

Proyecto Fin de Carrera
Ingeniería
Técnica de Telecomunicaciones
Sonido e Imagen

*Análisis de Técnicas de Diseño de Sonido en Entornos de Interacción
Audio-Imagen*

Autor:

Sergio Valero Garcia

Tutor:

José Manuel Iñesta Quereda



Universitat d'Alacant
Universidad de Alicante

Escola Politècnica Superior
Escuela Politécnica Superior

Febrero, año 2011

Agradecimientos

La realización de este proyecto ha sido un excitante reto para mi y me gustaría dar las gracias a aquellas personas que me han acompañado en el camino:

A la primera persona que quiero dar las gracias es a José Manuel Iñesta, por haber escuchado mis ideas y ofrecerme la posibilidad de realizar este trabajo, además, su ayuda y respeto durante la realización del mismo han sido fundamentales. Gracias.

A mi madre, mi padre y mi familia.

A todos mis amigos, en especial a Pilar, Pepe, Mj, Diego, Malaguita, Luis, Fran y Huellas.

A Salva y Martín por el equipo prestado.

A la música, por meterme en este mundo.

Y sobre todo a Ana, porque sin ti esto no hubiese sido posible.

Contenidos

Agradecimientos , i

Contenidos , ii

Índice de figuras , vi

Introducción	1
1.1 Motivación	1
1.2 Objetivos.....	3
1.3 Trabajos relacionados	3
1.4 Estado de la cuestión. Presente y futuro.....	4
Campos de actuación	7
2.1 Cine	7
2.1.1 Antecedentes	7
2.2.2 Aparición y necesidad de diseñar sonidos en el cine.....	8
2.2.3 La banda sonora	9
2.2.4 Jerarquías, metodología y arquitectura de proceso	11
2.2 Entornos programados: videojuegos, realidad virtual y entornos de aprendizaje	14
2.3 Análisis comparativo entre los diferentes campos	15
Análisis de técnicas para la generación de sonido 17	
3.1 Síntesis por modulación	17
3.1.1 Modulación de amplitud	19
3.1.2 Síntesis FM	19
3.2 Síntesis Sustractiva	21
3.2.1 Oscilador controlado por tensión VCO.....	21
3.2.2 Filtros	23
3.2.3 Etapa de amplificación	24

3.2.4 Conclusiones	25
3.3 Síntesis concatenativa por muestreo	27
3.3.1 Looping	27
3.3.1 Pitch-shifting	28
3.4 Síntesis Aditiva	28
3.5 Síntesis Granular	29
3.6 Revisión General. Taxonomía de las técnicas de síntesis	30
Análisis de la escena	35
5.1 Procedimiento	35
5.2 Objeto sonoro y evento sonoro	37
5.3 Paisaje sonoro y simulación de entornos ambientales inmersivos	38
5.4 Elementos psicológicos, cognitivos y emocionales	39
Técnicas de Análisis espectral de sonido	45
5.1 Enventanado y representación espectral	45
5.2 Perspectiva general de los métodos de análisis	46
5.3 Transformada de Fourier de tiempo corto STFT	47
5.4 El propósito del enventanado de tiempo corto	48
5.5 Inconveniente de en la resolución tiempo-frecuencia	49
5.6 Análisis basado en la Transformada Wavelet	50
5.6.1 La transformada Wavelet discreta (DWT)	51
5.7 Métodos de análisis de la envolvente espectral	54
5.7.1 Autoregresión (AR).....	54
5.7.2 Cepstrum	55
Modelo intuitivo de Análisis/diseño basado en las características físicas de la fuente	57
6.1 Introducción y planteamiento	57
6.2 Módulo de Imagen. Arquitectura	61
6.3 Módulo descriptor	64
6.3.1 Descriptores basados en la fuente-objeto	64

6.3.2	Eventos sonoros provocados por objetos vibratorios	66
6.3.3	Eventos sonoros provocados por líquidos	67
6.3.4	Eventos sonoros aerodinámicos	70
6.3.5	Complejidad temporal de los eventos sonoros	70
6.3.6	Análisis ontológico	74
6.3.7	Banco de muestras	75
6.3.7.1	Estructura	76
6.3.7.2	Metadatos asociados a cada muestra	78
6.4	Módulo de síntesis y procesado	79
6.4.1	Etapa de Análisis	81
6.4.2	Etapa de transformaciones	82
6.4.2.1	Aplicaciones y técnicas	82
6.4.2.2	Warping	83
6.4.3	Técnicas de resíntesis	84
6.4.3.1	Estructura de la etapa de resíntesis	87
6.4.5	El banco de filtros	87
6.4.6	Arquitectura del procesamiento para ETC	89
6.4.7	Procesado de localización especial.....	90
6.4.7.1	Aspectos teóricos sobre espacialización y reverberación a tener en cuenta para el diseño	91
6.4.7.2	Procesado de Reverberación por Convolución	95
6.4.7.3	Procesado sobre la dinámica y la intensidad del sonido	97
6.4.8	Módulos complementarios de síntesis	100
6.5	Sistema de mapeo	101
6.5.1	Descriptores de sonido	102
6.5.2	Diseño basado en la correlación de atributos	108
6.5.2	Estructura de la etapa de mapeo	109

Sonorización de varios clips de audio.

Experimentación y resultados 111

7.1	Equipamiento	112
7.2	Metodología y técnicas	113
7.3	Procedimiento y aplicación práctica del modelo mediante la sonorización de varios clips de video	114

7.3.1 Sonorización de la sección “Peripetics” [2:30:20 / 3:01:12]	115
7.3.2 Diseño de un objeto sonoro perteneciente a la secuencia Tierra compuesto por múltiples impactos	122
Conclusiones	127
8.1 Objetivos del proyecto. Revisión y valoración	127
8.1 Futuras líneas de trabajo	130
Bibliografía	133

Índice de figuras

2.1	Esquema jerárquico de los puestos de producción en el proceso de creación de una banda sonora. (a) Jerarquía de producción, (b) jerarquía de editorial y (c) jerarquía de mezcla	13
2.2	Sistema de producción y edición de sonidos Foley.....	13
2.3	Muestra de sonido etiquetada	14
2.4	Tabla resumen de actuación de las técnicas	16
3.1	Oscilador simple	18
3.3	Bloque de osciladores. Síntesis FM simple	20
3.4	Estructura de bloques de un sintetizador sustractivo	22
3.5	Envolvente de amplitud y cada una de sus partes	24
3.6	Análisis de relaciones para la síntesis sustractiva	26
3.7	Taxonomía de técnicas de síntesis de sonido propuesta por Smith.....	34
4.1	Diagrama de análisis de la escena	36
4.2	Composición de un tono complejo. Helmholtz	40
4.3	Relaciones subjetivas emocionales del sonido	41
4.4	Tamaño objetivo en función del tamaño real de una bola	42
5.1	(a)Ventana rectangular continua. (b)Ventana rectangular discreta. (c)Ventana Hanning continua (d)Ventana Hanning discreta (e)Ventana Hamming continua (f)Ventana Hamming discreta (g)Ventana Blackman continua (h)Ventana Blackman discreta.....	46
5.2	filtrado paso alto y paso bajo de la señal antes de aplicar la TW.	53
5.3	Submuestreo.	53
5.4	Sobremuestreo.	53
6.1	Arquitectura del modelo.....	59
6.2	Estructura del flujo del sistema.....	60
6.3	Diagrama de funcionamiento del módulo de imagen.	61

6.4	Fotogramas sobre los que se representan las unidades, los segmentos y las dimensiones.	63
6.5	Taxonomía propuesta para sonidos de impactos	65
6.6	Taxonomía propuesta para sonidos de rozamientos	66
6.7	La imagen refleja las cavidades formadas por un líquido al recibir un impacto y deformar su estructura superficial.	67
6.8	Taxonomía propuesta para sonidos de impactos en líquidos	68
6.9	Taxonomía propuesta para sonidos de flujo de líquido	69
6.10	Taxonomía propuesta para ETC.	72
6.11	Taxonomía propuesta para patrones por postergación.	73
6.13	Estructura de las muestras dentro del banco de sonidos	77
6.14	Alcance de transformación de cada muestra.....	78
6.15	Estructura metadatos asociados a cada muestra.....	79
6.16	Diagrama del módulo de síntesis y su flujo de datos	80
6.17	Diagrama de funcionamiento la etapa de análisis.	81
6.19	Transformaciones aplicadas sobre un proceso de análisis.	82
6.20	Efecto TW aplicado a una señal sinusoidal.	84
6.21	Señal reconstruida mediante el Método Overlap-add.	85
6.22	Banco de osciladores para la reconstrucción de la señal mediante síntesis aditiva.....	86
6.23	Estructura de la etapa de resíntesis.	87
6.24	Diagrama de bloques del filtrado por convolución.	88
6.25	Diagrama de flujo del sistema de retardo y construcción de patrones.	90
6.26	Filtrado por convolución.	96
6.27	Aplicación de procesadores de dinámica sobre una señal de audio.	99
6.28	Sistema de mapeo	101
6.29	Diagrama de la etapa de mapeo.	109
7.1	Diagrama del equipamiento utilizado	112
7.2	Secuencia de imágenes de los diferentes movimientos del objeto.	117
7.3	Muestras de la pista 1.	118
7.4	Ecuación aplicada a las muestras de la pista 1.....	118
7.5	Compresión aplicada a las muestras de la pista 1.	119
7.6	Generador de ruido del sintetizador y conexiones del vocoder.	119
7.7	Visualizador de análisis del vocoder	120
7.7	Parámetros de síntesis escogidos para el sonido de la inhalación.....	120

7.8	Imagen que muestra piedras cayendo sobre un montículo.	122
7.8	Imagen que muestra dos de las muestras finalmente escogidas.	124
7.9	Imagen de la ecualización aplicada a la muestra.	125

“El momento más excitante es cuando agrego el
sonido...(entonces) me estremezco”

Akira Kurosawa

1

Introducción

1.1 **Motivación**

Desde que la imagen y el sonido se unieron en el cine, surgió la necesidad de transmitir a través del sonido las mismas sensaciones que las imágenes causaban en los espectadores. Sin embargo la tecnología y técnicas de rodaje así como de captación de la imagen condicionaron el modo en el que el sonido directo era captado en la escena. Los micrófonos no siempre podían captar el sonido ambiente y los diálogos por dos cuestiones fundamentales. La primera es que

en el cine, el rodaje la película no se hace linealmente en el tiempo, si no que se ejecuta mediante varias tomas, las cuales serán seleccionadas convenientemente por el director y posteriormente montadas en el proceso de montaje. Este hecho deriva en que el sonido ambiental de la escena no presentará una continuidad lineal, ocasionado cortes no naturales en el sonido. La segunda vino derivada de la necesidad de doblar los diálogos, ya que perdíamos por completo el sonido ambiental de los otras tomas del rodaje. Estas complicaciones provocaron la aparición de una nueva disciplina que se ha ido imponiendo técnica y artísticamente: el diseño de sonido. Hay que destacar que no es el cine el único campo donde el diseño de sonido tiene un papel fundamental. La radio fue un medio donde se experimentó con la recreación de sonidos imitativos. Es de aquí de donde surgieron los primeros artistas Foleys: los compositores de la vanguardia contemporánea de música electro-acústica, concreta y electrónica. Entre otros destacaré Pierre Schaeffer, John Cage, Karlheinz Stockhausen, Luigi Russolo, Curtis Roads, Iannis Xenakis y laboratorios de investigación como el IRCAM, han llevado la experimentación técnica a la escala del arte.

Este proyecto nace con la intención de estudiar y analizar la parte del diseño de sonido menos arraigada a la música y al diseño de instrumentos musicales, aquella que trata de retratar los sonidos presentes en la naturaleza y nuestra vida cotidiana. Este tipo de sonidos responde a patrones sonoros impuestos por fuentes sonoras presentes en la la naturaleza, las cuáles muestran un comportamiento en algunos casos impredecible y con una estructura estocástica e inarmónica. La síntesis de este tipo de sonidos presenta gran complejidad, si lo que queremos conseguir es una simulación bastante real de estos. Por esta razón, la técnica más utilizada para conseguir este nivel de realismo ha sido la digitalización de muestras y su posterior manipulación y edición.

Este trabajo parte de la idea de que el diseño de sonido en entornos audiovisuales muestra unas características muy específicas de diseño que lo hacen extremadamente complejo para diseñadores que no sean grandes conocedores de esta disciplina. Además, las necesidades de competitividad de la industria del audiovisual han hecho que el sonido que se escucha en las películas y videojuegos sobrepase los límites de una visión realista y trate de conseguir texturas y formas sonoras cada vez más complejas.

1.2 **Objetivos**

Este trabajo presenta tres fases de desarrollo, la primera tiene como objetivo el análisis de las técnicas empleadas en el diseño de sonido, así como el estudio de la metodología característica en los diferentes campos que hacen uso de este. En la segunda fase se plantea un modelo para el diseño de sonido en entornos audiovisuales basado en un control intuitivo de los parámetros de diseño, por medio de un mapeo de descriptores basados en las características físicas de la fuente, extraídas mediante un análisis ontológico de la escena audiovisual. La tercera fase se plantea como la aplicación práctica al estudio del proyecto, en esta se lleva a cabo la sonorización de varios clips de video, en los que se pone en práctica las técnicas estudiadas en la fase uno y el modelo propuesto en la fase dos.

Destacaré que no se trata de una técnica de síntesis de sonido, sino que es un modelo que utiliza técnicas de síntesis, procesamiento de la señal y procesos cognitivos relacionados con la percepción para diseñar sonidos, de una manera intuitiva e iterativa.

A continuación estructuraremos en forma esquemática las principales líneas de desarrollo de este proyecto:

- Análisis de los campos relacionados con el tema
- Análisis de las técnicas de síntesis de sonido
- Análisis de técnicas de procesamiento de audio
- Planteamiento del modelo
- Desarrollo del modelo
- Aplicación práctica del modelo
- Conclusiones

1.3 **Trabajos relacionados**

Durante muchos años la mayoría de los esfuerzos que se han llevado a cabo en el desarrollo de aplicaciones dentro del campo del diseño de sonido, han estado mayoritariamente basados en sonidos musicales, es decir, en la implementación de instrumentos. Sin embargo la aproximación de la síntesis

de sonidos hacia la búsqueda de sonidos naturales realistas no ha tenido el mismo interés históricamente. En estos últimos años gracias a los estudios de unos de los pioneros dentro de la clasificación de sonidos cotidianos [1] y el trabajo con nuevas técnicas de análisis de sonido con wavelets [2][3][4], donde se han conseguidos grandes aproximaciones.

Por otro lado, es quizás en el campo de el análisis y resíntesis por modelos físicos donde se han conseguido los mejores resultados [5][6][7][8][9]. Quizás el hecho más preocupante es la escasez de trabajos que proponen técnicas de diseño intuitivas, es aquí donde encontramos un excelente campo de trabajo en el que destacamos [10], que propone un mapeo automático de transformaciones espectrales para sonidos de impactos resintetizados.

1.4 Estado de la cuestión. Presente y futuro

Si comparamos el diseño de sonido con la infografía, nos encontramos con algunas analogías ciertamente interesantes:

- La infografía trata de imitar formas, colores, texturas y movimientos de objetos de la naturaleza.
- Además, la infografía puede recrear nuevas formas, objetos y efectos, con características más abstractas e inusuales.
- La infografía es utilizada en los medios audiovisuales al igual que el diseño de sonido.
- Las técnicas de construcción que se usan guardan una relación directa con las técnicas usadas en diseño de sonido.
- En el campo del cine y la animación, se emplean técnicas híbridas de para la creación de texturas con el objetivo de proporcionar un mayor realismo a la imagen.
- También se analiza el comportamiento, movimiento y la física de los objetos en la naturaleza, para la recreación de gráficos.

- De la misma manera que en síntesis de sonido, se busca encontrar resultados muy realistas.

Por lo tanto al igual que la infografía, el diseño de sonido, esta sujeto a la mejora de tecnología y técnicas relacionadas con tres aspectos del diseño:

1. Desarrollo de herramientas de síntesis de sonido.
2. Desarrollo de herramientas de procesado de audio.
3. Desarrollo de herramientas edición.
4. Desarrollo de técnicas y herramientas de construcción y diseño

Los tres primeros forman parte del campo de la generación de sonido y el tratamiento de tratamiento audio. El desarrollo de nuevas aplicaciones es crucial a la hora de avanzar hacia unos resultados mucho mas creíbles y controlados. Por otro lado, el cuarto campo es donde menos técnicas e implementaciones se han desarrollado. Este campo debe avanzar hacia la automatización de procesos, el desarrollo de interfaces mas intuitivos, y dispositivos de interacción multimodal. Este trabajo pretende tratar algunos aspectos de este último campo, como son el análisis de la escena, o la creación de un mapa de relaciones fuente-sonido-usuario, que faciliten el diseño al usuario. Se ha hablado arriba de la necesidad de implementar interfaces intuitivos, que ayuden a usuarios profanos en la materia (o no tan expertos con los parámetros propios de la síntesis y el tratamiento de audio) a trabajar en el proceso de diseño de sonido sin encontrarse ante cientos de parámetros de control. En la actualidad este tipo de interfaces y automatizaciones se han desarrollado hacia el campo de la música, donde por ejemplo hay programas y sistemas capaces de imitar a los dj 's como Tractor Pro® o programas como Ableton Live®, que ayuda a los músicos a llevar canciones y actuaciones al directo, con un fantástico control a tiempo real. Por lo tanto, el futuro del diseño de sonido en el entorno audiovisual debe de pasar por los mismo cambios y avances.

