongoose');
const Schema = mongoose.Schema;

var audioSchema = new Schema({
  nombre:    { type: String },
  analisis:  {type: Object},
  statistics:  {type: Object},
  annoyid: {type: Number}
});
```

Medición de los puntos en un espacio multidimensional

En la librería de código Annoy, la medición de los puntos o índices del análisis

[124] Ver más información en https://mongoosejs.com/ [visitado 15 de octubre de 2010].
[125] Es un sistema de gestión de paquetes por defecto para Node.js, un entorno de ejecución para JavaScript.

de cada sonido conlleva a que la máquina muestre una posible agrupación de sonidos similares. Explicamos a continuación cómo se ejecuta este proceso.

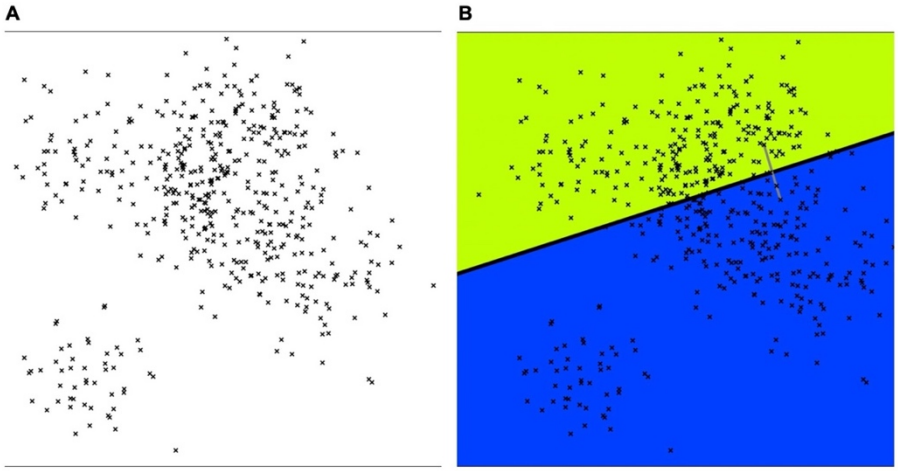
Recordemos que, del análisis que realizaron los descriptores acústicos, se derivó una serie de vectores que representan las características acústicas de cada sonido. Una vez que se obtiene esta información y que se ha subido a la base de datos (MongoDB), se convierte en índices. Dichos índices son colocados como puntos en un espacio de n-dimensiones; cada punto contiene las características acústicas en sus vectores. Cuando se concreta este proceso, Annoy procede de la siguiente manera:

1. Los puntos se colocan en un espacio de n-dimensiones. Ver figura 21A[126].
2. Annoy elige dos puntos aleatoriamente y luego los divide de forma equidistante por un hiperplano. En la figura 21B, se observa el hiperplano como una línea negra que divide el área de color verde del ámbito de color azul. La medición de los dos puntos está indicada por una línea de color gris.

[126] Debido que es imposible la representación gráfica de un espacio de n-dimensiones, aquí se hace una representación en dos dimensiones.

Figura 21

Representación en dos dimensiones de los puntos en el espacio

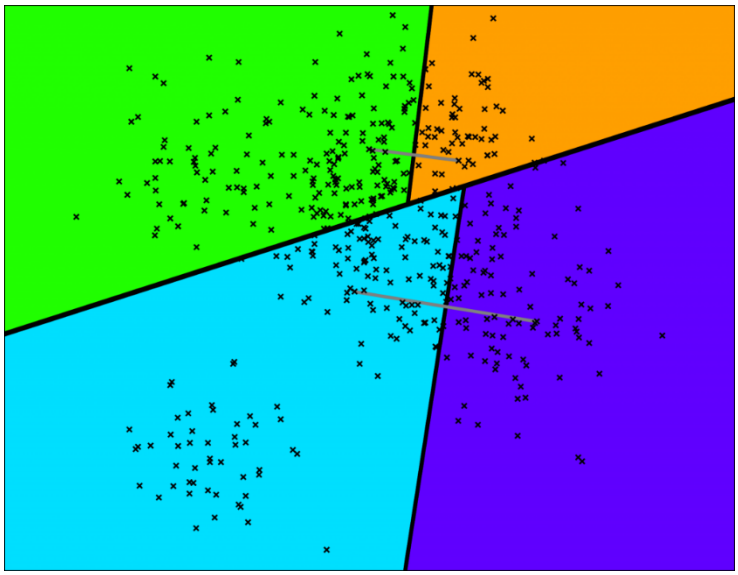


Tomado de Bernhardsson, 2020.

3. El siguiente paso es tomar otros dos puntos al azar de este espacio y volver a dividir a la mitad, de forma equidistante, como se observa en la figura 22.
4. El sistema informático va guardando la relación entre cada uno de los puntos (podemos verlo ejemplificado en la figura 22).
5. Las divisiones entre dos puntos que son atravesados por un hiperplano se repiten una y otra vez hasta completar los valores K que el usuario ha solicitado al sistema informático.

Figura 22

División a través de un hiperplano del espacio



Tomado de Bernhardsson, 2020.

6. Cada vez que *Annoy* hace una iteración, se realizan procesos como eliminación de datos duplicados y medición de las distancias de unos nodos con otros, con el fin de saber cuáles nodos son más cercanos entre sí.
7. El sistema informático encuentra posibles resultados de todas las iteraciones realizadas. En este punto es importante aclarar que, cuando la máquina detiene su búsqueda, es porque encontró un posible resultado óptimo. Es decir, que el sistema informático se detendrá siguiendo la función de disimilitud que afirma: "*Cuanto menor sea dicho valor, mayor será la coincidencia y, por lo tanto, la probabilidad de que se trate de un mismo objeto*" (Domínguez, 2021, p. 50). En esta investigación, se sigue la hipótesis de que los puntos más cercanos en un espacio multidimensional determinan una posible similitud tímbrica.

Texturizador

El texturizador se encarga de seleccionar, mediante la librería de código Annoy, los sonidos similares tímbricamente, los concatena utilizando la librería Sox[127] y devuelve una textura sonora a partir de este procesamiento. Dicho procesamiento consiste en hacer un *fade in* y *fade out*[128] en el ámbito de la amplitud, con el fin de que la concatenación tenga un acabado más homogéneo. Además de realizar una concatenación, yuxtapone cada segmento con una transición de 100 milésimas de segundo.

A continuación, se observa un fragmento del código que realiza esta función.

```
function applyFadeINOUT(fileInput, res) {
  var baseroot = __basedir + "/resources/static/assets/uploads/";
  getAudioDurationInSeconds(baseroot + fileInput).then((duration) => {
    console.log(duration);
    comando = "sox ";
    comando += baseroot + fileInput + " ";
    comando += baseroot + fileInput.slice(0, -4) + "_FADE.wav" + " ";
    comando += 'fade ' + 0.5 + ' ' + duration + ' ' + 0.5;
```

Script de Reaper

Como se explicó, Reaper da la opción al usuario de programar sus propios *scripts*. Para esta investigación, se creó un *script* en el lenguaje de programación Python, con el fin de permitir la comunicación entre el Reaper y el servidor.

Explicaremos a continuación la forma en que se usa este *script* en Reaper. Los pasos para seguir son:

---

[127] Ver más información en http://sox.sourceforge.net/ [visitado 10 de marzo de 2010].
[128] *Fade in* y *fade out* se pueden traducir en el idioma español como transiciones que van de muy poca a mucha amplitud, en el caso del *fade in,* y de manera inversa para el *fade out.*

1. Se abre la terminal y desde allí se escribe la siguiente instrucción:

   $node index.js

   En caso afirmativo, debe dar este mensaje de respuesta en la consola de la terminal:

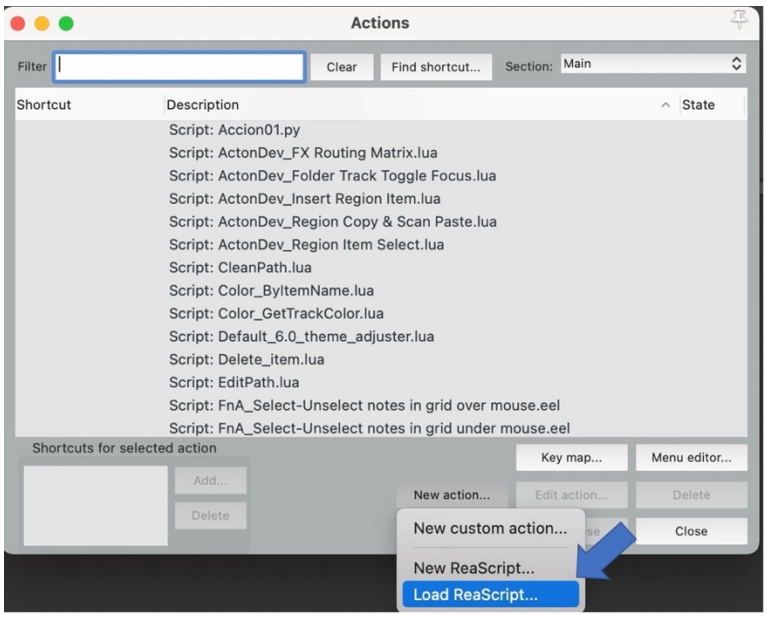
   Base de datos creada OK

   Running at localhost:3000

2. Se inicia el programa Reaper; se carga el script con el nombre *csp_render.py,* que viene en la carpeta llamada *2. Script_reaper*. Para cargar este script, se va al menú de Reaper con la ruta: *Actions>Load ReaScript* y se selecciona el *csp_render.py*, como se ve en la figura 23.

Figura 23

Captura de la pantalla del programa Reaper que muestra cómo cargar un script

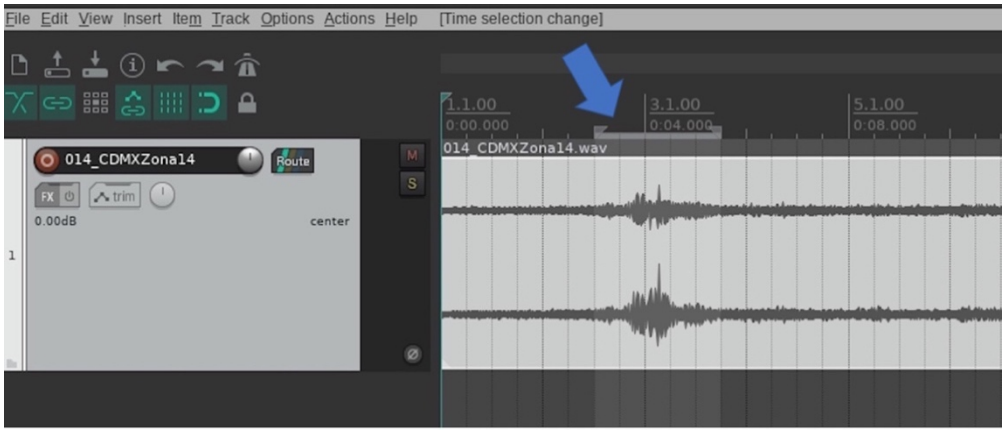


3. Se selecciona un fragmento del audio que se desea tomar como sonido de

referencia. Este segmento no puede ser mayor a 5 segundos.

Figura 24

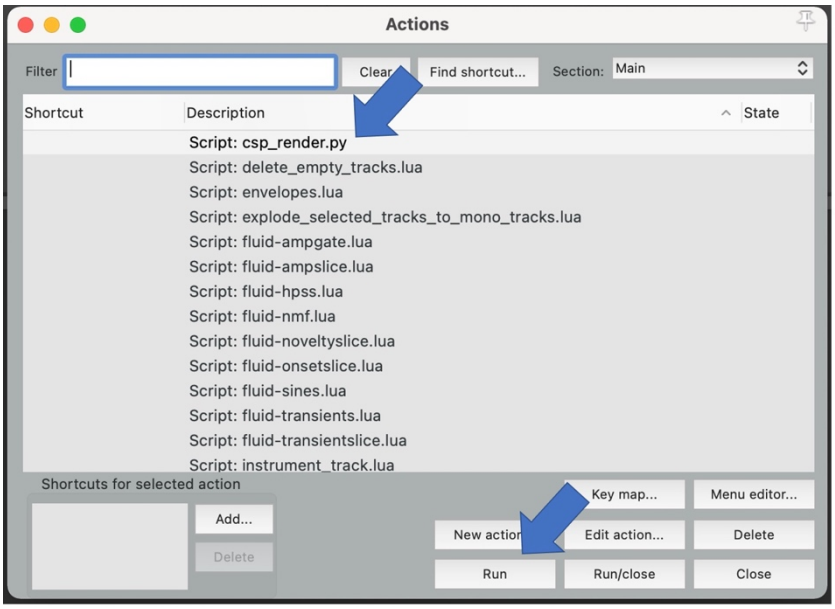
Selección de un segmento sonoro



4. Se selecciona el script *csp_reaper.py* y se ejecuta con el botón *Run*. Es importante anotar que el programa Reaper tiene atajos o *shortcuts* para automatizar la ejecución de scripts. Lo anterior puede tomarse en cuenta si se quiere que el proceso sea más rápido.

Figura 25

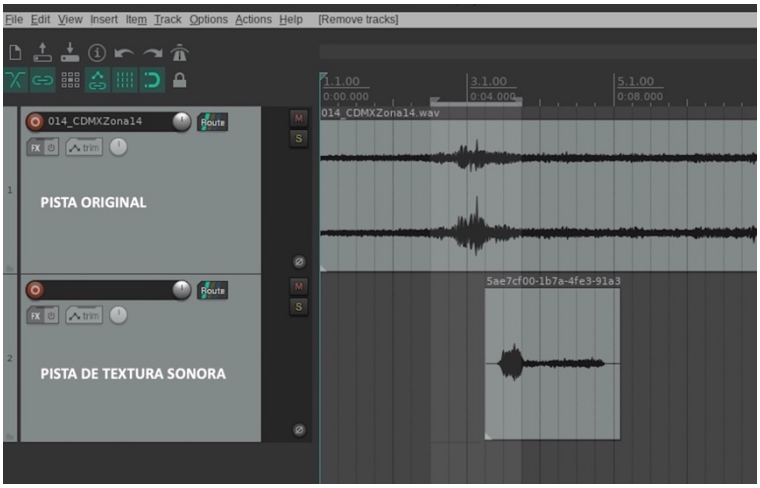
Selección y ejecución del script csp_reaper.py



5. Al ejecutar el *script*, esta llama al servidor, envía el segmento de audio para su análisis acústico y se realizan los procesos de búsqueda de similitud con otros segmentos. El servidor crea la textura sonora, ya concatenada con los diferentes segmentos de audio, y la monta como respuesta a su petición en el multipista de Reaper. En la figura 26, se observa la pista original, la cual proveyó el segmento de referencia, y la pista de textura sonora, que es la devuelta por el sistema.

Figura 26

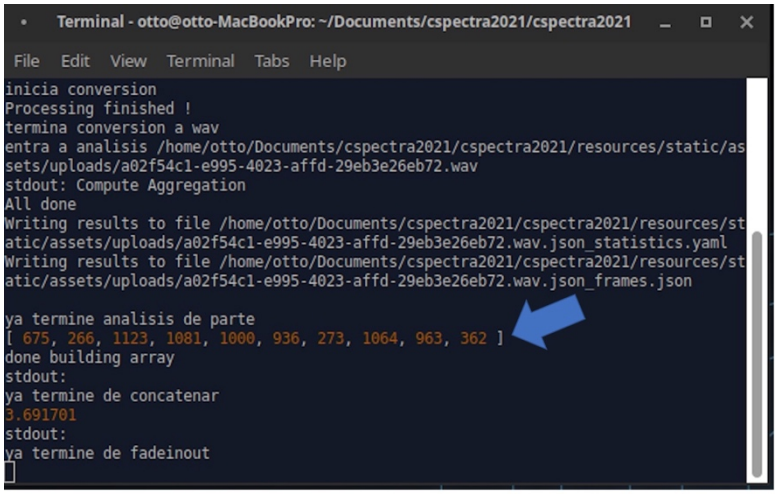
Pista original y textura sonora resultante



6. Después de dejar la textura sonora en el multipista de Reaper, el usuario puede ver la consola de la terminal para informarse acerca de con qué segmentos hizo la concatenación, como se ve en la figura 27.

Figura 27

Información desde la consola de la terminal con respecto a los segmentos seleccionados para conformar la textura sonora



En síntesis, a partir de este recorrido instructivo y de manejo operativo del *MAT2021* y del *CSpectra2021*, se concluye que el análisis espectral de un sonido mediante el uso de tecnología especializada para la creación musical brinda al compositor posibilidades y vinculaciones sonoras que de manera manual serían difíciles de lograr, debido a la velocidad, potencia y recursos de las computadoras actuales. Además, la identificación y organización de sonidos por su nivel de similitud espectral, a través de la mediación de la máquina, abre otras posibilidades de desarrollo de los materiales sonoros utilizados en una composición musical: puede favorecer otro tipo de taxonomías o evidenciar relaciones y vinculaciones diversas con los materiales sonoros en la microestructura y macroestructura de una pieza musical, que de otra manera pasarían inadvertidas para el compositor y, por tanto, no las implementaría. Así, estas nuevas formas facilitan un posicionamiento muy diferente a la identificación sonora humana y desarrollan otras posibilidades de

escucha que confrontan el proceso aural en el acto de componer. Esto, además, permite nuevas posibilidades de construcción del discurso en una obra sonora.

Capítulo 3:

Estrategias de composición musical

En el presente proyecto, las estrategias de composición se piensan desde las características tímbricas de los sonidos. A partir de estas, se establecen comparaciones, vinculaciones y desarrollos mediante procesos de análisis en dos vías: por un lado, el análisis espectral desde la máquina y, por otro, el análisis perceptual desde mi escucha, lo que permite el diálogo entre la máquina y mi percepción, para ser implementado en la creación sonora. Ambas vertientes benefician el proceso compositivo y la concreción de las obras. Más allá de los materiales obtenidos de los procesos maquinales, las decisiones tomadas con respecto a la articulación de los elementos son sumamente importantes dentro del proceso compositivo.

Así, por ejemplo, desde el ámbito de la máquina, la identificación de los sonidos a través de ciertos descriptores y procesos de comparación permite encontrar similitudes por medio de un sonido referente. Estos procesos arrojan resultados que, a su vez, son percibidos y catalogados por mi escucha, tratando de buscar aspectos tímbricos que permitan vincular algunos materiales sonoros con otros. De este modo, se obtienen relaciones que ayudan al desarrollo musical de las piezas compuestas en esta investigación. Al respecto, recomiendo al lector revisar el capítulo dos, en el cual se explica cómo la máquina realiza el proceso de identificación y comparación de los sonidos.

En el presente apartado nos enfocaremos en las estrategias de composición, desde el ámbito de mi percepción, totalmente subjetiva, retomando los materiales obtenidos de los procesos descritos en el capítulo dos.

Cuatro posibilidades de color tímbrico en las piezas musicales *Mosaico Sonoro, Anomalía Urbana y Texturas Urbanas*

Existe una gran cantidad de adjetivos que se han utilizado para describir, en la música, las diferentes calidades tímbricas del sonido. A modo de ejemplo, obsérvense las siguientes citas extraídas de tratados y libros canónicos de orquestación musical:

> El sonido se percibe vidrioso y metálico. (Piston, 1955, p. 19)
>
> Produce un ataque alegre, ágil y sobrearticulado. (Adler, 1989, p. 26).
>
> El sonido misterioso, velado y bastante agonizante de la trompeta Fa. (Berlioz y Macdonald, 1884/2004, p. 260)

Este tipo de adjetivos, utilizados en los métodos de orquestación y en la música en general, constituyen un recurso de identificación tímbrica con un objetivo pedagógico: mostrar a los estudiantes de composición musical las posibilidades de color, tanto de cada instrumento, como de sus combinaciones. También se utiliza como un método para facilitar la memoria en la clasificación taxonómica al momento de componer. Unido a esto, los adjetivos que describen los distintos timbres en la orquestación instrumental se usan con el fin de definir y combinar los recursos con los que los instrumentos cuentan en la composición de la pieza o como parte de un arreglo musical; es decir, como una posibilidad de “coloración” tímbrica.

Un acercamiento actual a los conceptos de la orquestación instrumental en su implementación en la música electroacústica es el que propone el compositor

argentino Mario Mary (1961-)[129], quien la ha designado como “orquestación acusmática” (2016). En el presente trabajo, la propuesta de Mary ha tenido una gran influencia; sin embargo, entendemos el timbre como un recurso de desarrollo y vinculación de los materiales sonoros, que abarca aspectos que Mary propone únicamente como posibilidades de distribución de los sonidos para lograr determinados colores en una obra sonora. Esto por cuanto, para Mary, el color tímbrico se contempla como una forma análoga en la orquestación instrumental, pero en el ámbito acusmático.

Ahora bien, el color tímbrico ha sido implementado convencionalmente en ámbitos como la composición musical, la enseñanza de la orquestación instrumental y la acusmática. Sin embargo, en esta investigación, se implementó el uso de adjetivos relacionados al timbre para establecer categorías perceptuales que identificaran los resultados obtenidos por la máquina. Dichas categorías permiten desarrollar los materiales al establecer niveles de similitud; y estas similitudes, desde mi experiencia en obras anteriores, se convierten en rasgos distintivos de cualidades tímbricas.

Así pues, para esta investigación se proponen cuatro categorías de identificación: brillo, opaco, rugoso y liso. Los primeros dos adjetivos remiten a una metáfora visual, mientras que los dos últimos, a una metáfora táctil. Las categorías se establecen a partir de mi percepción subjetiva, antes, durante y después de que

[129] Ver más información en http://ipt.univ-paris8.fr/mmary/bios.html [visitado el 22 de agosto de 2021]

la máquina realice su proceso de identificación automática y arroje sus resultados sonoros.

Los resultados brindados por el sistema informático constituyen el inicio de un proceso compositivo, en donde necesito tomar diferentes decisiones sobre la selección, el montaje, la yuxtaposición, la edición y el procesamiento de los sonidos, con el fin de lograr una dirección musical de la pieza en construcción. De allí que la herramienta informática sea una asistente en la composición, nunca un sistema autónomo de creación de obras sonoras.

A continuación, se definen las cuatro categorías que propongo a nivel perceptual, desde mi escucha y experiencia en la composición de obras electroacústicas:

Sonido brillante: es aquel material sonoro que se percibe como anguloso, incisivo, penetrante, sobre todo debido al reforzamiento que tiene en sus frecuencias agudas. Un ejemplo de material sonoro brillante es el siguiente:

Ejemplo sonoro 3.1.

Material sonoro brillante en Anomalía Urbana de 3'53" a 4'07"

https://ottocastro.com/o/146

Sonido opaco: es aquel que se percibe velado, poco definido en su contorno sonoro y ronco; da la impresión de ser un registro grave. En el ejemplo 3.2, se escuchan series de sonidos de motores de autobús que fueron grabados de manera muy

cercana. Estos sonidos fueron yuxtapuestos entre ellos para lograr una continuidad en su morfología, a manera de un pedal armónico.

Ejemplo sonoro 3.2.

Material sonoro oscuro en Texturas Urbanas de 0'00" a 0'30"

https://ottocastro.com/o/064

Sonido rugoso: en el caso de los sonidos rugosos, estos, para mí, se perciben de manera estriada y granulosa. Presentan un comportamiento rítmico aleatorio interno; cuando se escuchan dan la sensación de "arenosidad" sonora, una especie de disonancia entre sus materiales internos, con un componente inarmónico presente. Dan la percepción de algo cortado, segmentado o inacabado. Un ejemplo de rugosidad se muestra en el ejemplo 3.3.

Ejemplo sonoro 3.3.

Material sonoro rugoso en Mosaico Sonoro de 0'00" a 0'33"

https://ottocastro.com/o/113

Sonido liso: los materiales sonoros lisos son armónicos; desde mi percepción, dan la sensación de plenitud y continuidad. Se escuchan pastosos y suaves, poco impulsivos y percutidos.

Ejemplo sonoro 3.4.

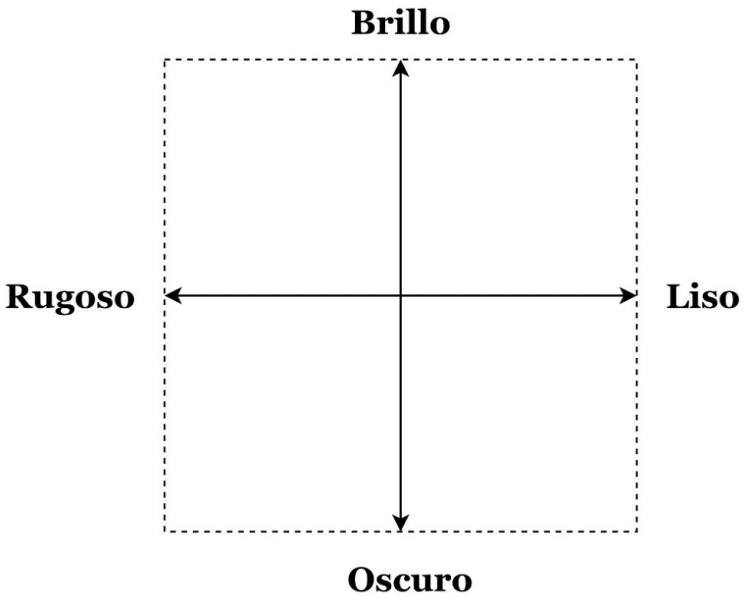
Material sonoro liso en Anomalía Urbana de 2’08” a 2’10”

https://ottocastro.com/o/145

Ahora bien, a partir de las cuatro categorías propuestas (brillo, opaco, rugoso y liso), elaboré una representación gráfica con el fin de situar cada sonido en un contexto comparativo de similitud y establecer una escala que los relacione. Esta representación tiene el propósito de facilitar el entendimiento de los materiales en el proceso de composición de la pieza musical.

Figura 28

Disposición de las cuatro categorías perceptuales utilizadas en las tres piezas compuestas



Dado que la máquina entrega sus resultados de similitud de forma automática, sin ninguna valoración estética, a partir de las peticiones que el

compositor le solicita, es este último quien debe establecer un plan que dé dirección a estos materiales para lograr musicalidad entre ellos. Para esto, una de las primeras estrategias que utilicé como método para comprender qué tipo de vinculación podría tener cada sonido fue categorizar de forma subjetiva y aproximada, en este plano gráfico, los resultados que había obtenido de la máquina y que deseaba vincular.

Si bien es cierto, en ocasiones las características perceptuales de los sonidos eran evidentes, en la mayoría tuve que utilizar este método de representación, pues las características no eran tan claras. Así, con este método, se puede determinar no solo qué posibilidades de similitud de un parámetro existen con respecto a los otros materiales sonoros, sino también el nivel de contraste que hay entre ellos. La similitud es importante como recurso de continuidad, pero también el contraste es fundamental en la composición musical como elemento de tensión.

Ahora bien, en relación con la similitud, podemos apreciar el ejemplo 3.5, en donde la similitud de la frecuencia fundamental de cada sonido sugiere un desarrollo de los materiales. En esta, el oído puede establecer una continuidad entre los sonidos, a pesar de que provienen de diversas fuentes sonoras.

Ejemplo sonoro 3.5.

Continuidad a través de la similitud en Anomalía Urbana en 1'08" a 1'25"

https://ottocastro.com/o/092

Por otro lado, el recurso de contraste tímbrico se aprecia en el ejemplo sonoro 3.6: aunque los dos materiales sonoros que se escuchan pertenecen a la categoría de rugosidad, el primero contiene más características de opacidad que el segundo, el cual es mucho más rugoso y brillante tímbricamente. Como se mencionó, el contraste puede ser utilizado como un recurso de tensión musical; es evidente que, con este, se logra una separación entre los materiales.

Ejemplo sonoro 3.6.

Tensión y separación de los materiales sonoros a través del contraste tímbrico en Anomalía Urbana en 7'50" a 8'00"

https://ottocastro.com/o/036

El cuadro de representación que se propone en este trabajo tomando los resultados de la máquina y llevando a cabo una categorización desde mi escucha es útil en los casos en que los sonidos pertenecen a una categoría, pero pueden poseer características de otras categorías perceptuales. Por ejemplo, en las muestras 3.7 y 3.8 se escuchan dos sonidos que pertenecen a la categoría de rugosidad; pero el 3.7 es más brillante tímbricamente que el 3.8.

Como ejemplos contrastantes, se presentan las muestras sonoras 3.9 y 3.10. El ejemplo 3.9 presenta más lisura que los anteriores, pero mantiene características de rugosidad y más brillo que el ejemplo sonoro 3.10, el cual, a su vez, es más oscuro, liso y rugoso que las muestras 3.7 y 3.8. Todo esto puede observarse de forma gráfica en la figura 29.

Ejemplo sonoro 3.7.

Mosaico Sonoro de 0'19" a 0'26"

https://ottocastro.com/o/148

Ejemplo sonoro 3.8.

Mosaico Sonoro de 1'29" a 1'38"

https://ottocastro.com/o/149

Ejemplo sonoro 3.9.

Anomalía Urbana de 2'09" a 2'22"

https://ottocastro.com/o/176

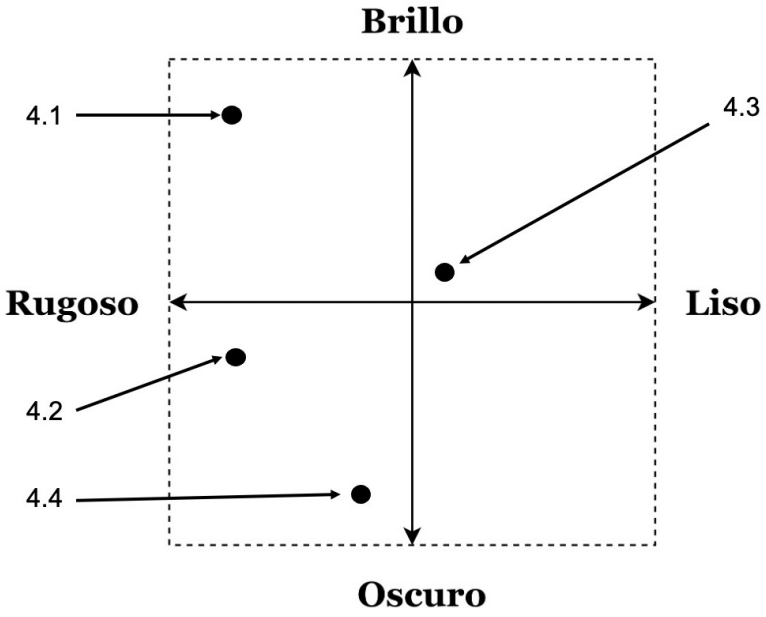
Ejemplo sonoro 3.10.

Anomalía Urbana de 8'55" a 9'03"

https://ottocastro.com/o/151

Figura 29

Representación de cuatro fragmentos sonoros a través de las cuatro categorías perceptuales



Similitud y contraste tímbrico en las obras: *Mosaico Sonoro, Anomalía Urbana y Texturas Urbanas*

En este apartado, haremos referencia a dos conceptos estéticos: la similitud y el contraste tímbricos.

Para efectos de la presente investigación, la similitud tímbrica se considera un recurso de continuidad, mientras el contraste tímbrico es un recurso de separación. De estas dos categorías, en las obras *Mosaico Sonoro, Anomalía Urbana y Texturas Urbanas,* la primera funcionó como un principio de desarrollo, mientras la segunda, como una manera de establecer tensión musical.

Ahora bien, la similitud tímbrica se define como la aparente cercanía perceptual entre dos sonidos; además, puede ser una forma de lograr vinculación entre dos fuentes sonoras diferentes a través de uno o más aspectos similares entre ambos sonidos. Cuando nos referimos a similitud, esta no implica que los sonidos sean iguales, sino que tienen elementos parecidos que los vincula perceptualmente entre sí. Por ejemplo, en el ámbito sonoro, dos sonidos similares podrían ser la sirena de una ambulancia y el maullido de un gato, pues es posible agruparlos por su brillo espectral, su frecuencia modulante y su color tímbrico nasal, entre otras características perceptuales que los relacionan. Así, aunque los dos sonidos son muy diferentes semánticamente, pueden corresponderse entre ellos por algunas de sus características acústicas.

Algunas veces la similitud entre una y otra categoría perceptual no es del todo evidente, ya sea porque los dos sonidos contienen otras categorías más acentuadas o porque se enmascaran con otro elemento acústico. Pero cuando existen categorías perceptuales parecidas, se percibe en los sonidos mayor fluidez o flujo sonoro, lo que permite la continuidad y el desarrollo de los materiales cercanos a la estética de un pedal armónico. Es decir, la similitud produce una sensación de fluidez, mientras que el contraste, una percepción de rítmica interna de los materiales sonoros.

A continuación, se presentan algunos ejemplos de continuidad tímbrica. Para categorizar el nivel de similitud entre los sonidos, se establece una escala de 1 a 10, donde 10 es el valor más alto y 1 el valor más bajo de similitud entre los materiales sonoros. Estos valores son subjetivos y sirven para ubicarlos

comparativamente. La designación de los valores correspondientes parte de mi escucha como compositor.

Ejemplo sonoro 3.11.

Continuidad tímbrica 10, Texturas Urbanas de 1'00" a 1'21"

https://ottocastro.com/o/152

Ejemplo sonoro 3.12.

Continuidad tímbrica 8, Anomalía Urbana de 2'27" a 2'43"

https://ottocastro.com/o/087

Ejemplo sonoro 3.13.

Continuidad tímbrica 6*, Anomalía Urbana de 1'46" a 2'01"*

https://ottocastro.com/o/153

Ejemplo sonoro 3.14.

Continuidad tímbrica 2, Anomalía Urbana de 7'29" a 7'41"

https://ottocastro.com/o/073

A partir de los ejemplos anteriores, puedo concluir, desde la subjetividad de mi escucha, que la continuidad tímbrica se logra por la similitud entre sonidos de un mismo tipo de fuente sonora o similar, como en los ejemplos 3.11 y 3.12. Por otro lado, es posible lograr otra sensación de continuidad más abstracta a partir de un

mayor número de fuentes sonoras de diversos orígenes, como en los ejemplos 3.13 y 3.14.

Asimismo, de manera complementaria, en las piezas de esta investigación, el contraste tímbrico fue utilizado para conseguir mayor realce en el desarrollo continuo y como descanso auditivo de las partes de flujo sonoro más similar, delimitando las zonas de continuidad tímbrica. Por tanto, se usó en la composición de las piezas como un recurso estético para evidenciar la separación entre los materiales sonoros.

Por su parte, el contraste se ha utilizado en la música en muchas ocasiones para lograr un punto de giro en la dirección musical o como recurso para incrementar la atención del oyente, así como permitir una mayor diversidad estética con el fin de enriquecer el discurso musical.

A continuación, se muestran cuatro ejemplos de contraste tímbrico. En el ejemplo 3.15, el contraste se utiliza para acrecentar el desarrollo de los materiales a través de la similitud tímbrica; además, sirve como un desvío de dirección musical, dentro de la construcción de la pieza para favorecer el interés de la escucha. Este recurso se utiliza a menudo con ese mismo objetivo en las tres piezas compuestas para esta investigación.

Ejemplo sonoro 3.15.

Texturas Urbanas de 2’49” a 2’59”

https://ottocastro.com/o/055

Ejemplo sonoro 3.16.

Texturas Urbanas de 8’18” a 9’17”

https://ottocastro.com/o/043

Ejemplo sonoro 3.17.

Mosaico Sonoro de 1’52” a 2’00”

https://ottocastro.com/o/154

Ejemplo sonoro 3.18.

Mosaico Sonoro de 6’58” a 7’03”

https://ottocastro.com/o/155

En los ejemplos 3.15 y 3.16, el contraste tímbrico se lleva a cabo a través de un cambio de brillo en disminución; además, la frecuencia fundamental se mantiene. Por su parte, en los ejemplos 3.17 y 3.18, el cambio sustancial se da de sonidos armónicos a inarmónicos, donde el contraste muta de opaco a brillante. En estos ejemplos, no existe una frecuencia fundamental que englobe los fragmentos, de manera contraria a los ejemplos 3.15 y 3.16.

El enmascaramiento como recurso compositivo en la pieza *Texturas Urbanas*

El *enmascaramiento tímbrico* se produce cuando existen simultáneamente dos sonidos con determinadas categorías tímbricas, y uno de los dos envuelve de

manera enfática al otro, haciendo difícil y compleja su claridad. El compositor Rodrigo Sigal define enmascaramiento como un "[...] *método útil para disfrazar un sonido*" (2014, p. 64). Este recurso se utilizó en las piezas con el fin de provocar unión de las categorías, pues el enmascaramiento provoca una especie de fusión de los colores tímbricos presentes.

La situación más usual es el enmascaramiento de un sonido a otro a través de su amplitud, pero también puede darse mediante la constitución tímbrica de uno hacia otro u otros. Por ejemplo, un sonido muy brillante tímbricamente, sobre otros más oscuros o rugosos, puede provocar enmascaramiento, o también puede darse mediante la superposición de sonidos muy rugosos encima de otros sonidos de textura lisa. En el ejemplo 3.19, se escucha la presencia de tres categorías: brillante, rugosa y lisa; esta última es la que produce el enmascaramiento con respecto a las otras dos. Unido a ello, existe una especie de pedal armónico que ayuda a que la amalgama de las categorías se perciba con más intensidad. En este caso, no se escucha ningún plano sonoro, sino un solo plano embadurnado y mezclado sin ninguna dirección, por lo cual se percibe un conglomerado de sonidos y un factor alto de densidad tímbrica.

Ejemplo sonoro 3.19.

Texturas Urbanas de 3'31" a 3'41"

https://ottocastro.com/o/156

Otro ejemplo de enmascaramiento se muestra en el ejemplo 3.20, en donde una textura sonora con doble categoría (en este caso, rugosa-opaca) envuelve poco a poco su categoría lisa, que además se caracteriza por tener algunas frecuencias bien definidas.

Ejemplo sonoro 3.20.

Texturas Urbanas de 3'44" a 3'58"

https://ottocastro.com/o/052

Atenuación y reforzamiento del color tímbrico en las obras *Anomalía Urbana y Texturas Urbanas*

La atenuación y el reforzamiento son recursos que se usaron en las piezas compuestas para dar énfasis o suavizar una categoría perceptual utilizada.

El reforzamiento tímbrico se incorporó cuando se deseaba enfocar o resaltar un gesto musical determinado, dando fuerza a la intención del color tímbrico. En el ejemplo 3.21, se puede escuchar mayor brillo, provocado por la acumulación de fragmentos tímbricamente brillantes con el objetivo de llegar a un pulso que coincide con un cambio contrastante. Es decir, esta acumulación de brillo ayudó estéticamente a lograr un punto de llegada.

Ejemplo sonoro 3.21.

Texturas Urbanas de 1'58" a 2'10"

https://ottocastro.com/o/157

También es posible obtener una demarcación tímbrica general utilizando categorías tímbricas contrastantes. En el ejemplo 3.22, se dan dos categorías muy evidentes y separadas en cuanto a sus características tímbricas: lo rugoso y lo liso. Se escuchan de manera muy distanciadas una de otra, con lo que se acrecienta más el contraste sonoro como resultado final.

Ejemplo sonoro 3.22.

Texturas Urbanas de 7'34" a 8'18"

https://ottocastro.com/o/047

Por otro lado, la atenuación fue un recurso utilizado en la composición de las piezas para suavizar sonidos con determinada categoría tímbrica. Este tipo de recurso es cercano al enmascaramiento, en el sentido de que una categoría envuelve a otra o a varias, pero la diferencia radica en que la atenuación le agrega a una categoría, como la rugosa, cierta sensación de suavidad. En el ejemplo 3.23, el sonido de los motores, que es de carácter filoso y muy presente, se escucha difuminado por la intervención del sonido continuo y vaporoso de la categoría lisa.

Ejemplo sonoro 3.23.

Anomalía Urbana de 7'34" a 8'18"

https://ottocastro.com/o/158

La densidad espectral como principio de tensión musical en la pieza musical *Mosaico Sonoro*

En esta investigación, se entiende la densidad espectral como un recurso estético relacionado con la cantidad de energía espectral existente en el ancho de frecuencia y su distribución. Sin embargo, no deseamos ser específicos en la forma en que se adquiere dicha energía en una sección determinada; más bien nos interesa el concepto de densidad espectral desde un ámbito general, ya sea si su energía está ubicada de manera segmentada o, por el contrario, si cubre todo el ancho de frecuencia. La cantidad de esta densidad se utiliza para efectos de dirección musical y resultados estéticos en una pieza musical.

Ahora bien, para efectos de esta investigación, la estrategia de uso de la densidad espectral para componer obras se implementa desde una apreciación subjetiva, debido a que su resultado es abordado perceptualmente y determina el nivel de coloración en una sección de la obra. Así, en las tres piezas compuestas, los niveles de coloración tímbrica se lograron agregando varias capas tímbricas en una sección determinada. Además, el control de los niveles de saturación permitió medir muy subjetivamente el grado de empaste o engrosamiento del color que se requería. Al yuxtaponer las capas, se consiguió, auditivamente, un plano sonoro con un acabado fuerte y un empastado grueso.

La función de la saturación tímbrica, por un lado, consiste en colorear la sección por trabajar y, por otro, en provocar un mayor interés en el contraste, al contraponerse a las otras secciones musicales con una presencia más tenue tímbricamente. Con este recurso compositivo se logran puntos de marcación y de

atención, además de nodos que pueden aprovecharse para giros en la construcción musical o implementarse como lugares de llegada con respecto a otras secciones con menor empaste tímbrico.

Para comprender la densidad espectral, se muestran dos ejemplos sonoros contrastantes: en el 3.24 se aprecia mucha densidad espectral por categorías de brillo, opaco y rugoso, mientras que en el ejemplo 4.25, se muestra poca densidad espectral, constituida por categorías rugosas y opacas.

Ejemplo sonoro 3.24.

Mosaico Sonoro de 4'40" a 4'56"

Densidad fuerte

https://ottocastro.com/o/159

Ejemplo sonoro 3.25.

Mosaico Sonoro de 0'00" a 0'30"

Muy poca densidad

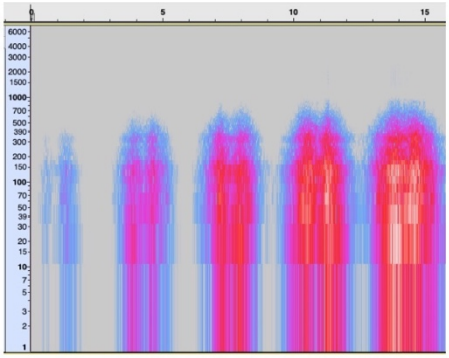
https://ottocastro.com/o/160

En la figura 30, se observan dos espectrogramas, correspondientes a los ejemplos sonoros 3.24 y 3.25. En estos se aprecian dos ejes: el eje Y, para la representación de sus frecuencias de forma logarítmica y el eje X, para la representación del tiempo en minutos y segundos.

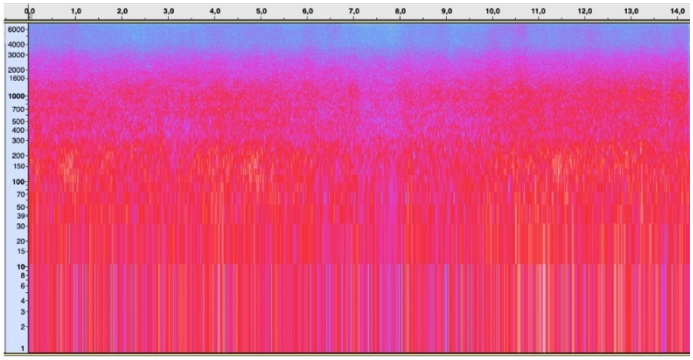
En el espectrograma correspondiente al ejemplo 3.24, puede verse una difusa y estrecha densidad representada por la intensidad de los colores rojo y blanco, que va de 1 Hz a 500 Hz de manera segmentada, con concentraciones en ciertas bandas de frecuencia. Por su parte, en el espectrograma del ejemplo 3.25, podemos observar que el color está dispuesto de manera distribuida y continua, extendiéndose desde 1 Hz hasta 4000 Hz.

Figura 30

Representación de la densidad espectral



Ejemplo 3.24.



Ejemplo 3.25.

En otros momentos, la densidad espectral es aprovechada como un efecto de multiplicación del color, como sucede en el ejemplo sonoro 3.26, en donde el mismo sonido es repetido y yuxtapuesto, logrando así un enjambre sonoro.

Ejemplo sonoro 3.26.

Mosaico Sonoro de 2'59" a 3'28"

https://ottocastro.com/o/161

También se utilizó como un recurso de contraste tímbrico entre una densidad espectral alta e, inmediatamente después, una densidad espectral baja. Estos cambios tímbricos dentro de la estructura de la obra crean giros rítmicos importantes que favorecen la atención en el oyente.

Ejemplo sonoro 3.27.

Mosaico Sonoro de 3'21" a 3'35"

https://ottocastro.com/o/102

Como se escucha, en el ejemplo 3.27, se produce un cambio abrupto no solo de intensidad, sino de yuxtaposición de capas.

Finalmente, la última forma de utilizar la densidad espectral en las tres piezas fue como parte de una tensión acumulativa, sumando una misma o varias categorías tímbricas de manera secuencial, tal y como sucede en el ejemplo sonoro 3.28.

Ejemplo sonoro 3.28.

Mosaico Sonoro de 1'16" a 1'24"

https://ottocastro.com/o/162

Posibilidades de composición a través de las capas tímbricas y los planos sonoros en las tres piezas musicales

Para esta investigación, las categorías tímbricas mencionadas se pueden percibir en las piezas musicales en dos unidades compositivas: la capa tímbrica y el plano sonoro: la capa responde a una escucha específica, determinada y particular, mientras que el plano, a una escucha general de la suma de las capas. Debido a esto, la capa sonora puede percibirse como un revestimiento de una sonoridad con otra, mientras que en el plano tenemos una percepción de totalidad sonora.

Las capas pueden percibirse claramente cuando tienen un espacio de frecuencia amplio y esto permite, a la vez, identificar cada categoría tímbrica contenida en ellas. Dichas categorías pueden presentarse de forma secuencial, combinadas, yuxtapuestas o fusionadas. Por el contrario, cuando se presentan en un espacio de frecuencia reducida (es decir, cohesionadas), su percepción a través de la escucha es difícil y la identificación de cada categoría de manera aislada resulta borrosa.

A continuación, se muestran algunos usos de capas tímbricas y planos sonoros que se implementaron en las tres piezas compuestas.

Fusión tímbrica de capas

El efecto de amalgama se presenta de dos maneras:

a. Fusión alternativa que va de una a cuatro categorías tímbricas, sin rasgos contrastantes entre sí, favoreciendo que se perciba cohesión entre ellas. Un ejemplo sonoro similar se puede escuchar en las muestras 3.29 y 3.30.
b. Por efecto de enmascaramiento de una categoría tímbrica sobre la otra, como en el ejemplo sonoro 3.31.

Ejemplo sonoro 3.29.

Texturas Urbanas de 5'00" a 5'22"

https://ottocastro.com/o/048

Ejemplo sonoro 3.30.

Anomalía Urbana de 8'55" a 9'03"

https://ottocastro.com/o/093

Ejemplo sonoro 3.31.

Texturas Urbanas de 7'10" a 7'33"

https://ottocastro.com/o/044

Como se desprende de los ejemplos anteriores, el recurso de fusión genera un nuevo color tímbrico híbrido, junto con las categorías tímbricas opaca, rugosa, lisa y brillante, por lo cual es una posibilidad estratégica y estética que brinda diversidad en el montaje y permite la construcción de las secciones musicales. Por su parte, con la táctica de enmascaramiento tímbrico, se tiene la sensación auditiva

de una especie de “compresión”, achatamiento de los timbres y poca identificación de cada categoría tímbrica separada.

Independencia de los planos tímbricos

La independencia entre los planos sonoros es un recurso importante para entender auditivamente las categorías tímbricas como elementos musicales que forman parte de un entretejido. La claridad con que se escucha cada plano tímbrico, beneficiará su continuidad como desarrollo de la composición.

Los planos sonoros pueden escucharse separados o comprimidos. Esta característica dependerá de la densidad tímbrica que se presente en cada situación. Dos muestras de planos separados son los ejemplos sonoros 3.32 y 3.33, con diferencias entre ellos en cuanto a niveles de densidad, brillo y rugosidad. Por su parte, el ejemplo 3.34 corresponde a un plano estrecho.

Ejemplo sonoro 3.32.

Plano tímbrico separado con muy poca densidad espectral, Mosaico Sonoro de 1’52” a 2’01”

https://ottocastro.com/o/106

Ejemplo sonoro 3.33.

Plano tímbrico separado con poca densidad espectral, Anomalía Urbana de 2’15” a 2’32”

https://ottocastro.com/o/088

Ejemplo sonoro 3.34.

Plano tímbrico comprimido con alta densidad espectral, Mosaico Sonoro de 4'40" a 4'56"

https://ottocastro.com/o/099

En algunas ocasiones, una categoría tímbrica puede contener una frecuencia fundamental que le brinda una ubicación en el ancho de frecuencia. Se puede apreciar esta situación en los dos ejemplos sonoros siguientes: las categorías pueden mostrarse muy juntas o totalmente distantes. Así, en el ejemplo 3.35, tenemos la sensación de escuchar un pedal armónico opaco que aparece solo y poco después se escucha un sonido liso, que se yuxtapone al pedal. En el ejemplo 3.36, tenemos una percepción contraria, pues se presentan simultáneamente dos categorías tímbricas opaco y brillo; el sonido brillante tiene una frecuencia fundamental aguda que produce la percepción de distancia entre los timbres, con contornos sonoros angulosos.

En síntesis, los ejemplos mencionados muestran efectos diferentes: en el 3.35, estrecho y en el 3.36, separado.

Ejemplo sonoro 3.35.

Texturas Urbanas de 0'44" a 1'10"

https://ottocastro.com/o/174

Ejemplo sonoro 3.36.

Texturas Urbanas de 4’38” a 4’47”

https://ottocastro.com/o/175

A modo de refuerzo, se presentan a continuación dos ejemplos más extremos en esta percepción de sonido separado o comprimido en las categorías tímbricas. En el ejemplo 3.37, se escuchan dos sonidos, uno opaco y otro brillante, los cuales, por sus características, dan la sensación de estar muy separados. Por otro lado, en el ejemplo 3.38, debido a su densidad espectral, dos sonidos, uno rugoso y otro opaco, dan la sensación de tanta cercanía que casi se puede escuchar un tercer color tímbrico.

Ejemplo sonoro 3.37.

Mosaico Sonoros de 0’47” a 1’05”

https://ottocastro.com/o/110

Ejemplo sonoro 3.38.

Texturas Urbanas de 3’44” a 3’58”

https://ottocastro.com/o/164

Por otro lado, las categorías tímbricas pueden presentarse de manera yuxtapuesta sincrónica, secuencial, acumulativa o sola, como se mencionó. La distribución secuencial y acumulativa en el tiempo permite una gran riqueza sonora

y favorece la atención de los colores tímbricos. En el ejemplo sonoro 3.39, se muestra una categoría tímbrica opaca sola, con el fin de facilitar la escucha de las sutilezas de esta categoría. Luego, en los ejemplos sonoros 3.40 y 3.41, se aprecia cómo la alternancia de las diferentes categorías crea una mayor profundidad de los elementos sonoros.

Ejemplo sonoro 3.39.

Anomalía Urbana de 0'00" a 0'30"

https://ottocastro.com/o/094

Ejemplo sonoro 3.40.

Texturas Urbanas de 2'10" a 2'19"

https://ottocastro.com/o/059

Ejemplo sonoro 3.41.

Anomalía Urbana de 0'22" a 0'50"

https://ottocastro.com/o/165

Ahora bien, en el ejemplo 3.42, se muestra la presentación regular de las categorías tímbricas rugoso-opaco. Debido a su regularidad rítmica y continuidad, se produce un estatismo a nivel de dirección musical. En el ejemplo 3.43, en cambio, tenemos cuatro categorías tímbricas: brillo, opaco, rugoso y liso, las cuales se

presentan en parejas, creando especies de bloques tímbricos regulares, que, también por su repetición rítmica, crean estatismo.

Ejemplo sonoro 3.42.

Anomalía Urbana de 4'50" a 5'08"

https://ottocastro.com/o/166

Ejemplo sonoro 3.43.

Texturas Urbanas de 2'49" a 2'59"

https://ottocastro.com/o/056

El ejemplo 3.44 muestra una presentación secuencial utilizando la acumulación como recurso estético. En este, las categorías tímbricas dan mayor tensión musical al ir sumando sus niveles de brillo, rugosidad y opacidad.

Ejemplo sonoro 3.44.

Anomalía Urbana de 0'22" a 0'50"

https://ottocastro.com/o/165

Por su parte, en el ejemplo 3.45, la presentación de las cuatro categorías tímbricas se muestra de manera irregular, lo cual crea un movimiento musical aleatorio e intenso. Podemos decir, en este último ejemplo, que la aparición de las categorías tímbricas es tan rápida que todo el conjunto se escucha como un

conglomerado sonoro: más que percibir detalle de las características de cada categoría, este ejemplo las contiene a todas como una amalgama sonora.

Ejemplo sonoro 3.45.

Texturas Urbanas de 3'04" a 3'19"

https://ottocastro.com/o/167

Reflexión acerca de los resultados estéticos logrados con la versión final del *CSpectra2021,* en la pieza *Ejes Urbanos*

El *CSpectra2021* es una herramienta pensada para la asistencia en la composición musical. La palabra "asistencia" se entiende para esta investigación, como un acompañamiento en la labor de la creación del compositor, que provee de distintas posibilidades de texturas sonoras, en cada consulta que éste realice al sistema. Esto con el fin de seleccionar la mejor opción que se adapte a sus necesidades estéticas. Desde esta perspectiva, el compositor tiene el trabajo de articular y relacionar los materiales sonoros que obtenga, para lograr un contexto musical favorable, que propicie la comunicación de las diferentes partes en la estructura de una pieza musical.

El *CSpectra2021* es el resultado de las versiones anteriores, como son el *MAT2021a* y *MAT2021b*. El lector puede conocer más acerca de ellas visitando el capítulo dos de esta tesis. Es por lo anterior, que en este apartado nos

concentramos en los resultados estéticos obtenidos por la herramienta en una cuarta composición musical desarrollada, bajo el título: *Ejes Urbanos*[130].

Algunos aspectos técnicos generales

Antes de pasar a mostrar los resultados estéticos, es conveniente mencionar algunos aspectos técnicos que se incorporaron para la composición de la pieza *Ejes Urbanos*. Los aspectos técnicos que aquí se mencionan son importantes porque denotan la capacidad de procesamiento de la versión final de la herramienta.

Para *Ejes Urbanos* se utilizó un repositorio de sonidos integrado con 96870 fragmentos, los cuales van con una duración de 1 a 10 segundos. El peso total del repositorio fue de 18 GB, que corresponde a 2.79 GB de peso en la base de datos *MongoDB*. De esta forma, *Ejes Urbanos* fue la pieza que más cantidad de posibles archivos de sonido utilizó para su creación, en comparación con las otras composiciones musicales.

Estructura general de la pieza

Antes de pasar a mencionar las estrategias de composición utilizadas en esta pieza, podemos hacer un análisis general de las categorías implementadas. *Ejes Urbanos* está constituida por 9 secciones. Estas secciones tienen dos

---

[130] Esta obra nace a partir de la imagen que tuve como espectador, ante los ejes urbanos que atraviesan a la Ciudad de México. Algunos ejes son contrastantes entre sí, otros por el contrario parecen continuar el mismo paisaje urbano de forma similar a otros ejes. El sonido por lo general de varios de estos ejes urbanos, en mi percepción, es continuo, borroso y los sonidos tienden a perderse ante una multiplicidad de otros sonidos que se funden, creando un pedal armónico de flujo constante.

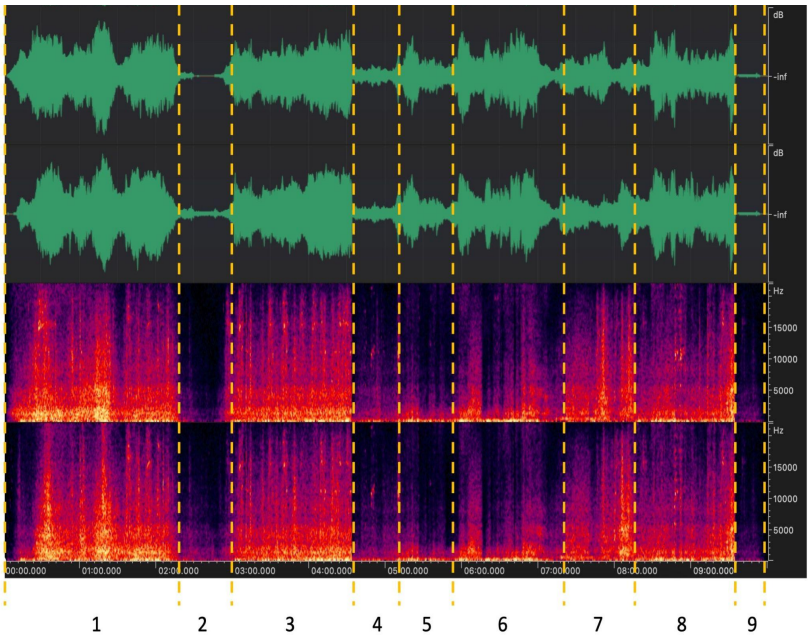
características en la manera en que se distribuyen los materiales obtenidos por el *CSpectra2021*:

- Una establece la continuación de la categoría tímbrica anterior, es decir, desarrolla el gesto anterior.
- La otra implementa contraste con la categoría presentada, es decir, establece una tensión de color, con el fin de que el oyente descanse su atención en el color tímbrico, para apreciar la nueva categoría tímbrica.

Podemos observar en la figura 31, la estructura general en la conformación de las secciones de la pieza y en la tabla 6, la descripción de cada de dichas secciones.

Figura 31

Estructura general de la pieza *Ejes Urbanos*



Enlace de la pieza musical: *Ejes Urbanos* (10’00”)

https://ottocastro.com/o/ejesurbanos

Tabla 6

Descripción de las secciones de la pieza *Ejes Urbanos*

#	Tiempo	Materiales sonoros utilizados
1	0’00” - 2’19”	Muy brillante
2	2’20” - 2’53”	Medio opaco
3	2’54” - 4’34”	Brillante
4	4’35” - 5’10”	Opaco
5	5’11” - 5’56”	Muy opaco + rugoso
6	5’57” - 7’23”	Opaco + rugoso
7	7’24” - 8’16”	Medio brillante
8	8’17” - 9’36”	Opaco
9	9’37" - 10’00”	Poco brillante

Durante esta investigación se ha ido documentando todo lo referente al uso, implementación y etapas de los procesos involucrados, en sinergia con las decisiones del compositor. Por esto, es preciso definir dos términos utilizados en esta investigación: *textura sonora* y *capa*. La *textura sonora* es aquella que ha sido devuelta por el sistema*,* y que está conformada por varios segmentos de sonido que han sido concatenados para conformar un sólo sonido entendido como una textura sonora. La *capa* puede estar conformada por una o varias texturas sonoras yuxtapuestas de manera simultánea. También puede existir el término de *textura sonora compleja,* que es aquella conformada por la suma de varias texturas sonoras.

Con relación a las texturas sonoras trabajadas en la composición con el último sistema, por lo general se utilizó el resultado obtenido tal y como el *CSpectra2021* lo dispuso en el multipistas. En ocasiones se incorporaron algunas de las técnicas básicas como: corte, estiramiento, filtraje o panoramización, las cuales fueron útiles para lograr ciertos resultados en la construcción estética de la obra. Estas particularidades serán explicadas más adelante.

Tres tipos de búsquedas para generar texturas sonoras con el CSpectra2021

Antes de explicar las estrategias que se implementaron con las categorías sonoras, es conveniente abordar las etapas de búsqueda de dichas categorías con el *CSpectra2021* para crear texturas sonoras. Con el fin de que otros usuarios puedan replicar los resultados obtenidos, se enumeran a continuación los tipos de búsqueda que se utilizaron.

Búsqueda básica: La manera más básica de generar texturas sonoras con el *CSpectra2021* es seleccionar un evento sonoro y solicitar su búsqueda, el sistema devolverá una concatenación de sonidos similares presentados como una textura sonora, que puede variar en su extensión de 3 segundos a 2 minutos. Esto dependerá del conjunto de sonidos que en el repositorio encuentre como similares tímbricamente.

Búsqueda con transposición: Es preciso mencionar que también es posible transponer en altura el segmento de referencia, seleccionado anteriormente, gracias a esa transposición el sistema puede encontrar otras texturas posibles. Para *Ejes Urbanos* se probó con 1, 2, 5 y 7 semitonos de diferencia, ya fuera hacia abajo o hacia arriba. En todos los casos se obtuvieron resultados diferentes al primer evento encontrado.

Búsqueda por capas fusionadas: Otro recurso implementado en la presente pieza, fue el utilizar como segmento de referencia la fusión de texturas sonoras ya encontradas. Esta posibilidad se probó con 2 y 3 capas mezcladas, obteniendo una textura sonora de la cual se seleccionó un segmento, para que el sistema buscara sonidos similares. Ya con 4 capas el sistema da error, porque no sabe qué requiere encontrar. El recurso de búsqueda de capas fusionadas se utilizó con el fin de acrecentar la densidad espectral del sonido referente.

Proceso de composición

A continuación, se explica cómo fue el proceso general de composición utilizando la herramienta *CSpectra2021*:

1. *Selección de gesto sonoro*: En esta primera etapa es preciso seleccionar un evento o gesto prominente, que contenga alguna característica acústica evidente. Algunos ejemplos que ilustran este tipo de eventos pueden ser: el sonido de un claxon brillante, el chillido de las llantas en el pavimento, que ejemplifica una categoría sonora rugosa, el encendido de un motor de autobús con una característica de sonido opaco, el sonido de varios autos en una autopista con un predominio de brillo o lisura, dependiendo del caso. Debe de tomarse en cuenta que es necesario seleccionar un fragmento no mayor de cinco segundos, con lo cual debe de ser un evento con fuertes características acústicas, en nuestro caso corresponde a las categorías sonoras de interés en esta investigación. De lo contrario, *CSpectra2021* brindará eventos sonoros difusos y muy generales. Una vez que se seleccione el segmento determinado, este pasará a ser el sonido referente con el cual el sistema informático tratará de buscar texturas sonoras similares acústicamente[131].
2. *Construcción del conjunto de texturas sonoras*: Una vez que se obtiene la primera textura sonora, se procede a repetir la operación una y otra vez,

[131] El lector puede revisar el capítulo dos de esta investigación, con el fin de conocer qué características acústicas son entendidas como similares.

con el fin de obtener varias de estas y conformar así, un conjunto de ellas. Para *Ejes Urbanos* se requirieron de 15 a 20 texturas sonoras, para integrar un repositorio de estos materiales. Este conjunto le servirá al compositor para construir un contexto musical, en el cual pueda relacionar unas texturas con otras. Esto no sólo funcionará para desarrollar los materiales a partir de una o varias categorías tímbricas, sino para crear diferentes planos sonoros, que incluso permitan recrear los gestos musicales en el espacio.