2

Campos de actuación

2.1 Cine

2.1.1 Antecedentes

Es en 1928 cuando se estrena en EE.UU. “ The lighs of New York” , la primera película norteamericana totalmente sonora, la cual incorporaba diálogos, música, ambientes y efectos de sonido. Atrás quedaban los años en los que el cine era silente o solamente estaba acompañado por música, la cual se tocaba en vivo en la propia sala de reproducción de la película por músicos

profesionales. Hubo otros sistemas posteriores basados en el gramófono que acompañaron al celuloide en sus primeros años de vida. Sin embargo, estos sistemas carecían de las cualidades técnicas necesarias para una buena producción industrial, así como la capacidad de difusión para una gran cantidad de espectadores. La imposibilidad de una buena sincronización entre el sonido y la imagen desacreditó a este sistema como un estándar tecnológico a implantar. Por otra parte, estos sistemas carecían de la potencia acústica necesaria para sonorizar salas de cine de la envergadura que se requería.

La llegada de la válvula triodo, bautizada por su inventor, Lee De Forest, como “audion” permitió a De Forest trabajar y desarrollar el sistema de grabación óptica al borde de la imagen en la película, sistema propuesto por Eugene Lauste. Este sistema fue descartado durante años debido a su principal problema: la falta de amplificación. Resuelto el problema de la amplificación, De Forest se une a Theodore Case y crean la compañía Phonofilm, destinada a la creación de películas sonoras.

La capacidad de sincronización perfecta del sonido con la imagen que ofrecía este nuevo sistema, junto a la capacidad de grabación de audio, proveniente tanto de sonido directo, como de sonidos recreados con posterioridad en un estudio, efectos o música grabada, acercaron a la grabación en la banda lateral de la película, a un estándar tecnológico desarrollado y evolucionado hasta la actualidad.

2.2.2 Aparición y necesidad de diseñar sonidos en el cine

Con la llegada del sonido al cine se abrió un nuevo mundo de posibilidades y los directores de cine no tardaron en empezar a explotar todo su potencial. En los primeros días del cine sonoro, el principal objetivo era la captación de sonido directo, es decir, que los diálogos de los actores se escucharan de una manera correcta. Los rodajes se empezaron a hacer en estudios insonorizados para que la captación del sonido tuviese la mejor calidad posible. Esto, junto a la inquietud de algunos directores de incorporar a través del sonido elementos narrativos y artísticos a las películas favoreció la incorporación de sonidos postproducidos.

Estos sonidos comenzaron siendo grabaciones de elementos sonoros de la escena como podrían ser las campanadas de la torre de una iglesia, el canto de unos pájaros o el ruido de una locomotora pasando a toda velocidad. La evolución tecnológica que se ha desarrollado entorno al cine, la tecnología de grabación sonora y sobre todo en el campo de la síntesis y el tratamiento de señales de audio han facilitado la creación de bandas sonoras de una extrema complejidad, las cuales incorporan sonidos diseñados especialmente para la película, recreando ambientes y paisajes sonoros antes imposibles de imaginar.

Hay que señalar que el papel que ejerce el sonido en el cine viene determinado por la imagen. Esta nos proporciona a nuestro sistema perceptivo tal cantidad de información que es determinante analizarla para poder diseñar el sonido correspondiente a cada elemento sonoro. No siempre la imagen nos ofrecerá una información sonora aprendida por ejemplo, en la película Parque Jurásico el sonido de rugidos de los dinosaurios fue diseñado, a pesar de que nadie sabe con certeza qué clase de rugidos emitían los dinosaurios. Sin embargo observando las propiedades físicas del cuerpo de los dinosaurios, la actitud o personalidad aplicada al “personaje” y manteniendo una sincronización adecuada entre audio-imagen fue posible recrear artificialmente esos rugidos. Estos son aceptados como reales por el espectador e inmersos en la imagen.

2.2.3 La banda sonora

Se definirá la banda sonora de una película como la o las pistas de audio que acompañan a la película y que están en sincronía con la imagen. Se compone de varias partes destacadas dentro del sonido en el cine, la más popular y por la que el término banda sonora es más conocido, es la parte destinada a la música de la película; tendremos además diálogos, sonido de sala y de ambiente y efectos de sonido. Todos ellos conforman la banda sonora de una película. Este proyecto de centra en el análisis de las técnicas que actualmente se utilizan para producir artificialmente los efectos de sonido y el audio ambiental.

Actualmente se ha conseguido automatizar muchos de los procesos que forman parte de la postproducción de sonido, sin embargo sigue siendo un

proceso costoso. Requiere la participación de un equipo humano importante y por otro lado, la calidad o la garantía de un buen trabajo viene determinada por la utilización de tecnología de muy alta gama. El problema fundamental lo encontramos en que la tecnología actual para la generación de sonidos sintetizados o editados se centra principalmente en la búsqueda de nuevos sonidos musicales.

La técnica más utilizada en el cine para recrear sonidos sigue siendo la grabación de muestras o pistas de audio. Son grabadas en salas acústicamente preparadas, utilizando equipos de grabación convencionales. A estos sonidos se les llama Foleys, en honor a Jack Foley quien popularizó esta técnica en los primeros años del cine sonoro. Esta técnica trata de recrear los sonidos que forman parte del ambiente sonoro de la escena y que no han sido grabados junto a los diálogos. Hay que tener en cuenta que no son efectos de sonido.

Como ejemplo de sonidos Foley tenemos pisadas, timbres, objetos en movimiento, golpes de objetos, apertura y cierre de puertas, etc. El principal problema que presentan estos sonidos es la gran sensación de realismo que requieren, ya que el espectador está muy familiarizado con este tipo de sonidos y es capaz de distinguir entre unas pisadas con una zapatilla de deporte sobre gravilla y otras pisadas producidas con el mismo zapato sobre un suelo de parque. Nos encontraremos ante dos sonidos de pisadas, que pueden estar perfectamente sincronizadas con la imagen, pero nuestra memoria sonora analiza y distingue con gran precisión los cambios en el timbre de un sonido. El cine sigue utilizando esta técnica por otra razón fundamental: la dificultad de recrear sintéticamente sonidos naturales con una alta calidad imitativa. Los movimientos, los golpes, en definitiva la interacción de los objetos sonoros en la escena se produce de una manera aleatoria, y con diferentes intensidades en la mayoría de los casos. Estos objetos sonoros [11] provienen de la naturaleza y pueden presentar comportamientos y patrones de diversos tipos. Nos podemos encontrar sonidos con una componente oscilatoria o amortiguada, la cual presenta patrones sencillos y fáciles de sintetizar, pero por otra parte encontramos sonidos de componente estocástica y caótica.

Basados en el campo de la síntesis por modelos físicos podemos encontrar diversos estudios dedicados al modelado y síntesis automática de Foleys [8]. Estos aproximan de una manera realista y muy próxima el objetivo, tanto el timbre de los objetos sonoros a sintetizar, como su transición en el tiempo.

Los efectos de sonido, comúnmente llamados SFX, conforman otra parte fundamental de la banda sonora. Son sonidos recreados para resaltar ciertas partes de la película y otorgarle cierto dramatismo, así como simular paisajes sonoros reales o imaginarios, sonidos de elementos que interfieren en la escena a través de infografía u objetos sonoros clásicos del cine de ciencia ficción como disparos, láseres, flechas, naves espaciales, etc.

Los resultados más eficientes suelen aparecer con la combinación de varias técnicas, ya que la mayoría de estos sonidos se compone de varias capas de audio compuestas por superposición y concatenación. Para ejemplificar esta técnica imaginemos un coro vocal cantado por varias personas. El coro está compuesto por 5 voces diferentes que se superponen una con la otra formando un conjunto coral; por otro lado tenemos la voz de cada uno de los cantantes, compuesta por fonemas que se unen entre sí, es decir, se concatenan para formar las palabras.

Por lo tanto, el diseño de sonido se basa en una combinación de síntesis, edición y procesado de señales de audio. La síntesis nos proporciona la unidades o elementos más simples. Con la edición superponemos y concatenamos esas unidades simples y a través del procesado de la señal situamos el objeto en un espacio físico adecuado.

2.2.4 Jerarquías, metodología y arquitectura de proceso

El diseño de sonido en cine difiere de otros entornos audiovisuales por estar jerarquizado entorno a la procesos de producción predefinidos por la industria. El modelo americano de cine obliga a trabajar de una manera diferente que el resto del cine, debido a la necesidad de cumplimiento de plazos y presupuestos. Los procesos de postproducción en el modelo americano se realizan al mismo tiempo que el rodaje, a diferencia de otras industrias cinematográficas que realizan el rodaje de toda la película y dejan la postproducción para el final. La jerarquía que podemos encontrar para la realización de una película es la que se muestra en las figuras 2.1(a)(b)(c)

Podemos destacar ciertos pasos del sistema jerárquico anterior a distinguir con relevancia para este estudio. Debemos estudiar la metodología y tecnología empleada en cada una de las siguientes fases: supervisión y diseño de sonido, edición de SFX, técnico de grabación, edición de Foley, artista Foley y

mezclador principal. El proceso de edición de diálogos y el de composición musical, así como la grabación y producción de ambos, no es objeto de estudio para este proyecto.

El proceso de recreación de los Foley se lleva a cabo entre el artista Foley, encargado de recrear los sonidos mediante objetos, diferentes pavimentos, etc. y el editor de Foley, cuya función es la de grabar y editar los sonidos recreados por el artista, ver figura 2.2. Este proceso es muy costoso, se necesitan entre 2 o 3 artistas Foley, 1 o 2 editores y a esto hay que sumarle la tecnología empleada. Un estudio Foley requiere de varios pavimentos, para recrear pisadas, puertas y toda clase de objetos, además de microfónica y todo el sistema para grabación y edición de estos sonidos. Como esta tecnología no está siempre al alcance de todas las producciones cinematográficas, podemos encontrar librerías comerciales de sonidos Foley. Estos sonidos presentan una gran ventaja a la hora de recortar gastos de producción y tiempo. Por otra parte, carecen de la posibilidad de manipulación y transformación.

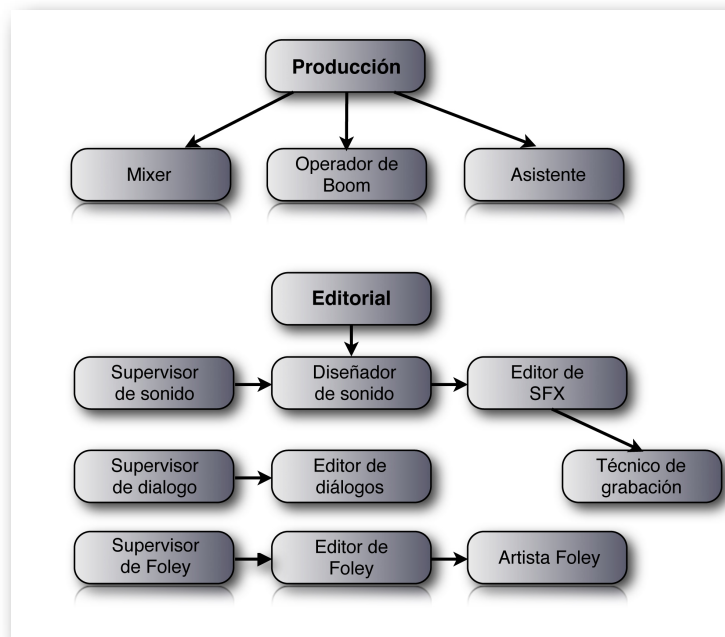


Figura 2.1(a)(b)

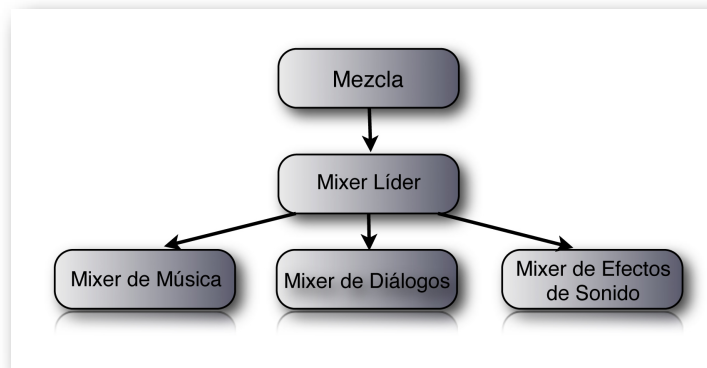


Figura 2.1 (c)

Figura 2.1 Esquema jerárquico de los puestos de producción en el proceso de creación de una banda sonora. (a) Jerarquía de producción, (b) jerarquía de editorial y (c) jerarquía de mezcla

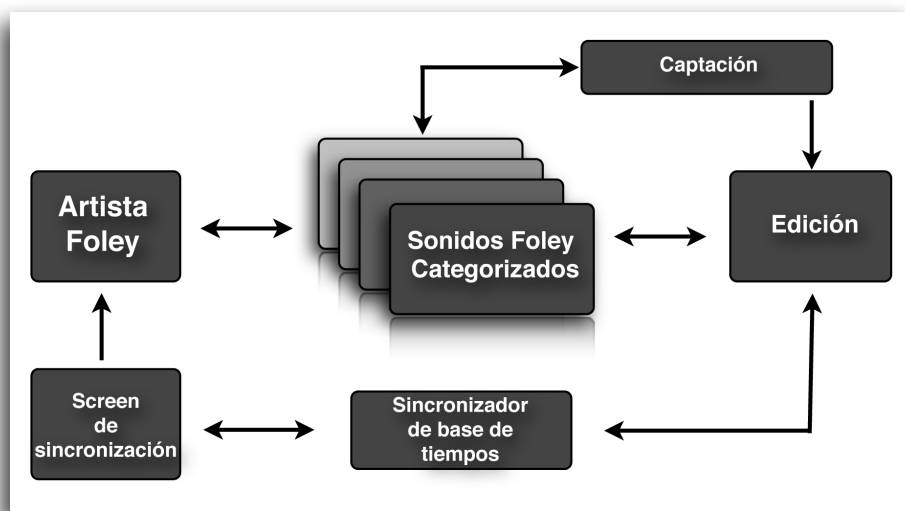


Figura 2.2 Sistema de producción y edición de sonidos Foley.

La última fase, la edición, nos da como resultado las muestras que serán utilizadas para recrear el sonido de sala perdido durante el rodaje. Estas muestras pasan por un proceso de etiquetado, donde se clasifican y distinguen según categoría, título, duración, ... Es decir, están sujetos a unos descriptores, figura 2.3.

Estos descriptores nos permiten hacer un primer análisis de bajo nivel, además nos facilitan gran cantidad de información para realizar un proceso inverso de síntesis y diseño.

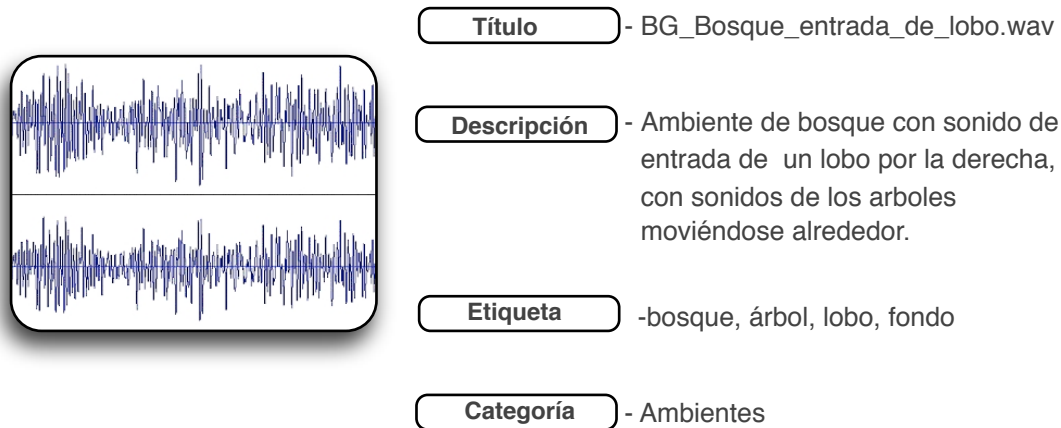


Figura 2.3 Muestra de sonido etiquetada

2.2 Entornos programados: videojuegos, realidad virtual y entornos de aprendizaje

En estos entornos audiovisuales el audio es implementado mediante un lenguaje de programación. Entre los más destacados están los videojuegos, realidad virtual, entornos de aprendizaje y entornos multimedia/web. Esto supone que la disposición espacial del sonidos, mezcla, intensidad, reverberaciones, etc. es programada mediante algoritmos. El personaje o usuario interactúa tridimensionalmente con el sonido, y este, a su vez, con él.

A diferencia del cine, donde el audio es grabado sobre un formato magnético u óptico, aquí es necesaria una digitalización, compresión y farmateo del audio. Por lo tanto, las plataformas que alojarán el sonido determinarán la calidad de este así como el número de canales y la capacidad de procesamiento. Quizás el ejemplo más claro está en los videojuegos. En los comienzos de la era del videojuego, la escasa capacidad de procesamiento y almacenaje de las primeras videoconsolas y computadoras, no permitían procesar el audio con una profundidad de bits superior a 8 bits, lo que hacía imposible recrear sonidos realistas. Las actuales plataformas de videojuegos,

permiten hasta 256 canales de audio, sonido Dolby, DTS y WMA Pro multi-canal, y una calidad de hasta 48 kHz. Por lo que el sonido actual en los videojuegos ha adquirido una importancia considerable dentro de estos, y no actúa sólo con motivos estéticos, sino que adquiere suma importancia en la jugabilidad y en la interacción con el usuario, al que abstrae y guía hacia los objetivos del juego.

Podemos decir que el diseño de sonido en el campo de los videojuegos comparte grandes similitudes con el cine, incluso supera a este en complejidad, ya que los videojuegos se diseñan sobre un espacio virtual tridimensional mucho más amplio que el del cine. Además se le añaden más dimensiones dependiendo de la actividad e interacción del usuario.

La metodología empleada es muy parecida a la del cine, incluso es posible grabar Foleys para los movimientos humanos y de objetos más complejos. La gran diferencia es que aquí el diseñador de sonido suele trabajar sobre vídeos que todavía están poco depurados, en las primeras fases de creación. Por otro lado, la sincronización no depende solo de esta fase, sino que es también importante en el proceso de implementación (programación), así como la localización espacial.

2.3 Análisis comparativo entre los diferentes campos

Una vez estudiados los campos audiovisuales donde el sonido juega un papel fundamental, podemos sacar algunas conclusiones que diferencian estos campos. La principal está ligada al dinamismo del medio. El sonido en el cine se presenta como un ente estático, sin posibilidad de manipulación ni transformación. Por el contrario, en los videojuegos y los entornos de realidad virtual, es sumamente importante la interacción con el medio, cualidad que obliga a que el sonido también interactúe con el usuario y el medio. Esto nos lleva a la siguiente conclusión: las técnicas de grabación y edición de muestras de audio no son las más adecuadas para este tipo de entornos. Está claro que esta técnica se ha depurado mucho y las muestras que se usan son de un realismo sorprendente que impacta al usuario, sin embargo carecen del dinamismo que se necesitaría en entornos de este tipo.

campos	Técnicas	interacción	Realismo	costes asociados	Nivel experto
cine	mixtas	estática	alto	económicos temporales artísticos	medio/ alto
Video juegos	mixtas	dinámica	medio	económicos temporales artísticos innovación	Alto
RV	mixtas	muy dinámica	muy alto	temporales computacional	Muy alto

Figura 2.4 Tabla resumen de actuación de las técnicas

Se ha avanzado en el dinamismo a la hora de la localización espacial del sonido, ya que la mayoría de las plataformas de desarrollo de juegos y RV disponen de procesadores de reverberación y control espacial de los objetos. Esto hace posible que sean programados dentro del juego o el entorno ciertos parámetros de procesamiento espacial. Sin embargo, en el campo de la síntesis se hace todavía muy difícil ajustar parámetros del sonido que modificarían su timbre y en tiempo real. Un ejemplo comercial de este tipo de plataformas es Wwise® de AudioKinetic®.

3

Análisis de técnicas para la generación de sonido

3.1 **Síntesis por modulación**

El concepto de modulación se basa en la idea de alterar algún parámetro de una onda sonora en razón del valor de la otra onda. Las formas más comunes de modulación son modulación de amplitud o AM, que consiste en alterar la

amplitud de una señal portadora en función del valor de otra señal llamada moduladora y la modulación de frecuencia o FM, que consiste en variar la frecuencia de una portadora en función de otra señal moduladora. Se consideran dentro del grupo de las síntesis no lineales.

Para explicar el concepto de modulación, consideremos primero un oscilador, como ilustra la figura 3.1 que genera una onda sinusoidal pura si todos los parámetros de entrada (amplitud, frecuencia y fase inicial) son constantes.

La síntesis sonora por modulación permite evitar uno de los problemas más graves de la síntesis aditiva, que es la necesidad de producir una gran cantidad de datos para generar timbres que sean suficientemente ricos y complejos.

La síntesis sonora por modulación ofrece una alternativa interesante, ya que permite generar espectros ricos con pocos recursos, tanto por la especificación de los datos de control como por el tiempo de cálculo necesario para generar el sonido sintético. Musicalmente, una modulación a baja frecuencia de la amplitud corresponde a un trémolo, mientras que una modulación a baja frecuencia de la frecuencia corresponde a un vibrato.

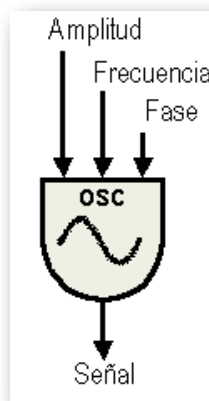


Figura 3.1 Oscilador simple

3.1.1 Modulación de amplitud

La modulación de amplitud es una de las técnicas de modulación más antiguas. Ha sido muy utilizada en la música electrónica analógica. En este tipo de técnica de modulación la amplitud de la señal portadora varía en consecuencia de una señal moduladora, siendo la señal portadora una onda bipolar y la onda señal moduladora una onda unipolar. Sus características principales son que solo necesita dos osciladores para funcionar y que, debido a su no linealidad, se hace más difícil el control intuitivo de los sonidos comparado con las técnicas de síntesis lineales. Más que un método de síntesis rico en generación de señales complejas, se presenta como una buena técnica de modulación. La síntesis AM es muy poco utilizada porque presenta importantes limitaciones para la creación de espectros ricos en parciales.

3.1.2 Síntesis FM

La modulación de frecuencia supuso una etapa importante en el desarrollo de los métodos modernos de síntesis de sonidos. Este método de síntesis sonora explota los mismos principios que se utilizan en las transmisiones radiofónicas FM y fue inventado por John Chowning en 1967 [12] mientras experimentaba con diferentes tipos de vibrato. Descubrió que modulando la frecuencia de un oscilador sinusoidal modulador se podían obtener cambios de timbre. En los años siguientes continuó investigando el fenómeno, hasta formular un modelo sencillo, eficiente y elegante para la generación de espectros complejos y dinámicos mediante frecuencia modulada. Una variante de este tipo de síntesis es la que se implementa en los famosos sintetizadores DX7 de Yamaha, quien compró la patente a Chowning. Este sintetizador es famoso por ser el primer sintetizador completamente digital.

La técnica más básica de síntesis FM consiste en que un oscilador que genera una señal sinusoidal portadora que es modulada en frecuencia por otra sinusoidal que actúa como moduladora. En la figura 3.3 podemos observar una configuración básica de síntesis FM. La simplicidad de su arquitectura y su capacidad para generar una gran variedad de diferentes timbres hacen de

esta técnica de síntesis una herramienta muy adecuada para ser implementada y utilizada en una gran cantidad de aplicaciones. Como hemos dicho, esta técnica es capaz de generar sonidos y texturas muy dinámicos, debido a que es una técnica no lineal. La cualidad de poder crear variaciones temporales en el espectro de un sonido es de suma importancia a la hora del diseño. Esto es muy sencillo de lograr, basta hacer que el índice de modulación varíe en el tiempo. El índice de modulación se define como la relación de la desviación pico de frecuencia dividida entre la frecuencia de la señal moduladora:

$$m = A_m / f_m \quad (3.1)$$

La amplitud de ambos osciladores es controlada por respectivos generadores de envolvente; el de la portadora determina la envolvente de amplitud general del sonido y el de la moduladora, la evolución del índice de modulación. En [12] se muestra cómo utilizando relaciones c/m (relación entre la frecuencia de la portadora y de la moduladora), envolventes de amplitud y de modulación diferentes, este algoritmo capaz de producir timbres muy variados, desde instrumentos de viento de metal y madera, como clarinetes y fagotes, hasta campanas y otros sonidos de percusión.

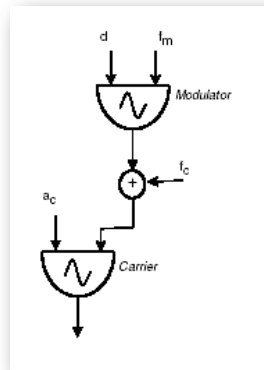


Figura 3.3 Bloque de osciladores. Síntesis FM simple

Las principales ventajas de la síntesis FM son:

- economía y eficiencia: con sólo cuatro operadores (dos osciladores y dos generadores de envolvente), se pueden lograr espectros complejos y dinámicos.
- flexibilidad: un único algoritmo es capaz de producir timbres muy diversos.
- control y predecibilidad: la técnica tiene una fundamentación matemática subyacente clara, que permite predecir sus resultados con alto grado de precisión.
- sencillez y elegancia: el manejo de unos pocos parámetros permite controlar el comportamiento global de una síntesis de resultados complejos.

Nos encontramos ante un método de síntesis muy potente y que aporta condiciones suficientes para ser empleado a la hora de generar sonidos continuos con una dinámica espectral variable. Este tipo de sonidos son conocidos dentro del campo del diseño sonoro como texturas, y en la naturaleza este tipo de texturas sonoras son fundamentales para el reconocimiento de espacios y lugares. El sonido continuo del mar, el murmullo de un río o el ruido del tráfico son ejemplo de este tipo de texturas sonoras.

3.2 Síntesis Sustractiva

La síntesis sustractiva se basa en el tratamiento de una fuente sonora mediante el uso de filtros. La fuente está generada mediante un oscilador que proporcionará trenes de impulsos simples o bien generadores de ruido. La salida del oscilador nos proporcionará un espectro muy rico en frecuencia, que será modelado mediante el uso de filtros. Estos filtros se encargan de modificar o suprimir frecuencias del espectro de la señal, esculpiendo este para conseguir sonidos naturales o sintéticos. La fase siguiente al filtrado es la de amplificación. Permite manipular parámetros sujetos a la envolvente de volumen. Su arquitectura viene descrita en la figura 3.4.

3.2.1 Oscilador controlado por tensión VCO

El oscilador se encarga de generar pulsos periódicos, dando como resultado una señal de audio primaria. Esta señal es generada en el dominio del tiempo, y

presenta una componente de amplitud y otra de tiempo, determinada por la longitud de onda. Por lo que podemos definir la altura de la señal, es decir, su frecuencia fundamental en esta primera fase.

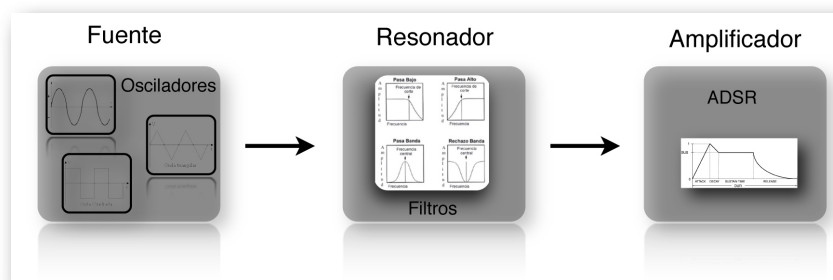


Figura 3.4 Estructura de bloques de un sintetizador sustractivo

La síntesis sustractiva es la más utilizada de las técnicas de síntesis analógica, ya que fue uno de los primeros métodos de síntesis en ser implementado en los primeros sintetizadores. En los primeros años de la síntesis sustractiva los osciladores generaban tres tipos de ondas: triangular, diente de sierra y cuadrada. Hoy, esta síntesis es implementada digitalmente y se consiguen formas de onda más complejas. Otras muy utilizadas son las ondas rectangulares moduladas en anchura de pulso (pulse width modulation, PWM),

$$D = \tau / T \quad (3.2)$$

generadores de ruidos blanco y rosa, pulsos, y otras. La forma de onda de estos pulsos determina el número y armonicidad de los parciales que tendrá la señal.

Realizando la transformada de Fourier a las señales generadas por los pulsos podemos observar la riqueza de su espectro, así también, determinaremos su frecuencia fundamental. Las ondas sinusoidales sólo presentan un armónico y no contienen otros parciales. El sonido que genera este tipo de ondas no se encuentra en la naturaleza, pero nos será muy útil a la hora de diseñar sonidos con una fuerte carga en bajas frecuencias. La onda

cuadrada generada por un oscilador, solo existe en dos estados: activo y no activo. esta onda genera parciales impares y depende del diseño de oscilador y de su precisión, ya que estos parciales están sujetos a la transición entre estados de la onda. Formas de onda como diente de sierra o triangular se caracterizan por generar sonidos ricos en parciales, produce el doble que la onda cuadrada, lo cual está relacionado con el brillo del sonido. Los generadores de ruido no nos aportarán una frecuencia fundamental, pero son utilizados para crear sonidos percusivos, así como sonidos con una fuerte componente estocástica, como son los sonidos de viento, mar, etc.

3.2.2 Filtros

Actúan directamente en el dominio de la frecuencia, atenuando o amplificando regiones del espectro. De esta manera consiguen esculpir la envolvente del este. El oscilador solo no es capaz de recrear un espectro complejo como es el de la mayoría de los sonidos encontrados en la naturaleza o en instrumentos musicales, por lo que el filtrado ejerce un papel fundamental a la hora de diseñar sonidos si utilizamos este tipo de síntesis. Destacaremos algunos de tipos de filtros más usados en la síntesis sustractiva. Estos son los Filtros Paso Alto (HPF), Filtros Paso Bajo (LPF), Filtros Paso Banda (BPF) y Filtros Elimina Banda (Notch, Peine). Entre estos filtros destaca el BPF por su capacidad de resonar a una cierta frecuencia, dando lugar a representar sonidos naturales expuestos a un resonador. así como su capacidad para eliminar gran parte del espectro de la señal y seleccionar frecuencias de una manera más precisa. Por otro lado, se encuentran los filtros elimina banda que son capaces de sustraer un rango de frecuencias preciso. Como se ha dicho, es la combinación de estos filtros lo que permite modelar la señal fuente y convertirla en otra más adecuada.

Los filtros pueden ser combinados en serie o paralelo, dependiendo del resultado que queramos conseguir. En la combinación en serie la salida de cada filtro entra en otro filtro, por ejemplo si la señal es modificada con un LFP, y pasa a través de un filtro Notch, modificar las frecuencias que el LPF no ha sustraído. En paralelo los filtros son aplicados individualmente y la salida de cada uno de ellos es sumada al resto.

3.2.4 Conclusiones

La arquitectura de la síntesis sustractiva proporciona ciertos parámetros de control en cada una de sus etapas. Estos parámetros están sujetos a propiedades físicas del sonido, que permiten el diseño de ciertos sonidos dependiendo de los valores otorgados a cada parámetro. En la etapa de generación del sonido, el oscilador controla tres parámetros fundamentales: frecuencia fundamental y riqueza espectral, esta última determinada por la forma del pulso utilizado:

- Oscilador : frecuencia fundamental, riqueza espectral.

La etapa de filtrado actúa en el dominio la frecuencia, sus parámetros de diseño son:

- Filtrado: atenuación, amplificación, selección, discriminación espectral.

Por ultimo, la señal entra en el amplificador que modificará su envolvente temporal:

- Amplificador: Amplitud global, ataque, caída, sostenimiento y liberación.

Este análisis sobre la síntesis sustractiva demuestra la gran capacidad de diseño que esta técnica ofrece. Su arquitectura de excitador-resonador favorece el diseño de gran cantidad de sonidos de instrumentos reales, ya que la mayoría de estos están compuestos por la misma arquitectura. Además son muchos los sonidos naturales que presentan esta configuración de excitador - resonador, entre ellos los sonidos vocales. Por otro lado, utilizando generadores de ruido como excitadores, podemos modular la envolvente espectral y la temporal y así crear gran cantidad de sonidos de procedencia estocástica, que no requieren una componente determinista. Si a esta técnica le sumamos la posibilidad de variación temporal de filtros y envolventes, podemos modelar sonidos con patrones de comportamiento más orgánicos.

Parámetro	Etapa	Dominio	Lineal	complejidad	variación/ tiempo	magnitud
Amplitud	oscilador	Tiempo	si	baja	no	Volt/dB
Frecuencia Fundamental	oscilador	Tiempo	no	alta	si	Hz
Riqueza espectral	oscilador	Frecuencia	no	alta	si	parciales/ OCT
Atenuación	filtrado	Frecuencia	no	media	si	dB/Q
Amplificación	filtrado	Frecuencia	no	media	si	dB/Q
Selección	filtrado	Frecuencia	si	baja	si	Hz
Discriminación	filtrado	Frecuencia	si	baja	si	Db/Q/Oct
Ataque	amp	Tiempo	si	baja	si	ms
Caída	amp	Tiempo	si	baja	si	ms
Sostenimiento	amp	Tiempo	si	baja	no	ms
Liberación	amp	Tiempo	si	baja	si	ms

Figura 3.6 Análisis de relaciones para la síntesis sustractiva

Es la posibilidad de control en tiempo real la que nos otorgará una complejidad de diseño superior. De esta manera, la interacción con la imagen se realiza de una manera más intuitiva.

Por otra parte, señalaremos la dificultad que esta técnica de síntesis presenta, ya que requiere un buen control sobre los pulsos excitados. Estos nos proporcionan movilidad sobre la frecuencia fundamental, pero no sobre sus parciales. Así, necesitaremos varios osciladores si lo que pretendemos es recrear sonidos con varias formantes que varían temporalmente, o el uso de diferentes bancos de filtros. Uno de los problemas que presenta esta técnica es que no realiza ningún Feedback entrada-salida, cualidad presente en muchos instrumentos y sonidos naturales. Los controles sobre el espectro de la

frecuencias afectan a la amplitud y al tono, sin embargo carece de una capacidad de manipulación espectral compleja. Por lo que la búsqueda de sonidos muy realistas es difícil y por lo general requiere un control muy preciso de los parámetros.