3. *Contexto musical*: En el repositorio de texturas sonoras obtenidas, cada una presentará al compositor diferentes calidades de una característica tímbrica. Esta diversidad de texturas sonoras puede recrear un contexto musical provechoso, que vincule unas texturas con otras. Precisamente, el trabajo del compositor, en este punto, es el de encontrar confluencias o momentos de vinculación con otras texturas, creando de esta manera un contexto que permita la relación entre las partes. Por otro lado, las texturas sonoras son devueltas por el *CSpectra2021* al multipistas en el formato de audio mono. Esto podría parecer una limitación, sin embargo, esta característica se convierte en una potencialidad, debido a que provee la posibilidad de controlar en cada textura, su gestualidad en el espacio. Para la obra *Ejes Urbanos* se colocaron todas las texturas sonoras, pensando en una difusión estéreo, pero sin problema, puede aplicarse para una obra pensada en cuatro, cinco, seis u ocho canales.

4. *Planos sonoros:* Las texturas sonoras montadas a manera de capas pueden disponerse para resaltar el carácter gestual y de dirección espacial, con el fin de beneficiar el desarrollo de los materiales sonoros. La suma de las diferentes capas sonoras puede ser útil para efectos compositivos que resalten alguna categoría tímbrica, para dar mayor densidad espectral al conjunto o distribución espacial de las capas, así como para dar una mayor coloración tímbrica a la sección. En *Ejes Urbanos* se trabajaron varios aspectos, a fin de dar mayor realce a las características tímbricas y como parte del desarrollo de los materiales sonoros, esto se mencionará más adelante, cuando se trate el tema de las estrategias implementadas en esta pieza.

Una de las principales metas en la composición de esta pieza, fue la de implementar las cuatro categorías perceptuales: brillante, opaco, rugoso y liso, de la misma manera que se implementó en las piezas anteriores. A través del uso de la versión final del *CSpectra2021,* en la composición de *Ejes Urbanos,* se percibieron desde mi escucha otras categorías intermedias, entre las cuatro categorías mencionadas anteriormente. Es así, que se agregaron términos tales como: translúcido, transparente, densidad espectral baja, media y alta. Estos adjetivos perceptuales desde mi escucha, en el proceso de composición, ayudaron a ubicar mejor ciertos resultados sonoros, obtenidos a través de la yuxtaposición de las diferentes texturas sonoras arrojadas por el *CSpectra2021*.

Debido a que la máquina busca en su análisis, categorías tímbricas generales. Para mejorar estas búsquedas es necesario lograr un mejor contraste en los sonidos de referencia que se desea utilizar. Por lo cual se toma el sonido referente y se resalta a través de una mayor ecualización de sus frecuencias agudas o graves. Los resultados sonoros que produce este proceso son mucho más enfáticos en categorías como brillante u opaco.

Ejemplo sonoro 3.46.

Sonido de referencia sin procesamiento y el resultado obtenido.

https://ottocastro.com/o/346

Ejemplo sonoro 3.47.

Sonido de referencia en el cual se resaltan las frecuencias agudas (+2000 Hz) y el resultado obtenido.

https://ottocastro.com/o/347

Ejemplo sonoro 3.48.

Sonido de referencia en el cual se atenúan las frecuencias agudas (-1000 Hz) y el resultado obtenido.

https://ottocastro.com/o/348

Ahora bien, es preciso explicar las categorías intermedias que se citaron anteriormente, que fueron encontradas en el procesamiento de los materiales sonoros con el CSpectra2021.

Translúcido: Se presenta cuando dos categorías contrastantes como brillante y opaco, se encuentran superpuestas en una *textura sonora compleja*, en la cual las categorías involucradas se fusionan. Por ejemplo, una categoría opaca puede tener un porcentaje de brillo, dando como resultado un sonido opaco más tenue. De la misma forma, un sonido rugoso puede contener características de lisura y viceversa.

A continuación, se pueden escuchar dos ejemplos sonoros, en el 3.49 tenemos un ejemplo de una textura translúcida, más opaca que en el 3.50, la cual es una textura también translúcida, pero más brillante.

Ejemplo sonoro 3.49.

Textura translúcida con más opacidad. Ejes Urbanos (4'42" a 4'54")

https://ottocastro.com/o/349

Ejemplo sonoro 3.50.

Textura translúcida con más brillo. Ejes Urbanos (1'15" a 1'27")

https://ottocastro.com/o/350

Transparente: Se presenta cuando las categorías se mezclan entre sí, dando como resultado las siguientes categorías: 1. brillo liso, 2. rugoso opaco, 3. liso opaco, ó 4. brillo rugoso. La diferencia con el comportamiento de la textura *translúcida* es que en ella las dos categorías se fusionan, mientras que, en el comportamiento de la

textura transparente, las dos características acústicas de sonidos se pueden percibir de forma independiente.

Ejemplo sonoro 3.51.

Textura transparente: categoría brillante rugoso. Ejes Urbanos (1'57" a 2'03")

https://ottocastro.com/o/351

Ejemplo sonoro 3.52.

Textura transparente: categoría opaco liso. Ejes Urbanos (6'19" a 6'26")

https://ottocastro.com/o/352

Algunas estrategias como parte del proceso de composición

Con el objetivo de aprovechar al máximo las posibilidades del *CSpectra2021* se utilizaron una serie de posibilidades que se mencionan y detallan a continuación:

- Continuidad
- Contraste
- Acumulación de densidad espectral
- Variedad en los planos sonoros
- Heterofonía tímbrica

Continuidad: Consiste en la incorporación de dos o más texturas sonoras que pueden ayudar en la percepción de una categoría tímbrica, debido a que ellas

comparten alguna de las características predominantes. Para hacerlo se deben unir o yuxtaponer las texturas sonoras, con una categoría tímbrica similar. Esta estrategia ayuda a fortalecer el color tímbrico deseado.

Ejemplo sonoro 3.53.

Textura sonora de continuidad tímbrica. Ejes Urbanos (6'41" a 6'58")

https://ottocastro.com/o/353

Contraste: Consiste en la contraposición entre una categoría tímbrica y otra, de forma abrupta. Esta no sólo evidencia la potencia diferenciada entre una y otra, sino que ayuda a crear tensión en la construcción del discurso musical. Además, colabora con un descanso de atención en la escucha.

Ejemplo sonoro 3.54.

Textura sonora de contraste tímbrica. Ejes Urbanos (4'26" a 4'46")

https://ottocastro.com/o/354

Acumulación de densidad espectral: Consiste en la yuxtaposición de varias texturas sonoras que incrementan la densidad tímbrica. Puede sumar sus elementos simultáneamente, o ir sumándolos poco a poco para mayor eficacia en un resultado estético más evidente. Este recurso puede ser utilizado como parte de una estrategia de desarrollo de los materiales sonoros, a través de la acumulación de la tensión en el color tímbrico.

Ejemplo sonoro 3.55.

Textura sonora por acumulación de densidad espectral. Ejes Urbanos (3'54" a 4'34")

https://ottocastro.com/o/355

Variedad en los planos sonoros: La combinación de dos o más texturas sonoras da como resultado diferentes planos sonoros. Estos pueden permitir una mayor profundidad espacial y disponer de mayores posibilidades para el control de la direccionalidad de sus sonidos. Para la composición *Ejes Urbanos*, se implementaron diferentes disposiciones de paneo con respecto a las capas sonoras, con el fin de conseguir diferentes planos complejos. También fue posible crear "modulaciones espaciales", en donde se partía de disponer todas las capas en una posición de paneo determinada, e ir abriendo en posiciones diferentes cada capa. Esto, a fin de lograr una imagen estéreo, con una sensación de multiplicidad en las capas sonoras. Pero también haciéndolo de manera inversa.

Con el objetivo de resaltar o atenuar las diferentes categorías tímbricas, se utilizó la posibilidad de presentar los materiales en una disposición espacial compleja o con un carácter de percepción monoaural y plano.

Ejemplo sonoro 3.56.

Modulación espacial de varios planos sonoros. Ejes Urbanos (3'19" a 3'50")

https://ottocastro.com/o/356

Heterofonía tímbrica: Un gesto musical, en una textura sonora, puede contener una característica tímbrica predominante. Esta característica vinculada al gesto puede resaltarse y desarrollarse a través de la suma de otras texturas sonoras que enriquezcan su categoría tímbrica, a manera de una polifonía tímbrica o estratificación de las diferentes capas. En la pieza *Ejes Urbanos* este recurso fue muy utilizado, implementando para esto diferentes técnicas que fueron:

- Duplicación de la misma categoría tímbrica en una intensidad menor, para crear la percepción de un sonido más lleno.
- Acortamiento o alargamiento del mismo evento sonoro, con el fin de lograr desfases que enriquecen el gesto musical principal.
- Transposición en altura de un gesto principal, a medio o un tono de diferencia, o incluso a décimas de tono, con el objetivo de crear variaciones al gesto original y enriquecerlo. Por lo general, para esta posibilidad se alargaba o acortaba la capa secundaria con el fin de provocar desfases en la ejecución del gesto principal. Siempre estas segundas capas se colocaron con una intensidad menor y buscando que no estuvieran en la misma posición de paneo que el gesto original, para que pudieran ser identificadas y que no se confundieran con la textura sonora original.

Ejemplo sonoro 3.57.

Textura sonora de heterofonía tímbrica (1’57” a 2’17”)

https://ottocastro.com/o/357

Otras estrategias utilizadas

Dirección armónica: La mezcla de las diferentes texturas sonoras, por lo general, producen una ambigüedad en cuanto a la dirección armónica. En algunas secciones de *Ejes Urbanos* se procuró dar dirección armónica, a través de pedales que fueron conduciendo las diferentes texturas sonoras que se iban montando. Lo anterior se logró, colocando algunas texturas de sonidos continuos, en condición de intensidad, atenuando las frecuencias superiores de 3000 Hz, a través de un filtro pasa bajos. Estos pedales obtenidos, por lo general, se disponían en diferentes alturas con el proceso de transposición, para lograr una especie de pedales armónicos, que beneficiarán el control en la construcción del discurso de la pieza musical.

Ejemplo sonoro 3.58.

Textura sonora con pedales armónicos. Ejes Urbanos (1'57" a 2'17")

https://ottocastro.com/o/358

Fusión de capas: En algunos momentos la disposición de cada textura sonora en el paneo, permite una mejor identificación de cada capa. Sin embargo, en otras ocasiones se requiere que estas se fusionen en una sola masa sonora. Para esto, un recurso que se utilizó fue el de duplicar alguna textura sonora y estirar su longitud. Esto provocó una ambigüedad espacial y tímbrica importante, que favorece la fusión de las otras capas.

Ejemplo sonoro 3.59.

Fusión de texturas sonoras. Ejes Urbanos (4'21" a 4'30")

https://ottocastro.com/o/359

En conclusión, la propuesta de cuatro categorías tímbricas que organizan los sonidos urbanos constituye una nueva forma de utilizar este tipo de materiales sonoros para la composición musical. Dichas categorías, entendidas a través de la similitud tímbrica, representan una manera novedosa de distribuir los sonidos en el tiempo y de establecer relaciones que conlleven al desarrollo de una metodología artística diferente en el acto de componer. Asimismo, la valoración de la similitud tímbrica en los materiales sonoros ayuda a identificar los diferentes grados de contraste entre ellos, los cuales pueden ser utilizados como recursos de tensión musical.

Por otro lado, las posibilidades de color que se pueden generar con los materiales sonoros obtenidos por el sistema informático, mediante las diferentes yuxtaposiciones y encadenamientos de las categorías tímbricas, brindan otro enfoque creativo y estético del tratamiento de los sonidos en la composición de las piezas musicales generadas a partir de sonidos urbanos.

Finalmente, el recurso de la utilización del color tímbrico como estrategia de composición, más allá de ser valorado para obtener sutileza y embellecimiento sonoro, se convierte en una oportunidad de conocer otras posibilidades de los sonidos, además de sus características acústicas, en un contexto de sintaxis en combinación con otros sonidos.

Consideraciones finales

Al finalizar la presente investigación, es importante recordar el origen que motivó este camino: un interés por el desarrollo de los materiales sonoros a partir del timbre para la construcción de las piezas musicales.

Desde inicios de mi carrera, cuando era estudiante de Composición, me interesaron los empastes sonoros a partir de la yuxtaposición de diferentes sonidos instrumentales que provocaban cambios en los colores tímbricos obtenidos. Así, desarrollé un gusto particular por el control de la combinación de los colores tímbricos, más allá de lo rítmico o lo melódico. Sobre todo, este interés estaba relacionado con la teoría de la orquestación y, a través de ella, con la manera en que se lograba una economía de los colores.

Toda esta ruta marcó una manera de trabajar, percibir y concebir la composición; sin embargo, la presente investigación, aunque continuó con mi interés en el timbre, me hizo replantear los abordajes pasados y me colocó en una posición que “deconstruyó” mis acercamientos a la composición. Así, involucrarme con métodos de procesos de aprendizaje automático me llevó a conocer otro tipo de herramientas y posibilidades totalmente diferentes a las que usaba anteriormente. Procesos como segmentar, identificar y agrupar sonidos por diferentes tipos de algoritmos, me llevaron a pensar las metodologías de creación musical de una manera distinta, pero también a reflexionar acerca de mi percepción y mi escucha. El uso de estas tecnologías posibilitó un yo extendido en continuo diálogo, con una nueva percepción, lo que provocó un resultado estético diferente al de mis trabajos anteriores.

Ahora bien, el objetivo general planteado para esta investigación fue crear obras sonoras a través de la colaboración entre la escucha de máquina y la escucha humana. Dicho objetivo fue cumplido mediante la composición de tres piezas musicales, las cuales fueron desarrolladas a lo largo de diferentes etapas, en las que se intercalaba la participación de la herramienta diseñada para una escucha tecnológica especializada y la participación de la escucha subjetiva del compositor (es decir, mi propia escucha), en una continua sinergia que enriquecía la labor dialéctica.

Así, por un lado, en la configuración de la escucha de máquina, se construyó la herramienta tecnológica *CSpectra2021* y el prototipo anterior *MAT2021*. Dichas herramientas se crearon utilizando lenguajes de programación como Python y JavaScript y librerías de código como Essentia y Annoy. Con estas se obtuvieron resultados a lo largo de todo el proceso, los cuales influyeron estéticamente de manera distinta en cada una de las piezas musicales. Además, durante la confección de estas herramientas, se exploraron programas cercanos que pudieran dar resultados similares a los que se esperaba obtener.

Por otro lado, en el proceso de concepción inicial, exploración sonora de un entorno, registro y configuración de las obras, también se estableció un diálogo entre la precisión de la máquina y la experiencia subjetiva del compositor. Esto potenció la reflexión que rige todo el proceso de la escucha de máquina y que se enriquece precisamente de una escucha diferenciada de la propia. De este modo, se favorece una escucha mixta, que es notoria en los resultados sonoros, en la

efectividad del tiempo de búsqueda de los materiales sonoros y en la gran cantidad de sonidos ensamblados en las piezas musicales resultantes.

Uno de los objetivos específicos de esta investigación fue proponer una taxonomía delimitada por las características tímbricas, desde la escucha subjetiva, para ordenar los materiales sonoros resultantes de la escucha de máquina. Precisamente, en este punto converge la escucha mixta, pues interviene el análisis y la comparación espectral de la máquina, el cual, luego, es categorizado subjetivamente desde la escucha del compositor, mediante la percepción del timbre. En ese proceso, se concretaron las categorías taxonómicas *brillante, oscuro, rugoso y liso*, como una estrategia que funcionó de forma efectiva como elemento conector o contrastante en el desarrollo de los materiales sonoros de la obra.

En síntesis, este proyecto planteó un desarrollo creativo que involucró procesos maquinales de segmentación, identificación tímbrica y comparación espectral. Dichos procesos, unidos a aspectos subjetivos de escucha y a estrategias de organización de los materiales resultantes, enriquecieron el acto de componer. En esa línea, se formularon inicialmente varias preguntas que se desarrollaron durante todo el trabajo y que retomaremos a continuación:

La primera interrogante fue: ¿de qué forma es necesario adecuar la segmentación a características determinadas para obtener un material sonoro específico para componer? La segmentación fue un paso fundamental en diferentes momentos del proceso de escucha humana y de máquina. Se establecieron dos etapas importantes en las que medió la segmentación: una fue el proceso de

grabación y exploración de los sonidos urbanos; la otra se desarrolló mediante el procesamiento de la máquina. Así, se determinó que, en la etapa de grabación, en la que mediaba la escucha humana, las variables de duración, calidad de grabación y variedad tímbrica eran importantes para resolver aspectos prácticos de aparición de eventos sonoros y de empatía con el sistema informático desarrollado, en cuanto a variedad tímbrica de las grabaciones.

Por otra parte, los aspectos determinantes para que la máquina realizara una segmentación efectiva fueron logrados con algoritmos enfocados en el contraste tímbrico. Esto proveyó segmentos irregulares de gran diversidad de duración, con mayor potencial para producir texturas sonoras ricas y diversas, que pudieron aprovecharse más en la composición musical. En resumen, las etapas de segmentación desde la escucha humana y la de máquina son muy diferentes entre ellas y complementarias en el trabajo compositivo.

Otra interrogante que surgió en la construcción de este trabajo fue: ¿qué características acústicas/perceptuales pueden definirse para encontrar niveles de similitud tímbrica? Para encontrar similitudes entre el timbre de los sonidos, fue necesario implementar una categorización vinculada a adjetivos que partían de mi percepción. Dicha categorización sirvió para establecer un ordenamiento no solo perceptual, sino también conceptual. Para ello, se partió de observar el comportamiento de los descriptores acústicos que son especialmente sensibles a ciertas características acústicas. Descarté una gran cantidad de ellos (aquellos que no me eran útiles para trabajar el “color” tímbrico) y utilicé particularmente los que

analizaban el espectro del sonido. A partir de esta observación, se determinaron las siguientes características perceptuales: brillante, opaco, rugoso, liso.

Considero que la implementación de las categorías mencionadas es el mayor logro de este proceso, porque plantea una metodología de desarrollo de los materiales con un componente indivisible entre las dos escuchas establecidas para esta labor. Es decir, la categorización permitió definir una ruta de trabajo con los materiales.

Por otra parte, una vez resuelto el ordenamiento y la categorización de los sonidos, quedaba pendiente otra pregunta: ¿qué estrategias de composición podrían derivarse de materiales similares tímbricamente? El desarrollo tecnológico de esta investigación se orientó al análisis y la identificación de segmentos espectralmente similares, a partir de los cuales mi oído y mi experiencia definieron categorías. Estos agrupamientos contenían sonidos similares entre sí, pero que podían ser contrastantes entre ellos. Tal contraste tímbrico se convirtió en un principio estético dentro de las estrategias de ensamble de sonidos en la composición. Además, se concretaron recursos compositivos que fueron eficientes en las estrategias de creación: el enmascaramiento, la atenuación y el reforzamiento del color tímbrico, la densidad espectral como principio de tensión musical y la incorporación de capas tímbricas y planos sonoros, dispuestas en fusión o independencia de los planos tímbricos.

Ahora bien, al finalizar el proceso de creación y vinculación con las herramientas tecnológicas es importante preguntarse: ¿qué resultados estéticos pueden ser obtenidos a través de un sistema informático de identificación sonora y

de la creación del compositor, a partir de su experiencia subjetiva? Los resultados estéticos obtenidos en las piezas desarrolladas en esta investigación se alejan semánticamente del tipo de trabajo artístico que yo solía hacer. Aunque se le dio al sistema un sonido de referencia, la máquina produjo montajes de texturas con características muy artificiales (al decir “artificiales”, no lo hago de forma peyorativa; por el contrario, fue un resultado diferenciador con respecto a lo que se puede obtener de manera manual). Esta encontró sonidos que no estaban dentro de mi escucha y que ni siquiera imaginaba cómo podían sonar; es decir, el sistema informático proponía recursos compositivos como “serialización”, iteración, y repetición, con resultados que iban más allá del montaje humano. Dicha potencialidad de la máquina fue usada para lograr ciertos acabados estéticos que no se utilizaban en mi obra anterior.

Después de esta experiencia de investigación, considero que voy a componer de manera diferente. El atravesar por este camino me ha hecho escuchar de otra forma e incorporar nuevas estrategias propuestas por la máquina, con características de búsqueda muy finas y poco usuales, que yo ni siquiera escucho. Así, a partir de la escucha de máquina, he aprendido que puede haber otros lugares de creación y aspectos acústicos.

Con respecto a la composición con sonidos urbanos, el trabajo con la máquina me motivó a interesarme en el error como un recurso de trabajo; por ejemplo, el error de proximidad microfónica y la imprecisión en los resultados que podía arrojar la máquina en la comparación de sonidos similares se convirtió en un material interesante.

Por todo lo anterior, este trabajo constituye una invitación a que otros creadores utilicen las herramientas y estrategias planteadas para “deconstruir” su escucha y descolocar los hábitos comunes de creación sonora, siempre y cuando tengan presente que esta ruta tenderá a alejarse del ámbito semántico, en el reconocimiento de las fuentes sonoras, para buscar otros universos sonoros.

Bibliografía

Abbado, A. (1988). Perceptual Correspondences of Abstract Animation and Synthetic Sound. (Tesis de Maestría) MIT, Estados Unidos.

Acoustical Society of America Standards Secretariat (1994). Acoustical Terminology ANSI S1.1–1994 (ASA 111-1994). American National Standard. ANSI / Acoustical Society of America.

Adkins, M. (1999). Acoustic Chains in Acousmatic Music. In: *Imaginary Space: Proceedings of the 1999 Australasian Computer Music Conference*. University of Wellington, New Zeland.

Adler, S. (1989). *The study of orchestration* (2.ª ed.). W.W. Norton.

Aguirre, F. (2017). *Desarrollo y análisis de clasificadores de señales de audio* (Tesis Maestría) Universidad Politécnica de Valencia, España.

Aikaterini, T. (2011). *Into the sounding environment. A compositional approach*. [Tesis de Doctorado, Montfort University].

Akkermans, V., Font, F., Funollet, J., Roma, G., Toglias, S., y Serra, X. (2011). Freesound 2: An Improved Platform for Sharing Audio Clips.

Akkermans, V., Serrà, J., y Herrera, P. (2009) Shape-based spectral contrast descriptor. *Sound and Music Computing Conference* (SMC'09), 143–148.

Aletta, F., Kang, J., y Axelsson, Ö. (2016). Soundscape descriptors and a conceptual framework for developing predictive soundscape models. *Landscape and Urban Planning*, *149*, 65-74.

Alvarez, J. (1990). *Mambo à la Braque* [3:02]. Sonogram. https://electrocd.com/en/oeuvre/14261/Javier_%C3%81lvarez/Mambo_%C3%A0_la_Braque

Alvelos, H. (2015) *Faith*. [CD] Londres: Touch Records.

Amarel, S., Webber, B.L., y Nilsson, N. (eds.) (1985). *Readings in artificial intelligence: a* collection *of articles*. Los Altos, Calif: Morgan Kaufmann, Publishers.

American National Standards Institute (1973). *American National Psychoacoustical Terminology.* S3.20. New York: American National Standards Association.

Anderson, J., Michalski, R., Carbonell, J., y Mitchell, T. (eds.) (1983). *Machine learning: an artificial intelligence approach*. Los Altos, California, Estados Unidos: M. Kaufmann.

Anderson, J. (2000). A provisional history of spectral music. *Contemporary Music Review*, *19*(2), 7-22.

ANSI (1973). American National Standard—Psychoacoustical Terminology (American National Standards Institute, New York).

Antoni, J. (2006, febrero). The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, *20*(2).

Appleton, J., y Perera, R. (eds.) (1975) *The Development and Practice of Electronic Music*. Universidad de Michigan, Estados Unidos: Prentice-Hall.

Arce, C., de Francisco, C. y Arce, I. (2010). Escalamiento Multidimensional. Concepto y aplicaciones. *Papeles del Psicólogo, 31*(1), 46-56.

Arendt, W., y Schleich, W. P. (Eds.). (2009). *Mathematical Analysis of Evolution, Information, and Complexity*. Wiley-VCH.

Arfib, D., Keiler, F., y Zölzer, U. (2002). Time-Frequency Processing. En Zölzer, U. (ed.) DAFX—Digital Audio Effects. West Sussex, UK: Wiley, pp. 237–298.

Ariza, C. (2005) Navigating the landscape of computeraided algorithmic composition systems: A definition, seven descriptors, and a lexicon of systems and research. In Proc. of the Int. Computer Music Conference, Barcelona, pp. 765-772.

Ariza, J. (2008). *Las imágenes del sonido: una lectura plurisensorial en el arte del siglo XX*. Cuenca: Ediciones de la Universidad de Castilla-La Mancha.

Arkette, S. (2004). Sounds Like City. *Theory, Culture & Society*, 21 (1), 159–168.

Arnheim, R. (2009). Art and Visual Perception: A Psychology of the Creative Eye; the New Version. Berkeley, California, Estados Unidos: University of California Press.

Arnold Cathalifaud, M., y Osorio, F. (1998). Introducción a los conceptos básicos de la teoría general de sistemas. *Cinta de moebio*, abril, (*3*).

Asterios, Z. (2013). *Musical timbre: Bridging perception with semantics.* (Doctoral Thesis). Queen Mary University of London.

Astor, M. (2002). Aproximación fenomenológica a la obra musical de Gonzalo Castellanos Yumar (1. ed). Comisión de Estudios de Postgrado, Facultad de Humanidades y Educación-Universidad Central de Venezuela.

Aubio Library. (s.f) *Herramientas para análisis de audio y música*. [Sitio Web] Recuperado el 3 de marzo del 2019 en https://github.com/aubio/aubio/tree/master/python

Aucouturier, J.-J., Defreville, B., y Pachet, F. (2007). The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. *The Journal of the Acoustical Society of America*, 122 (n), 881–891.

Aucouturier, J.-J., y Pachet, F. (2006). Jamming with Plunderphonics: Interactive concatenative synthesis of music. *Journal of New Music Research*, *35*(1), 35-50.

Audioseg (2006, 13 octubre) [Herramienta de segmentación de audio y clasificación de transmisiones de audio]. Recuperado el 8 de noviembre del 2019 en http://audioseg.gforge.inria.fr

Audioguide. (s. f.). GitHub. https://github.com/benhackbarth/audioguide

Augoyard, J. F. (1997). La sonorización antropológica del lugar. En Amerlinck, M.J. (ed.), *Hacia una antropología arquitectónica* (pp.176190). Guadalajara, México: Universidad de Guadalajara.

Augoyard, J. F., McCartney, A., Torgue, H., y Paquette, D. (Eds.). (2006). *Sonic experience: A guide to everyday sounds.* McGill-Queen's University Press.

Axelsson, Ö., Nilsson, M. E., y Berglund, B. (2010). A principal components model of soundscape perception. *The Journal of the Acoustical Society of America*, *128*(5), 2836-2846.

Barchiesi, D., Giannoulis, D., Stowell, D., y Plumbley, M. (2015). Acoustic Scene Classification: Classifying environments from the sounds they produce. *Signal Processing Magazine, IEEE*, 32 (n), 16-34.

Baron, R. A., y Kalsher, M. J. (1995). Psychology (3rd Edition) (3rd ed.). Allyn & Bacon.

Bartra, R. (2014). *Antropología del cerebro: Conciencia, cultura y libre albedrío* (2. ed). México, D.F: Fondo de Cultura Económica.

Battier, M. (2007). What the GRM brought to music: from musique concrète to acousmatic music. *Organised Sound*, 12(03), 189-202.

Bauer, N. (2017). Segmentation. En D. Jannach, G. Rudolph, C. Weihs, & I. Vatolkin (Eds.), *Music Data Analysis. Foundations and Applications.* (p. 411). Taylor & Francis.

Beauchamp, J. (ed.) (2007). *Analysis, synthesis, and perception of musical sounds: the sound of music. Modern acoustics and signal processing*. New York, Estados Unidos: Springer.

Beazley, D. M., y Jones, B. K. (2013). Python cookbook (Third edition). O'Reilly.

Beller, G., Schwarz, D., Hueber, T., y Rodet, X. (2005). Hybrid Concatenative Synthesis On The Intersection of Music and Speech. *Journée d'Informatique Musicale.*

Bellman, Ri. (1978). An introduction to artificial intelligence: can computers think? San Francisco: Boyd & Fraser Pub. Co.

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., y Sandler, M. B. (2005). A tutorial on onset detection in music signals. IEEE Transactions on Speech and Audio Processing, 13(5), 1035–1047.

Bello, J.P., Duxbury, C., Davies, M., y Sandler, M. (2004) *On the use of phase and energy for musical onset detection in the complex domain*. IEEE Signal Processing Letters, 11: 553-556.

Bello, J. P., y Sandler, M. (2003). Phase-based note onset detection for music signals. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *5*, 441-444.

Ben Hackbarth: AudioGuide – Creating new tools for authoring music [26.03.2019]. (2019, 1 mayo). [Vídeo]. YouTube. https://www.youtube.com/watch?v=BEwTCZxx9r0

Benesty, J., Sondhi, M., y Huang, Y. (eds.) (2008). *Springer handbook of speech processing*. Berlin; London: Springer.

Berg, P. (2009). Composing Sound Structures with Rules. *Contemporary Music Review*, *28*(1), 75-87.

Bernhardsson, E. (2020) Vecinos más cercanos y modelos vectoriales. Recuperado el 15 de junio de 2020 en https://erikbern.com/2015/10/01/nearest-neighbors-and-vector-models-part-2-how-to-search-in-high-dimensional-spaces.html

Berlioz, H., y Macdonald, H. (2004). *Berlioz's orchestration treatise: A translation and commentary*. Cambridge University Press.

Bernades, G. (2014) earGram. *earGramv0.20c.zip*. Dropbox https://www.dropbox.com/s/6c8vnyacfto52ck/earGramv0.20c.zip?dl=0

Bertalanffy, L. (1976). *Teoria General De Los Sistemas: fundamentos, desarrollo, aplicaciones* (1.ª ed.). Fondo de Cultura Económica.

Bickle, J. (2015). Marr and Reductionism. *Topics in Cognitive Science,* 7 (2), 299-311.

Bishop, C.M. (2007) *Pattern Recognition and Machine Learning*. New York, Estados Unidos: Springer.

Blackburn, M. (2011). The Visual Sound-Shapes of Spectromorphology: An Illustrative Guide to Composition. *Organised Sound,* 16 (01), 5-13.

Boden, M., y Pérez, I. (2017). *Inteligencia artificial*. Madrid, España: Turner.

Bogdanov, D., Wack N., Gómez E., Gulati S., Herrera P., Mayor O., et al. (2013). ESSENTIA: an Audio Analysis Library for Music Information Retrieval. International Society for Music Information Retrieval Conference (ISMIR'13). 493-498.

Bones, O., Cox, T.J., y Davies, W.J. (2018). Sound Categories: Category Formation and Evidence-Based Taxonomies. *Frontiers in Psychology,* 9 (julio): 1277.

Botteldooren, D., De Coensel, B., y De Muer, T. (2006). The temporal structure of urban soundscapes. *Journal of Sound and Vibration*, *292*(1-2), 105-123.

Boulez, P. (1968) Notes of an Apprenticeship. Trans. Herbert Weinstock. New York: Alfred A. Knopf.

Bountourakis, V. (2015*) Análisis semántico de los sonidos ambientales a través de la extracción de características de audio y el uso de métodos de aprendizaje automático.* Universidad Aristóteles de Salónica.

Boy, G. A. (2011). *The Handbook of Human-machine Interaction*. Ashgate.

Branchi, W. (1983). The State of Anxiety. *Computer Music Journal*, *7*(1), 8-10.

Bregman, A. (1990) *Auditory Scene Analysis*. Cambridge: MIT Press.

Bregman, A. S. (2005). Auditory Scene Analysis and the Role of Phenomenology in Experimental Psychology. *Canadian Psychology/Psychologie canadienne*, *46*(1), 32-40.

Brown, E., Brown, L., Jean H., y Pendray, E. (eds.) (1930). *City noise*: *the report of the commission appointed by Dr. Shirley W. Wynne, commissioner of health, to study*

noise in New York city and to develop means of abating it. Nueva York, Estados Unidos: Comisión de reducción del Ruido.

Brown, A., Kang, J., y Gjestland, T. (2011). Towards standardization in soundscape preference assessment. *Applied Acoustics*, 72 (6), 387-392.

Burger, S., Jin, Q., Schulam, P.F., y Metze, F. (2012). *Noisemes: Manual Annotation of Environmental Noise in Audio Streams*.

Burgess, A., y Leal. A (2018). *La naranja mecánica*. Barcelona: Minotauro.

Caclin, A., McAdams, S., Smith, B. K. y Winsberg, S. (2005) Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. J*ournal of the Acoustical Society of America, 118*(1), 471–482.

Cage, J. (1961). Silence: Lectures and writings (19. pr). Wesleyan University Press.

Calvo-Manzano Ruiz, A. (2004). *Acústica físico-musical*. Real Musical.

Camilleri, L., y Smalley, D. (1998). The Analysis of Electroacoustic Music: Introduction. *Journal of New Music Research*, 27 (1-2), 3-12.

Çakır, E., Parascandolo, G., Heittola, T., Huttunen, H., y Virtanen, T. (2017) Convolutional recurrent neural networks for polyphonic sound event detection. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 25 (6).

Carretero, A. (2014). Composición musical bio-inspirada. Punto Rojo Libros.

Carter, M. (2007). *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence.* Edinburgh: Edinburgh Univ. Press.

Casey, M. (2001). Herramientas de reconocimiento de sonido MPEG-7. Transacciones IEEE sobre circuitos y sistemas para tecnología de video, 11 (6), 737–747.

Castro, O. (2016). Paisaje sonoro urbano: Nuevas estrategias de transición para la composición acusmática. [Tesis de Maestría] UNAM.

Castro, O. (2016). *Paisaje sonoro urbano: nuevas estrategias de transición para la composición acusmática*. [Tesis de Maestría, Universidad Nacional Autónoma de México].

Castro, O. (2017, 7 noviembre). *Puntos en común en los sonidos de la ciudad. ¿Cómo explorar ese entorno? Algunas reflexiones*. La oreja (in)culta. https://laorejainculta.net/2017/11/07/puntos-en-comun-en-los-sonidos-de-la-ciudad-como-explorar-ese-entorno-algunas-reflexiones/

Castro, O. (2018). La ciudad como fuente de sonidos para la creación sonora. *Resonancias: Revista de investigación musical*, 42, 143-150.

Chadabe, J. (1984) Interactive composing. Comput. Music J. 8(1), 22-27.

Chadabe, J. (1997) *Electric Sound: The Past and Promise of Electronic Music*. Universidad de Michigan. Estados Unidos: Prentice Hall.

Chargueron, C. (2015) Agnioètiomix. [Sound Cloud] Recuperado el 10 de diciembre del 2019 en https://soundcloud.com/carole-chargueron/agnioetiomix

Charles, J.-F. (2008). A Tutorial on Spectral Sound Processing Using Max/MSP and Jitter. *Computer Music Journal*, *32*(3), 87-102.

Chaudhuri, S., y Raj, B. (2013). Unsupervised hierarchical structure induction for deeper semantic analysis of audio. *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference*, pp. 833–837. IEEE.

Chen, Yung-Chang, Longwen Chang, y Chiou-Ting Hsu, (eds.) (2002). *Avances en el procesamiento de información multimedia, PCM 2002: Tercera Conferencia IEEE Pacific Rim sobre Multimedia, Hsinchu, Taiwán, 16-18 de diciembre de 2002: actas*. *Apuntes de informática 2532. Berlín*. New York, Estados Unidos: Springer

Chion, M. (1991). El arte de los sonidos fijados. Centro de Creación Experimental.

Chion, M. (1994). *Guide des Objects Sonores*. Paris, Francia: Buchet Chastel.

Chion, M. (1999) El corte. El Sonido: música, cine, literatura. Pp. 247-277. Barcelona, España: Paidós.

Chion, M., Dack, J., y North, C. (2009). Guide to sound objects.

Chion, M., y Steintrager, J. A. (2011a). Dissolution of the Notion of Timbre. *Differences* 22 (2-3), 235-39.

Chion, M., y Steintrager, J. A (2011b). Let's Have Done with the Notion of "Noise". *Differences* 22 (2-3), 240-48.

Chion, M., y Steintrager, J. A. (2016). *Sound: An Acoulogical Treatise*. Duke University Press Books.

Christensen, E. (1996). *The Musical Timespace: A Theory of Music Listening* (Vol. 1). Aalborg University Press.

Cipolla, R., y Pentland, A. (1998). *Computer Vision for Human-Machine Interaction* (1st ed.). Cambridge University Press.

Creasey, D. (2017). *Audio processes: Musical analysis, modification, synthesis, and control*. Routledge.

Coates A., y Ng A.Y. (2012) Representaciones de funciones de aprendizaje con K-Means. En: Montavon G., Orr GB y Müller KR. (eds.) Redes neuronales: trucos del oficio. Lecture Notes in Computer Science, vol 7700. Springer, Berlín, Heidelberg.

Cogan, R. (1985) *New Images of Musical Sound*. Cambridge, Massachusetts, Estados Unidos: Harvard University Press.

Connell, J. (2011). *Musical Mosaicing with High Level Descriptors* [Tesis de Maestría]. Pompeau du Fabra.

Couprie, P. (2017). Analyse et représentation du fantastique dans les musiques anecdotiques de Luc Ferrari. En Carayol, C., Castanet, P.-A., and Pistone, P. (eds.) *Le fantastique dans les musiques des XXe et XXIe siècles*. Paris, Francia: Delatour.

Davies, W. J., Adams, M. D., Bruce, N. S., Cain, R., Carlyle, A., Cusack, P., Hall, D. A., Hume, K. I., Irwin, A., Jennings, P., Marselle, M., Plack, C. J., y Poxon, J. (2013). Perception of soundscapes: An interdisciplinary approach. *Applied Acoustics*, 74 (2), 224-231.

Davis, S.B., y Mermelstein, P. (1980) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust. Speech Signal Process,* 28 (n), 357–366.

Dane, E., y Pratt, M. G. (2007). Exploring Intuition and its Role in Managerial Decision Making. Academy of Management Review, 32(1), 33-54.

Dannenberg, R., y Mont-Reynaud, B. (1987, August). Following an Improvisation in Real Time. International Computer Music Conference, 241–248.

Dean, R.T., y McLean, A. (eds.) (2018). *The Oxford handbook of algorithmic music. Oxford handbooks*. New York, NY: Oxford University Press.

Deutsch, H. (1993) *Electroacoustic Music: The First Century*. Universidad de Virginia, Estados Unidos: Belwin Mills.

De Poli, G., y Prandoni, P. (1997). Sonological models for timbre characterization. *Journal of New Music Research,* 26 (2), 170-97.

De Souza Britto, A. (Ed.). (2013). Proceedings of the 14th International Society for Music Information Retrieval C (ISMIR): November 4th-8th, Curitiba, Brazil /ed. by Alceu de Souza Britto.

Deza, M., y Deza, E. (2013). Encyclopedia of distances (Second edition). Springer.

Dikmen, O., y Mesaros, A. (2013) Sound event detection using non-negative dictionaries learned from annotated overlapping events. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*

Dixon, S. (2006). Onset Detection Revisited. Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada.

Domínguez Ruiz, A. (2014). Vivir con ruido en la Ciudad de México. El proceso de adaptación a los entornos acústicamente hostiles. *Estudios Demográficos y Urbanos*, 29 (1), 89-112.

Domínguez T. (2021). Visión artificial; aplicaciones practicas con OpenCV - Python. Madrid: Marcombo

Donnadieu. (2007). Mental Representation of the Timbre of Complex Sounds. En *Analysis, synthesis, and perception of musical sounds: the sound of music*. New York, Estados Unidos: Springer.

Dowsett, P. (2015). *Audio Production Tips*: *Getting the sound right at the source*. Amsterdam, Países Bajos: Focal Press.

Dufourt, H. (2014). *La musique spectrale: Une révolution épistémologique*. Delatour France.

Duxbury, C., Bello, J., Davies, M., y Sandler, M. (2003). Complex domain Onset Detection for Musical Signals. In *Proceedings of the 6th Conference on Digital Audio Effects (DAFx-03)*

EarGram. (2014). https://sites.google.com/site/eargram/

Emmerson, S. (1986a). The Relation of Language to Materials. En Emmerson, S. *The Language of Electroacoustic Music.* (pp. 17-40). London: Macmillan Press Ltd.

Emmerson, S. (ed.) (1986b). *The Language of Electroacoustic Music*. London: Macmillian Pr.

Emmerson, S. (ed.) (2000). *Music, electronic media, and culture*. Burlington, Estados Unidos: Ashgate.

Erickson, R. (1975). *Sound Structure in Music*. Berkeley/Los Angeles: University of California.

Ellis, D. (1996). *Prediction-driven computational auditory scene analysis*. (Unpublished doctoral dissertation) MIT Department of Electrical Engineering and Computer Science, MIT, Cambridge.

Ellis, D. (2001). Detecting alarm sounds. *Workshop Consistent and Reliable Acoustic Cues*.

Ellis, D., & Lee, K. (2004). Minimal-impact audio-based personal archives. *Workshop on Continuous Archival and Retrieval of Personal Experiences.*

Eronen, A., Peltonen, V., Tuomi, J., Klapuri, A., Fagerlund, S., y Sorsa, T. (2006). Audio-based context recognition. *IEEE Transactions on Audio. Speech and Language Processing*, 14 (n), 321–329.

Essentia (2021). *Algorithm reference: Dissonance — Essentia 2.1-beta6-dev documentation*. Universitat Pompeu Fabra. https://essentia.upf.edu/algorithms_reference.html

Fan, J., Thorogood, M., y Pasquier, P. (2017). *Emo-Soundscapes - A Dataset for Soundscape Emotion Recognition.* Séptima Conferencia Internacional sobre Computación Afectiva e Interacción Inteligente. San Antonio, Estados Unidos.

Fan, J., Thorogood, M., Riecke, B., y Pasquier, P. (2015). Automatic Recognition of Eventfulness and Pleasantness of Soundscape. En *Proceedings of the 10th Audio Mostly Conference*, Greece.

Fan, J., Thorogood, M., Riecke, B., y Pasquier, P. (2016). Automatic Recognition of Eventfulness and Pleasantness of Soundscape. *Journal of the Audio Engineering Society*. Special Issue (Intelligent Audio Processing, Semantics, and Interaction).

Fan J., Thorogood, M., Tatar, K., y Pasquier, P. (2018). *Quantitative Analysis of the Impact on Perceived Emotion of Soundscape Recordings*. Sound and Music Computing (SMC).

Fan, J., Tung, F., Li, W., y Pasquier, P. (2018*). Soundscape Emotion Recognition via Deep Learning*. Sound and Music Computing (SMC), 2018.

Farina, A. (2014). *Soundscape Ecology*. Springer Netherlands.

Feld, S. (1994) From Schizophonia to Schismogenesis. En Marcus, G y Myers, F. (eds). *The Traffic in Culture*. (pp. 96-126) University of California Press. Reimpreso por Keil, C. & Feld, S. Music Grooves. Estados Unidos: University of Chicago Press.

Fernández, J. D., y Vico, F. (2013). AI Methods in Algorithmic Composition: A Comprehensive Survey. *Journal of Artificial Intelligence Research*, *48*, 513-582.

Ferrari, L. (1995) Presque Rien No2. [Audios] https://www.maison-ona.com/catalog

Ferrari, L. (2018) Obra electrónica vol. 3. [Audios] Recuperado el 10 de febrero del 2019 en https://archive.org/details/ferrari3

Ferrari, L. (2018) Obra electrónica vol. 10. [Audios] Recuperado el 15 de febrero del 2019 en https://archive.org/details/ferrari10/01+L'escalier+des+aveugles.wav

Ferrari, L. (2020). *ANALYSIS / THOUGHTS*. [Sitio Web] Recuperado de http://lucferrari.com/en/analyses-reflexion/far-west-news/.

Finney, N. (2009). *Autonomous generation of soundscapes using unstructured sound databases.* (Tesis de Maestría). Universitat Pompeu Fabra, Barcelona.

Florentine, M., Popper, A. N., y Fay, R. R. (Eds.). (2011). Loudness. *Springer Handbook of Auditory Research*, *37*.

Fluid Corpus Manipulation. (s. f.). FluCoMa. https://www.flucoma.org/

Fonseca, E., Plakal, M., Font, F., Ellis, D., y Serra, X. (2019) Audio tagging with noisy labels and minimal supervision. Proceedings of the DCASE 2019 Workshop, NYC, US.

Fonseca, E., Plakal, M., Font, F., Ellis, D., y Serra, X. (2020, 20 enero). *FSDKaggle2019*. Zenodo. https://zenodo.org/record/3612637/export/json#.X30ynGhKjIU

Font, F., y Serra, X. (2011). Extending Sound Sample Descriptions through the Extraction of Community Knowledge. *User Modeling, Adaptation and Personalization*, 418-421.

Foote, J. (julio del 2000) Automatic audio segmentation using a measure of audio novelty. *Proc. IEEE Int. Conf. Multimedia and Expo* (ICME2000), New York, vol. I (n), 452–455.

Furui, S. (1986). Speaker-independent isolated word recognition based on emphasized spectral dynamics. *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 11.

Gabor, D. (1946) *Theory of communication*. Journal of the Institute of Electrical Engineers Part III, 93: 429-457.

Gainza, M., Coyle, E., y Lawlor, B. (2005). Onset detection using comb filters. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005*, 263-266.

Galard, J. (2000). La computadora en la música. *Anales del Instituto de Investigaciones Estéticas*, *22*(77), 263-267. Recuperado el 06 de abril de 2020, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-12762000000200010&lng=es&tlng=es

Garman, C.G. y Garman, J.F. (1992). *Teaching young children effective listening skills*. York,PA:William Gladden Foundation.

Gaver, W. (1993). What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception. *Ecological Psychology*, 5 (1), 1-29.

Gemmeke, J.F., Ellis, D.P.W., Freedman, D., Jansen, A., Lawrence, W., Moore, R.C., Plakal, M., y Ritter, M. (2017) Conjunto de audio: un conjunto de datos ontológicos y etiquetados por humanos para eventos de audio. *Conferencia Internacional de IEEE 2017 sobre Acústica, Procesamiento de Voz y Señal* (ICASSP) 776-780.

Gerhard, D., (2003). Audio signal classification: History and current techniques. Dept. of Computer Science, University of Regina.

Géron, A. (2017). *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly.

Giordano, N. (2010). College physics: reasoning and relationships. Student ed. Australia; United States: Brooks/Cole, Cengage Learning.

Gravier, G., Betser, M., y Ben, M. (2010) Audio Segmentation Toolkit, versión 1.2, 2010. Recuperado el 5 de mayo de 2019 en https://gforge.inria.fr/frs/download.php/25187/audioseg-1.2. pdf

Green, O., Tremblay, P. A., y Roma, G. (2019). Interdisciplinary Research as Musical Experimentation: A case study in musicianly approaches to sound corpora.

Grekow, J. (2018). *From Content-based Music Emotion Recognition to Emotion Maps of Musical Pieces*. Studies in Computational Intelligence.

Grey, J. (1977). Multidimensional Perceptual Scaling of Musical Timbres. *The Journal of the Acoustical Society of America,* 61 (5), 1270–77.

Grey, J. M., y Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *Journal of the Acoustical Society of America*, *63*, 1493-1500.

Griffey, J. (2020). Introduction to interactive digital media: Concept & practice. Routledge; Taylor & Francis Group

Grillmeyer, O. (2011). *Exploring Computer Science with Scheme*. New York; London: Springer.

Grisey, G. (1989). Vers una `écologie des sons´. *Entretemps 8*. (pp. 55-81) Vendôme, Francia: Presses Universitaires de France.

Gonzales, L. (2020) Árboles de decisión. Recuperado el 12 de mayo del 2020 https://aprendeia.com/arboles-de-decision-clasificacion-teoria-machine-learning/

Goodfellow, I., Bengio, Y., y Courville, A. (2016*) Deep Learning*. Cambridge: MIT Press.

Gouyon, F. (2005). *Audio features for the computation of rhythm periodicity functions and their use in tempo induction and music content processing.* (Doctoral Thesis) Universidad Pompeau du Fabra, Barcelona.

Ghosh, D., Sengupta, R., Sanyal, S., y Banerjee, A. (2017). *Musicality of Human Brain through Fractal Analytics (Signals and Communication Technology)*. Springer Berlin Heidelberg

Guastavino, C. (2007). Categorization of environmental sounds*. Canadian Journal of Experimental Psychology*, 61 (1), 54.

Guerrero, M. y Ramírez, J. M. (2002). El análisis de Escalamiento Multidimensional: una alternativa y complemento a otras técnicas multivariantes. Recuperado de http://goo.gl/Eb4bSQ.

Gulhane, S.R., Badhe, S.S., y Shirbahadurkar, S. (2018). Cepstral (MFCC) Feature and Spectral (Timbral) Features Analysis for Musical Instrument Sounds. *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*, 109-113.

Gygi, B. (2007) Similarity and categorization of environmental sounds. *Perception and psychophysics*, 69 (6), 839-855.

Hackbarth, B. (2019, May 1). *Ben Hackbarth: AudioGuide – Creating new tools for authoring music [26.03.2019]* [Video]. YouTube. https://www.youtube.com/watch?v=BEwTCZxx9r0

Hackbarth, B., Schnell, N., Esling, P., y Schwarz, D. (2013, abril). Composing Morphology: Concatenative Synthesis as an Intuitive Medium for Prescribing Sound in Time. *Contemporary Music Review*, *32*(1), 49-59.

Hall, D. A., Irwin, A., Edmondson-Jones, M., Phillips, S. y Poxon, J. E. W. (2013). An exploratory evaluation of perceptual, psychoacoustic and acoustical properties of urban soundscapes. *Applied Acoustics*, 74 (2), 248-254.

Hall, J. E. (2016). Guyton and Hall textbook of medical physiology (13th edition). Elsevier.

Hanbury, A., Kazai, G., Rauber, A., y Fuhr, N. (eds). (2015). *Advances in Information Retrieval*. Vol. 9022. Lecture Notes in Computer Science. Cham: Springer International Publishing.

Handelman, E. (1995). Robert Rowe Interactive music systems: Machine listening and composing. *Artificial Intelligence*, *79*(2), 349-359.

Harvey, J. (1980) *Mortous Plango, Vivos Voco*. [CD] Londres: Sargasso.

Herbordt, W. (2008). *Sound Capture for Human / Machine Interfaces: Practical Aspects of Microphone Array Signal Processing (Lecture Notes in Control and Information Sciences)* (2005th ed.). Springer.

Hoffman, E. y Ariza, C. (2005). Un diseño abierto para la composición musical algorítmica asistida por computadora: athenaCL.

Holzapfel, A., Stylianou, Y., Gedik, A. C., y Bozkurt, B. (2010, agosto). Three Dimensions of Pitched Instrument Onset Detection. *IEEE Transactions on Audio, Speech, and Language Processing*, 1517-1527.

Hoskinson, R., y Pai, D. K. (2007). Synthetic Soundscapes with Natural Grains. *Presence: Teleoperators and Virtual Environments*, *16*(1), 84-99.

Hsu, Y., y Kira, Z. (2016). Agrupación basada en redes neuronales utilizando restricciones por pares. *ArXiv, abs / 1511.06321*.

Huber, D. M., Runstein, R. E., & Furmankiewicz, E. (2011). Tecnicas Modernas de Grabación de Audio. Elsevier.

Hume, K. y Ahtamad, M. (2013). Physiological responses to and subjective estimates of soundscape elements. *Applied Acoustics*, 74 (2), 275-281.

ICASSP, y IEEE Signal Processing Society. (eds.) (2005). Actas de la *Conferencia Internacional IEEE 2005 sobre acústica, habla y procesamiento de señales: 18-23 de mayo de 2005, Centro de Convenciones, Hotel Marriot. Filadelfia, Pensilvania, Estados Unidos.* Nueva Jersey, Estados Unidos: IEEE Operations Center.

Iconoclasistas (2013). *Manual de mapeo colectivo: recursos cartográficos críticos para procesos territoriales de creación colaborativa*. Risler, J. y Ares, P. (eds). 1ª ed. Buenos Aires, Argentina: Tinta Limón.

Institute of Electrical and Electronics Engineers (ed.) (2013). *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2013): New Paltz, New York, USA, 20 - 23 October 2013*. Piscataway, NJ, Estados Unidos: IEEE.

Institute of Electrical and Electronics Engineers, y IEEE Signal Processing Society (eds). (19 - 24 April 2015) ICASSP 2015: Conferencia internacional sobre acústica, habla y procesamiento de señales. South Brisbane, Queensland, Australia: IEEE.

Instituto Nacional de Astrofísica, JAC-O., Martínez-Trinidad, J.F., Olvera, J.A., y Boyer, K.L (eds.) (2012) *Reconocimiento de patrones: Cuarta Conferencia Mexicana, MCPR 2012, Huatulco, México, 27-30 de junio del 2012. Actas*. Procesamiento de imágenes, visión artificial, reconocimiento de patrones y gráficos. 2012. Vol. 7329. New York: Springer-Verlag Berlín Heidelberg.

Interaction, ACM Special Interest Group on Computer-Human, y ACM Digital Library. (2014). Actas del noveno audio en su mayoría: una conferencia sobre interacción con el sonido. Association for Computing Machinery, Nueva York, Estados Unidos. http://dl.acm.org/citation.cfm?id=2636879

International Listening Asociation. (abril de 1995) An ILA Definition of listening. *The listening Post*, 53 (1), 4-5.

ISO 12913-1:2014 Acoustics-Soundscape-Part I: Definition and Conceptual Framework. International Association for Standarization. (2014). Recuperado de https://www.iso.org/standard/52161.html

Iverson, P., y Krumhansl, L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America, 94*, 2595-2603.

Jain, A., y Harris, H. (2004, abril). *Speaker identification using MFCC and HMM based techniques*. University of Florida.

James, W. L. (1991). Book review: Multidimensional scaling: Concept and applications: P. E. Green, F. J. Carmone, Jr., and S. M. Smith (Boston: Allyn and Bacon, 1989, viii+407 pp). *Applied Psychological Measurement*, *16*(2), 206–208.

Jeck, P., y Kirkegaard J. (2013) *Soaked*. [CD] Londres: Touch Records.

Jennings, P., y Cain, R. (2013). A framework for improving urban soundscapes. *Applied Acoustics*, 74 (2), 293-299.

Jiang, D., Lu, L., Zhang, H., Tao, J., y Cai, L. (2002). Clasificación del tipo de música por función de contraste espectral. *Actas. IEEE International Conference on Multimedia and Expo, 1* (n), 113-116.

Jiang, W., Liu, J., Zhang, X., Wang, S., y Jiang, Y. (2020). Analysis and Modeling of Timbre Perception Features in Musical Sounds. Applied Sciences, 10(3), 789.

Johansen, O. (1983). *Introducción a la teoría general de sistemas* (1.ª ed.). Editorial Limusa.

Kajihara, Y., Dozono, S., y Tokui, N. (2017). Imaginary Soundscape: Cross-Modal Approach to Generate Pseudo Sound Environments. *Imaginary Soundscape. Take a walk in soundscapes "imagined" by AI*. Presentado en la 31ª Conferencia sobre Sistemas de Procesamiento de Información Neural (NIPS 2017): Machine Learning for Creativity and Design. Long Beach, CA, Estados Unidos.

Kane, B. (2012). Acousmate : History and de-Visualised Sound in the Schaefferian Tradition. Organised Sound, 17 (2), 179-88.

Kang, J., y Schulte-Fortkamp, B. (2016). *Soundscape and the Built Environment*. Abingdon, Reino Unido: CRC press.

Kauppinen I. (2002) *Methods for detecting impulsive noise in speech and audio signals.* Proceedings of the 14th International Conference on Digital Signal Processing (DSP 2002), 2, 967-970.

Kendall, G. (2008) What is an Event? The Event Schema, Circumstances, Metaphor and Gist. In: *Proceedings of the 2008 International Computer Music Conference*, Belfast, UK.

Kendall, G. S. (2010). Spatial Perception and Cognition in Multichannel Audio for Electroacoustic Music. *Organised Sound* 15 (03), 228-38.

Kittler, F. (1999). *Gramophone, Film, Typewriter. Writing Science*. Stanford, California, Estados Unidos: Stanford University Press.

Klapuri, A. (1999). Detección de inicio de sonido mediante la aplicación de conocimientos psicoacústicos. *1999. Conferencia internacional de IEEE sobre acústica, habla y procesamiento de señales. Actas. ICASSP99 (Cat. No.99CH36258), 6* (n), 3089-3092.

Klapuri, A., y Davy, M. (eds.) (2006). *Signal processing methods for music transcription*. New York, Estados Unidos: Springer.

Koenig, G. M. (s.f). *Gottfried Michael Koenig – COMPOSER*. About Gottfried Michael Koenig. http://koenigproject.nl/

Krimphoff, J., McAdams, S., y Winsberg, S. (1994). Caracterisation du timbre des sons complexes: 11. Analyses acoustiques et quantification psychophysique. *Journal de Physique, 4*, 625-628.

Krumhansl, L. (1989). Why is musical timbre so hard to understand? En Nielzen, S. & Olsson, O. (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43-53). Elsevier.

Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*(1), 1–27.

Kryter, K. (1972) “Acoustical Society of America Policy on Noise Standards”. J. Acoustical Society of America, 51 (3), 803-806.

Kulkarni, N., y Bairagi, V. (2018). Role of Different Features in Diagnosis of Alzheimer Disease. En EEG-Based Diagnosis of Alzheimer Disease (pp. 37-46). Elsevier.

Kurosu, M. (Ed.). (2013). *Human-Computer Interaction. Towards Intelligent and Implicit Interaction*. Springer Berlin Heidelberg.

Lachenmann, H. (1979). *Cuatro aspectos fundamentales de la escucha musical*. (Traductor: Bernal, A.) Alemania. Breitkopf & Härtel, Wiesbaden. Recuperado el 28 de abril del 2020 en http://www.tallersonoro.com/anterioresES/07/Articulo2.html

Lachenmann, H., y Häusler, J. (1996). *Musik als existentielle Erfahrung: Schriften 1966-1995 (German Edition)* Breitkopf & Härtel: Insel.

Lakatos, S. (2000). "A Common Perceptual Space for Harmonic and Percussive Timbres". *Perception & Psychophysics, 62*(7), 1426–39.

Langley, P. (1996). Elements of machine learning. San Francisco, California, Estados Unidos: Morgan Kaufmann.

Law, E., y Ahn, V. L. (2011). *Human Computation (Synthesis Lectures on Artificial Intelligence and Machine Learning)* (1st ed.). Morgan & Claypool Publishers.

Lawless, W., Mittu, R., y Sofge, D. (2020). *Human-Machine Shared Contexts* (1st ed.). Academic Press.

Lazier, A., y Cook, P. (2003). *MoSievius: Feature Driven Interactive Audio Mosaicing*. 8–11. In Proceedings of the International Conference on Digital Audio Effects (DAFX), London, England.

Lemaitre, G. y Susini, P. (2019). Timbre, Sound Quality, and Sound Design. En K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper y R. R. Fay (Eds.), *Timbre: Acoustics, Perception, and Cognition*, 69, 245-272.

Lerdahl, F., y Jackendoff, R. (2003). *Teoría Generativa De La Música Tonal/ The General Theory of the Musical Tone* (J. Gonzalez-Castelao, Trad.). Akal.

Levack Drever, J. (2002) Soundscape composition: the convergence of ethnography and acousmatic music. *Organised sound*, 7(1), 21-27.

Leveau P., Daudet L., y Richard G. (2004) *Methodology and tools for the evaluation of automatic onset detection algorithms in music*. Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR). Barcelona, España.

Levitin, D. (ed.) (2002). *Foundations of cognitive psychology: core readings*. Cambridge, Mass: MIT Press.

Librosa — librosa 0.8.0 documentation. (s. f.). librosa. https://librosa.org/doc/latest/index.html

Lindsay, T., Burnett, I., Quackenbush, S., y Jackson, M. (2003) Fundamentals of Audio Descriptors, *Introduction to MPEG-7 Multimedia Content Description Interface*, West Sussex, Inglaterra.

López, F. (1988) *La Selva. Sound environments from a Neotropical rain forest* [CD] Holanda: V2_Archief.

Lopez, F. (2004). Profound listening and environmental sound matter. In: Cox, C, Warner, D, editors. *Audio culture: readings of modern music*. New York (NY): Continuum International Publishing Group. p. 82-87.

López, R. (1997). *Música Plurifocal: Conciertos de Ciudades de Llorenç Barber*. D.F., México: JGH Editores.

Lopez-Tello, C. (2016). *Acoustic Detection, Source Separation, and Classification Algorithms for Unmanned Aerial Vehicles in Wildlife Monitoring and Poaching*. Estados Unidos: University of Nevada, Las Vegas.

Loy, G. (2006) *Musimathics. The mathematical foundations of music*. (Vol. I) Cambridge Mit Press.

Lu, L., Zhang, H.J., y Jiang, H. (s.f). Análisis de contenido para clasificación y segmentación de audio. *IEEE Trans. en Audio, Speech and Language Processing*, 10 (7), 504.

Maestre, E., Ramírez, R., Kersten, S., y Serra, X. (2009). Expressive Concatenative Synthesis by Reusing Samples from Real Performance Recordings. *Computer Music Journal*, *33*(4), 23-42.

Magariños de Morentin, J. A. (2008). La semiótica de los bordes: Apuntes de metodología semiótica. Comunicarte.

Maka, T. (2016, septiembre). A comparative study of onset detection methods in the presence of background noise. *2016 International Conference on Signals and Electronic Systems (ICSES)*.

Malasky, J. (2005). Human machine collaborative decision making in a complex optimization system. Massachusetts Institute of Technology.

Malbrán, S. (2007). *El oído de la mente: teoría musical y cognición*. Madrid, España: Akal.

Malkin, R., y Waibel, A. (2005). Classifying user environment for mobile applications using linear autoencoding of ambient audio. *IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Manoury, P. (2013). Compositional Procedures inTensio. *Contemporary Music Review*, *32*(1), 61-97.

Maresz, Y. (2013). On Computer-Assisted Orchestration. *Contemporary Music Review*, *32*(1), 99-109.

Marks, L. E. (1974). On Associations of Light and Sound: The Mediation of Brightness, Pitch, and Loudness. The American Journal of Psychology, 87(1/2), 173.

Marozeau, J., de Cheveigne´, A., McAdams, S., y Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America, 114*(5), 2946–2957.

Marques, P.A., y De Araujo, C.B. (octubre del 2014). The need to document and preserve natural Soundscape recordings as acoustic memories. *Proceedings Invisible Places Conference*. Recuperado el 1 de junio de 2019 en http://invisibleplaces. org/pdf/ip2014-marques. pdf.

Marr, D. (2010). *Vision: a computational investigation into the human representation and processing of visual information*. Cambridge, Mass: MIT Press.

Marr, D. (n.d.). *Teoría de la visión: PSICOLOGÍA DE LA PERCEPCIÓN VISUAL*. Universidad de Barcelona. http://www.ub.edu/pa1/node/93

Martin, B. (2018). Soundscape Composition: Enhancing our understanding of changing soundscapes. *Organised Sound*, 23 (1), 20-28.

Martin, K., Scheirer, E., y Vercoe, B. (1998). Music content analysis through models of audition. *Proceedings of ACM Multimedia Workshop on Content Processing of Music for Multimedia Applications*.

Mary, M. (2004). *Techniques de contrôle de logiciels en informatique musicale*. (Tesis doctoral). Atelier National de Reproduction des Thèses. Lille, Francia.