3.3 Síntesis concatenativa por muestreo

Esta técnica no es, según la definición, una técnica de síntesis muy ortodoxa, sin embargo es quizás una de las mejores técnicas y de las más utilizadas a lo largo de la historia de la música electrónica y del diseño de sonido en audiovisuales. El concepto del muestreo es el de utilizar muestras de audio, *samples*. Los orígenes del muestreo los encontramos en la música concreta y en los experimentos que realizaron Pierre Schaeffer y Pierre Henry mediante la manipulación de grabadores de cinta. Se grababan sonidos mediante el uso de micrófonos, posteriormente se manipulaba la cinta, haciendo cortes y pegados de esta para conseguir objetos sonoros concretos y modificando la velocidad de reproducción de esta. Estos dos conceptos son el fundamento básico del muestreo y en la actualidad son conocidos como Looping y Pitch-shifting. Hoy en día esta técnica se ha perfeccionado mucho en cuanto a la arquitectura, procesamiento y administración de los samples por parte de los samplers (palabra con la que se llama a los sintetizadores de muestras). Sin embargo, las dos principales manipulaciones destacadas anteriormente son las que siguen siendo el corazón y motor de esta técnica.

3.3.1 Looping

El looping (bucles) es una técnica que consigue mantener la duración de una nota u otro sonido durante un tiempo ilimitado. Esto es posible siempre y cuando el sonido disponga de una fase estable. Esta se suele encontrar en la parte de sostenimiento la envolvente temporal de un sonido. Esta técnica funciona muy bien para sonidos musicales y más concretamente para la simulación de instrumentos reales no percusivos. Sin embargo no es esta forma de bucle la que más se adecúa a las condiciones de diseño de esta investigación. Si hacemos hincapié en el concepto de muestra y de bucle desde otra escala [13], nos encontramos, por un lado, llevando la escala a un

concepto macro, con muestras de larga duración y los bucles de estos determinan una forma precisa de diseñar entornos sonoros.

El diseño de paisajes sonoros se crea mediante la concatenación de este tipo de muestras cuya variación temporal requiere de una escala macro. En el lado opuesto, la escala micro, el sampleo y loopeo de microsonidos, logran mimetizar el concepto de síntesis en su forma más ortodoxa, esta forma de síntesis se llamada síntesis granular [14] la cual será estudiada en el apartado 3.6 de este capítulo. Por lo tanto, hay que destacar que los bucles de muestras no musicales serán la principal aportación de esta técnicas a este trabajo.

3.3.2 Pitch-shifting

La otra manipulación que ofrece la síntesis por muestreo es el desplazamiento de tono o pitch-shifting, que consiste en una transposición de la frecuencia fundamental de la muestra. Hay dos maneras de hacerlo: una es mediante el cambio de la frecuencia de reloj en el proceso de conversión y la otra es remuestreando el sonido con una frecuencia de muestreo diferente. Esta técnica nos permitirá ante sonidos musicales abarcar una gran cantidad de notas. Por supuesto, esto produce un estrechamiento o agrandamiento en la longitud de la muestra, por lo que para disponer de una buena cobertura de todas las notas deberemos muestrear el mayor numero de muestras posible, con el consecuente gasto de memoria que conlleva esto. Por otro lado, esta técnica llevada al extremo consigue resultados muy fidedignos de diseño de sonidos poco comunes, por lo que ha sido utilizada hasta la saciedad en el mundo del cine y de la música.

3.4 Síntesis Aditiva

La síntesis aditiva es uno de los primeros métodos utilizados para la obtención de espectros sonoros ricos que se aproxime al comportamiento natural de los sonidos. El análisis acústico de las señales de audio nos enseña que los sonidos naturales están, de hecho, compuestos de una multitud de componentes simples frecuencias (parciales), las cuales, en el caso de los sonidos periódicos con una altura determinada, tienen una frecuencia múltiplo de la frecuencia

fundamental. El concepto de síntesis aditiva apareció en la edad media, aplicado a los órganos que podían comprender un conjunto de tubos para cada altura (acompañamiento por varias quintas y octavas que pertenecen a la serie armónica de la fundamental).

Según el Teorema de Fourier, cualquier forma de onda puede expresarse como una suma de señales sinusoidales a diferentes frecuencias. Si dicha forma de onda es periódica, entonces las frecuencias de las sinusoides son múltiplos enteros de la frecuencia fundamental y entonces el sonido es armónico. Estas dos afirmaciones son la base de la síntesis aditiva.

El problema subyacente es que, para recrear sonidos naturales, necesitaremos una gran cantidad de parciales, con el consecuente coste computacional que esto conlleva. Por otro lado, los sonidos que encontramos en la naturaleza no son sonidos periódicos sino que varían con el tiempo. Esto nos obliga a crear una envolvente temporal manipulable para cada parcial de forma independiente.

3.5 Síntesis Granular

La idea detrás de la síntesis granular la propuso Dennis Gabor a través de la definición de “cuanto sónico” como una alternativa a la descripción temporal de las ondas sinusoidales. Estos “granos” son pequeños trozos de audio (duración < 50 ms), sobre los que se aplica una envolvente de amplitud. Iannis Xenakis fue el primero en aplicar los principios de Gabor en la música, en su obra *Analogique A-B* (1958-9). Él adoptó la idea de que todo sonido es concebido como la unión de un gran número de sonidos elementales adecuadamente dispuestos en el tiempo.

Si trasladamos este concepto micro a una escala un poco más grande (hay que tener en cuenta que estos granos tienen una duración de entre 10 y 50 ms) encontramos en la síntesis granular fundamentos de diseño sonoro que utilizaremos en este trabajo. Tratando el diseño de sonidos naturales bajo el análisis de sus fuentes a niveles elementales. Esta fabulosa técnica de síntesis se caracteriza por presentar cualidades excelentes para la creación de efectos

de sonidos y texturas, debido a la capacidad para crear espectros dinámicos que evolucionan con el tiempo.

Destacaremos que el procesamiento granular fue desarrollado en forma separada por Truax y Roads y que los granos fueron creados a partir de archivos de sonido grabado. Este concepto relaciona en cierta manera el muestreo con esta técnica. Los tipos de síntesis granular se pueden resumir dependiendo de la colocación temporal de los granos:

- **Síntesis Granular Síncrona:** Un flujo de granos con un período constante produce un envolvente periódico, esto es un caso especial de AM lo que produce bandas laterales discretas.
- **Síntesis Granular quasi-síncrona:** Un flujo de granos con un período variable dentro de un rango produce una envolvente no periódica; las bandas laterales se encuentran más espaciadas en frecuencia y el contenido espectral es mucho más rico, pareciéndose a la salida de un filtro resonante.
- **Síntesis Granular Asíncrona:** las propiedades de los granos varían en forma aleatoria, entre ciertos límites, de manera que producen una distribución estadística de los granos.

3.6 Revisión General. Taxonomía de las técnicas de síntesis

Un primer esquema básico de clasificación divide las distintas técnicas entre las que intentan reproducir un sonido existente y las que no. El primer grupo puede ser denominado síntesis “imitativa” (no el sentido de la música concreta, sino en el de reproducir algún sonido ya existente) y el segundo “abstracto”. Esta clasificación puede parecer confusa ya que hay técnicas de síntesis que no están basadas en ninguno de estos criterios y al mismo tiempo satisfacen ambos, como es el caso de la síntesis FM, la cual puede ser utilizada tanto para simular metales o campanas como también para producir sonidos electrónicos “abstractos” que responden a criterios matemáticos no necesariamente musicales. No obstante, esta primera clasificación resulta adecuada para discutir el problema concerniente a la fuente sonora, o el objeto

sonoro según Schaeffer, y nos provee de un claro acercamiento a estas técnicas.

Otro criterio de clasificación muy utilizado divide las técnicas en tres áreas principales: abstracta, modelos físicos y modelos espectrales. En el primer grupo se pueden incluir las técnicas de síntesis más clásicas que pretenden simplemente explorar algoritmos matemáticos para generar sonidos. Es así como en 1977, Moorer [15] ya propone que la forma natural de clasificación debiera considerar técnicas de síntesis aditiva, sustractiva y modulación (también llamada síntesis no lineal o de distorsión), y que en general son usadas en combinación. Esta clasificación es ratificada por Moore en su libro "Elements of Computer Music" [16]. Es importante aclarar que hoy en día la síntesis no lineal no se refiere exclusivamente a la modulación, sino que agrupa un gran número de técnicas basadas en ecuaciones matemáticas con comportamiento no lineal [17].

En el segundo grupo se incluyen todas las técnicas que se basan en el estudio de las propiedades físicas de los instrumentos musicales u otras fuentes sonoras y su posterior implementación y simulación en el computador. Existen autores que sólo toman en cuenta estos dos primeros grupos para clasificación. Por ejemplo, Borin et al. sugiere que la síntesis algorítmica de sonidos se puede dividir en dos clases: Síntesis algorítmica directa clásica, la que incluye transformaciones de síntesis aditiva, granular, sustractiva y no lineal y las técnicas de modelos físicos, en la cual se encuentran la familia de métodos que modelan la acústica de los instrumentos [18].

El tercer grupo consiste en técnicas que pretenden modelar el comportamiento del sonido directamente en su espectro de frecuencias. Este tipo de técnicas de síntesis se basa fuertemente en el mundo perceptual de los sonidos y no en el mundo físico. El sistema auditivo humano puede modelarse como un analizador de frecuencias, de forma muy similar al análisis de Fourier. Es por esto que este tipo de técnicas son más apropiadas para sintetizar sonidos naturales que instrumentales y usualmente requieren de una etapa previa de análisis, antes de la síntesis.

Existen otros esquemas de clasificación bastante divergentes de este modelo tripartito. En los últimos años, se han propuesto clasificaciones más detalladas

y sin grupos tan divisorios o fuertemente delimitados. Es así como en su famoso libro “The Computer Music Tutorial” [19], Curtis Roads propone la siguiente clasificación para la síntesis de sonidos:

- Sampling y síntesis aditiva.
- Síntesis mediante múltiples tablas de ondas, wave terrain, síntesis granular, y sustractiva.
- Síntesis por modulación : en anillo, AM, FM, distorsión de fase y forma de onda.
- Síntesis mediante modelos físicos.
- Síntesis por segmento de ondas, síntesis gráfica y estocástica.
-

En cambio, Dodge and Jerse [20] usan el siguiente esquema de clasificación:

- Síntesis utilizando técnicas no perturbativas: aditiva, AM, modulación en anillo.
- Síntesis utilizando técnicas perturbativas: FM, modelado de onda, síntesis usando fórmulas aditivas discretas.
- Síntesis sustractiva.

- Síntesis para el análisis de datos : voz, STFT, phase vocoder, wavelets.
- Síntesis granular.
- Modelos físicos.

En 1991, Smith propone una clasificación más detallada dividiendo los métodos de síntesis en cuatro grupos: algoritmos abstractos, grabación procesada, modelos espectrales y modelos físicos. Propone además la siguiente taxonomía para las técnicas de síntesis digital [21](ver figura 3.7). En 2004, Bank et al.[22] dividen las técnicas de síntesis de audio en tres grupos, unificando dos grupos de la clasificación propuesta por Smith. El primer grupo es la familia de métodos abstractos. Estos son diferentes algoritmos que generan fácilmente sonidos sintéticos. Métodos tales como FM y waveshaping pertenecen a esta categoría. El segundo grupo, denominado modelo de señales, es el que modela los sonidos de los instrumentos musicales. En este caso, la entrada al modelo es solo la forma de onda o un conjunto de ellas generadas por el instrumento y la física no es tratada en detalle. Los

métodos de síntesis tales como sampling y SMS (spectral modeling synthesis) pertenecen a esta categoría. Este grupo corresponde al procesamiento de sonidos pre-grabados en la taxonomía de Smith.

El tercer grupo, denominado modelos físicos, es el intenta reproducir el comportamiento físico de un instrumento o fuente sonora. Usualmente, los sistemas físicos (tales como una cuerda de un instrumento o la membrana de un tambor) pueden ser descritos resolviendo una ecuación de onda. Dada una excitación inicial (tal como tocar la cuerda o golpear el tambor) y estableciendo las condiciones de contorno para el problema en particular, las ecuaciones pueden ser resueltas y utilizadas como entradas al sistema por lo que la salida se espera sea cercana al sonido original. Un método conocido en esta categoría es la síntesis digital con guíasondas o waveguides, que modela eficientemente ondas unidimensionales y bidimensionales.

Procesado	Modelos espectrales	Modelos físicos	Algoritmos abstractos
Concreta	Tabla de Onda	Cuerdas de Ruiz	VCO, VCA, VCF
Tabla de Onda	Aditiva	Karplus-Strong Ext.	Karplus-Strong
Muestreo	Phase Vocoder	Ondas guías	FM
Vector	PARSHL	Modal	FM con retroalimentación
Granular	Sinosoides + Ruido (Serra)	Cordis-Anima	moldeado de onda
Componentes principales	Componentes principales	Mosaico	Distorsión de fase
Wavelet	Chant		
	VOSIM		

Procesado	Modelos espectrales	Modelos físicos	Algoritmos abstractos
	FM de Risset (bronces)		
	FM de Chowning (voz)		
	Sustractiva		
	LPC		
	FFT inversa (aditiva)		
	Clusters de líneas de Xenakis		

Figura 3.7 Taxonomía de técnicas de síntesis de sonido propuesta por Smith

4

Análisis de la escena

5.1 Procedimiento

En cualquier sistema de diseño se debe precisar un análisis previo con el fin de extraer el mayor número de características posibles de él. Esta es la información que utilizaremos para sistematizar nuestro diseño de la mejor manera posible. Desde este trabajo se propone el análisis de la escena bajo cinco puntos de análisis: Análisis espacial, análisis cinemático, análisis del objeto, análisis del entorno y por último un análisis psicoacústico. En la figura

4.1 se presenta un diagrama de los elementos pertenecientes a cada tipo de análisis según sus niveles de abstracción.

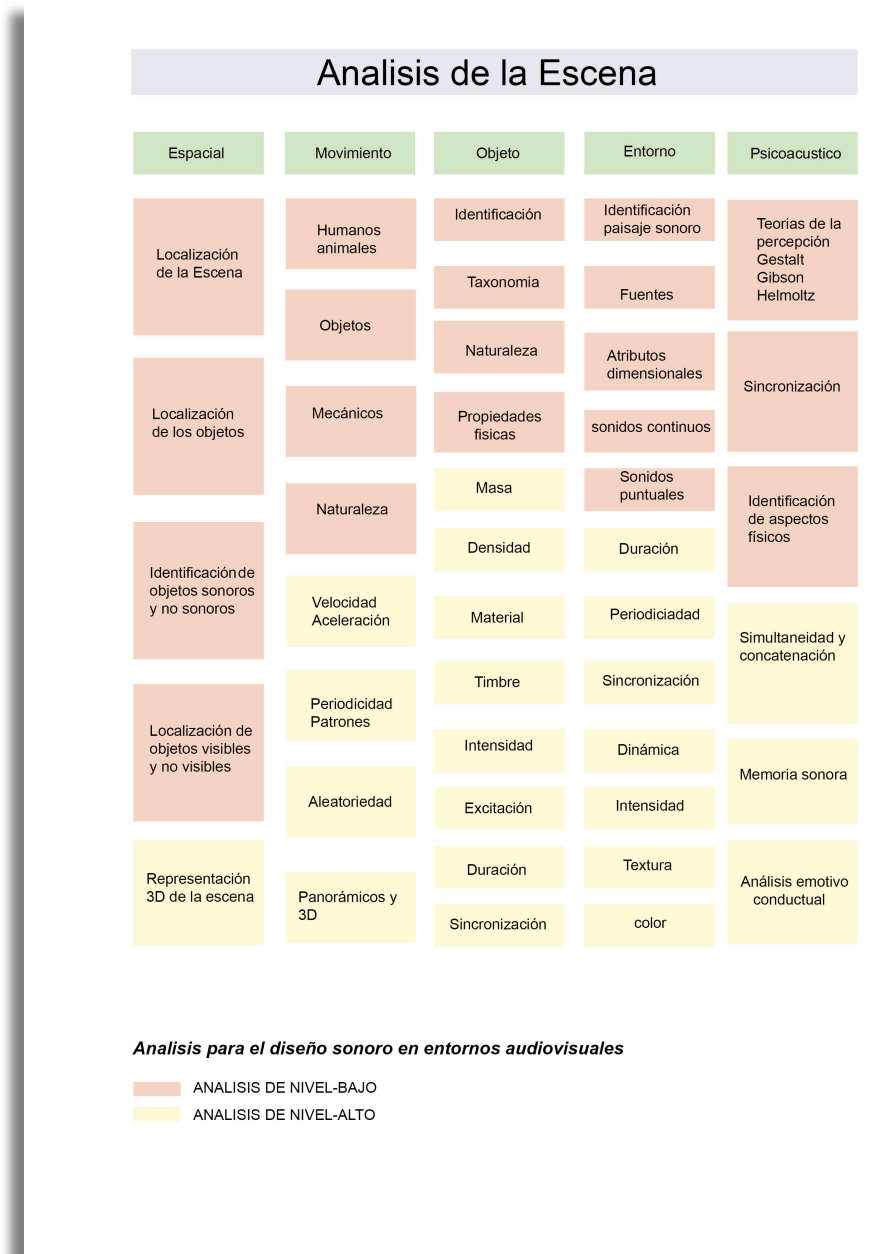


Figura 4.1 Diagrama de análisis de la escena. El diagrama se divide en cinco fases principales de análisis dispuestas en cinco columnas. Cada fase se divide en subfases de análisis más concretas que describen la información de análisis de la escena en orden descendente.