Mary, M. (2005). L'orchestration électroacoustique. En Vande Gorne, A. (Dirección), *Revue d'esthétique musicale, Colloque d'analyse perceptive des musiques électroacoustiques*. Ohain: Musiques & Recherches.

Mary, M. (2007). El sonido del siglo XXI. En Fessel, P. (comp.) *Nuevas poéticas de la música contemporánea argentina* (pp. 154-169). Buenos Aires, Argentina: Biblioteca Nacional.

Mary, M. (2016). De la idea a la obra: Los caminos de la creatividad. Orquestación electroacústica y polifonía del espacio. *Ideas sónicas*, 10 (19), 7.

McAdams, S., y Giordano, B. (2012). The perception of musical timbre. *Oxford University Press*, *1*.

McAdams, S. (2013). Musical Timbre Perception. *The Psychology of Music*, 35–67.

McAdams, S., y Giordano, B. (2016). *The Perception of Musical Timbre*. Hallam, S., Cross, I. y Thaut, M. (eds.) Oxford University Press.

McCartney, A. (2000) *Soundscape Composition and the Subversion of Electroacoustic Norms*. Canadá. Recuperado el 15 de febrero del 2020 en https://naisa.ca/radio-art-companion/soundscape-composition-and-the-subversion-of-electroacoustic-norms/

Merino de la Fuente, J. M. (2010). *Las vibraciones de la música*. España: Editorial Club universitario.

Merriam-Webster. (n.d.). *Dictionary.* The Merriam-Webster.Com Dictionary. Retrieved 2020, from https://www.merriam-webster.com/

Metze, F., Rawat, S., y Wang, Y. (2014) Improved audio features for large-scale multimedia event detection. *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. New York, Estados Unidos: IEEE.

Meyda (2021) *Descriptores acústicos* https://meyda.js.org/audio-features.html (visitado en marzo del 2019)

Meyer, A. (1972): "The Need for Standards on noise". *Journal Acoustic Society American*, 51 (3). 800-802.

Miller, J., y Carterette, E. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America*, *58*, 711–720.

Mindell, D. A. (2002). *Between Human and Machine: Feedback, Control, and Computing before Cybernetics (Johns Hopkins Studies in the History of Technology)* (1st ed.). Johns Hopkins University Press.

Middlebrooks, J. C., Simon, J. Z., Popper, A. N., y Fay, R. R. (Eds.). (2017). The Auditory System at the Cocktail Party. *Springer Handbook of Auditory Research*, *60*.

Misdariis, N., Smith, B., Presssnitzer, D., Susini, P., y McAdams, S. (1998). Validation of a multidimensional distance model for perceptual dissimilarities among musical timbres. *Journal of the Acoustical Society of America, 103*, 3005–3006.

Misra, A., Cook, P. R. y Wang, G. (2006) Musical Tapestries: Re-composing Natural Sounds. In Proceedings of the 2006 International Computer Music Conference (ICMC), New Orleans, U.S.

Moffat, D., y Reiss, J. (2015). An Evaluation of Audio Feature Extraction Toolboxes. DAFx-15, 8.

Mosaicr seagull. (2006). [Mosaic, image details and attributions are at Main image derived from See through you by Steve Jurvetson]. https://es.wikipedia.org/wiki/Archivo:Mosaicr_seagull.jpg

MoSievius: A System for Interactive Audio Mosaicing. (s. f.). https://soundlab.cs.princeton.edu/research/mosievius/.

Muda, L., Begam, M., y Elamvazuthu, I. (2010, marzo 22). Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. Recuperado de https://arxiv.org/abs/1003.4083

Müller, M. (2015) *Fundamentos del procesamiento de música: Audio, Análisis, Algoritmos, Aplicaciones*. Springer Berlin Heidelberg.

Murail, T. (1989). Questions de cible. *Entretemps 8*. (pp. 147- 174). Vendôme, Francia: Presses Universitaires de France.

N/A. (s.f) GuideEpilogue. [PDF] Recuperado el 10 de octubre del 2019 en http://www.iremus.cnrs.fr/sites/default/files/EARS-23-03-2016/08-GuideEpilogue.pdf

N/A. (s.f) *Sound listening and place the aesthetic dilemma.* [PDF] Recuperado el 21 de marzo de 2018 de https://www-cambridge-org.pbidi.unam.mx:2443/core/services/aop-cambridge-core/content/view/D3120A5CAE27670CD73E505BC71721FC/S1355771811000380a.pdf/sound_listening_and_place_the_aesthetic_dilemma.pdf.

Nancy, J. L. (2007). *A la escucha*. Buenos Aires, Argentina: Amorrortu.

Navarro, C., Zamora-Torres, A., y Cano, M. (2015). Análisis de escalamiento multidimensional del turismo de los países que conforman la APEC. *Turismo y Sociedad*, *16*, 17–30.

Neapolitan, R.E, y Jiang, X. (2018). *Artificial Intelligence: With an Introduction to Machine Learning*.

Neupert, M., y Gossmann, J. (2013). A remix instrument based on fragment feature-analysis. *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 1-5.

Nicol, A. (2005) *Development and Exploration of a Timbre Space Representation of Audio*. (PhD thesis), University of Glasgow.

Nilsson, N. (1986). *Principles of artificial intelligence*. Los Altos, Calif: Morgan Kaufmann Publishers.

Nilsson, N. (2010). *The quest for artificial intelligence: a history of ideas and achievements*. Cambridge; New York: Cambridge University Press.

Norman, K. (1996). Real-world music as composed listening. *Contemporary Music Review*, 15 (1): 1-27.

Normandeau, R. (2009). Timbre Spatialisation: The Medium Is the Space. *Organised Sound*, 14 (03): 277.

Norowi, N. M., Mustaffa, M. R., y Miranda, E. R. (2016). Using concatenation cost for unit selection of homosonic segments in concatenative sound synthesis. *2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP)*, 37-42.

North, C., y Dack, J. (2017). Treatise on musical objects: Essays across disciplines. University of California Press.

Noya, M. (2012). Ambiente sónico-sonidos del ambiente - De arte sonoro a la música en un canto de ave. Reflexiones y conceptualización personal en el proceso de

composición y realización de obra electroacústica. En Rodriguez, L. G., y Izarra, A. (Compiladores), *Canto electroacústico: Aves latinoamericanas en una creación colaborativa.* (pp. 119-149) España: Ariel. Fundación Telefónica.

Nuanáin, C. Ó., Herrera, P., y Jordá, S. (2018). k-Best Unit Selection Strategies for Musical Concatenative Synthesis. *Music Technology with Swing*, 76-97.

Ong, B.S, Serra, X., Universitat Pompeu Fabra, y Institut Universitari de l'Audiovisual. (2007). *Structural Analysis and Segmentation of Music Signals*. Barcelona: Universitat Pompeu Fabra. Recuperado el 10 de julio de 2019 en http://www.tdx.cat/TDX-0117108-190540/

Opitz, M. F., y Zbaracki, M. D. (2004). *Listen Hear! 25 Effective Listening Comprehension Strategies*. Heinemann.

Oracle (2020) Base de datos. [Sitio Web] Recuperado el 15 de marzo del 2020 https://www.oracle.com/mx/database/what-is-database/

Orchestron. (2009, 9 junio). *James Tenney Collage #1 («Blue Suede»)*. YouTube. https://www.youtube.com/watch?v=xC7sdH2XvbU

O'Shaughnessy, D. (1999). *Speech Communications: Human and Machine* (2nd ed.). Wiley-IEEE Press.

Oswald, J. (s.f). *Plunderphonic*. Album Notes. http://www.plunderphonics.com/xhtml/xnotes.html#DAB

Oswald, J., y McCartney, P. (2011, July 16). *John Oswald - Paul McCartney: Birth* [Video]. YouTube. https://www.youtube.com/watch?v=QWmTUse-sv8

Olswald, J. (2016, November 1). *Plunderphonics — DAB — the official video* (1988-9) [Video]. YouTube. https://www.youtube.com/watch?v=Rhck5dVP9AA

Paddison, M., y Deliège, I. (Eds.). (2010). Contemporary music: Theoretical and philosophical perspectives. Ashgate.

Pancoast, S., y Akbacak, M. (2012) Bag-of-audio-words approach for multimedia event classification. *Proceedings of Interspeech: 13th Annual Conference of the International Speech Comunication Association*, pp. 2105–2108. ISCA, Portland, OR, Estados Unidos.

Peeters, G., McAdams, S. y Herrera, P. (2000, setiembre) Instrument sound description in the context of MPEG-7. In Proceedings of the 2000 International Computer Music Conference, 166–169.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification). Technical Report: *CUIDADO IST Project Report*, *54*, 1–25.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., y McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. The Journal of the Acoustical Society of America, 130(5), 2902-2916.

Peltonen, V. (2001). *Computational auditory scene recognition*. (Tesis de mestría) Tampere University of Technology, Finlandia.

Penha, R., y Carvalhais, M. (2018) Will Machinic Art Lay Beyond Our Ability to Understand It? En Adebayo, R., Farouk, I., Jones, S. y Rapeane-Mathonsi, M. (eds.) ISEA 2018. Pp. 22-30. Durban: Faculty of Arts and Design, Durban University of Technology.

Penha, R., y Carvalhais, M. (2019) If machines want to make art, will humans understand it? *Aeon*.

Peña, D. (2002). *Análisis de Datos Multivariantes.* McGraw Hill.

Piczak, K.J. (2015) Environmental sound classification with convolutional neural networks. *IEEE: International Workshop on Machine Learning for Signal Processing*. Setiembre 17-20, 2015. Boston, Estados Unidos.

Pierce, A. D. (2019). *Acoustics: An introduction to its physical principles and applications* (Third edition). ASA Press, Springer.

Pietruszewski, M. (s. f.). *Fancyyyyy*. Bandcamp. https://fancyyyyy.bandcamp.com/

Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., Gage, S. H. y Pieretti, N. (2011). Soundscape Ecology: The Science of Sound in the Landscape. *BioScience*, 61 (3), 203-216.

Piston, W. (1955). *Orchestration*. Norton.

Plomp, R. y Levelt, W.J.M. (1965) Tonal Consonance and Critical Bandwidth. Journal of the Acoustical Society of América, 38(4), 548-568.

Plomp, R., Pols, L. C. W., y Van de Geer, J. P. (1967). Dimensional Analysis of Vowel Spectra. *The Journal of the Acoustical Society of America*, *41*(3), 707-712.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. En R. Plomp & G. F. Smoorenburg (Eds.), *Frequency Analysis and Periodicity Detection in Hearing* (pp. 397–414). Leiden.

Pohjalainen, J. (2017). *Methods of Automatic Audio Content Classification* (Master Thesis). Helsinki University of Technology.

Poole, D. L., y Mackworth, A. (2010). *Artificial intelligence: foundations of computational agents*. New York: Cambridge University Press.

Poturak, M., y Goksu, A. (2012, diciembre). *Usage of the multidimensional scaling in exploring a firm's image and Competitiveness.* Advanced Research in Scientific Areas Volume: INTERNATIONAL VIRTUAL CONFERENCE. https://doi.org/10.13140/2.1.2848.1923

Pratt, R., y Doak, P. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration*, *45*(3), 317–328.

Raimbault, M., y Dubois, D. (2005). Urban soundscapes: Experiences and knowledge. *Cities*, *22*(5), 339-350.

Ramírez, S., Torres Alcaraz, C., Cocho Gil, G., y Torres Nafarrate, J. (Eds.). (1999). Perspectivas en las teorías de sistemas: Por Carlos Torres Alcaraz... Coord. por Santiago Ramírez (1. ed). Siglo Veintiuno Ed.

Ras, Z. W., y Wieczorkowska, A. (2010). *Advances in Music Information Retrieval* (1.ª ed., Vol. 274). Springer Publishing.

Real Academia Española (2001). Ciudad. En *Diccionario de la Lengua Española.* Recuperado el 15 de setiembre de 2021, de https://www.rae.es/drae2001/ciudad

Reich, M. (2016). *Soundscape Composition as Environmental Activism and Awareness: An Ecomusicological Approach.* Estados Unidos: University of Puget Sound.

Rich, E., y Knight, K. (2004) *Artificial Intelligence*. Gardners Books.

Risset, J. C., y Wessel, D. (1982) *Indagine sul timbro mediante analisi e sintesi: Quaderno 2*. Limb/ Bienal de Venecia. Pp. 28-29.

Rivera, J., y Molero, V. (2010). El Escalado Multidimensional en la investigación de mercados. Usos y aplicaciones. *Investigación y Marketing*, *70*, 16–23.

Roads, C. (1996). The computer music tutorial. MIT Press.

Roads, C. (2001). *Microsound*. MIT Press.Robindoré, B. (s.f.) Interview with Curtis Roads. Recuperado de https://static1.squarespace.com/static/5ad03308fcf7fd547b82eaf7/t/5b8348994fa51af3fa01769f/1535330460627/03+Roads+Interview.pdf

Roca, M., Anguera, B., Domenjó, B., Molla, C., Rivas, G., de la Casa Rivas, G., y Civera, C. (2009). *Historia de la psicología*. Barcelona, España: Editorial UOC.

Rocchesso, D., y Fontana, F. (eds). (2003). *The Sounding Object.* Florencia, Italia: Edizioni di Mondo Estremo.

Rodríguez, L. G., y Izarra, A. (2012). *Canto electroacústico: Aves latinoamericanas en una creación colaborativa.* España: Ariel. Fundación Telefónica.

Roederer, J. G. (2008). *The physics and psychophysics of music: An introduction* (4ª ed.). Springer.Roederer, J. G., y Pozzati, G. D. (2000). *Acústica y psicoacústica de la música*. Buenos Aires, Argentina: Ricordi.

Roy, S. (2003). *L'analyse des musiques életroacoustiques: Modèles et propositions*. París, Francia: L'Harmattan.

Rudy, P. (2007). Timbral Praxis: When a Tree Falls in the Forest Is It Music? *Organised Sound,* 12 (01), 5.

Russolo, L. (1913/1986). El arte de los ruidos. Nueva York: Editorial Prendragon.

Russolo, L. (2018) *El arte de los ruidos*. Madrid, España: Dobra Robota Editorial.

Russomanno, S. (2009). La Selva [archivo de audio]. En Francisco López Through the Looking-Glass. *La Quinta de Mahler*. http://www.laquintademahler.com/shop/detalle.aspx?id=52873

Rychtáriková, M. y Vermeir, G. (2013). Soundscape categorization on the basis of objective acoustical parameters. *Applied Acoustics*, 74 (2), 240-247.

Saariaho, K. (1983) Using the Computer in a Search for New Aspects of Timbre Organisation and Composition. *International Computer Music Conference*

Proceedings. Ann Arbor, Michigan: University of Michigan Library editions, 1983 (n), 269-273.

Saariaho, K. (1984) Shaping A Compositional Network with Computer. *International Computer Music Conference Proceedings*. Ann Arbor, Michigan: University of Michigan Library editions, 1984 (n), 163-165.

Saariaho, K. (1987) Timbre and Harmony: Interpolations of Timbral Structures. Contemporary Music Review*. Overseas Publishers Association*, 2 (1), 93-133.

Salomon, J., y Bello, J.P. (2015) Unsupervised feature learning for urban sound classification. *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), pp. 171–175.

Salamon, J., MacConnell, D., Cartwright, M., Li, P., y Bello, J. (2017). Scaper: una biblioteca para la síntesis y el aumento del paisaje sonoro. En *2017, IEEE Workshop sobre aplicaciones de procesamiento de señales para audio y acústica, WASPAA 2017* (Vol. 2017-octubre, pp. 344-348). Instituto de Ingenieros Eléctricos y Electrónicos Inc. New Paltz, Estados Unidos.

Salcedo, F. (2007). *Modelos ocultos de Markov: del reconocimiento de voz a la música*. S.l.: Lulu.

Salesa Batlle, E., Perelló Scherdel, E., y Bonavida Estupiñá, A. (2008). *Tratado de Audiología*. Masson.

Sanchez, B. (2010). Spectrum and Scale. Worcester Polytechnic Institute.

Sánchez, R. C., y Cardemil, J. A (1994). Montaje cinematográfico: Arte de movimiento. UNAM, Centro Universitario de Estudios Cinematográficos, Coordinación de Difusión Cultural.

Schaeffer, P. (1967/2008). *Tratado de los objetos musicales*. Madrid, España: Alianza Editorial S.A.

Schafer, R.M. (1969). *El nuevo paisaje sonoro*. Buenos Aires, Argentina: Ricordi Americana S.A.

Schafer, R.M. (1977a*). The soundscape. Our sonic environment and the tuning of the world*. Vermont, Canadá: Destiny Books.

Schafer, R.M. (1977b). *The tuning of the world. New York: Knopf. reprinted as The soundscape: our sonic environment and the tuning of the world*. Rochester, VT: Destiny Books 1994.

Schafer, R.M. (1992). *A sound education: 100 exercises in listening and sound-making. Indian River*. Ontario, Canadá: Arcana Editions.

Schafer, R.M. (2016). El paisaje sonoro y la afinación del mundo. (Traductor: Cazorda, V. G.) España: Intermedio. (1993).

Scharf, B. (1978). Loudness. En M. P. Friedman & E. C. Carterette (Eds.), *Hearing* (pp. 187-242). Amsterdam University Press.

Schlüter, J. (2011) Unsupervised Audio Feature Extraction for Music Similarity Estimation. (Tesis de maestría) Facultad de Informática, Universidad de Munchen, Alemania. Recuperado el 10 de setiembre del 2019 en http://www.ofai.at/~jan.schlueter/pubs/msc.pdf

Schoder, J. (2019). *Music Similarity Analysis Using the Big Data Framework Spark* (Thesis Master) Universidad JENA, Alemania.

Schoenberg, A. (2004). *Tratado de armonía*. Las Matas, Madrid: Real Musical.

Schouten, J.F. (1968). The Perception of Timbre. En Kohasi, Y. (ed) *Reports of the 6th International Congress on Acoustics, Tokyo,* GP-6-2, 6 vols., 35–44, 90. Tokyo: Maruzen; Amsterdam: Elsevier.

Schrader, B. (1982) *Introduction to Electroacoustic Music.* Universidad de Michigan, Estados Unidos.

Schwarz, D. (2004). Data-Driven Concatenative Sound Synthesis. (Doctoral dissertation) Université Paris 6 Pierre et Marie Curie.

Schwarz, D. (2005). *Current Research in concatenative sound synthesis.* Proceedings of the International Computer Music Conference.

Schwarz, D. (2006). Concatenative sound synthesis: The early years. *Journal of New Music Research*, *35*(1), 3-22.

Schwarz, D., Beller, G., Verbrugghe, B., y Britton, S. (2006, septiembre). *Real-time Corpus-based concatenative synthesis with Catart*. 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada. http://www.ircam.fr/anasyn/schwarz

Schwarz, D. (2007). Corpus-Based Concatenative Synthesis. *IEEE Signal Processing Magazine*, *24*(2), 92-104.

sCrAmBlEd?HaCkZ! (2006, March 20). [Video]. YouTube. https://www.youtube.com/watch?v=eRlhKaxcKpA

Segura, S. (2014). Lexicón etimológico y semántico del latín y de las voces actuales que proceden de raíces latinas o griegas. Universidad de Deusto.

Sethares, W. A. (1998). Consonance-Based Spectral Mappings. *Computer Music Journal*, 22(1), 56.

Sethares, W. A. (2005). *Tuning, timbre, spectrum, scale*. (2nd ed). Springer.

Shepard, R. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika, 27*(2), 125–140.

Sheu, P.C.Y (Ed.). (2018). *Semantic computing*. World Scientific.

Shifres, F., Burcet, M. I., y ProQuest. (2013). Escuchar y pensar la música. Editorial de la Universidad de La Plata.

Shukla, N. (2017). *Machine Learning with TensorFlow*. San Jose, Costa Rica: Manning Publications.

Siedenburg, K., Saitis, C., McAdams, S., Popper, A. N., y Fay, R. R. (Eds.). (2019). Timbre: Acoustics, Perception, and Cognition. *Springer Handbook of Auditory Research*, *69*.

Sigal, R. (2012). Una reflexión sobre los centros de creación de música electroacústica en Latinoamérica. En *Canto electroacústico: Aves latinoamericanas en una creación colaborativa* (pp. 26-34). Madrid, España: Ariel.

Sigal, R. (2014). *Estrategias compositivas en la música electroacústica: generación de materiales y creación de un lenguaje musical eficaz. Música y Ciencia*. Bernal, Argentina: Universidad Nacional de Quilmes Editorial.

Silvers, R. (2000). *Photomosaics*. Robert Silvers: Inventor of the Photomosaic Process and Original Artist. https://www.photomosaic.com/statement.html

Simon Fraser University. (s.f) *World Soundscape Project*. [Página Web] Recuperado el 18 de marzo de 2020 en https://www.sfu.ca/sonic-studio/worldsoundscaperoject.html

Skowronski, M. D., y Harris, J. G. (2002). Human factor cepstral coefficients. *The Journal of the Acoustical Society of America*, *112*(5), 2279.

Skowronski, M. D., y Harris, J. G. (2003, mayo). *Improving The Filter Bank Of A Classic Speech Feature Extraction Algorithm*. IEEE Intl Symposium on Circuits and Systems, vol IV, pp 281-284. Bangkok, Thailand.

Smalley, D. (1986). Spectro-morphology and Structuring Processes. En S. Emmerson (Ed.), *The Language of Electroacoustic Music* (pp. 61-93). Palgrave Macmillan UK.

Smalley, D. (1994). Defining Timbre — Refining Timbre. *Contemporary Music Review,* 10 (2), 35-48.

Smalley, D. (1995). La spectromorphologie. Une explication des formes du son. En Poissant, L. (ed.), *Esthétique des arts médiatiques*, vol 2, p. 126. Presses de l'Université du Québec.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound*, *2*(2), 107-126.

Smalley, D. (2020). *Espectromorfología. Una explicación de las formas del sonido*. Ars Sonora.

Smith, L. (1994). Sound segmentation using onsets and offsets. *Journal of New Music Research*, 23 (n),11–23.

Sobrino, J. A., Sobrino, J. A., Raissouni, N., y Kerr, N. (2000). Teledetección. Universitat de València.

Solis, H. (2018). SONODE; a collaborative sound map of Mexico City. *Universidad Autonoma Metropolitana, Unidad Lerma*, 169–170. http://hugosolis.net/media/pages/publications/sonusgo_icmc/867913556-1591008852/bbp2372.2018.032.pdf

Solis, H. (2018b). *El proceso en la producción de arte electrónico* (1st ed.). Universidad Autónoma Metropolitana. http://hugosolis.net/media/pages/publications/produccion_arte_electronico/979975346-1591008796/ElProcesoEnLaProduccionDeArteElectronico.pdf

Sonic Visualiser (2006–2008). *Sonic Visualiser: A Brief Reference*. [App] University of London. https://www.sonicvisualiser.org/doc/reference/1.2/en/

Sononym (1.1.1). (2020). [A sample browser that offers a fresh perspective on how sounds can be explored and organized.]. Sononym UG. https://www.sononym.net/

Staff, W. (2017, June 4). *The Man Behind Scrambled Hackz*. WIRED. https://www.wired.com/2006/04/the-man-behind-scrambled-hackz/

Steintrager, J., y Chow, R. (eds.) (2019). *Sound objects*. Durham: Duke University Press.

Stennett, C. (2013). *Concatenative Sound Synthesis and Microsound: Establishing a New Aesthetic*. https://www.academia.edu/20211019/Concatenative_Sound_Synthesis_and_Microsound_Establishing_a_New_Aesthetic

Stowell, D., Giannoulis, D., Benetos, E., Lagrange, M., y Plumbley, M. D. (2015). Detection and classification of acoustic scenes and events. IEEE Transactions on Multimedia, 17 (10), 1733-1746.

Stroppa, M. (1991). Un orchestre synthétique: Remarques sur une notation personnelle. En Barrière, J.B. (ed.). *Le Timbre, métaphore pour la composition* (pp. 485-537). París, Francia: Christian Bourgois.

Sturm, B. L. (2006). Adaptive Concatenative Sound Synthesis and Its Application to Micromontage Composition. *Computer Music Journal*, *30*(4), 46-66.

Suárez, R., y Carrasco, M. (2018, July 7). *La geometría del taxista en 3 minutos (Taxicab geometry)* [Video]. YouTube. https://www.youtube.com/watch?v=Z7Y_s-ON05o

Subramanian, H. (2004, noviembre) Audio Signal Classification. *M.Tech. Credit Seminar Report, Electronic Systems Group*, EE. Dept, IIT Bombay.

Suchman, L. (2006). *Human-Machine Reconfigurations: Plans and Situated Actions (Learning in Doing: Social, Cognitive and Computational Perspectives)* (2nd ed.). Cambridge University Press.

Szendy, P., y Nancy, J.L. (2003). *Escucha: una historia del oído melómano*. Barcelona, España: Paidós.

Tardieu, D. (2008). *Modèles d'instruments pour l'aide à l'orchestration* (Unpublished doctoral dissertation). Université Pierre & Marie Curie, Paris 6

Tarkovskii, A., y Bustos, M. (2013). *Esculpir el tiempo*. D.F, México: UNAM. Centro Universitario de Estudios Cinematográficos.

Taube, H. K. (2004). *Notes From The Metalevel: An Introduction To Algorithmic Music Composition* (Pap/Cdr ed.). Taylor & Francis.

Tenney, J. (2009, June 9). *James Tenney Collage #1 ("Blue Suede")* [Video]. YouTube. https://www.youtube.com/watch?v=xC7sdH2XvbU

Thibaud, J.P. (2003) The Sonic Composition of the city. The Auditory Culture Reader. Oxford: Berg, 2003. pp. 329-341.

Thibaud, J.P., Kalogianni, D., y Wolfgang T., et al. (2014) Invisible places sounding cities. Sound, urbanism and sense of place. IV Edición. Portugal: Jardins Efémeros: Viseu.

Thomas, J. (1983) Collage/Space/Montage. En Parisier, J. (ed) *Collage*. New York: New York Literary Forum.

Thoresen, L. (1979). Spectromorphology: explaining soundshapes. *Organized Sound*, vol. 2.

Thoresen, L. (1985a). Un model d'analyse auditive. Analyse Musicale No. 1 (Traducción al francés del artículo publicado en Semiotica).

Thoresen, L. (1985b). Auditive Analysis of Musical Structures. A summary of analytical terms, graphical signs and definitions. *Proceedings from ICEM Conference on Electroacoustic Music Stockholm, Sweden, 25 –27 September.*

Thoresen, L. (1987). An Auditive analysis of Schubert's Piano Sonata Op. 42. *Semiotica,* 66-1/3.

Thoresen, L., & Hedman, A. (2007). Spectromorphological analysis of sound objects: An adaptation of Pierre Schaeffer's typomorphology. *Organised Sound*, 12(2), 129-141.

Thorogood, M., Fan, J., y Pasquier, P. (julio del 2016) Soundscape Audio Signal Classification and Segmentation Using Listeners Perception of Background and Foreground Sound. *Journal of the Audio Engineering Society*. *Special Issue (Intelligent Audio Processing, Semantics, and Interaction)*, 64 (7/8), 484-492.

Thorogood, M., Fan, J., y Pasquier, P. (octubre del 2015). *BF-Classifier: Background/Foreground Classification and Segmentation of Soundscape*

Recordings. En Actas de la 10ma conferencia Audio Mayormente 2015 sobre Interacción con sonido (AM ’15). (Artículo 32, 1-6) Association for Computing Machinery, Nueva York, Estados Unidos.

Thorogood, M., y Pasquier, P. (2013a). *Computationally Generated Soundscapes with Audio Metaphor*. En la 4ta Conferencia Internacional sobre CReatividad Computacional. Junio-2013 Sydney, Australia.

Thorogood, M., y Pasquier, P. (2013b). *Impress: A Machine Learning Approach to Soundscape Affect Classification for a Music Performance Environment*. Proceedings of the 13th International Conference on New Interfaces for Musical Expression, Daejeon + Seoul, Korea Republic.

Thorogood, M., Pasquier, P., y Eigenfeldt, A. (2012). *Audio Metaphor: Audio Information Retrieval for Soundscape Composition*. En Actas de la IX Conferencia de Computación de Sonido y Música.

Tindale, A., Kapur, A., Tzanetakis, G., y Fujinaga, I. (2004). Retrieval of percussion gestures using timbre classification techniques. *Proceedings of the International Symposium on Music Information Retrieval (ISMIR 2004),* Audiovisual Institute Pompeu Fabra University, pp. 541–544.

Toral, R. (2001) Violence of discovery and calm of Acceptance. [CD] Londres: Touch Records.

Torcoli, M. (2019). An Improved Measure of Musical Noise Based on Spectral Kurtosis. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) 90-94. SigPort. http://sigport.org/4868

Torgeson, S. (1958). *Theory and methods of scaling*. Wiley.

Torigoe, K. (1982) *A Study of the World Soundscape Project*. (M.F.A. thesis). Toronto: York University.

Truax, B. (1984). *Acoustic communication*. 2nd ed. Ablex Publishing.

Truax, B. (1994). The inner and outer complexity of music. *Perspectives of New Music*, 32 (1), 176-193.

Truax, B. (1995). Sound in context: Acoustic communication and soundscape research at Simon Fraser University. *The Journal of the Acoustical Society of America*, *97*(5), 3407.

Truax, B. (1996). Paisaje sonoro, comunicación visual y composición con sonidos ambientales. *Contemporary Music Review*, 15 (1), 49-65.

Truax, B. (1999). *RIVERRUN*. http://www.sfu.ca/%7Etruax/river.html

Truax, B. (1999). *Handbook for Acoustic Ecology*. Cambridge Street Publishing.

Truax, B. (2000). The aesthetics of computer music: A questionable concept reconsidered. *Organised Sound*, 5 (3), 119-126.

Truax, B. (2001). *Acoustic communication*. 2nd ed. Westport, Conn: Ablex.

Truax, B. (2002). Genres and techniques of soundscape composition as developed at Simon Fraser University. *Organised Sound*, 7(1), 5-14.

Truax, B. (2008). Soundscape Composition as Global Music: Electroacoustic music as soundscape. *Organised Sound*, 13(2), 103-109.

Truax, B. (2012). Sound, Listening and Place: The aesthetic dilemma. *Organised Sound*, 17 (3), 193-201.

Truax, B. (2017a). Acoustic space, community and virtual soundscapes. En Cobussen, M., Meelberg, V., y Truax, B. (eds.*) The Routledge companion to sounding art.* pp. 253–263. New York: Routledge.

Truax, B. (2017b). Editorial. *Organised Sound*, 22 (1), 1–3.

Truax, B. (2020) *World Soundscape Project*. Recuperado el 18 de marzo de 2020 en https://www.thecanadianencyclopedia.ca/en/article/world-soundscape-project

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59 (n), 433-460.

Turing, A. M. (1996). Intelligent Machinery, a Heretical theory. *Philosophia Mathematica*, 4 (n), 256-260.

Tzanetakis, G., & Cook, P. (1999, June 14). *A framework for audio analysis based on classification and temporal segmentation*. In Proceedings of Euromicro99, Workshop on Music Technology and Audio processing 1999.

Tzanetakis, G., & Cook, P. R. (2000) *MARSYAS: A Framework for Audio Analysis*. Organized Sound 4(3), Cambridge University Press.

Tzanetakis, G., & Cook, P. (2002) Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process*, 10 (5), 293-302.

Umesh, S., Cohen, L., & Nelson, D. (1999). Fitting the Mel scale. En *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Proceedings. ICASSP 99 (Cat. No.99CH36258), 217-20 vol.1. Phoenix, Estados Unidos: IEEE.

Universidad Austral de Chile y Otondo, F. (2018). Paisajes sonoros reales e imaginarios. *Resonancias: Revista de investigación musical*, 42, 131-141.