5.2 Objeto sonoro y evento sonoro

El concepto de objeto sonoro fue descrito por Pierre Shaeffer como "todo fenómeno sonoro que se perciba como un conjunto, como un todo coherente, y que se oiga mediante una escucha reducida que lo enfoque por sí mismo, independientemente de su procedencia o de su significado...". Este enfoque fue descrito en torno a la música concreta en los años 50 y 60, y proponía la idea de que todo objeto sonoro estaba formado por diversos parámetros, como timbre, frecuencia, altura, ritmo... Este modo de analizar mediante características describibles y medibles a sonidos que han sido grabados y formateados determina parte del proceso del análisis propuesto en este trabajo. El punto de vista del objeto sonoro propuesto por los músicos concretos no es definitivo y es libre de interpretaciones. En este trabajo se pretende comprender, analizar, sintetizar y diseñar este tipo de objetos, y para ello debemos asumir tres premisas que proporcionarán una base de análisis más clara:

- El objeto sonoro forma parte de un todo. El todo describe al objeto y el objeto describe al todo.
- El objeto adquiere significado completo.
- Un objeto sonoro se puede descomponer en unidades sonoras más pequeñas.

Estas tres premisas nos ayudarán a diseñar sonidos partiendo de la base de que pueden ser descritos desde tres formas diferentes: el plano espacial al que pertenecen, el plano conceptual y significativo que forman y el plano micro del cual se componen/descomponen.

La definición de evento sonoro tiene unas connotaciones más simbólicas o semánticas que la de objeto sonoro. Este atiende a una interpretación cognitiva de la fuente que origina el sonido, proporcionando información acerca de ella. Esta información está relacionada con las condiciones físicas de la fuente.

Es, por tanto, esencial poder disponer de estos dos conceptos con el fin de definir de una manera concisa qué tipo de características son inherentes a cada concepto y proponer un buen análisis.

5.3 Paisaje sonoro y simulación de entornos ambientales inmersivos.

Tanto en el cine, los videojuegos o la realidad virtual, se crean entornos o espacios que necesitan ser recreados para que dispongan de un sentido más completo que el que la imagen por sí sola puede llegar a proporcionar. Partiendo del concepto definido como paisaje sonoro (soundscape) propuesto por R. Murray Schafer [23] podemos hacernos una idea de cómo el entorno sonoro o acústico es esencial para la comunicación o transmisión de información en cualquiera de los campos citados. Hay que tener en cuenta que se entiende como paisaje sonoro tanto al ente acústico en espacios exteriores como en interiores, así como de mundos reales o ficticios. Otra idea de paisaje sonoro es la propuesta por Trevor Wishart [24] que se acerca con su definición a la de evento sonoro descrita en el apartado 5.2. Aquí aparece la idea de paisaje sonoro como fuente imaginaria o conceptual del entorno, proponiendo el concepto de imagen sonora.

A través de estas dos importantes ideas, el plano ambiental se presenta de nuevo desde dos formas de análisis diferentes. Una describirá cada uno de los elementos que componen el espacio acústico y sus cualidades acústicas, y la otra tratará de dar sentido perceptivo global a ese espacio.

Desde este trabajo se propone la condición de dimensionalidad de este concepto. Como ente dimensional presenta diferentes planos perceptivos que deben ser tomados en cuenta a la hora del diseño. Uno de ellos es el plano espacial vertical-horizontal, otro el punto de escucha, y por último el plano territorial. Empezaré explicando este último ya que es el más importante desde el punto de vista cognitivo. Este plano nos permitirá ubicar la escena en un determinado lugar, sea exterior o interior, imaginario o real, de un tiempo futuro o pasado, etc., nos facilita la personalidad de la escena y nos ayuda a describir los elementos internos del paisaje sonoro.

En cuanto punto de escucha, en entornos audiovisuales hay que pensar en la imagen como la que marca ese punto de escucha. No escucharemos más allá, ya que nuestra percepción del entorno está delimitada por su enfoque. Es sólo en entornos inmersivos de RV o videojuegos donde el punto de escucha

puede ser controlado por el usuario o persona encargada de controlar el sistema.

Hay que destacar la importancia del punto de escucha ya que determina la relación existente entre el entorno acústico y la persona. Así un mal diseño del punto de enfoque será un error fatal a la hora de pretender ubicar a la persona dentro de la escena. Imaginemos la la escasa o nula sensación de dinamismo que nos ofrecerá una escena en la que la imagen se va acercando hacia un río y el murmullo de este no se hace cada vez más presente, o la desubicación espacial que tendríamos si se produce una conversación desde otra habitación y no se considera la atenuación-filtro que provoca la pared en la voz.

El último plano a señalar es el espacial vertical-horizontal. Con esto nos referimos a la idea de verticalidad y horizontalidad que los sonidos presentan dentro del espacio sin que el punto de escucha se mueva o cambie.

5.4 Elementos psicológicos, cognitivos y emocionales.

La relación y efectos que el sonido provoca en el cerebro fueron estudiadas por Helmholtz (1863), quien asumía que estaban basados en mecanismos psicológicos asociados al sistema de audición humano. Esta idea influenció las grandes investigaciones musicales, relacionadas con cómo la disonancia, armonía o tonalidad tienen un gran impacto en nuestro aparato sensorial. Otros estudios más recientes como [25][26][27] investigan qué palabras son las más utilizadas para describir cualidades tímbricas del sonido. Además, las teorías de la percepción de los psicólogos de la Gestalt sobre la organización y la cognición dan sentido a las relaciones estructurales que se presentan en la composición o el análisis de patrones o formas. En la figuras 4.2a y 4.3 podemos observar algunas de estas relaciones.

La percepción de la sincronía del sonido con la imagen está ligada a un mecanismo cerebral a través del cual todas las sensaciones tienden a fundirse en un mismo ente con un único significado individual. Esta cualidad ha permitido la incorporación de la música al cine y a los videojuegos. La ley de la continuidad de la Gestalt expresa bien este hecho.

<i>Composición de un tono complejo</i>	<i>Sensación relacionada</i>
Fundamental solo	Suave
Fundamental + 1º armónico	Melosa
Fundamental + Varios armónicos Ancha o Plena	Ancha o plena
Fundamental + armónicos altos	Brillante
Fundamental intenso, armónicos - intensos	Plena
Fundamental - intenso, armónicos intensos	Hueca
Armónico impar predominante	Nasal
Razones de frecuencia 16:15,9:8,15:8,7:5,7:6	Aspera o o de Rechinado

Figura 4.2 Composición de un tono complejo. Helmholtz

Cuando se pierde esta sincronización entre la imagen y el objeto sonoro se crea una rara sensación de desconcierto. Por otro lado la disposición de ciertos sonidos en sincronismo con determinadas partes de una secuencia de video es una atractiva técnica utilizada para crear movimiento, excitación... Otra forma de sincronismo muy importante es la que se produce con elementos que no aparecen es la escena debido a los cambios de cámara o plano, pero que ayudan a mantener la sensación de continuidad.

Cuando analizamos el comportamiento de los sonidos, cadenas y racimos de sonidos observamos que hay una serie de elementos que evolucionan con el tiempo, provocando cierta entonación o ritmos a la escena. Este comportamiento hace referencia a la concatenación. Por otra parte, la cantidad de elementos simultáneos es lo que entendemos como densidad sonora o simultaneidad. La atención que prestamos a los sucesos viene ligada al fenómeno de la concatenación. Esta será más significativa ante sonidos con características de intensidad determinadas. Por otra parte, demasiados sonidos emitidos simultáneamente conforman una masa sonora en la que no podemos diferenciar sus componentes, esto está asociado a las leyes de la proximidad y la semejanza de la Gestalt.

Otro aspecto que debemos de tener muy en cuenta a la hora del análisis y el diseño de sonido es la capacidad cognitiva de cerebro para diferenciar parámetros físicos asociados a los objetos sonoros.

<i>Características del sonido</i>		<i>Se asocian con</i>
Volumen	Sonidos altos	Grande, fuerte, autoritario, poderoso, enérgico, excitante, activo.
	Sonidos suaves	Pequeño, tranquilizante, delicado, poca energía.
Acentuación Agudos y Graves	Agudos realzados	Excitante, ligero, frágil, estimulante.
	Graves realzados	Potente, pesado, profundo, solemne, siniestro.
Calidad tímbrica tonal	Pura, tenue	Pureza, debilidad sencillez
	Rica	Riqueza, grandeza, plenitud, complejidad, ruidoso, fuerza, vitalidad.
	Metálica	Frío, estridente, alegre.
	Prolongada	Dulzura, melancolía, nostalgia.
	Transitorio y percusivos Tono alto Tono bajo	Emocionante, excitante. Dramáticos, poderosos.
Velocidad y ritmo	Lento	Serio, importante, noble, poderoso.
	Rápido	excitante, alegre, ágil, trivial.
	Sencillo	Sin complicación, intencionado.
	Complejo	Complicación, elaboración.
	Constante	Uniformidad, monótono.
	Cambiante	Vigor, irregularidad, incertidumbre, salvaje.
	Aceleración, ralentización	Incremento o disminución de excitación, energía o fuerza.
Cambios de volumen	Tremolo	Incertidumbre, timidez acción inminente.
Duración	Breve	Interés, Excitación, energía.
	Sostenido	Persistencia, monotonía.
Reverberación	Acústica muerta	Intimidad, cercanía, compresión.
	Acústica viva	Apertura, vivacidad, espacio, magnitud, distancia.

Figura 4.3 Relaciones subjetivas emocionales del sonido

Este aspecto es de suma relevancia en esta investigación. Los estudios relacionados con este asunto son varios [28][29][30][31][32][43][44] y demuestran la capacidad que tenemos para distinguir el tamaño de una bola cuando cae al suelo, el material del que está hecha o el objeto sobre el que cae y alguna de sus propiedades; distinguimos la densidad de los líquidos al derramarse y somos capaces de intuir el caudal de un río escuchando su murmullo.

En la figura 4.4 vemos los resultados de un experimento realizado en [28], en el que los encuestados podrían determinar el tamaño y forma de un objeto al caer sobre platos de diferente tamaño.

De este experimento podemos postular las siguientes conclusiones:

- Es posible evaluar métricamente el tamaño del objeto que menos vibra.
- El tamaño del objeto con menos sonoridad estará condicionado por la sonoridad del objeto.
- La potencia del sonido es el descriptor más poderoso descrito por los encuestados.

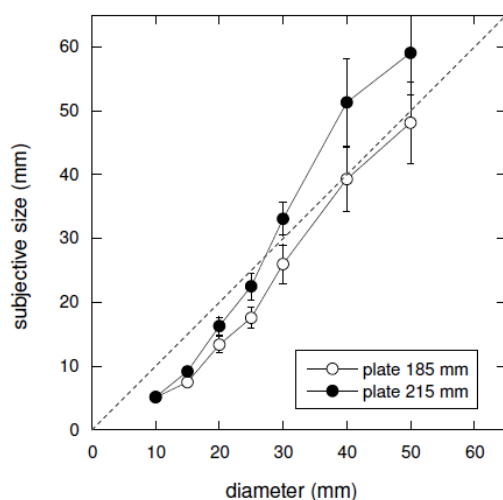


Figura 4.4 Tamaño objetivo en función del tamaño real de una bola. Los puntos blancos representados sobre la línea muestran el tamaño subjetivo de las bola cuando son soltadas sobre el plato pequeño. Los puntos negros representados sobre la línea muestran el tamaño subjetivo de las bola cuando son soltadas sobre el plato grande. Las barras verticales representan la desviación estándar.

Esta capacidad para distinguir formas, materiales y demás tipos de cualidades físicas viene precedida de la asombrosa capacidad del cerebro para memorizar sonidos. Podemos memorizar para siempre el timbre de cientos de voces y los músicos son capaces de tocar de memoria inmensas cantidades de notas sin equivocarse. Así aprendemos a reconocer fuentes sonoras de todo tipo, por lo que a la hora de analizar la escena audiovisual debemos de tener en cuenta si las fuentes a diseñar son codificadas como códigos naturales en nuestra mente y necesitarán ser tratadas de una manera diferente, ya que somos capaces de distinguir las con gran precisión entre otras.

5

Técnicas de Análisis espectral de sonido

5.1 **Eventanado y representación espectral**

La digitalización de las señales de audio proporciona una representación discreta de estas en el dominio del tiempo. Para convertir la señal al dominio de la frecuencia debemos realizar un análisis de la misma, el cual pasa por un primer paso llamado eventanado o segmentación. El uso de ventanas limita

la señal en el tiempo y dota a la misma de una envolvente. Podemos definir la ventana como un tipo específico de función envolvente que selecciona cortos segmentos de tiempo de la señal para el análisis espectral. La duración temporal de estas ventanas varía entre 1 ms y 1 segundo, además los segmentos en ocasiones pueden solaparse.

Tras el proceso de segmentación y enventanado, la señal es analizada y como resultado obtendremos una representación de la distribución de energía en función de la frecuencia. Cada frecuencia es representada mediante su amplitud y fase. Esta visualización característica de la señal en el dominio de la frecuencia es lo que denominamos espectro.

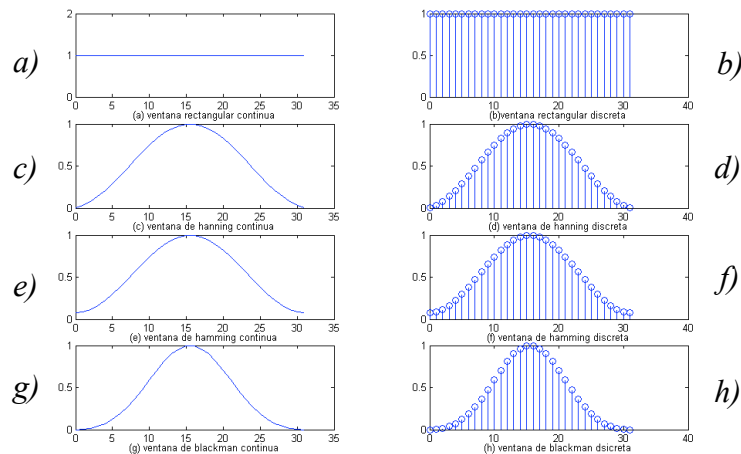


Figura 5.1 (a) Ventana rectangular continua. (b) Ventana rectangular discreta. (c) Ventana Hanning continua (d) Ventana Hanning discreta (e) Ventana Hamming continua (f) Ventana Hamming discreta (g) Ventana Blackman continua (h) Ventana Blackman discreta.

5.2 Perspectiva general de los métodos de análisis

Es a través del análisis frecuencial de Fourier como se ha propiciado el mayor número de técnicas de análisis. Una forma práctica de aplicación de esta técnica a las señales de audio es la Short-Time Fourier Transform (STFT). Esta técnica consiste en segmentar la señal de audio en fragmentos muy cortos a los que se les aplica la transformada de Fourier en su versión computacional (DFT o Transformada Discreta de Fourier), que devuelve para cada

frecuencia asociada a la señal su amplitud y fase. A través de este método se han llevado cabo desde hace más de 4 décadas numerosas implementaciones y tratamientos sobre la señal de audio. A diferencia de otras técnicas de análisis, esta mantiene uniforme el plano tiempo-frecuencia.

El análisis por Wavelets u ondetas, es otro tipo de análisis por inventanado espectral que, a diferencia del STFT, divide el plano tiempo-frecuencia en regiones no uniformes. La región de frecuencias bajas debe ser más grande en el tiempo y más estrecha en frecuencia, mientras que para las frecuencias altas, ocurre lo contrario, es más estrecha en el tiempo y más ancha en frecuencia.

Otros tipos de estimación y análisis que destacaremos son aquellos que actúan sobre la envolvente espectral del sonido. La envolvente espectral es una función en el dominio de la frecuencia que enfrenta amplitud contra frecuencia. La sensibilidad del oído humano a pequeños cambios en la envolvente es la que le otorga la importancia que tiene esta en el análisis de sonido y la resíntesis. Describiremos tres métodos de análisis de la envolvente espectral, por Autoregresión (AR), por el cálculo de Cepstrum, y Cepstrum Discreto.

5.3 Transformada de Fourier de tiempo corto STFT

Los sonidos cotidianos que son analizados en este estudio, no presenta un patrón periódico, es decir estas señales presentan una forma de onda que varía con el tiempo. Esta técnica prepara la señal para ser analizada mediante Fourier, para ello realiza un inventanado sobre la señal. Para el caso que nos ocupa, ya que utilizamos señales discretas, definiremos la STFT como:

$$\text{STFT} \{x[n]\} \equiv X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n} \quad (5.1)$$

Donde, $x[n]$ es una señal y $w[n-m]$ es la ventana. De nuevo, el índice de tiempo discreto m es normalmente considerado como un tiempo "corto" y usualmente no se expresa con tan alta resolución como con el tiempo n . La

salida $X(m, \omega)$ es la transformada de Fourier de la señal enventanada en cada tiempo discreto m para cada banda de frecuencias discreta. La ecuación presenta valores de n desde menos infinito a más infinito, es decir, para una señal de entrada de longitud arbitraria. Para señales discretas prácticas, utilizamos un enventanado de tiempo corto, por lo que los límites de n dependerán de la longitud de la ventana N . Además, k es el índice para cada banda de frecuencia. A través de la siguiente fórmula podemos relacionar la frecuencia correspondiente a cada banda:

$$f_k = k \frac{f_s}{N} \quad (5.2)$$

Siendo f_s la frecuencia de muestreo. De la siguiente relación deduciremos que para una frecuencia de muestreo de 44.1 kHz, un tamaño de ventana N de 1024 muestras, y un índice de banda de frecuencia de $k=1$, el ancho de banda de frecuencia f_k es de 43 Hz.