Universidad de Korea (2014). TRECVID: Multimedia Event Detection. Universidad de Korea. [Material Audiovisual] Recuperado el 6 de enero del 2020 en https://www.youtube.com/watch?v=Z4RO3S7wScU

Universidad de Korea (2015). Sound Source Separation Realtime. [Material Audiovisual] Recuperado el 7 de enero del 20020 en https://www.youtube.com/watch?v=rG9oufs865g

Universidad Nacional de la Plata. Programa de Comunicación y Cultura. (2004). *Trampas de La Comunicación y La Cultura: Revista Del Programa de Comunicación y Cultura de La Facultad de Periodismo y Comunicación Social de La UNLP.*, *25*. https://books.google.com.mx/books?id=iPMuAAAAYAAJ

Universidad Nacional Autónoma de México, & M.M Castro, O. (2018). La ciudad como fuente de sonidos para la creación sonora. *Resonancias: Revista de investigación musical*, 42 (junio), 143-50.

Universitat Pompeu Fabra. (2000). *Algorithm reference: Dissonance — Essentia 2.1-beta6-dev documentation*. Essentia. https://essentia.upf.edu/reference/streaming_Dissonance.html

Vaggione, H. (1994). Timbre as syntax: A spectral modeling approach. *Contemporary Music Review*, *10*(2), 73-83.

Valentinuzzi, M. E. (2004). *Understanding the Human Machine*. World Scientific.

Valle A., Lombardo V., & Schirosa M. (2010) Simulando el paisaje sonoro a través de una metodología de análisis / resíntesis. En: Ystad S., Aramaki M., Kronland-Martinet R., Jensen K. (eds) *Auditory Display. CMMR 2009, ICAD 2009. Lecture Notes in Computer Science*, vol 5954. 330-357. Springer, Berlín, Heidelberg.

Vamp Plugins. (s. f.). The Vamp audio analysis plugin system. https://www.vamp-plugins.org/

Varèse, E. (1917) The liberation of sound (compiled and edited by Chou Wen-cheng). In B. Boretz and E. T. Cone (eds.) Perspectives on American Composers. New York: W. W. Norton (1971).

Virtanen, T., Plumbley, M. & Ellis, D. (eds.) (2017) Computational analysis of sound scenes and events. New York, Estados Unidos: Springer Science.

Voegelin, S. (2010). *Listening to noise and silence: towards a philosophy of sound art*. New York, Estados Unidos: Continuum.

Voorvelt, M. (1997). The Enviroment Element in Barry Truax´s Composition. *Journal of New Music Research.* 26 (1997), 48 – 69.

Von Helmholtz, H. (1954/1877) On the Sensations of Tone as a Physiological Basis for the Theory of Music. 4ta edición. Ellis, A. (trad. Inglés) Dover.

Wang, D. L., & Brown, G. (eds.) (2006). *Computational auditory scene analysis: principles, algorithms, and applications*. Estados Unidos: Wiley-IEEE Press.

Wang, W. (Ed.). (2011). *Machine Audition: Principles, algorithms, and systems*. Information Science Reference.

Wapnick, J., & Freeman, P. (1980). Effects of Dark-Bright Timbral Variation on the Perception of Flatness and Sharpness. Journal of Research in Music Education, 28(3), 176-184.

Ward, D. (1970) Musical Perception. En Tobias, J. (ed*) Foundations of Modern Auditory Theory* vol. 1. New York: Academic Press.

Watson, C. (2010) *A journey South*. [CD] Londres: Touch Records.

Watson, C. (2011) *El tren fantasma*. [CD] Londres: Touch Records.

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrument timbre. *Scandinavian Journal of Psychology, 13*, 228- 240.

Wessel, D. L. (1973). Psychoacoustics and music: A report from Michigan State University. *PACE: Bulletin of the Computer Arts Society*, *30*, 1–2.

Wessel, D. (1979). Low Dimensional Control of Musical Timbre. Computer Music Journal, 3 (n), 45–52.

Westerkamp, H. (2002). Linking soundscape composition and acoustic ecology. *Organised Sound*, 7 (1), 51–56.

Westerkamp, H. (2002/2015). *Bauhaus y estudios sobre el paisaje sonoro. Explorando conexiones y diferencias* [en línea]. Disponible en: http://www.eumus.edu.uy/eme/ps/txt/westerkamp.html [visitado el 13 de abril de 2019].

Wiil, U. (ed.) (2005). Modelado y recuperación de música por computadora: Segundo Simposio Internacional, CMMR 2004, Esbjerg, Dinamarca, 26-29 de mayo de 2004, documentos revisados. Sistemas de información y aplicaciones, incl.. Internet / Web y HCI, 3310. New York: Springer-Verlag Berlín Heidelberg.

Winderen, J. (2013) *Aquaculture*. [CD] Londres: Touch Records.

Winderen, J. (2014) *Out of range*. [CD] Londres: Touch Records.

Windsor, L. (2000). Through and around the acousmatic: the interpretation of electroacoustic sounds. En Emmerson, S. (comp) Music, electronic media, and culture. 7-35. Aldershot; Burlington, Estados Unidos: Ashgate.

Wishart, T. (1985) *On Sonic Art*. NewYork: Imagineering Press.

Wishart, T. (1986) Sound symbols and landscapes. En Emmerson, S. (ed) *The Language of Electroacustic Music*. London: Macmillan.

Wishart, T. (1992) *Red Bird / Anticredos*. [CD] Londres: October Music.

Wishart, T. (1994) *Tongues Of Fire Orpheus*. [CD] Londres: The Pantomime.

Wishart, T. (2002) *Journey Into Space*. [CD] Londres: Paradigm Discs.

Wissmann, T. (2014). Geographies of urban sound. Ashgate Publishing.

Wold, E., Blum, T., Keislar, D. & Wheaten, J. (1996). Clasificación, búsqueda y recuperación de audio basada en contenido. IEEE Multimedia, 3 (3), 27–36.

Wolvin, A. (ed.) (2010). *Listening and human communication in the 21st century*. Chichester, U.K.; Malden, MA: Wiley-Blackwell.

Xambó, A., Lerch, A., & Freeman, J. (2019) Music Information Retrieval in Live Coding: A Theoretical Framework. *Computer Music Journal*, 42 (4), 9-25.

Xenakis, I. (1992). *Formalized Music: Thought and Mathematics in Composition.* (S. E. Kanach, Ed.) Amsterdam University Press.

Xu, R., & Wunsch, D. (2009). *Clustering. IEEE Press series on computational intelligence*. Hoboken, N.J. Estados Unidos: IEEE Press.

Zacharakis, A. (2013). *Musical timbre: bridging perception with semantics*. England: Queen Mary University of London.

Zátonyi, M. (2002). Una estética del arte y el diseño de imagen y sonido. CP67.

Zils, A., & Pachet, F. (2001). Musical Mosaicing. Proceedings of the COST G-6 Conference on Digital Audio Effects. U.K., University of Limerick. 39-44.

Anexo

Ejemplos sonoros de las piezas

Anomalía Urbana, Mosaico Sonoro y Texturas Urbanas

Indicaciones de nomenclatura en los ejemplos