Al conjunto de datos proporcionados por el análisis de la STFT se le denomina frame o cuadro, cada frame va formando una cadena similar a la que forma una película, de ahí la utilización de la palabra frame. En cada uno de los frames hay representadas dos magnitudes espectrales, una estará formada por la amplitud para cada componente de frecuencia analizada y la otra representa el valor inicial de la fase para cada componente de frecuencia.

5.4 El propósito del enventanado de tiempo corto

Si disponemos de una señal de audio de una duración considerablemente grande (estaríamos hablando de una señal de más de un minuto, hasta 30 o 40 minutos aunque en teoría esta podría ser de cualquier longitud) y esta es analizada mediante la transformada de Fourier, podemos resintetizar el segmento de audio analizado, sólo si es una señal estacionaria. Por lo tanto, ¿cuál es el propósito de la multi-segmentación de la señal en ventanas de tiempo reducido? Destacaré tres razones: precisión de análisis y posibilidad de analizar señales no estacionarias, velocidad de procesamiento y ahorro de memoria.

- Precisión de análisis y posibilidad de analizar señales no estacionarias: Si analizamos una señal de audio como puede ser una canción de unos 5 minutos tendremos como resultado mas de trece millones de puntos en el espectro. Podremos saber qué frecuencias componen la canción pero no cuándo estas ocurren. Esto es debido a que si el segmento de audio que nosotros cogemos es muy grande perderemos precisión en el tiempo. La solución a esto es la micro segmentación. Al usar una ventana de duración muy corta ganamos en precisión temporal a costa de perder resolución frecuencial. Este aspecto será estudiado en sección 5.5.
- Ahorro de memoria: Segmentando la señal en ventanas de tiempo corto estaremos ahorrando memoria de almacenaje a la hora de calcular la transformada.
- Velocidad de procesamiento: Calculando la transformada de segmentos de audio muy pequeños se consigue obtener los resultados mucho más rápido que si lo hacemos de una señal mucho más larga.

5.5 Inconveniente de en la resolución tiempo-frecuencia

El análisis por enventanado espectral está sujeto a una condición de incertidumbre entre resolución temporal y resolución frecuencial. El principio de resolución lineal [33] formula que si queremos una alta resolución en el tiempo estaremos sacrificando resolución en frecuencia. Por lo tanto cuando utilizamos ventana para realizar nuestro análisis tendremos que llegar a un compromiso dependiendo de cuál sea nuestro objetivo, ya que no llegaremos a conocer con precisión una componente de frecuencia y tiempo en el que ocurre. El número de bandas en las que podemos dividir el espectro audible viene determinado por la relación $N/2$, siendo N el número de muestras que conforman la longitud de la ventana. Por lo tanto, para una frecuencia de muestreo $f_s = 44.1$ kHz nuestro espectro audible constará de un rango de 22.05 kHz y, con una ventana de longitud de 1024 muestras o lo que es lo mismo 23 ms, dispondremos de 512 bandas. Cada banda tiene un ancho frecuencial de $f_k = 43$ Hz. Con un enventanado de 1 ms (46 muestras) dispondremos de 23 bandas en el espectrograma, a razón de un ancho de algo

menos de 1000 Hz de ancho de banda por cada una de las bandas representadas.

Las limitaciones de resolución tiempo-frecuencia impuestas por el enventanado hacen necesario llegar a un compromiso o compensación en la resolución del análisis: a mayor resolución en frecuencia menor en tiempo y viceversa. La alternativa es utilizar métodos de análisis multiresolución para poder compensar este problema.

5.6 Análisis basado en la Transformada Wavelet

En este capítulo nos ocuparemos de la llamada transformada Wavelet^[40], herramienta matemática desarrollada a mediados de los años 1980.

Esta transformada es eficiente para el análisis local de señales no estacionarias y de rápida transitoriedad y, al igual que la transformada de Fourier con ventanas, mapea la señal en una representación de tiempo-escala. Preserva el aspecto temporal de las señales. La diferencia está en que la transformada Wavelet provee análisis de multiresolución con ventanas dilatadas. El análisis de las frecuencias de mayor rango se realiza usando ventanas estrechas y el análisis de las frecuencias de menor rango se hace utilizando ventanas anchas.

Las Wavelets, funciones bases de la transformada Wavelet, son generadas a partir de una función Wavelet básica, mediante traslaciones y dilataciones. Estas funciones permiten reconstruir la señal original a través de la transformada Wavelet inversa. La transformada Wavelet no es solamente local en tiempo, sino también en frecuencia.

La transformada Wavelet se presenta como una alternativa a la transformada de Fourier que propone una resolución tiempo frecuencia fija. La TW utiliza ventanas moduladas, pero de dimensión variable, ajustada a la frecuencia de oscilación. Mas precisamente, que mantiene un mismo numero de oscilaciones en el dominio de la ventana. Esto sugiere, naturalmente, contar con una única ventana modulada y generar una completa familia de funciones elementales mediante sus dilataciones o contracciones y traslaciones en el tiempo :

$$\psi(t) \implies \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad (5.3)$$

donde a (siendo distinto de 0) y b son los parámetros de escala y de traslación. Se preserva la energía de las funciones mediante un factor de normalización. La función $\psi(t)$ debe verificar ciertas condiciones de admisibilidad y se denomina wavelet madre, el resto de las funciones generadas, simplemente, wavelets. Denotemos las mismas como :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad (5.4)$$

Entonces, dada una señal $s(t)$, de energía finita la Transformada Wavelet Continua de s se define como :

$$W_{\psi,s}(a,b) = \int_{-\infty}^{\infty} s(t) \psi_{a,b}(t) dt \quad (5.5)$$

5.6.1 La transformada Wavelet discreta (DWT)

Cuando la función $S(t)$ es continua y las wavelets son continuas con factor de escala y traslación discretas, la transformada Wavelet resulta en una serie de coeficientes wavelets, y es llamada la descomposición en series Wavelet.

La función $S(t)$ puede ser reconstruida desde los coeficientes wavelets discretos $W_f(s,\tau)$, de la siguiente manera:

$$S(t) = A \sum_s \sum_{\tau} W_f(s,\tau) \psi_{s,\tau}(t), \quad (5.6)$$

donde A es una constante que no depende de $S(t)$.

A estas funciones wavelets continuas con factores de escala y traslación discretos se las denomina wavelets discretas. Los factores de escala y traslación de las wavelets discretas pueden ser expresados como:

$$s = s_0^i \quad \text{y} \quad \tau = k\tau_0 s_0^i, \quad (5.7)$$

El factor de traslación τ depende del paso de dilatación s . Entonces, a partir de la Ec.5.4 y con la Ec.5.7, las correspondientes wavelets discretas quedan expresadas como:

$$\psi_{i,k}(t) = s_0^{-i/2} \psi(s_0^{-i} (t - k\tau_0 s_0^i)) = s_0^{-i/2} \psi(s_0^{-i} t - k\tau_0) \quad (5.8)$$

A través de la ecuación anterior, la transformada Wavelet de una función continua es realizada a frecuencias y tiempos discretos que corresponden a muestreos con distintas traslaciones (tiempo) y distintas dilataciones (o cambios de escala).

El paso de muestreo en tiempo es pequeño para el análisis utilizando wavelets de pequeña escala, mientras que es grande para el análisis con wavelets de gran escala. La posibilidad de variar el factor de escala s permite usar wavelets de escala muy pequeña para concentrar el análisis en singularidades de la señal. Cuando sólo los detalles de la señal son de interés, unos pocos niveles de descomposición son necesarios. Por lo tanto, el análisis wavelet provee una forma más eficiente de representar señales transitorias. Además sirve para agilizar el proceso, sin tanta memoria requerida, y además se ha encontrado que la eficiencia se puede mantener utilizando escalas diádicas (escalas y posiciones en potencias de 2).

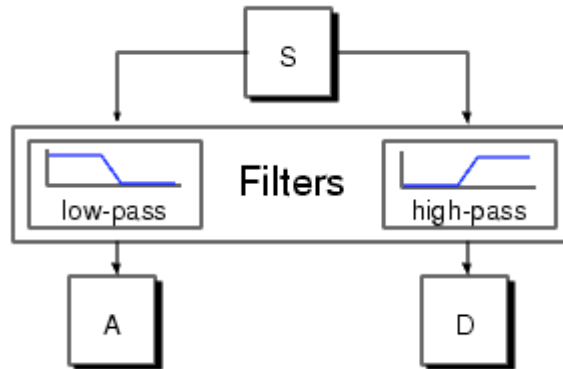


Figura 5.2 filtrado paso alto y paso bajo de la señal antes de aplicar la TW.

Una manera de implementar la DWT es utilizando filtros, lo que lleva a la transformada rápida de wavelets; una caja a la que entra una señal y de la que salen coeficientes. El proceso básico se ve en la figura 5.2.

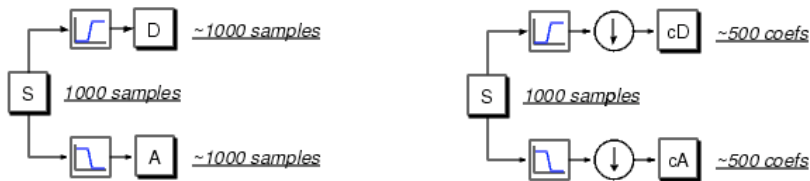


Figura 5.3 Submuestreo.

Mediante un submuestreo (downsampling) (figura 5.3) se elimina la duplicación de los datos. Mientras que el análisis de wavelets involucra filtraje y submuestreo, la reconstrucción involucra sobremuestreo (upsampling) y filtrado. El sobremuestreo (figura 5.4) es el proceso de alargar la señal componente insertando ceros entre muestreos.

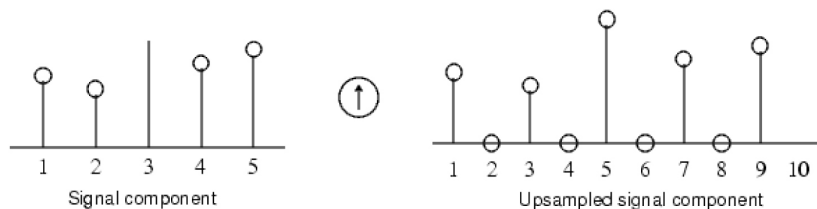


Figura 5.4 Sobremuestreo.

5.7 Métodos de análisis de la envolvente espectral

Una manera de estimar la variación temporal de la envolvente espectral de un sonido es a través del inventanado de tiempo corto de la envolvente espectral $S_i(\omega)$. Hay varios métodos que podemos utilizar [39], de los cuales, unos funcionarían mejor para una aproximación de la señal a componentes sinusoidales y otros métodos presentarían mejores características para componentes no sinusoidales de la señal, caracterizada por ser la parte continua del espectro. La estimación de la envolvente espectral requiere del cumplimiento de algunas propiedades, entre las que destacamos :

- **Exactitud:** La envolvente espectral debe corresponder con exactitud con la máxima amplitud asociada a cada parcial. El grado de precisión requerida viene determinada por las características frecuenciales del sistema auditivo humano, el cual es más sensible a los cambios en la envolvente del sonido para componentes de baja frecuencia que para los de frecuencias altas.
- **Robustez:** El método elegido debe permitirnos abarcar un amplio número de señales, siendo capaz de ser útil para señales con componente determinista estocástica o ambas.
- **Suavizado:** La envolvente espectral requiere de que esta presente una curva relativamente suave, sin muchas fluctuaciones, ya que debe de dar una idea general de la distribución de energía de la señal respecto a la frecuencia.

5.7.1 Autoregresión (AR)

Este método es uno de los más conocidos y ha sido usado para la transmisión y compresión de voz bajo el nombre de codificación por predicción lineal LPC. Este análisis representa cada muestra de la señal $s(n)$ en el dominio del tiempo por una combinación lineal de valores anteriores $s(n-p-1)$ a través de $s(n-1)$. P se considera el orden del modelo AR. El valor estimado de $\hat{S}(n)$ se calcula de los valores anteriores usando los coeficientes AR o coeficientes LPC , a_i :

$$\hat{s}(n) = \sum_{i=1}^p a_i s(n-i) \quad (5.9)$$

Para cada cuadro de análisis, los coeficientes son calculados con el fin de minimizar el error definido por $e(n) = \hat{S}(n) - s(n)$. Cuando la señal residual $e(n)$ es minimizada, un filtro de análisis dado por la función de transferencia de la transformada Z :

$$A(z) = 1 - \sum_{i=1}^p a_i z^{-i} \quad (5.10)$$

intenta atenuar los componentes frecuenciales de la señal de entrada $s(n)$ que presentan la mayor amplitud para realizar un ajuste lo más fino posible.

Uno de los inconvenientes que presenta este método es que para señales con un número poco acusado de parciales dominantes, la señal decae entre los huecos que dejan estos parciales y baja hasta el nivel de ruido residual.

5.7.2 Cepstrum

Consideraremos la señal como la convolución de una señal que será la fuente por la respuesta al impulso de un filtro, este filtro podemos considerarlo como la señal de envolvente espectral que necesitamos:

$$s(n) = h(n) * x(n) \quad (5.10)$$

y en el dominio de la frecuencia esta convolución puede verse como la multiplicación de las respectivas transformadas de Fourier:

$$S(\omega) = H(\omega) \cdot X(\omega) \quad (5.11)$$

Aplicando logaritmos a la ecuación:

$$\log |S(\omega)| = \log |H(\omega)| + \log |X(\omega)| \quad (5.12)$$

Se define el CEPSTRUM (CS) real de una señal como la transformada inversa de Fourier (IFT) del módulo del espectro en escala logarítmica (en belios) de esa señal, es decir:

$$c(t) = I/F[\log(S(\omega))] \quad (5.13)$$

De la ecuación se concluye que el CS de una señal es la suma del CS de la excitación y el CS del filtro (precisamente la respuesta impulsiva del filtro). Las bajas componentes cepstrales corresponden a variaciones lentas de las componentes espectrales y por tanto contienen información de la envolvente del espectro, la cual se relaciona con la respuesta en frecuencia del filtro que modela la señal fuente.

Este método presenta el mismo inconveniente que el análisis por autoregresión. Además este método lleva a cabo un filtrado paso bajo de la magnitud del espectro interpretada como señal, promediando las fluctuaciones de esta curva, el efecto se traduce en una caída de la curva después de cada pico.

Una manera de solucionar los problemas inherentes en los dos métodos anteriores es mediante un análisis cepstral discreto, este calcula puntos discretos espaciados no uniformemente en el dominio de la frecuencia. Estos puntos corresponden con los picos del espectro. Este análisis une los picos de la señal mediante una curva muy suavizada evitando los problemas que generaban los métodos anteriores.

6

Modelo intuitivo de Análisis/diseño basado en las características físicas de la fuente

6.1 Introducción y planteamiento

La complejidad presente en el diseño de sonido en entornos audiovisuales e inmersivos viene condicionada a la necesidad de ofrecer el mismo realismo en el plano sonoro como el que ofrece la imagen. Es de esta de donde extraemos la información que nos permitirá definir los parámetros de diseño. El modelo

que se presenta en este documento viene sujeto a una etapa de análisis de la escena. Los objetos sonoros que aparecen en la escena deben ser clasificados y descritos. Estos hacen referencia a tres planos:

1. El plano de la física del objeto sonoro
2. El plano físico-acústico
3. El plano cognitivo

La segunda etapa es la de diseño, conformada por un sistema de síntesis híbrida, basada en técnicas de resíntesis, además de los conceptos de diseño que propone la síntesis granular y la síntesis concatenativa. En esta etapa se aplican técnicas de procesamiento de la señal a los sonidos obtenidos a través de la síntesis y que ayudan a conformar un diseño más complejo.

La técnica para diseñar sonido en este tipo de entornos es la grabación de muestras, pero esta técnica carece de la capacidad de transformación de sonido para adaptarse a la imagen, lo que hace que en muchos casos sea demasiado costosa y poco reutilizable. Por otro lado, las técnicas de modelos físicos son capaces de aproximar el comportamiento acústico y mecánico de un instrumento. El proceso pasa por dos fases:

- Fase de análisis: es la encargada de estudiar el comportamiento del instrumento, resonador, cuerda, membranas u objetos capaces de vibrar y transmitir un sonido audible.
- Fase de implementación o síntesis: se encarga de aproximar mediante modelos matemáticos el comportamiento sonoro del objeto.

Esta técnica es capaz de producir sonidos realistas a la vez que puede transformar las dimensiones físicas de los objetos o instrumentos, provocando frecuencias antes incapaces de conseguir. Un ejemplo de esto sería la implementación de un tambor cuya membrana tuviese un diámetro de 100 m. El análisis e implementación de las características físicas de objetos se debe de plantear como una alternativa a la grabación de muestras. Sin embargo, la numerosa cantidad de objetos sonoros que pueden aparecer en una escena y no digamos en una película, videojuego, etc., harían necesario el modelado de

cientos de objetos resonantes. Una buena aproximación es la que se expone en [5].

Las dos técnicas expuestas en las líneas anteriores presentan inconvenientes para el caso de estudio de este proyecto. A su vez, la grabación y digitalización de muestras de audio se puede plantear como una alternativa al modelado matemático de cientos de objetos sonoros, siendo el análisis espectral, transformación y resíntesis de muestras digitalizadas de audio, la técnica planteada en este proyecto para modelar los objetos en relación a sus propiedades físicas.