Cada ejemplo sonoro que se muestra a continuación, representa un fragmento tomado de una de las tres piezas compuestas para esta investigación. Con el fin de que el lector pueda saber a qué pieza corresponde, las tres piezas se enumeran como se muestra posteriormente. El número inicial de cada ejemplo indica la pieza de la cual se tomó. En dichas piezas se utilizó la siguiente numeración:

1. *Anomalía Urbana*
2. *Mosaico Sonoro*
3. *Texturas Urbanas*

La siguiente lista corresponde a los ejemplos sonoros mencionados:

Rugoso

1.1. Rugoso y brillante con comportamiento aleatorio (0'19" a 0'26")

https://ottocastro.com/o/168

2.1. Rugoso-oscuro con sonidos de motores de autos (3'09" a 3'20")

https://ottocastro.com/o/085/

2.2. Rugoso-brillante, densidad y acumulación tímbricamente (3'53" a 4'07")

https://ottocastro.com/o/170

2.3. Rugoso-brillante con pulsación rítmica fuerte (8'55" a 9'03")

https://ottocastro.com/o/067/

Brillo

3.1. Brillo y fusión tímbrica creciente (3’51” a 4’06”)

https://ottocastro.com/o/051/

1.2. Brillo en *Anomalía Urbana* (1'48" a 1'53")

https://ottocastro.com/o/030/

2.4. Brillo en Mosaico Urbano (5'55" a 6'02")

https://ottocastro.com/o/021/

3.2. Brillo en Texturas Urbanas (5'51" a 6'03")

https://ottocastro.com/o/171

Oscuro

2.5. Oscuro tímbrico con modulación de amplitud. Sonidos de motores de autos (0’00” a 0’30”)

https://ottocastro.com/o/094/

3.3. Oscuro tímbrico con pedal armónico en crescendo (0’00” a 0’30”)

https://ottocastro.com/o/064/

2.6. Oscuro en *Anomalía Urbana* (1'37" a 1'45")

https://ottocastro.com/o/027/

3.4. Oscuro en Mosaico Urbano (3'30" a 3'40")

https://ottocastro.com/o/025/

Liso

2.7. Lisa con comportamiento continuo (6’12 a 6’24”)

https://ottocastro.com/o/035/

3.5. Liso armónico con comportamiento continuo (1’00” a 1’21”)

https://ottocastro.com/o/034/

2.8. Liso en *Anomalía Urbana* (4'29" a 4'35")

https://ottocastro.com/o/028/

Diferentes perspectivas sonoras de Barry Truax

1.3. Perspectiva sonora fija y posición suelo. Sonido de llantas con superficie ondulada (1’38” a 2’09”)

https://ottocastro.com/o/139/

1.4. Perspectiva sonora móvil de autos con procesamiento (4’15” a 4’39”)

https://ottocastro.com/o/100/

2.9. Perspectiva sonora móvil continua, armónica (6’26” a 6’34”)

https://ottocastro.com/o/075/

1.5. Perspectiva sonora móvil y proceso de acumulación de brillo tímbrico (4’55” a 5’02”)

https://ottocastro.com/o/173

1.6. Perspectiva sonora de movimiento con sonido de generador eléctrico

https://ottocastro.com/o/003

1.7. Perspectiva sonora fija con sonido de generador eléctrico

https://ottocastro.com/o/139

Perspectivas sonoras suelo

1.8. Perspectiva sonora de suelo con efecto de proximidad (2’59” a 3’28”)

https://ottocastro.com/o/103/

2.10. Perspectiva sonora plano suelo (4’55” a 5’08”)

https://ottocastro.com/o/078/

Atenuación

3.6. Atenuación por categoría: lisa a rugosa (6’07” a 6’30”)

https://ottocastro.com/o/032/

Sonidos armónicos

3.7. Sonidos armónicos de tonos cercanos en el metro (2’40” a 2’44”)

https://ottocastro.com/o/057/

3.8. Pedal armónico (5’43” a 6’10”)

https://ottocastro.com/o/046/

3.9. Sonidos armónicos de registro medios similares de cláxones estirados (6’11” a 6’35”)

https://ottocastro.com/o/045/

2.11. Secuencia de sonidos armónicos de metro (1’08” a 1’25”)

https://ottocastro.com/o/057/

Densidad tímbrica

1.9. Densidad tímbrica con cambio abrupto misma categoría tímbrica (3’21” a 3’35”)

https://ottocastro.com/o/102/

1.10. Densidad tímbrica con múltiples planos rugosos y brillantes (4’40” a 4’56”)

https://ottocastro.com/o/099/

1.11. Densidad tímbrica por saturación. Presencia de tres categorías (4’55” a 5’02”)

https://ottocastro.com/o/098/

2.12. Densidad tímbrica por saturación 2 (2’09” a 2’22”)

https://ottocastro.com/o/089/

3.10. Densidad inarmónica. Alta saturación. Claxon aparece en 3’50” (3’44” a 3’58”)

https://ottocastro.com/o/052/

1.12. Densidad tímbrico comprimido (0’47” a 1’05”)

https://ottocastro.com/o/110/

Enmascaramiento

3.11. Enmascaramiento de categoría tímbrica brillante sobre una categoría: rugosa-opaca (3’31 a 3’42”)

https://ottocastro.com/o/081/

3.12. Enmascaramiento de sonidos de cláxones (fa5 mi4) (1’20” a 1’30”)

https://ottocastro.com/o/019/

3.13. Enmascaramiento de capa lisa sobre capa opaca (7’10” a 7’33”)

https://ottocastro.com/o/044/

Planos tímbricos

3.15. Plano tímbrico de dos capas: oscuro y brillante. Sonidos estirados (0’44” a 1’10”)

https://ottocastro.com/o/063/

3.16. Plano tímbrico por acumulación de la densidad (1’58” a 2’10”)

https://ottocastro.com/o/060/

3.17. Plano tímbrico con acumulación de la misma categoría tímbrica (7’34” a 8’18”)

https://ottocastro.com/o/047/

3.18. Plano tímbrico de múltiples capas con sonidos armónicos 1 (2’10” a 2’19”)

https://ottocastro.com/o/059/

2.13. Plano tímbrico de múltiples capas con sonidos armónicos 2 (5’40” a 6’05”)

https://ottocastro.com/o/076/

3.19. Plano tímbrico con capas yuxtapuestas: oscuro-rugoso (2’44” a 2’53”)

https://ottocastro.com/o/055/

3.20. Plano tímbrico con cuatro capas con comportamiento azaroso (3’04” a 3’19”)

https://ottocastro.com/o/054/

3.21. Plano tímbrico con dos a tres capas alternadas (4’07” a 4’36”)

https://ottocastro.com/o/050/

3.22. Plano tímbrico dos capas brillante-opaco (4’38” a 4’47”)

https://ottocastro.com/o/049/

3.23. Plano tímbrico con capas similares y continuidad de la misma frecuencia (5’00” a 5’22”)

https://ottocastro.com/o/048/

2.14. Plano tímbrico con tres capas tímbricas alternadas: oscuro, rugoso, brillante (0’41” a 1’12”)

https://ottocastro.com/o/093/

2.15. Plano tímbrico con secuencia de capas armónicas. Sonidos del metro (1’08” a 1’25”)

https://ottocastro.com/o/092/

2.16. Plano tímbrico con varias capas tímbricas, sobresaliendo la rugosa (4’50” a 5’08”)

https://ottocastro.com/o/079/

2.17. Plano tímbrico con tres capas brillante, rugosa, opaca (8’37” a 8’52”)

https://ottocastro.com/o/068/

Contraste tímbrico de transición

2.18. Contraste tímbrico de transición entre categoría: brillante a opaca (1’28” a 1’42”)

https://ottocastro.com/o/091/

1.13. Contraste tímbrico de transición entre categoría rugosa de brillante a opaca (1’52” a 2’01”)

https://ottocastro.com/o/106/

2.19. Contraste tímbrico abrupto a través de percutido sonido de autobús (1’46” a 2’01”)

https://ottocastro.com/o/090/

2.20. Contraste tímbrico de transición por medio de enmascaramiento. Categoría: rugosa a brillante (2’27” a 2’43”)

https://ottocastro.com/o/087/

2.21. Contraste tímbrico de transición. Categoría rugosa (3’24” a 3’32”)

https://ottocastro.com/o/084/

2.22. Contraste tímbrico por acumulación. Categoría: rugoso-brillante a opaco-rugosa (2’56” a 3’07”)

https://ottocastro.com/o/086/

2.23. Contraste tímbrico de transición con sonidos rugosos opaco a brillantes (1’52” a 2’00”)

https://ottocastro.com/o/033/

Contraste tímbrico abrupto

3.24. Contraste tímbrico abrupto de brillante a opaco (8’18” a 9’17”)

https://ottocastro.com/o/043

1.14. Contraste tímbrico abrupto opaco a brillante (6’58” a 7’01”)

https://ottocastro.com/o/097/

1.15. Contraste tímbrico abrupto densidad alta a poca densidad (7’50” a 8’00”)

https://ottocastro.com/o/036/

3.25. Contraste tímbrico abrupto entre categoría: rugoso-inarmónico versus liso-armónico (2’49” a 2’59”)

https://ottocastro.com/o/018/

1.16. Contraste tímbrico abrupto de categoría: rugosa-brillante a rugosa-lisa (1’52” a 2’00”)

https://ottocastro.com/o/122

1.17. Contraste tímbrico abrupto de categoría: rugosa-oscuro a rugosa-brillante (6’58” a 7’03”)

https://ottocastro.com/o/123

1.18. Contraste tímbrico abrupto de poca densidad a alta densidad tímbrica y cambios de intensidad (3’33” a 3’45”)

https://ottocastro.com/o/101/

1.19. Contraste tímbrico abrupto entre bloques sonoros (7’04” a 7’14”)

https://ottocastro.com/o/1s18

2.24. Contraste tímbrico abrupto de categoría: rugosa-brillante con comportamiento aleatorio (9’49” a 9’57”)

https://ottocastro.com/o/065/

2.25. Contraste tímbrico abrupto de categoría: brillante-rugosa a opaca-rugosa (2’15” a 2’32”)

https://ottocastro.com/o/088/

Otros recursos musicales para reforzamiento de dirección musical

2.26. Tono electrónico pedal con caída de frecuencia (4’20” a 4’40”)

https://ottocastro.com/o/080/

2.27. Sonido electrónico de línea de bajo. Cadencia (8’25” a 8’31”)

https://ottocastro.com/o/069/

1.20. Sonido electrónico con fragmentos sonoros de motor. Efecto de proximidad (2’29” a 2’43”)

https://ottocastro.com/o/104/

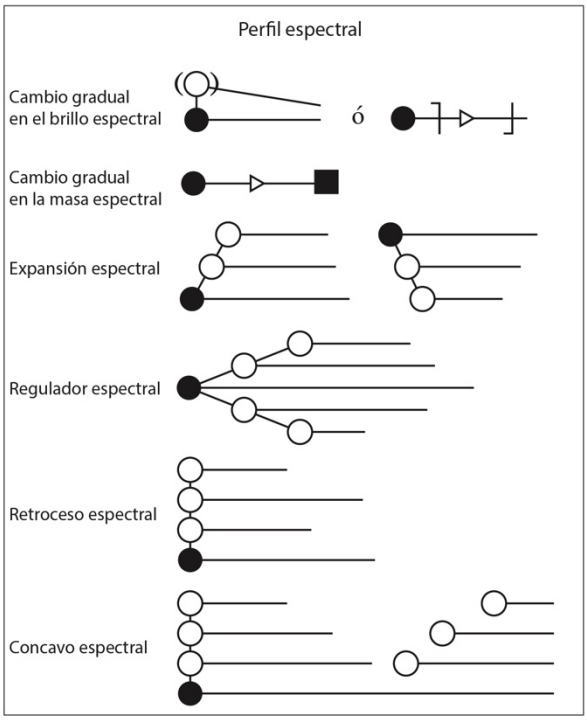
Representación visual en formato de notación

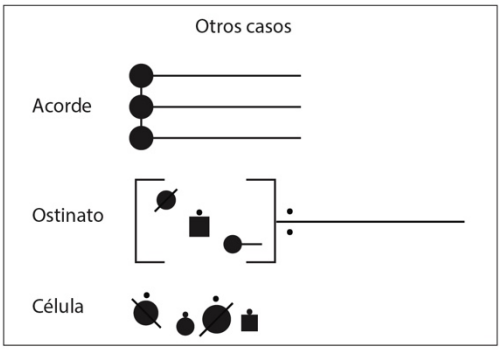
de *Lasse Thoresen* de las piezas musicales: *Mosaico Sonoro, Anomalía*

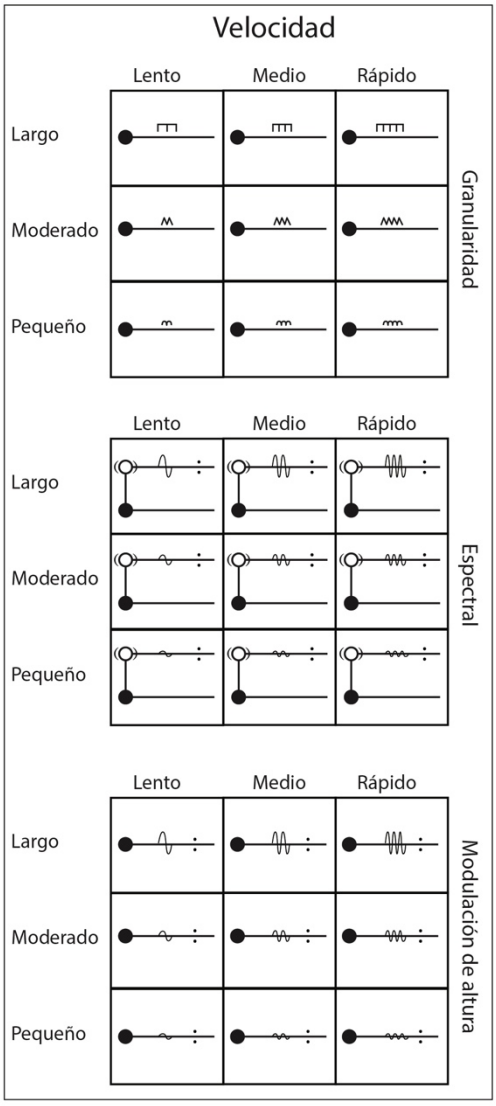
Urbana y Texturas Urbanas

Glosario sonoro de la notación de *Thoresen*

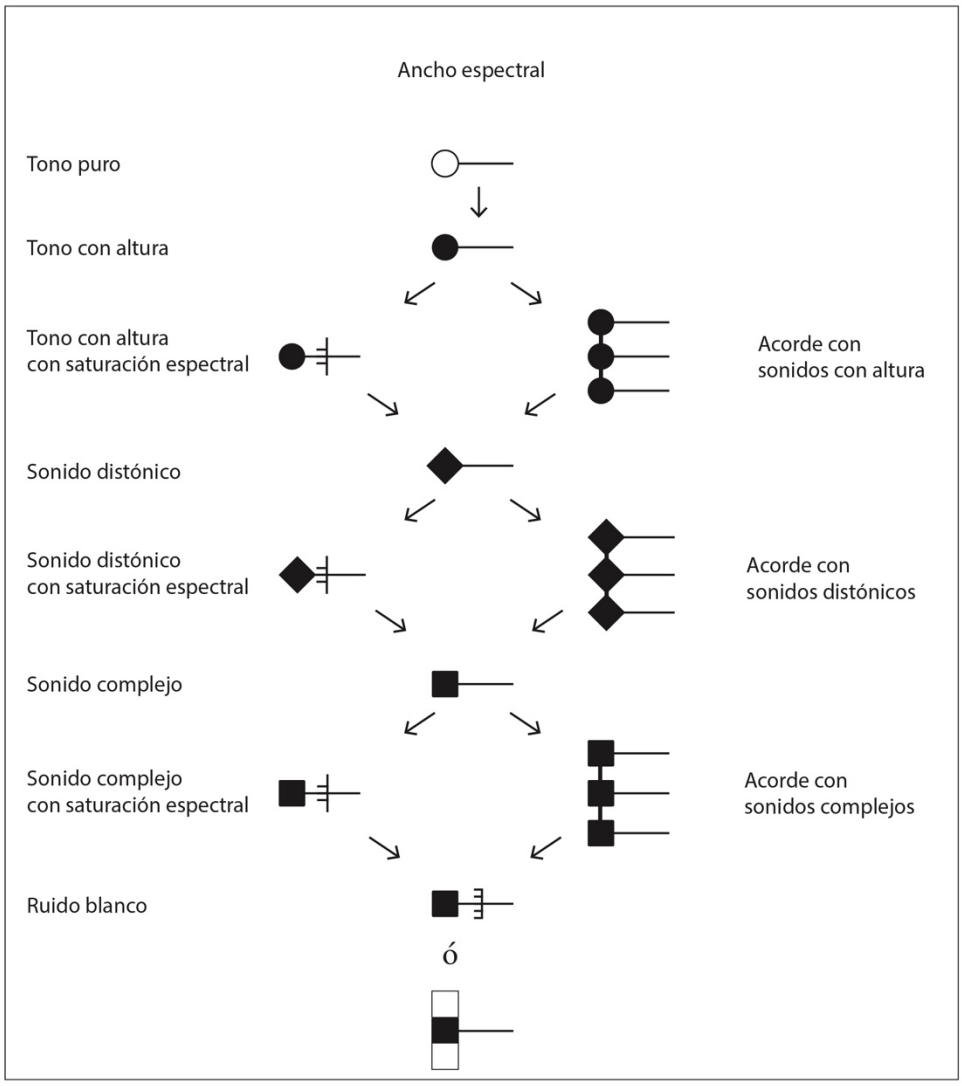
Glosario gráfico*

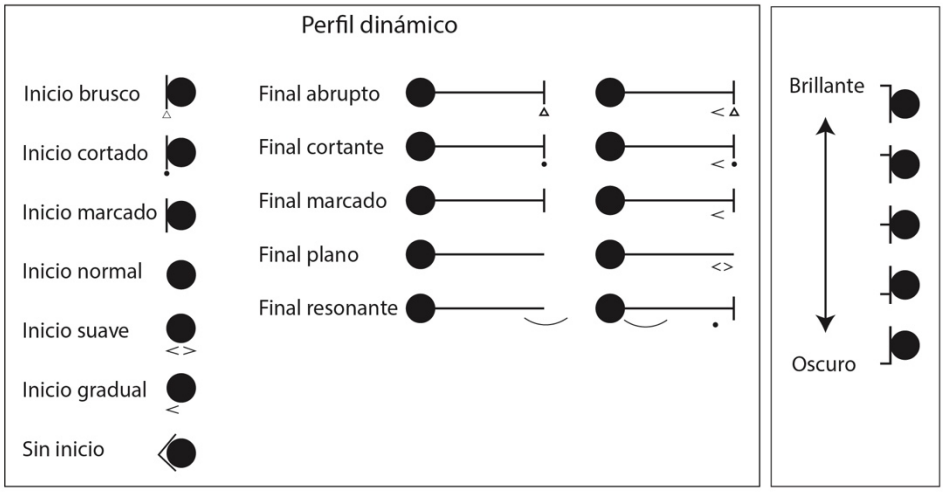




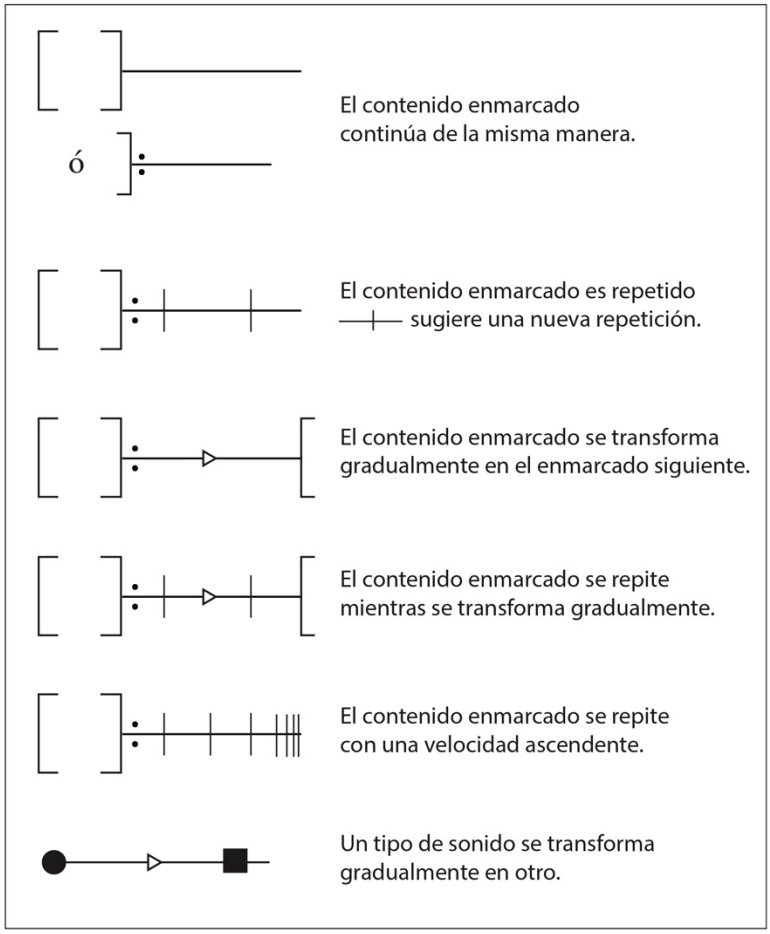


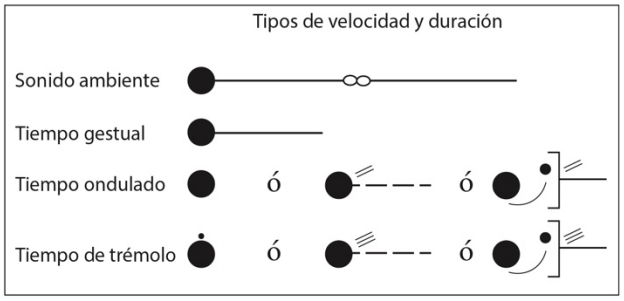
* Signos de Lasse Thoresen

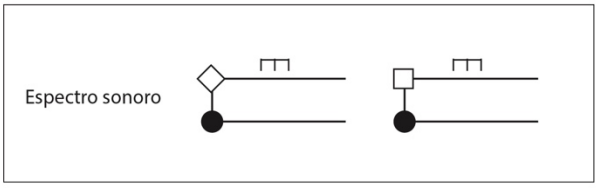




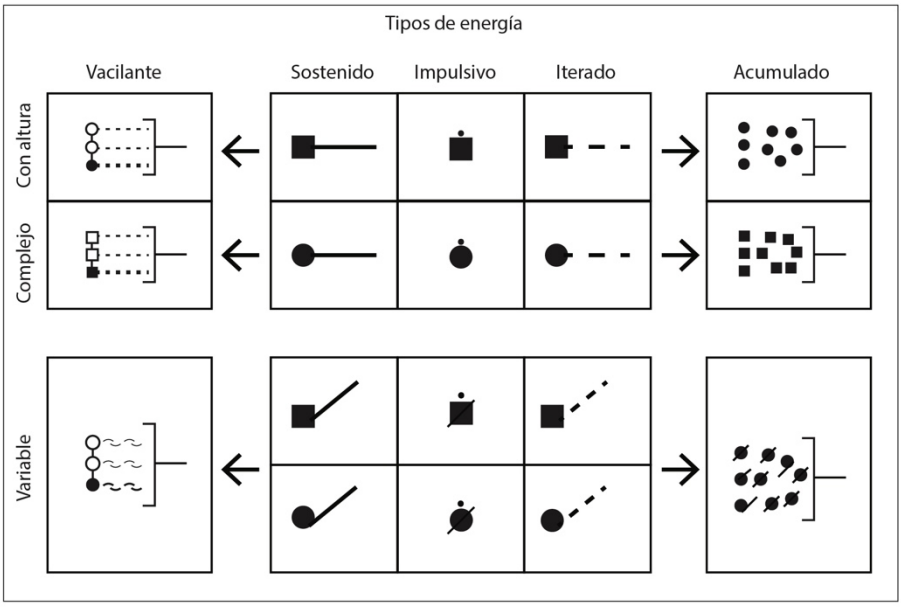
* Signos de Lasse Thoresen

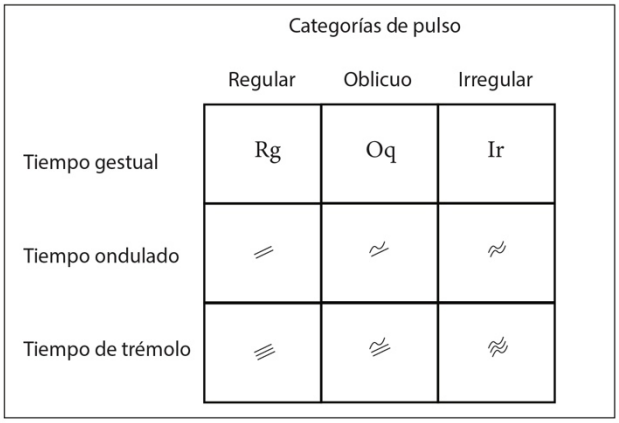






* Signos de Lasse Thoresen





* Signos de Lasse Thoresen

Otros signos del glosario gráfico **

Iteración

Irregular

d_1 d_2 d_3 d_4 d_5 Niveles de densidad


Crescendo desde silencio absoluto


Decrescendo a silencio absoluto


Silencio muy corto


Silencio corto


Silencio medio


Silencio largo


Comportamiento irregular

p sub: Piano súbito

mp sub: Mezzopiano súbito

** Signos nuevos
incorporados por Otto Castro

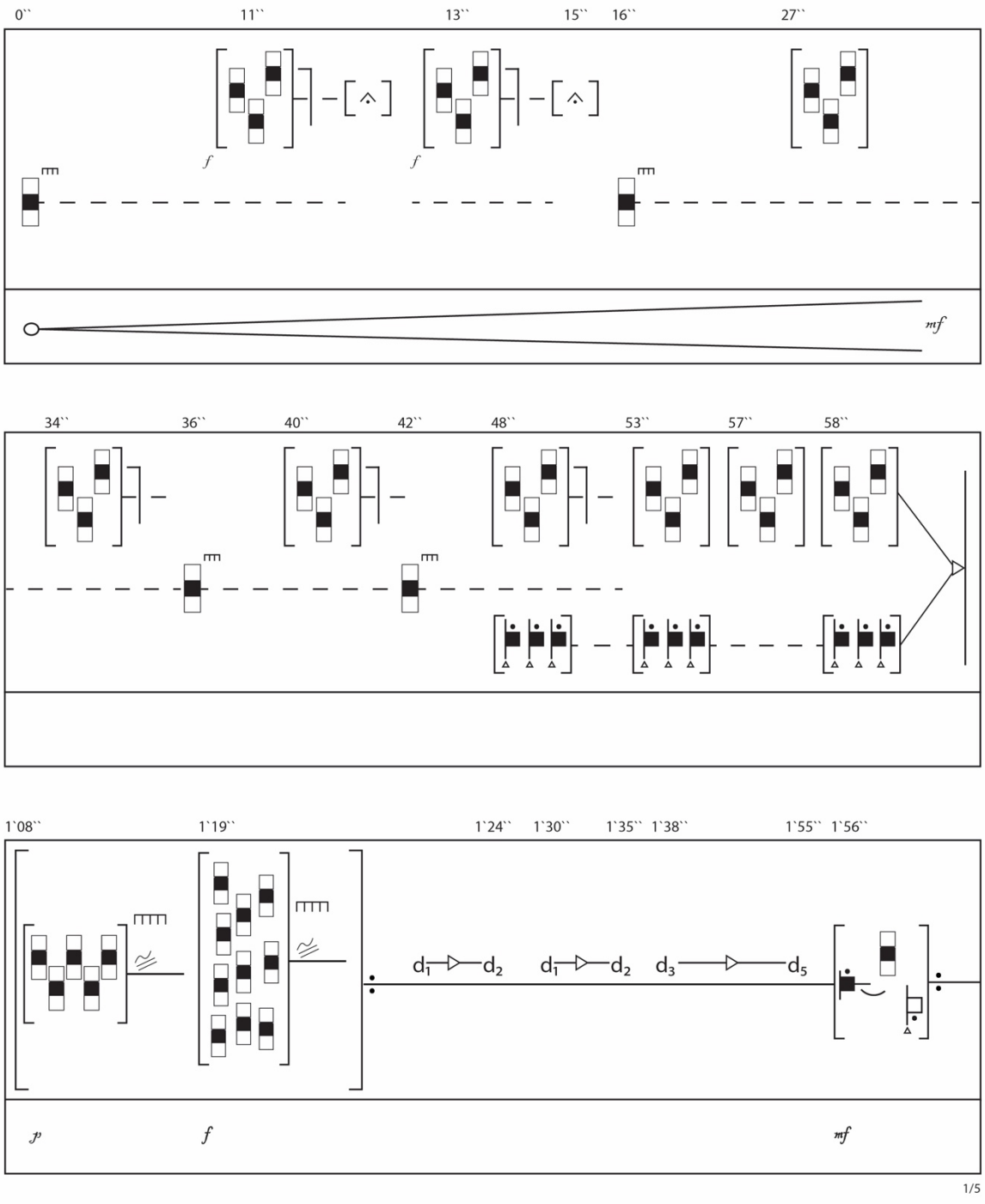
Mosaico Sonoro

Noviembre, 2019
Coyoacán, México

MOSAICO URBANO

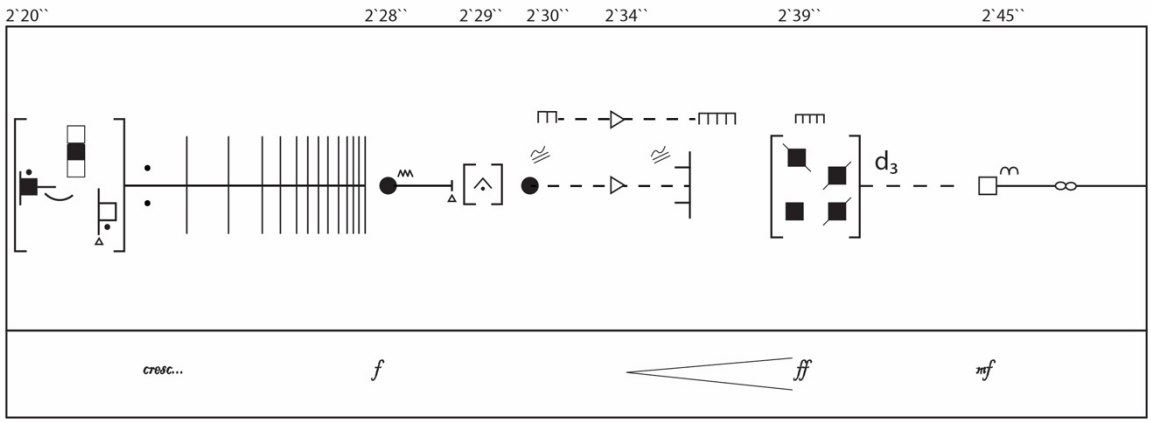
Pieza musical desarrollada a través de la investigación Doctoral:
"Estrategias de composición a través de la similitud tímbrica
mediante la sinergia entre compositor y máquina"

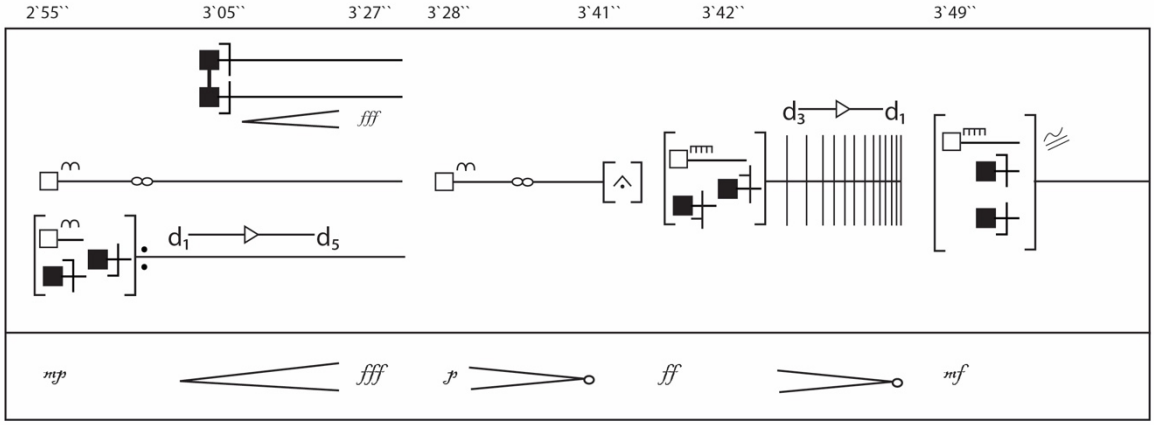
Compositor: Otto Castro Solano
Transcripción: Alejandro Cabezas Morera

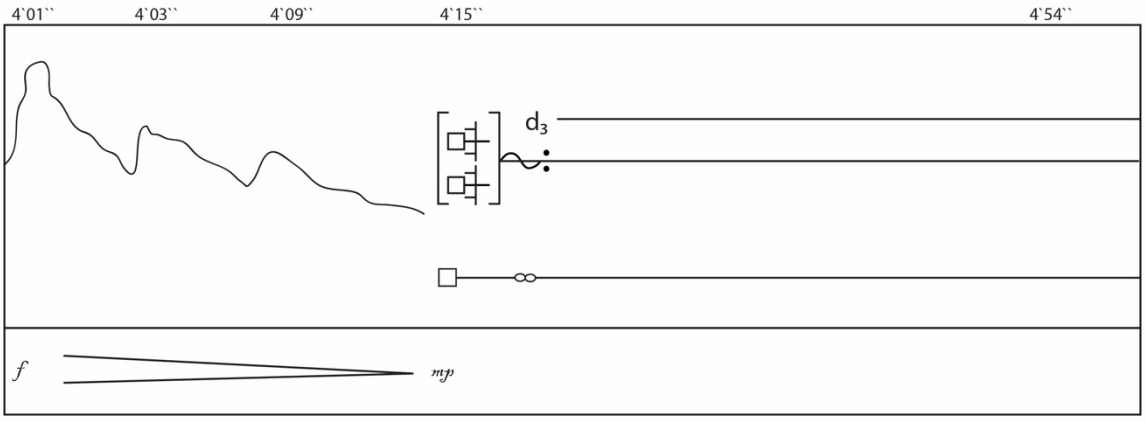


1/5

Mosaico sonoro, 2/5

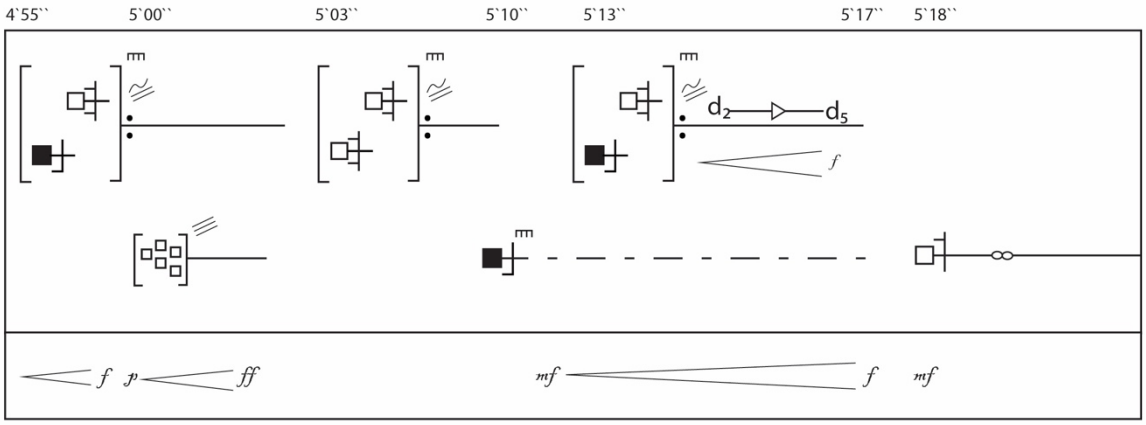


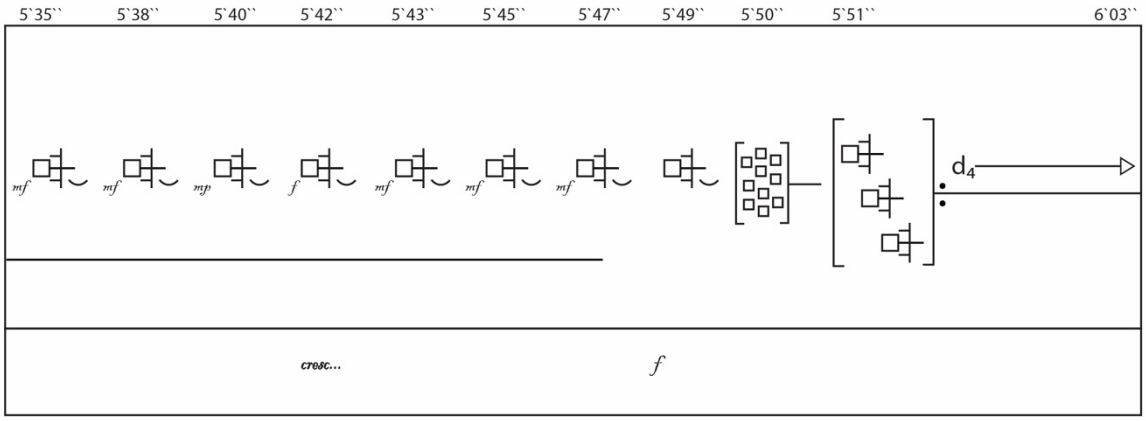


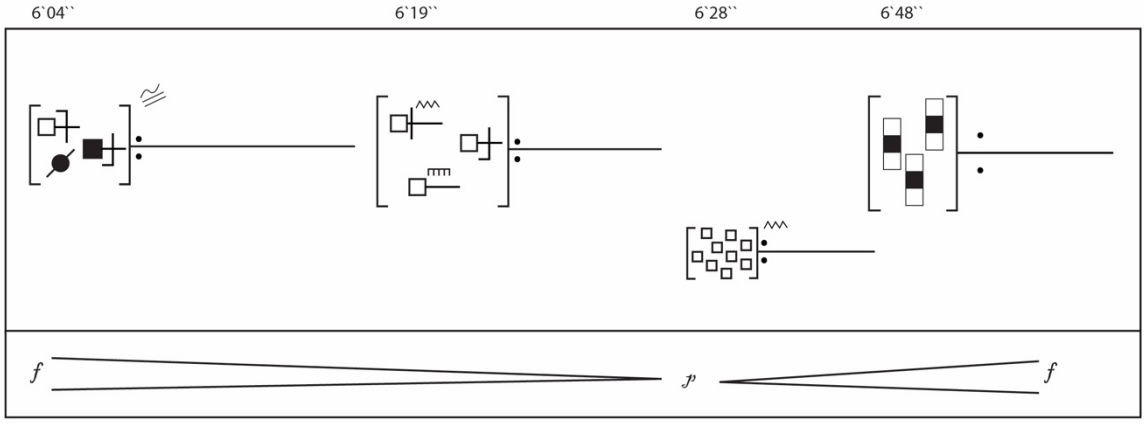


2/5

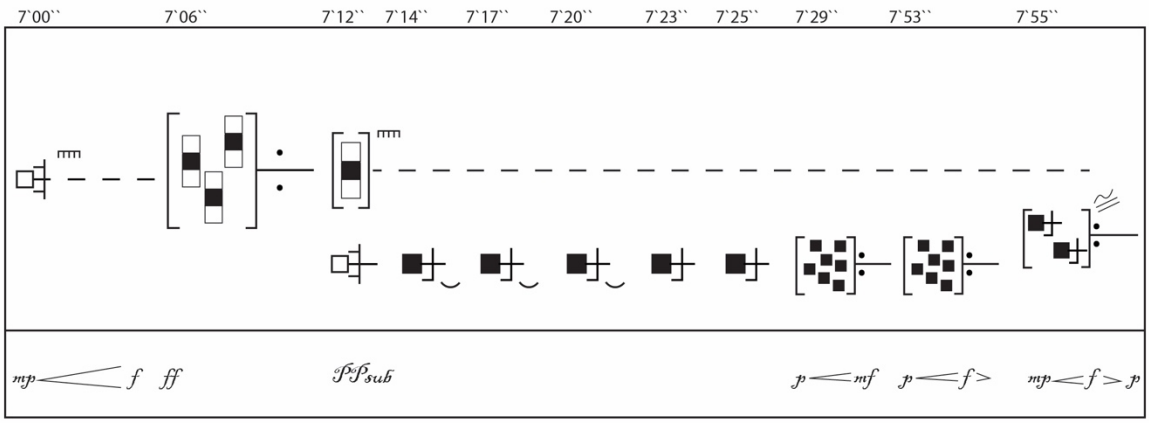
Mosaico sonoro, 3/5

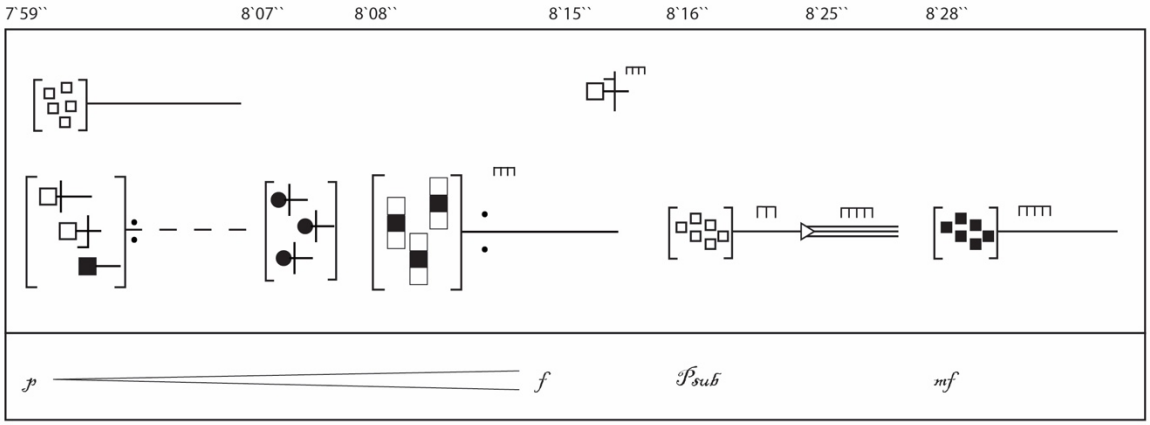


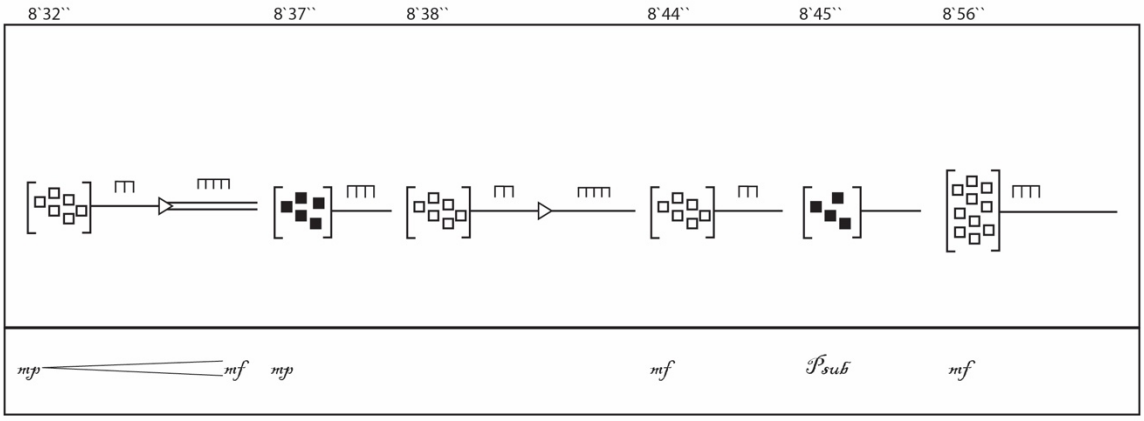




Mosaico sonoro, 4/5

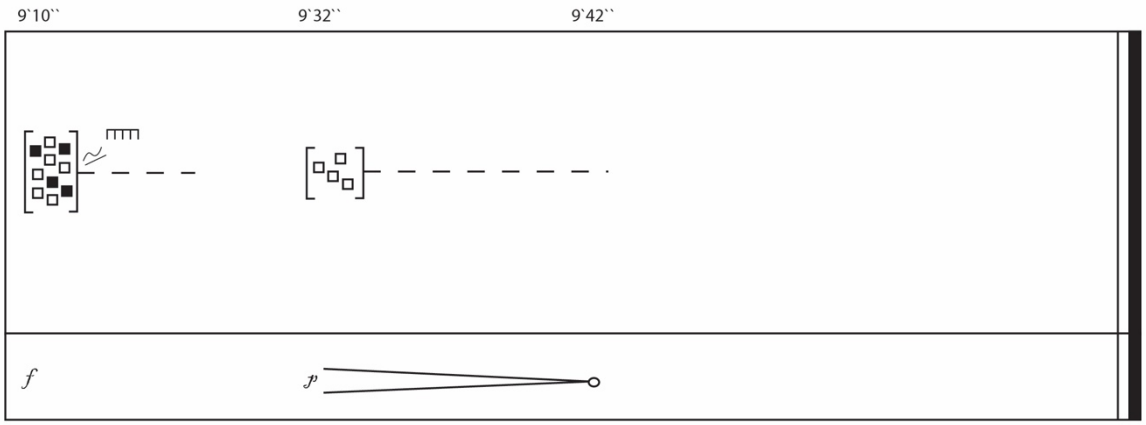






4/5

Mosaico sonoro, 5/5



5/5

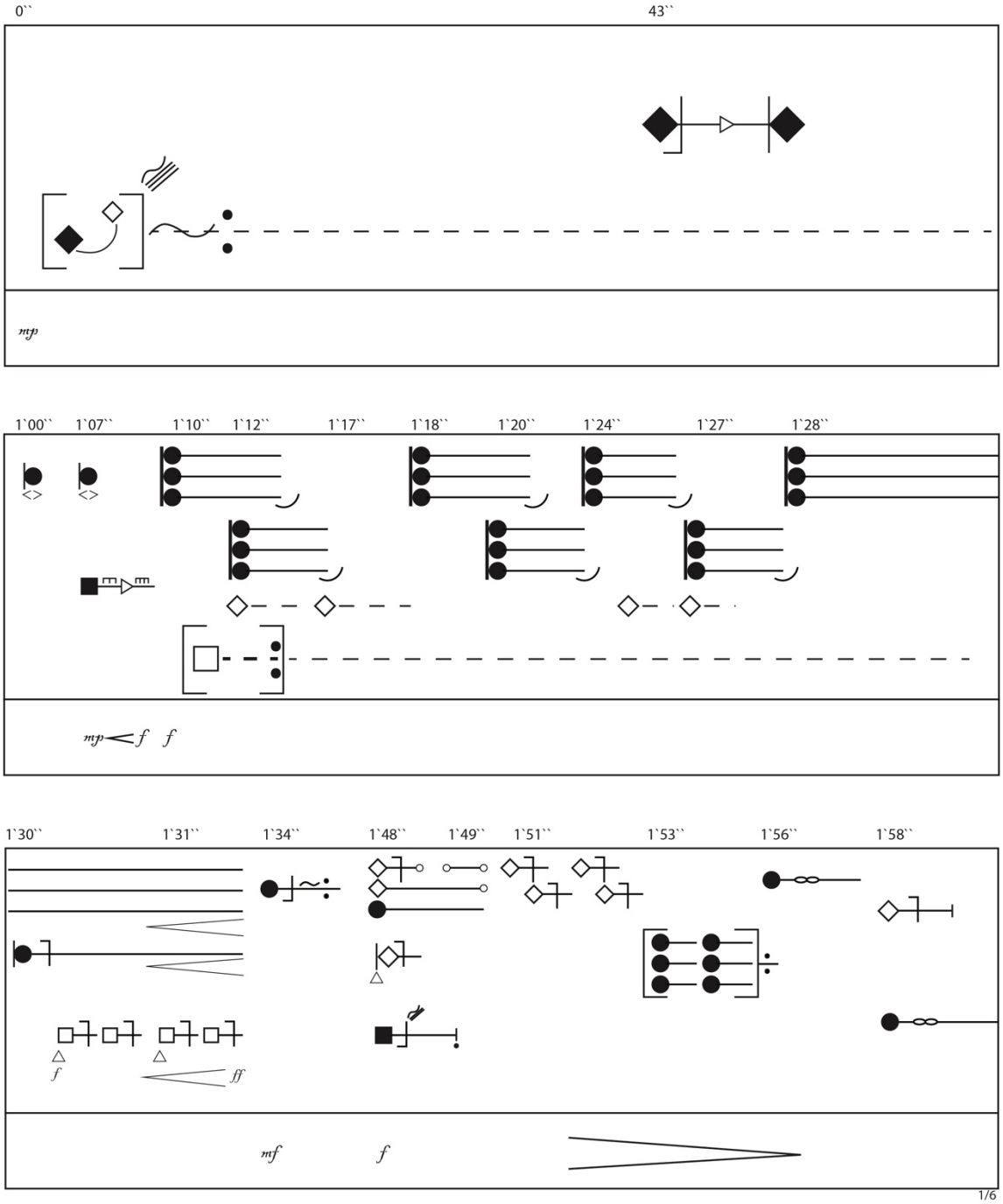
Anomalía Urbana

Septiembre, 2020
Coyoacán, México

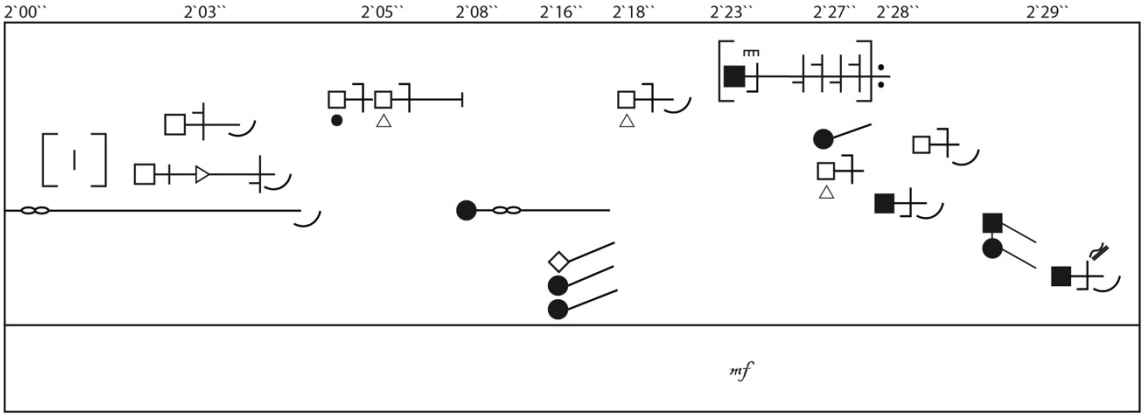
ANOMALÍA URBANA

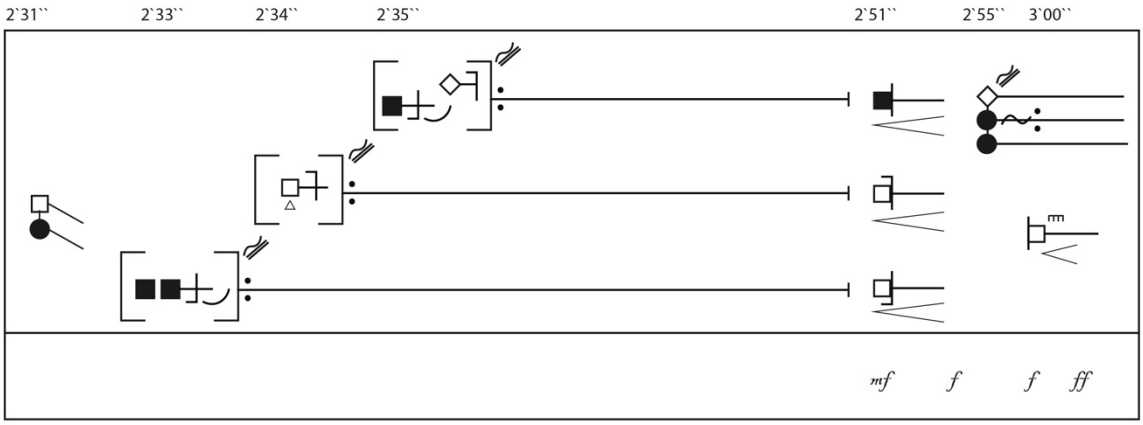
Pieza musical desarrollada a través de la investigación Doctoral:
"Estrategias de composición a través de la similitud tímbrica
mediante la sinergia entre compositor y máquina"

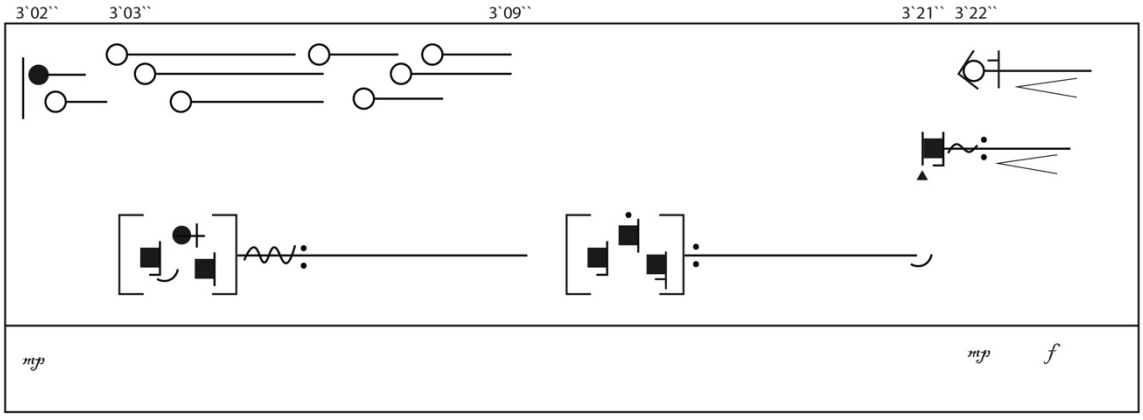
Compositor: Otto Castro Solano
Transcripción: Alejandro Cabezas Morera



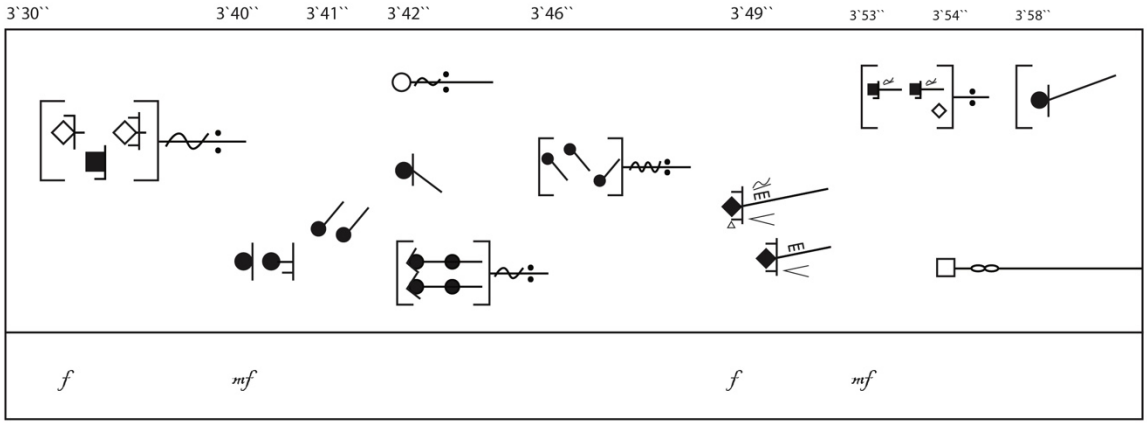
Anomalía urbana, 2/6

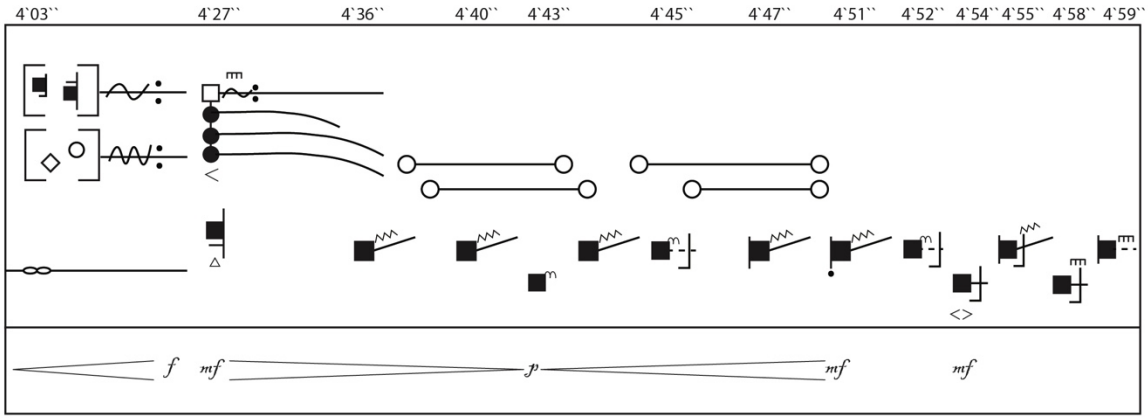


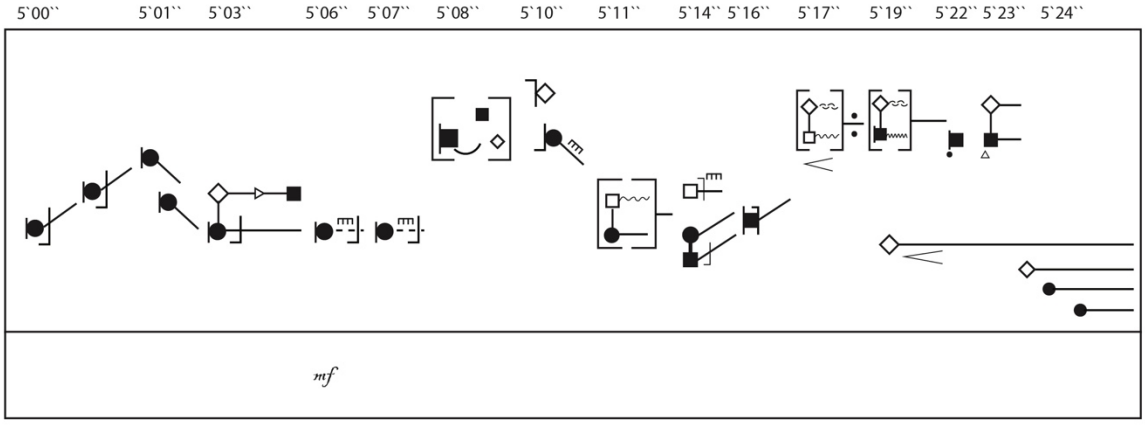




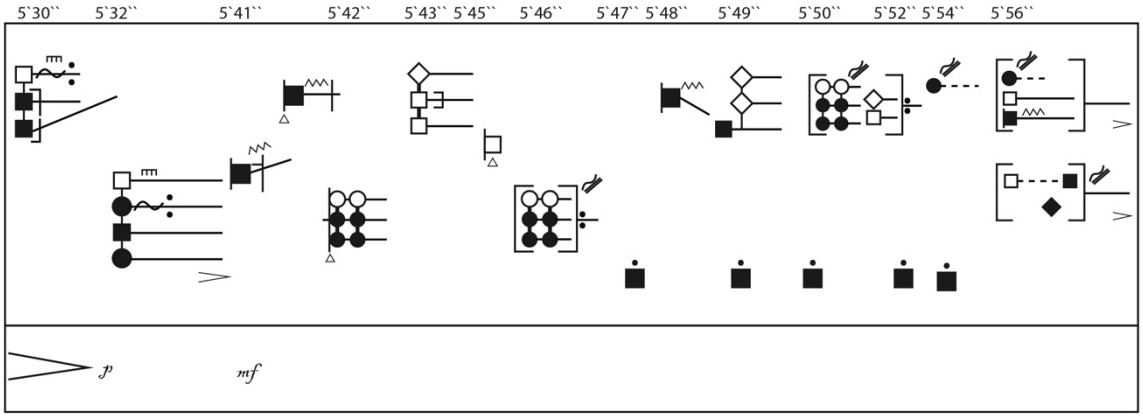
2/6

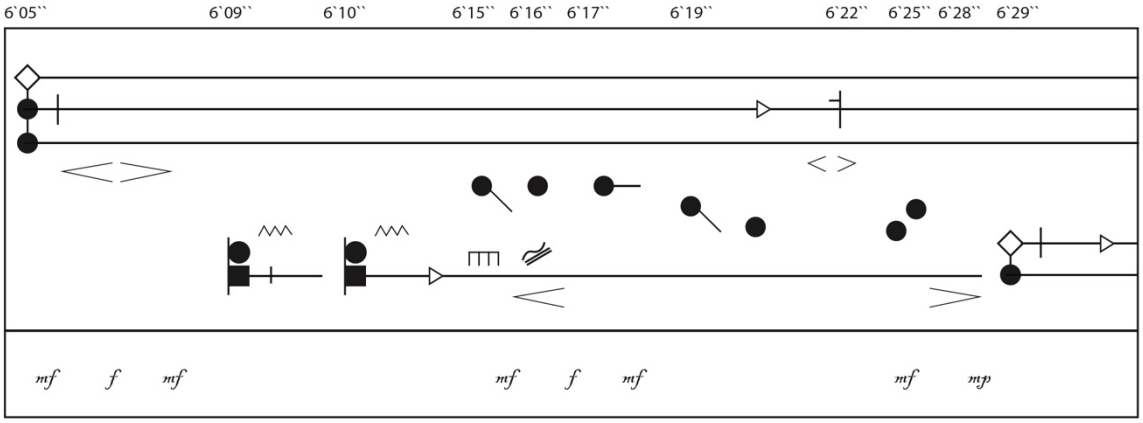

3`30``
3`40``
3`41``
3`42``
3`46``
3`49``
3`53``
3`54``
3`58``
f
mf
f
mf

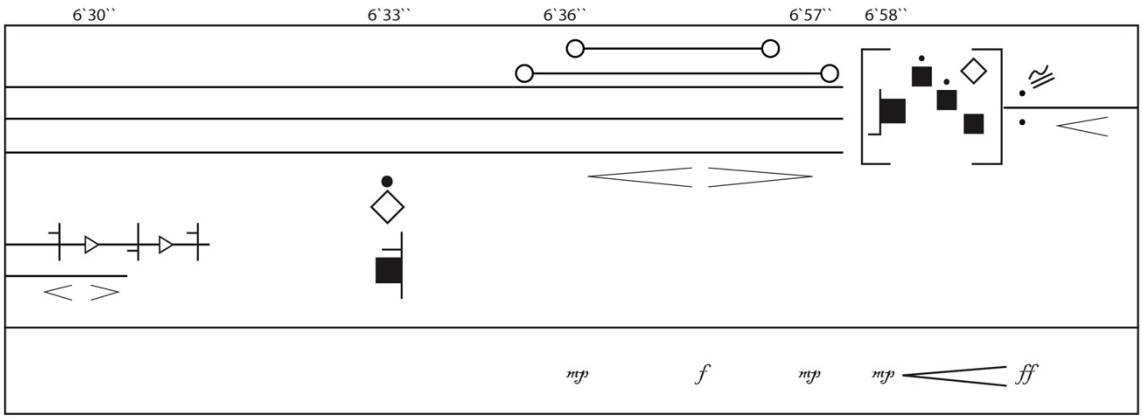

4`03``
4`27``
4`36``
4`40``
4`43``
4`45``
4`47``
4`51``
4`52``
4`54``
4`55``
4`58``
4`59``
f
mf
p
mf
mf


5`00``
5`01``
5`03``
5`06``
5`07``
5`08``
5`10``
5`11``
5`14``
5`16``
5`17``
5`19``
5`22``
5`23``
5`24``
mf

Anomalía urbana, 4/6

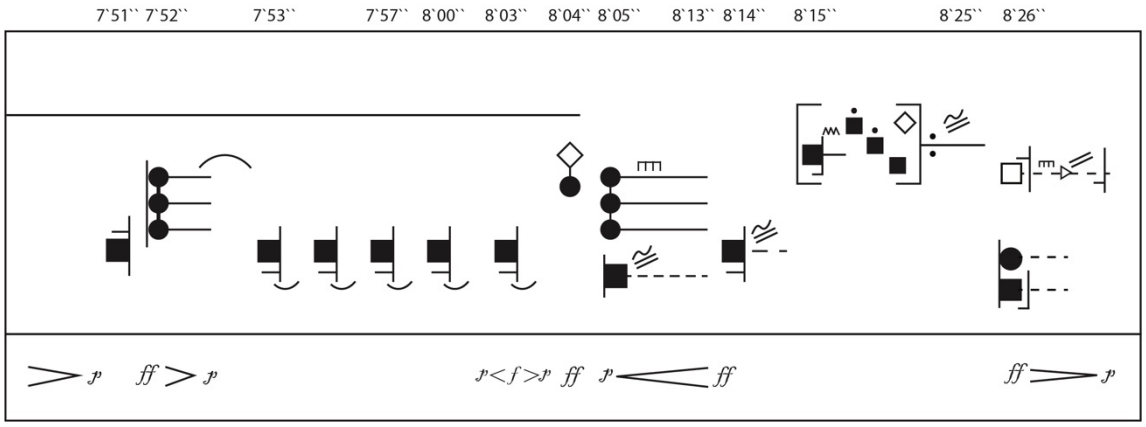


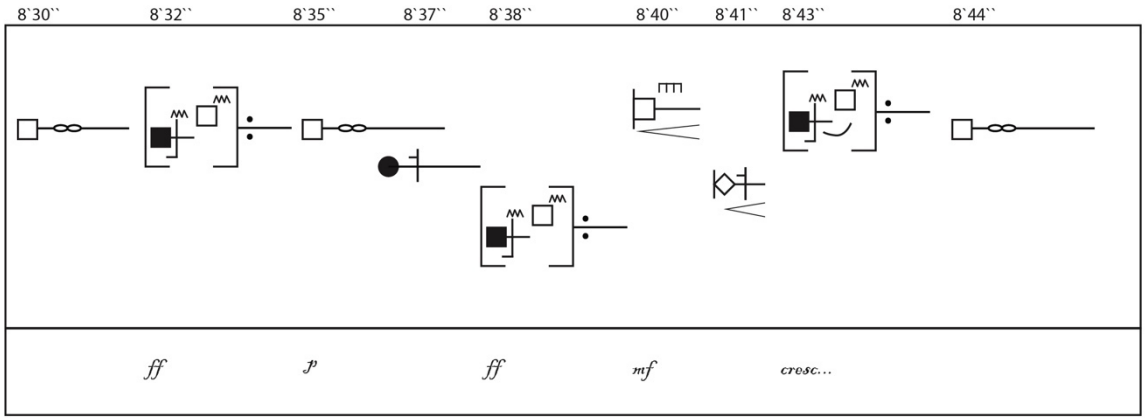


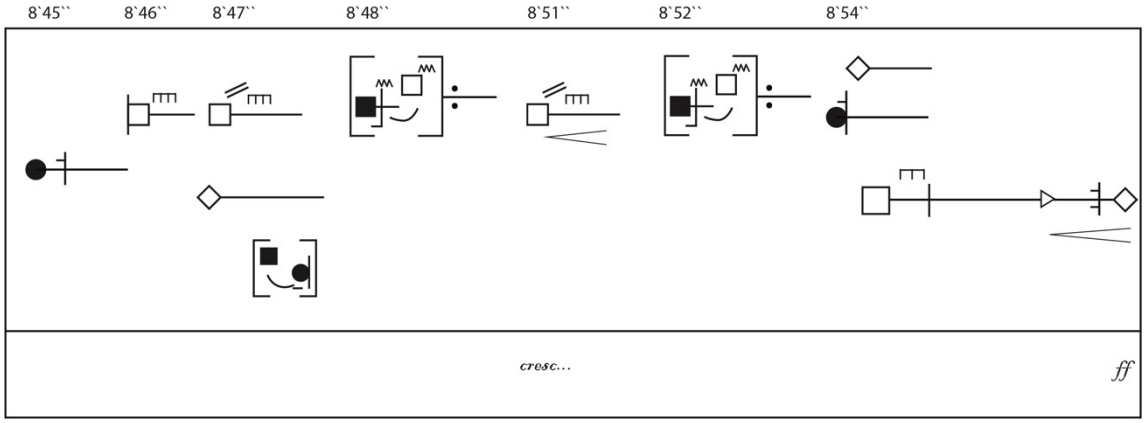


4/6

Anomalía urbana, 5/6

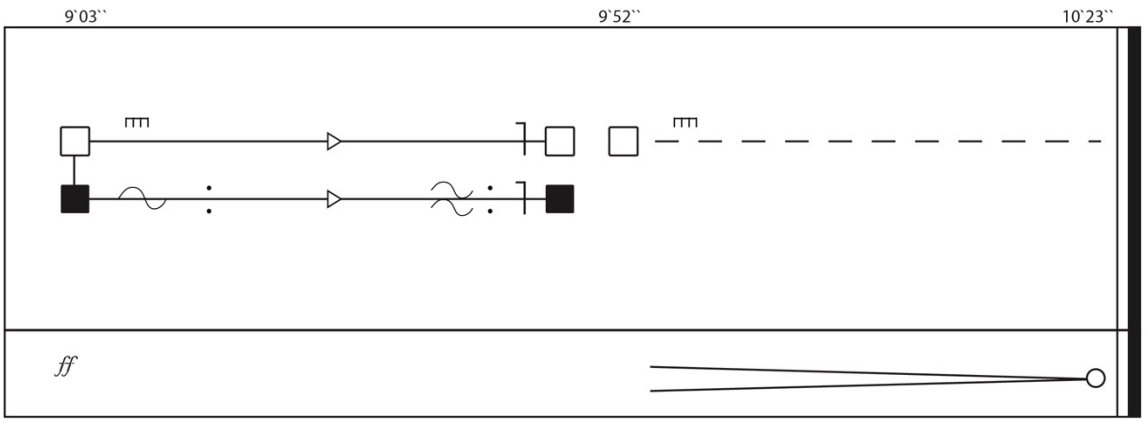






5/6

Anomalía urbana, 6/6



6/6

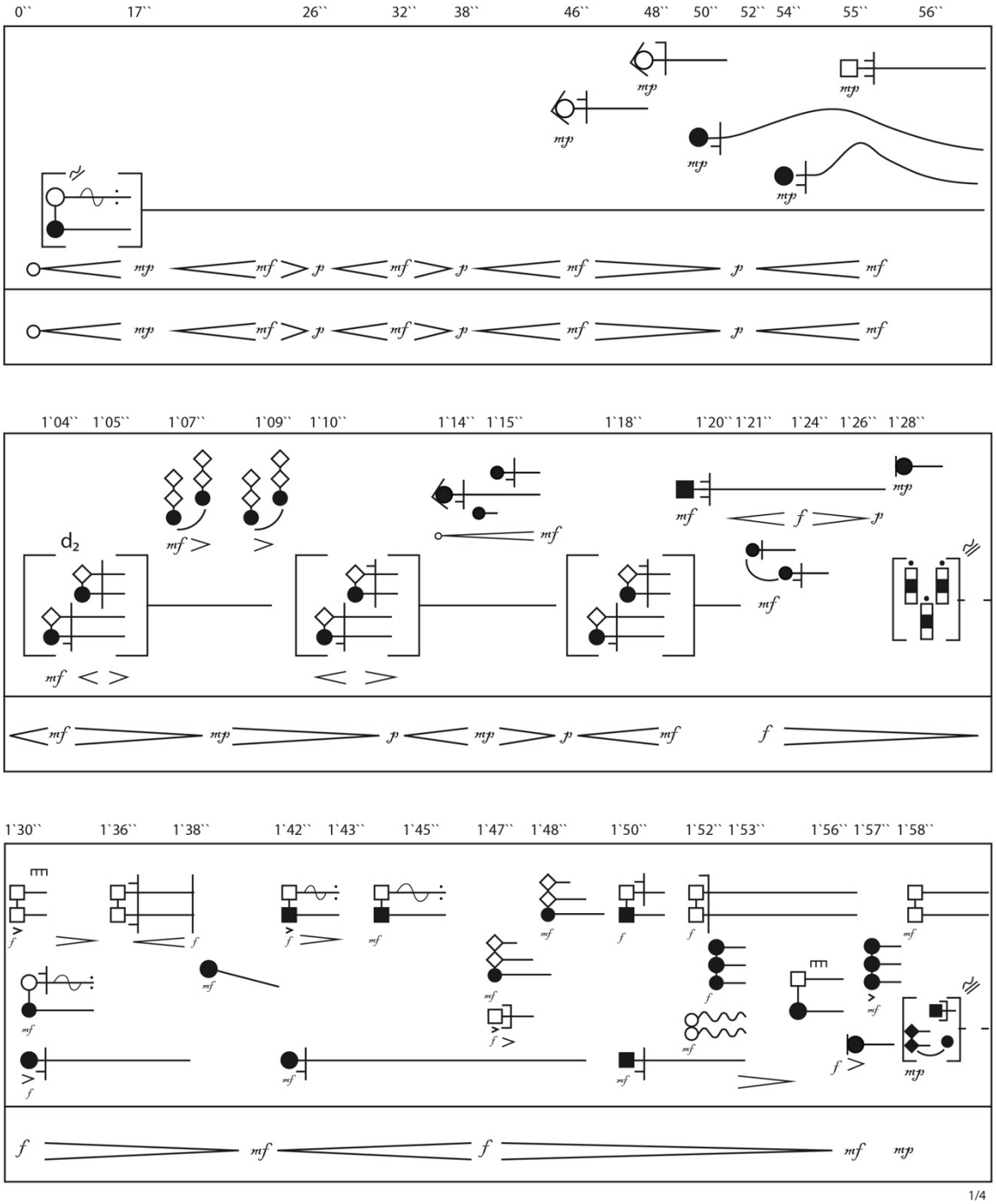
Texturas Urbanas

Mayo, 2021
Coyoacán, México

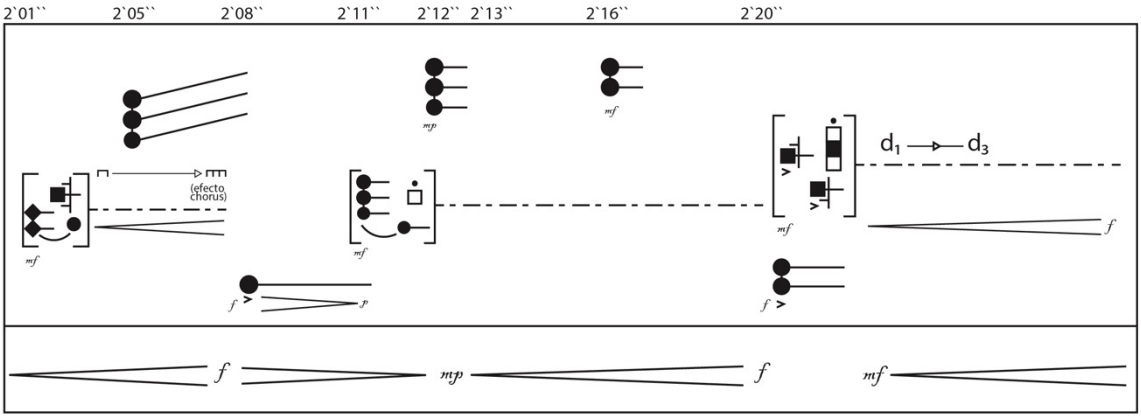
TEXTURAS URBANAS

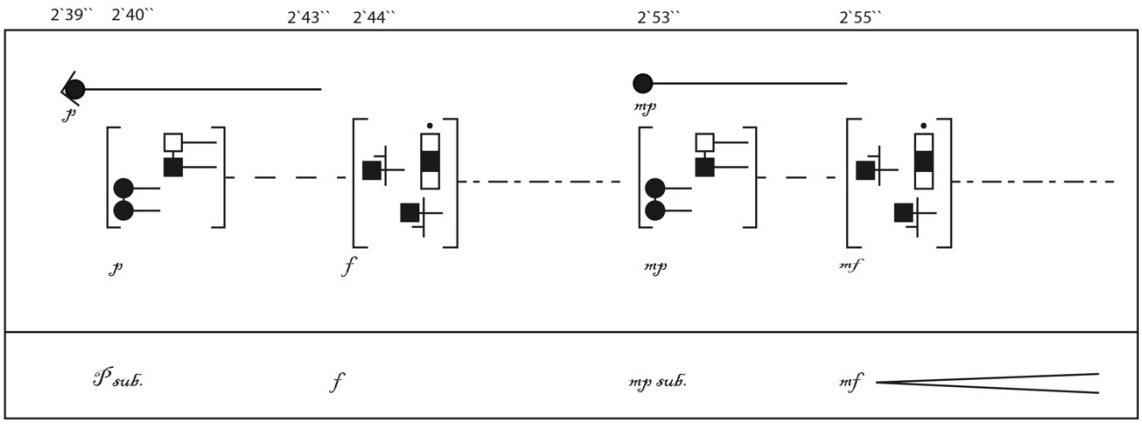
Pieza musical desarrollada a través de la investigación Doctoral:
"Estrategias de composición a través de la similitud tímbrica
mediante la sinergia entre compositor y máquina"

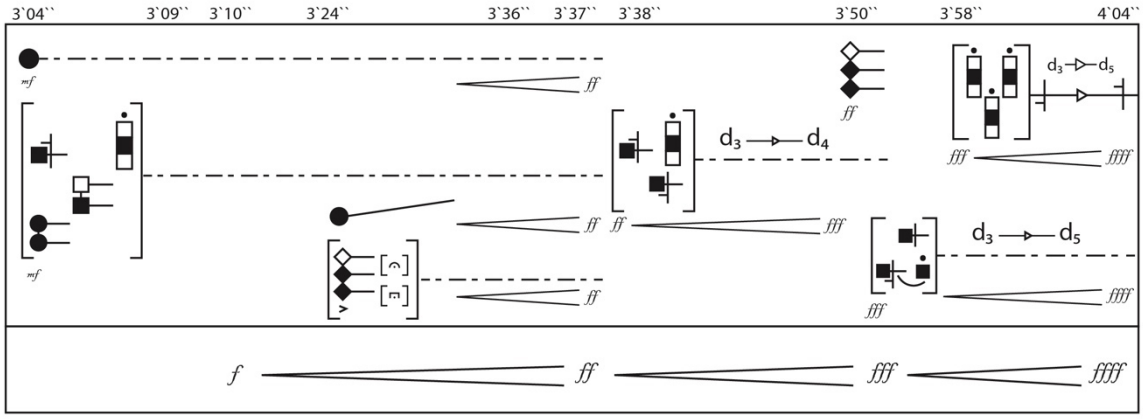
Compositor: Otto Castro Solano
Transcripción: Alejandro Cabezas Morera



Texturas Urbanas, 2/4

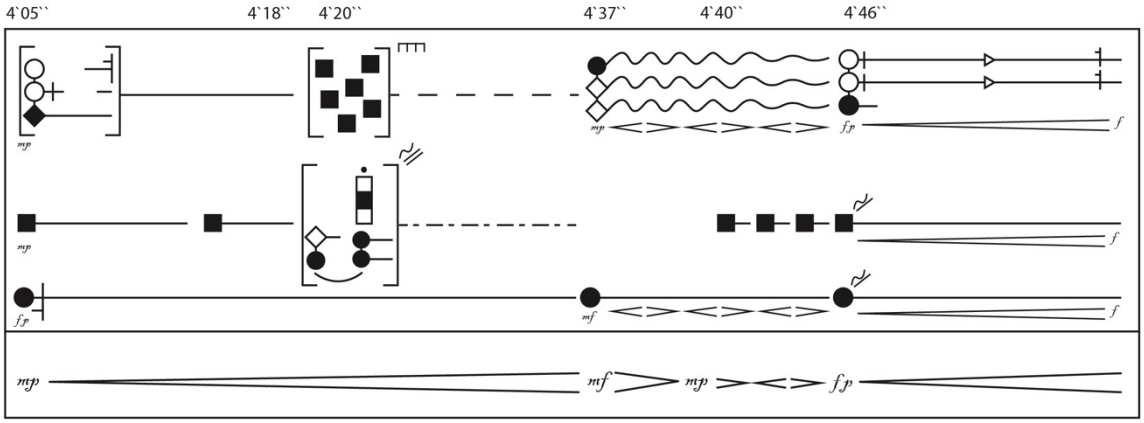


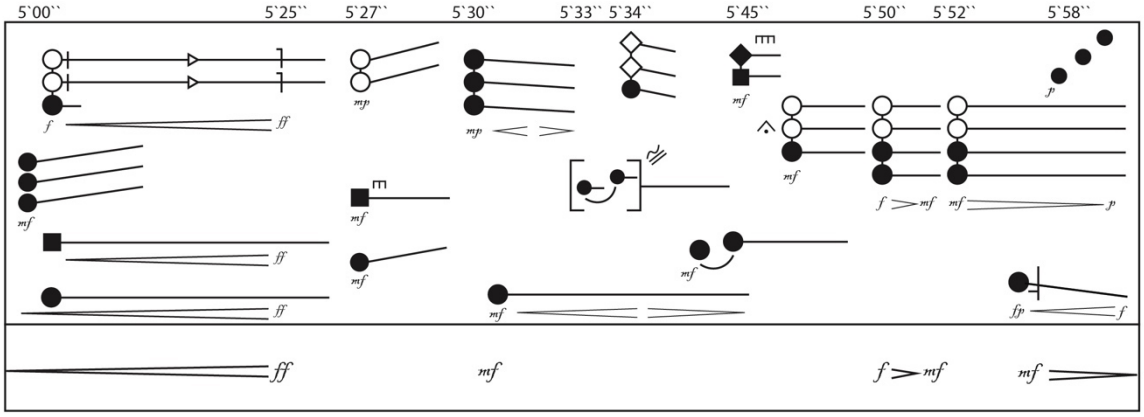


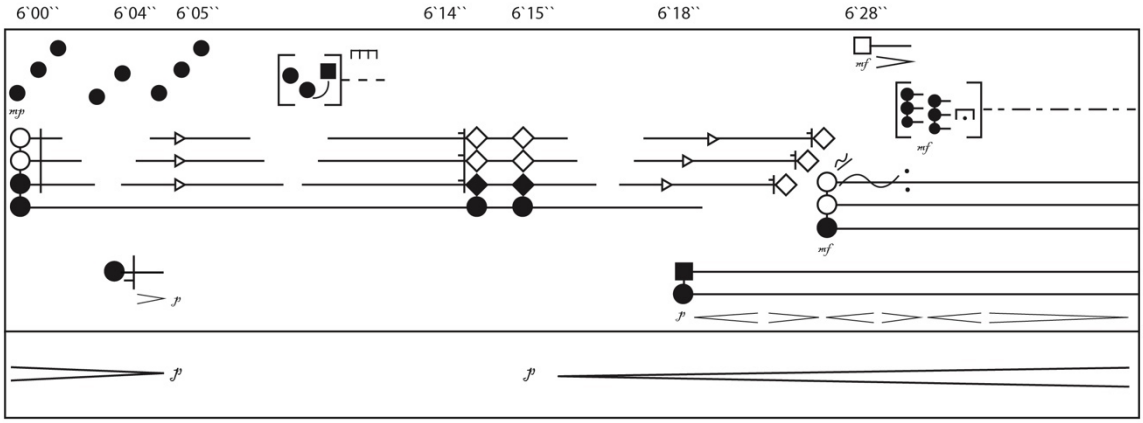


2/4

Texturas Urbanas, 3/4

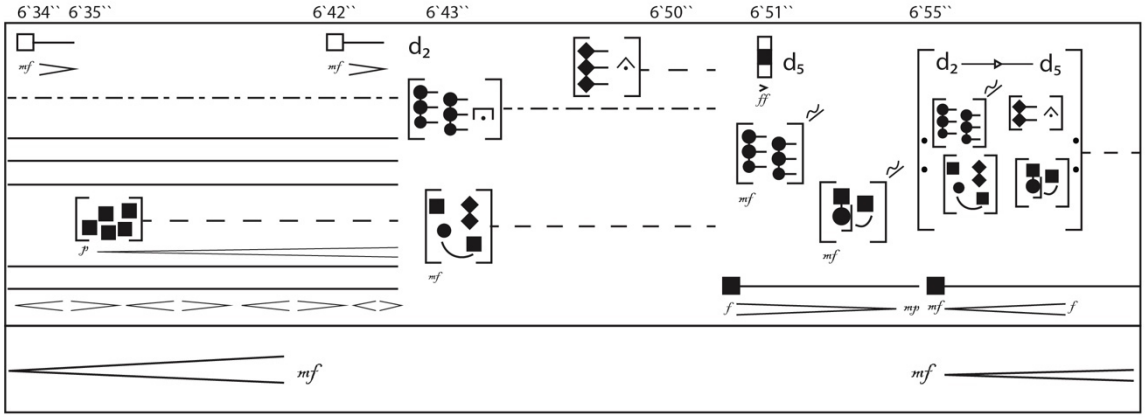


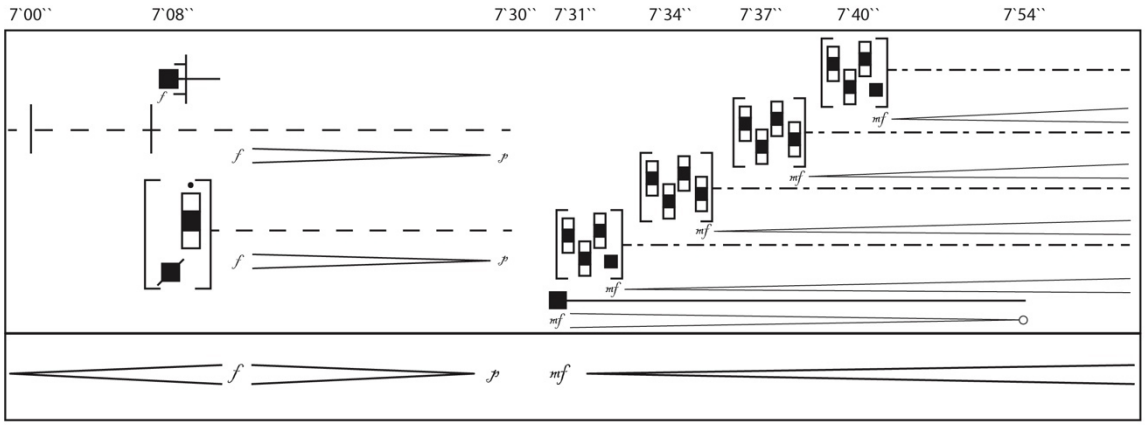


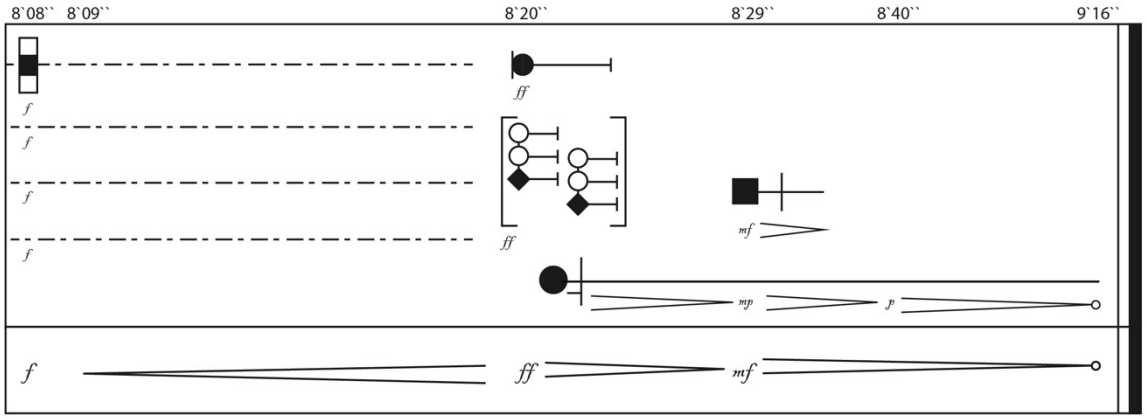


3/4

Texturas Urbanas, 4/4







4/4