La arquitectura del modelo se presenta en la figura 6.1. Está formada por el módulo de imagen, módulo descriptor, módulo de síntesis, módulo de procesado, bases de datos y módulo de decisión de parámetros.

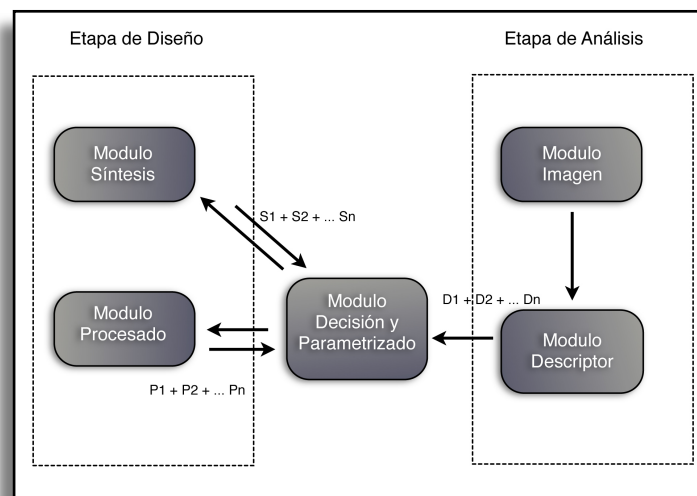


Figura 6.1 Arquitectura del modelo.

El nexo de unión entre la etapa de análisis y la de diseño es el módulo de decisión y parametrizado. Este determina mediante un mapeo en dos fases:

1. Parámetros de control y transformación.
2. Tipo de diseño que más se ajusta a las necesidades propuestas.

Estas decisiones son propuestas a través de un análisis ontológico derivado de la taxonomía sonidos cotidianos analizada y que es el motor de análisis del módulo descriptor. Es debido a este modulo por lo que este trabajo se presenta como modelo intuitivo de diseño de sonido, destacando que no se plantea como una técnica de síntesis de sonido. En la figura 6.2 se presenta al estructura de flujo de datos del sistema.

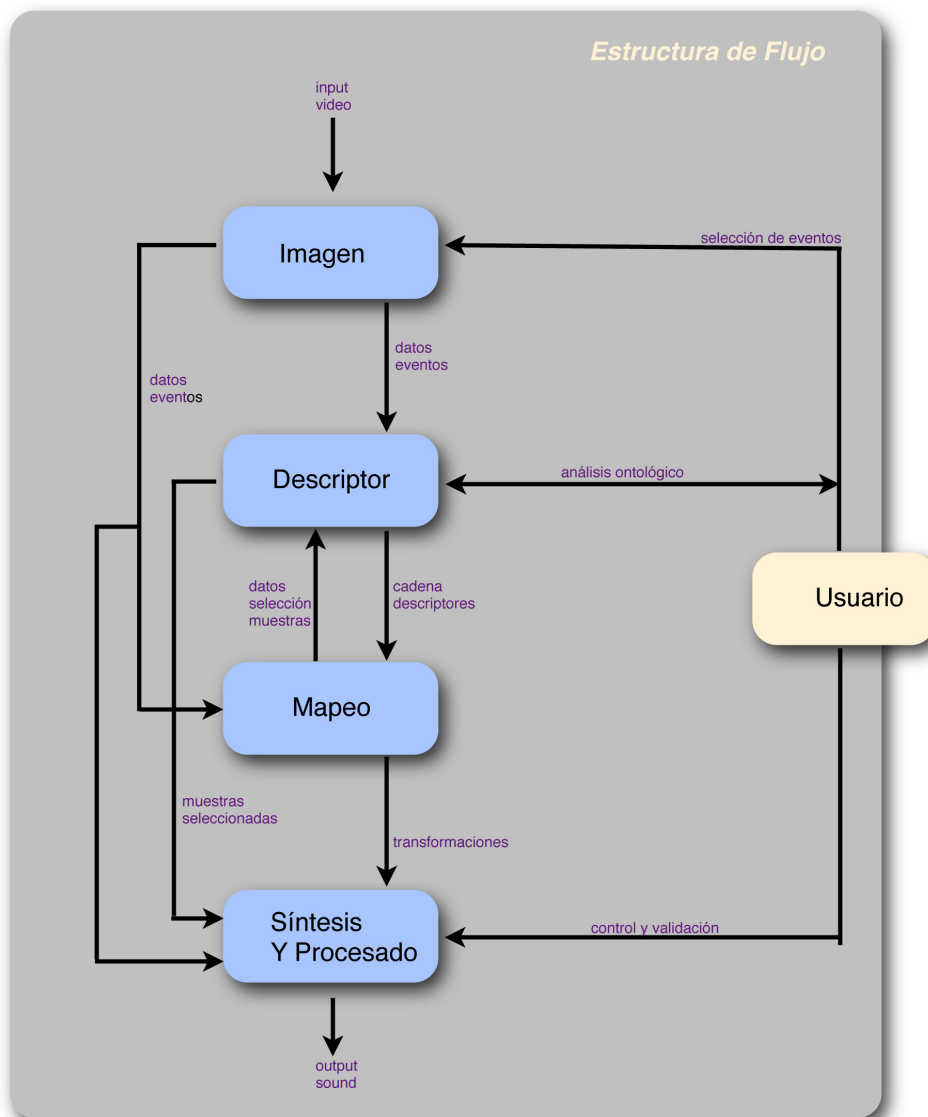


Figura 6.2 Estructura del flujo del sistema

6.2 Módulo de Imagen. Arquitectura

Este módulo está dedicado a la recepción de la señal de imagen y a la localización y marcación de los elementos críticos. La señal de imagen Input, contiene la información de tiempo que se utilizará para la correcta sincronización audio-imagen. La sincronización temporal se realiza mediante un sincronizador de base de tiempos, que nos permite trabajar en diferentes formatos, como Frames por segundo, tiempo (ms). Esta señal de reloj se utiliza para determinar la duración del audio así como la localización de dimensiones, segmentos y unidades.

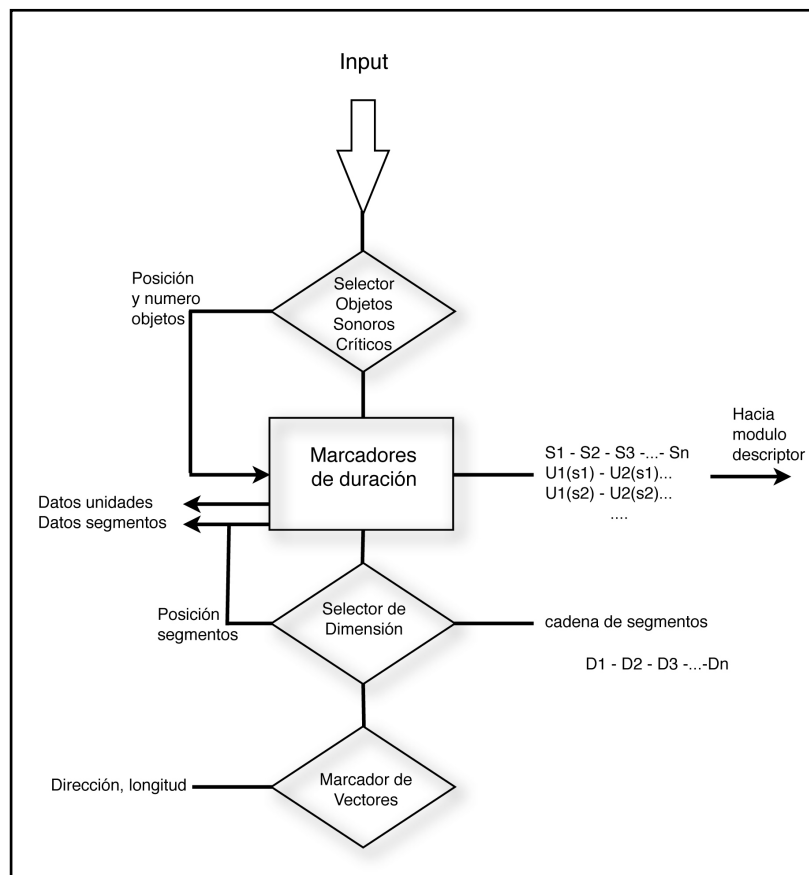


Figura 6.3 Diagrama de funcionamiento del módulo de imagen.

El análisis de la escena comienza con la descomposición de esta en estados, cada estado proporcionará un tipo de información descriptiva acerca del objeto sonoro seleccionado. La figura 6.3 representa el diagrama de funcionamiento del modulo de imagen. Podemos diferenciar los siguientes estados en el sistema:

1. Selección de los objetos sonoros críticos. Esta acción viene determinada por el usuario quien se encarga del proceso de selección bajo criterios artísticos sujetos al guión o a la subjetividad del mismo. Estos objetos sonoros pueden aparecer en la escena de dos maneras:
 - Directa: relaciona visualmente la imagen con el objeto sonoro.
 - Indirecta: el sonido aparece en escena pero no en la imagen. El análisis de este se debe hacer mediante la contextualización de los elementos colindantes.
2. Marcadores de duración. Mediante el uso de marcadores de tiempo seleccionamos segmentos, a su vez, cada segmento es dividido en unidades. Un segmento se define como un espacio sonoro completo que está dentro de una dimensión sonora perteneciente a la escena. Por unidad entenderemos cada una de las subdivisiones en las que podemos descomponer un segmento. Cada unidad puede guardar relación con segmentos contiguos o con otras unidades del mismo segmento al que pertenecen, así como la capacidad de superposición espacial. En esta acción se devuelve la duración de cada segmento y unidad.
3. Selector de dimensión. Los segmentos son agrupados en dimensiones siguiendo una serie de criterios propuestas por los psicólogos de la Gestalt y por Michel Chion [34] y James Gibson [35] sencillos de agrupación y percepción. Podemos diferenciar entre planos o dimensiones:
 - Plano diegético y extradiegético. El sonido diegético es el que es generado en la escena, por ejemplo en una película el sonido de dos objetos al golpearse, pasos de los personajes, etc. El extradiegético es

aquel que no es producido en la escena, como ciertos efectos especiales de sonido, música, etc.

- Proximidad. Los elementos que están próximos entre sí, se agrupan conjuntamente.
- Similitud. los elementos que parecen similares se agrupan entre sí.
- Destino común. Los elementos que parecen moverse conjuntamente se agrupan entre si.
- Localización espacial. Los elementos que pertenecen a un mismo espacio se pueden agrupar.
- Sonido Ambiente. Sonido ambiente que rodea una escena y habita su espacio.
- Foleys. Sonidos de sala.
- SFX. Sonidos de efectos especiales

En la figura 6.4 se muestra un ejemplo de marcación y selección de de los elementos que determinan el funcionamiento del sistema.

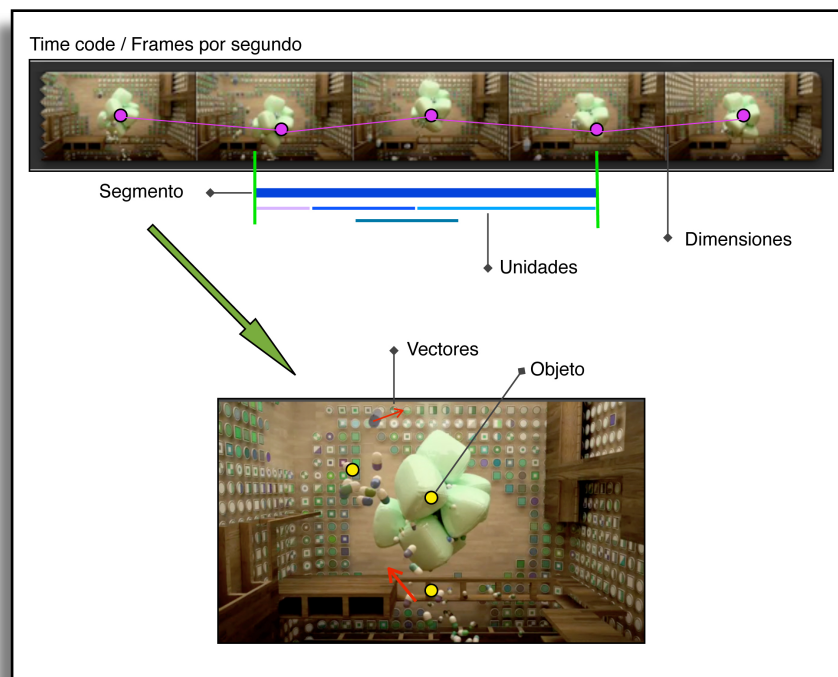


Figura 6.4 Fotogramas sobre los que se representan las unidades, los segmentos y las dimensiones.

4. Marcador de vectores de movimiento. Los objetos sonoros que se mueven en la escena son identificados mediante un vector, cuya dirección determina la dirección del movimiento del objeto dentro del espacio tridimensional de la escena. Este proceso está sujeto al tipo de formato multicanal que estemos utilizando o bien, a un sistema estéreo.

6.3 Módulo descriptor

Este módulo está formado por una base de datos de descriptores destinados a etiquetar dimensiones, segmentos y unidades. Estos descriptores están conformados en tres categorías principales, la primera basada en la física de la fuente, la segunda en la física-acústica del sonido emitido y la tercera en la psicoacústica y características cognitivo-perceptivas. Cada dimensión, segmento y unidades son descritas mediante una cadena de descriptores que se unen para formar cadenas más complejas. Estas cadenas guardan sentido completo por sí solas, pero a su vez comparten características con otras cadenas, siempre y cuando estas cadenas formen parte de una misma clase jerárquica, por esto, entendemos que cadenas de descriptores de unidades pertenecientes al mismo segmento pueden relacionarse entre sí y, a su vez, las cadenas de segmentos pertenecientes a la misma dimensión, también guardarán características comunes. Esta identificación de elementos comunes entre varias cadenas facilita la aplicación de los mismos parámetros de diseño a diferentes elementos sónicos que deben guardar relación en común, como puede ser el procesado de reverberación que se le debe aplicar a ciertos objetos que estén en el mismo espacio.

6.3.1 Descriptores basados en la fuente-objeto

Los descriptores basados en la fuente-objeto proporcionan información sobre el tipo de interacción entre objetos, así como atributos físicos que condicionarán el diseño del objeto sonoro. Como estamos actuando en el campo de los entornos audiovisuales, la información debe ser extraída directamente de la escena visual, entendiendo que la escena visual caracteriza y condiciona al

conjunto a los objetos de la escena. El procedimiento utiliza un análisis ontológico en racimo para formar las cadenas de descriptores.

Los descriptores provienen de una categorización primaria propuesta por W. Gaver sobre el tipo de interacción entre materiales. Estos son clasificados en eventos vibratorios, sonidos aerodinámicos y sonidos de líquidos.

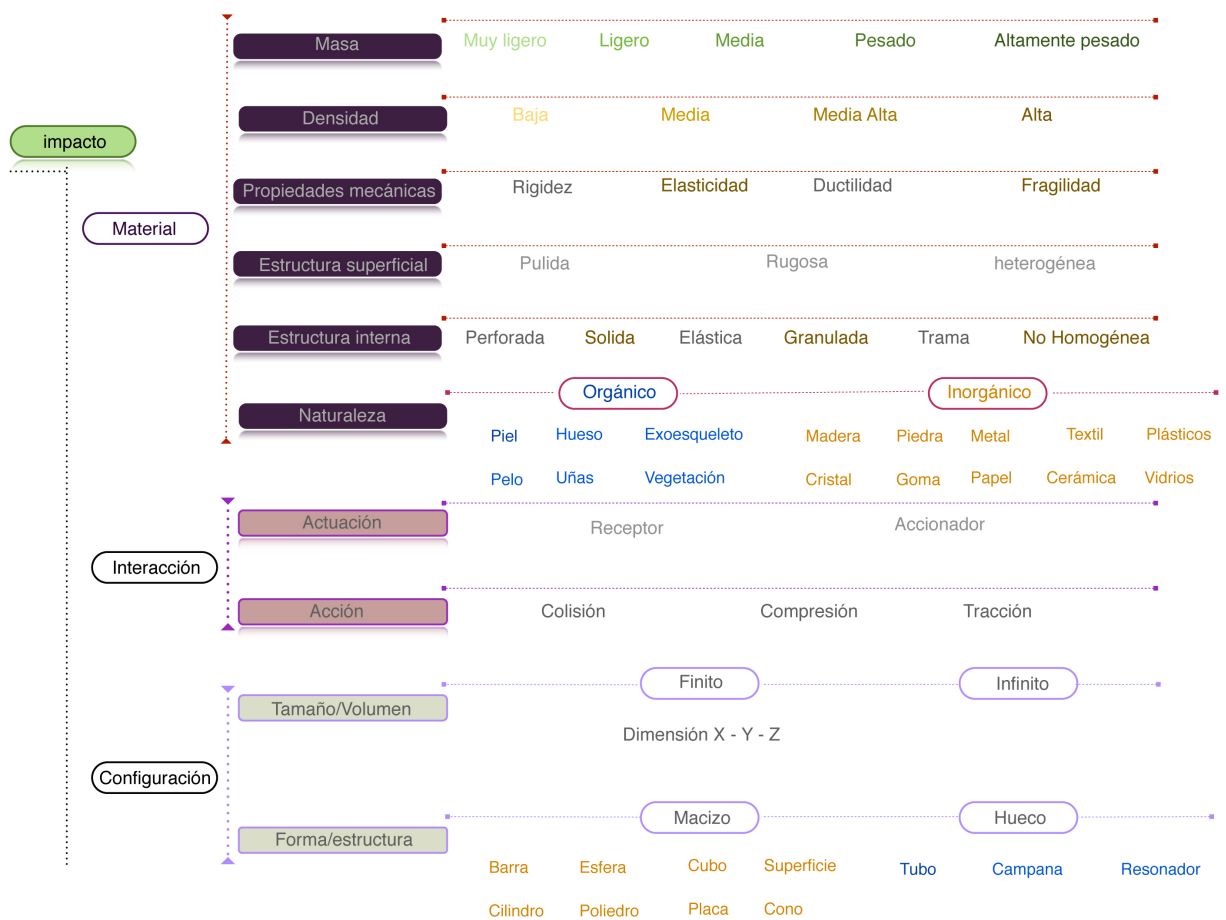


Figura 6.5 Taxonomía propuesta para sonidos de impactos

6.3.2 Eventos sonoros provocados por objetos vibratorios

Estos eventos sonoros son provocados por la interacción de fuerzas entre dos objetos que provoca la deformación del sistema y trasforma esta en energía potencial. El sistema trata de volver a su equilibrio provocando una oscilación que es traducida al aire en forma de vibración.

En este estudio se proponen dos tipos de eventos vibratorios, impactos y rozamientos, descritos en las figuras 6.5 y 6.6, los cuales son escogidos por ser capaces de describir un número muy elevado de eventos sonoros primitivos.

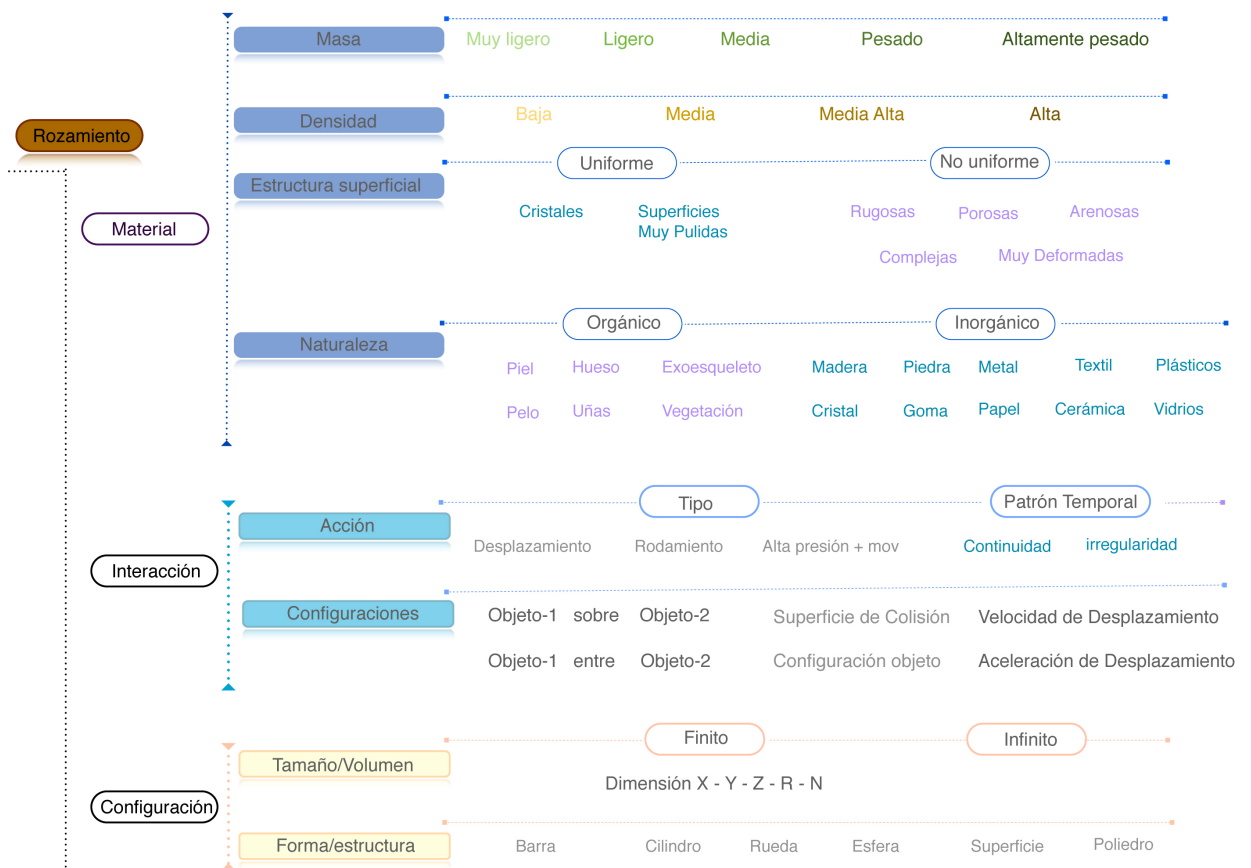


Figura 6.6 Taxonomía propuesta para sonidos de rozamientos

Los atributos físicos escogidos son aquellos que según el tipo de evento, impactos o rozamientos, son útiles para definir un patrón sonoro del objeto a diseñar. Estos son: tipo de material, tipo de interacción y configuración.

6.3.3 Eventos sonoros provocados por líquidos

A diferencia de los eventos sonoros provocados por la deformación de un objeto que entra en vibración y transfiere su energía al aire, los sonidos provocados por líquidos se producen gracias a las cavidades resonantes que se forman en la superficie de los líquidos.

Un claro ejemplo de esto lo encontramos cuando un objeto cae a un líquido, este se deforma y crea una cavidad antes de volver a su estado inicial, es la energía liberada en el proceso de formación y deformación de esta cavidad la que se transfiere al aire y podemos oír. Como se puede deducir lo que nuestro cerebro entiende por sonidos de líquidos no son sino deformaciones en la estructura de estos que crean cavidades de aire, como se muestra en la figura 6.7.



Figura 6.7 La imagen refleja las cavidades formadas por un líquido al recibir un impacto y deformar su estructura superficial.

Además en la mayoría de los casos, estos sonidos están formados por cientos de cavidades para dar lugar a un evento sonoro. Este se caracteriza por tener

sentido completo y único y ser, a su vez, indivisible. Es el caso del flujo de un riachuelo, el cual asimilamos como el sonido del agua fluyendo a través de un canal a una velocidad determinada. Si analizamos este evento sonoro lo podemos identificar rápidamente con las deformaciones y perturbaciones que sufre el agua por toda su superficie, creando, como citábamos anteriormente, cavidades resonantes de distinta consideración, que en su conjunto identificamos como un riachuelo o agua canalizada.

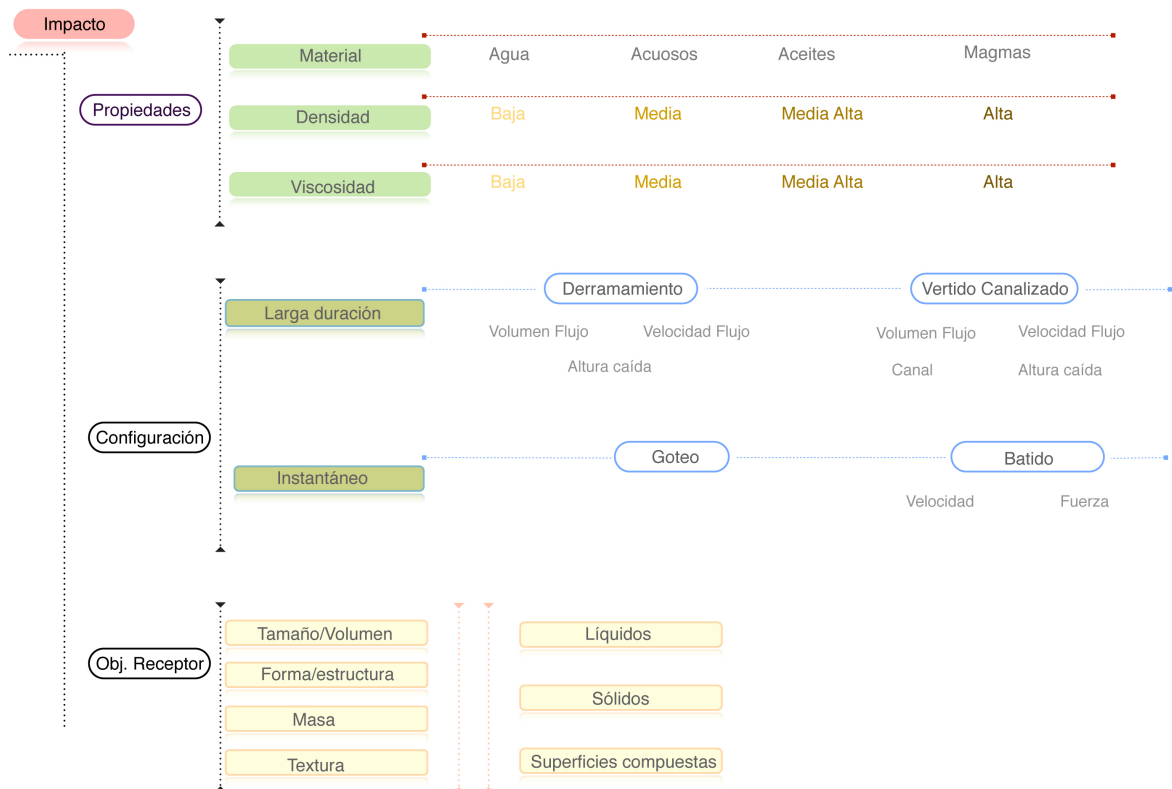


Figura 6.8 Taxonomía propuesta para sonidos de impactos en líquidos

Otro aspecto importante que caracteriza a los sonidos de líquidos es su variación temporal, la cual es determinante para diferenciar unos sonidos de

otros. En el ejemplo propuesto anteriormente del riachuelo decíamos que está formado por cientos de cavidades resonantes, estas serán más o menos grandes y se abrirán más o menos rápido liberando energía. Serán cientos o miles de ellas dependiendo de si nos encontramos ante un río con mucho caudal o en el que el agua fluye muy rápido. Pero la cualidad que nos proporciona la idea o imagen de que se trata de un riachuelo nos la da la variación temporal y su dinámica en el tiempo.

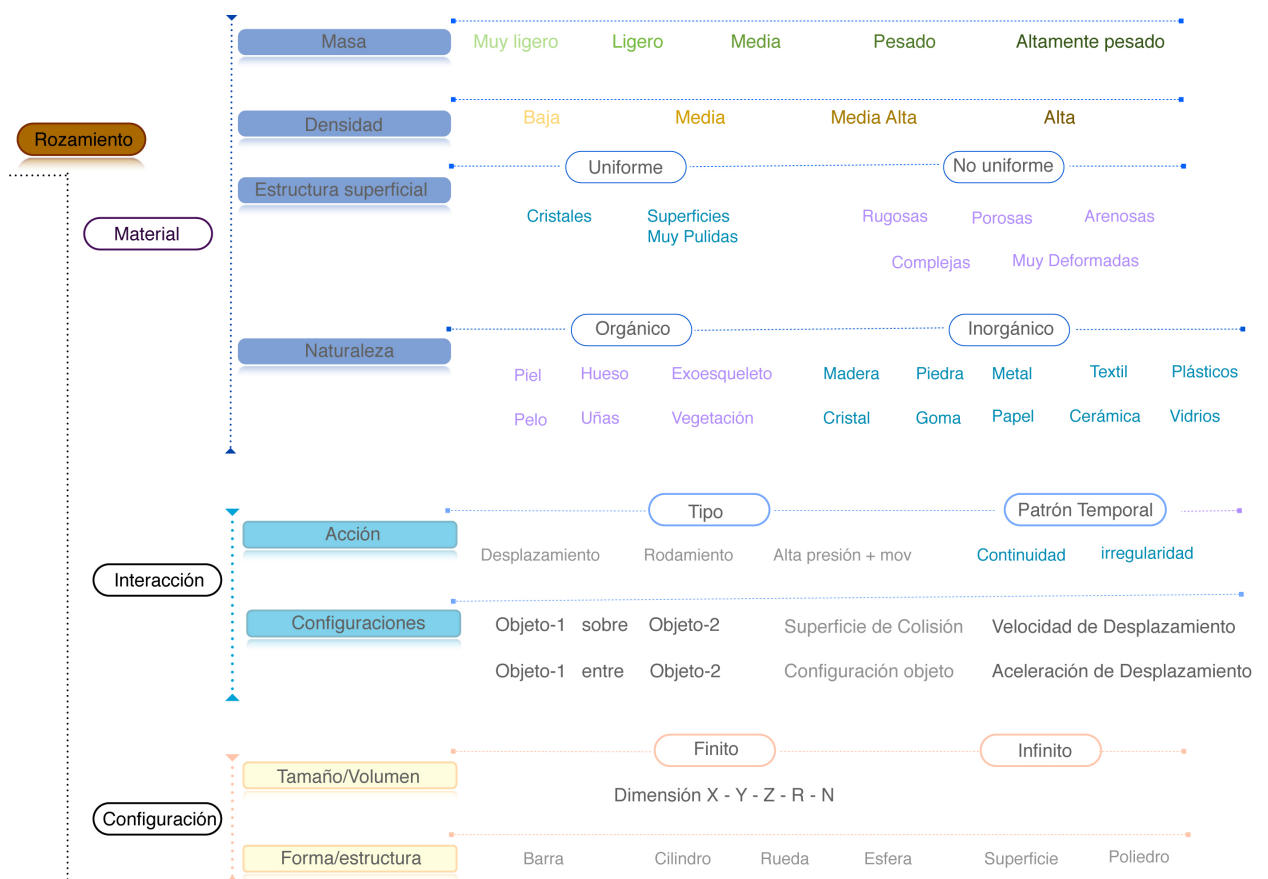


Figura 6.9 Taxonomía propuesta para sonidos de flujo de líquido

Es por esto que en nuestro análisis partiremos de dos grandes grupos de sonidos de líquidos:

- los producidos mayoritariamente por impactos (figura 6.8)
- y los que se caracterizan por el flujo de líquidos con continuidad temporal (figura 6.9).

6.3.4 Eventos sonoros aerodinámicos

Este tipo de evento sonoro se produce cuando un gas o, en el caso más común, en el aire, hay un cambio de presión, una variación en la presión que propaga una onda que, si se produce a una velocidad suficiente, puede ser oída. Un ejemplo simple lo encontramos cuando un globo explota, el aire que hay dentro de él se libera rápidamente a la atmósfera. Las explosiones dan lugar a una gran variedad de sonidos aerodinámicos. Por otro lado se han analizado dos tipos de eventos más, los sonidos aerodinámicos continuos y los que se producen por simpatía (ver figura ext.2, colocada al final de la memoria):

- Los primeros son aquellos derivados de la acción de un elemento que mueve el aire haciendo que este varíe su presión y sea audible, como por ejemplo un ventilador, choque del aire con superficies o cavidades resonantes o con objetos en movimiento, como un coche a gran velocidad.
- En los sonidos aerodinámicos que se producen por simpatía, no es el gas el que produce el cambio de presión necesario para ser escuchado, sino que es el aire el que transmite su energía a un elemento que puede entrar en vibración gracias a la acción del aire, como puede ser un cable tensado o una placa ligera.

6.3.5 Complejidad temporal de los eventos sonoros

Cuando analizamos un sonido proveniente de la naturaleza casi nunca se trata de un sonido aislado, sino que identificamos un evento sonoro como parte de una composición de sonidos primitivos que forman un patrón temporal

determinado. Es esta disposición temporal la que nos otorga significado. Es el caso de las pisadas de una persona, sin un patrón temporal que nos indique que se trata de una persona caminando nos resultaría complicado distinguirlas de simples impactos contra el suelo.

Por lo tanto, tan importante es la identificación física de la fuente como el patrón temporal de un cierto evento sonoro. Cuando un balón cae al suelo lo identificamos porque la sucesión de los impactos de este contra el suelo genera un patrón temporal amortiguado que rápidamente es identificado como tal.

La combinación de sonidos primitivos compuestos y dispuestos en un determinado orden son otro ejemplo de patrón temporal complejo, debido a que su concatenación o su simple orden de aparición temporal proporciona una clara identificación del objeto sonoro. A su vez, estos eventos compuestos pueden presentar patrones en su conjunto, que aportan información adicional o determinante sobre el objeto. Un ejemplo es el sonido producido por una ventana abriéndose y cerrándose, primero oímos el rozamiento de las bisagras, luego silencio, otra vez las bisagras y el impacto de la ventana contra el marco, es una composición temporal de varios sonidos primitivos, pero todos pertenecientes al mismo objeto, que es la ventana. En su conjunto lo identificamos con que la ventana se ha abierto y cerrado sola. Si ahora repetimos el evento con un patrón de rebote, en el cual en cada rebote modificamos intensidad y forma de cada evento primitivo asociado a este patrón, podemos transformar el significado de ventana a otro: una ventana en un día con mucho viento.

La figura 6.10 muestra el análisis de un gran número de eventos temporales. Estos son divididos dependiendo de si son simples, si parten de un mismo evento sonoro primitivo, o son compuestos por la concatenación o superposición de varios eventos primitivos.

Para los eventos temporalmente complejos (ETC) compuestos, se debe identificar cuáles serán los atributos modificables en cada evento sónico primitivo, posteriormente actuará siguiendo un patrón simple.

Del análisis de la figura 6.11 sobre patrones postergados es posible ejemplificar la metodología expuesta. Mediante este análisis conseguimos extraer suficiente información para el diseño de algoritmos que creen patrones

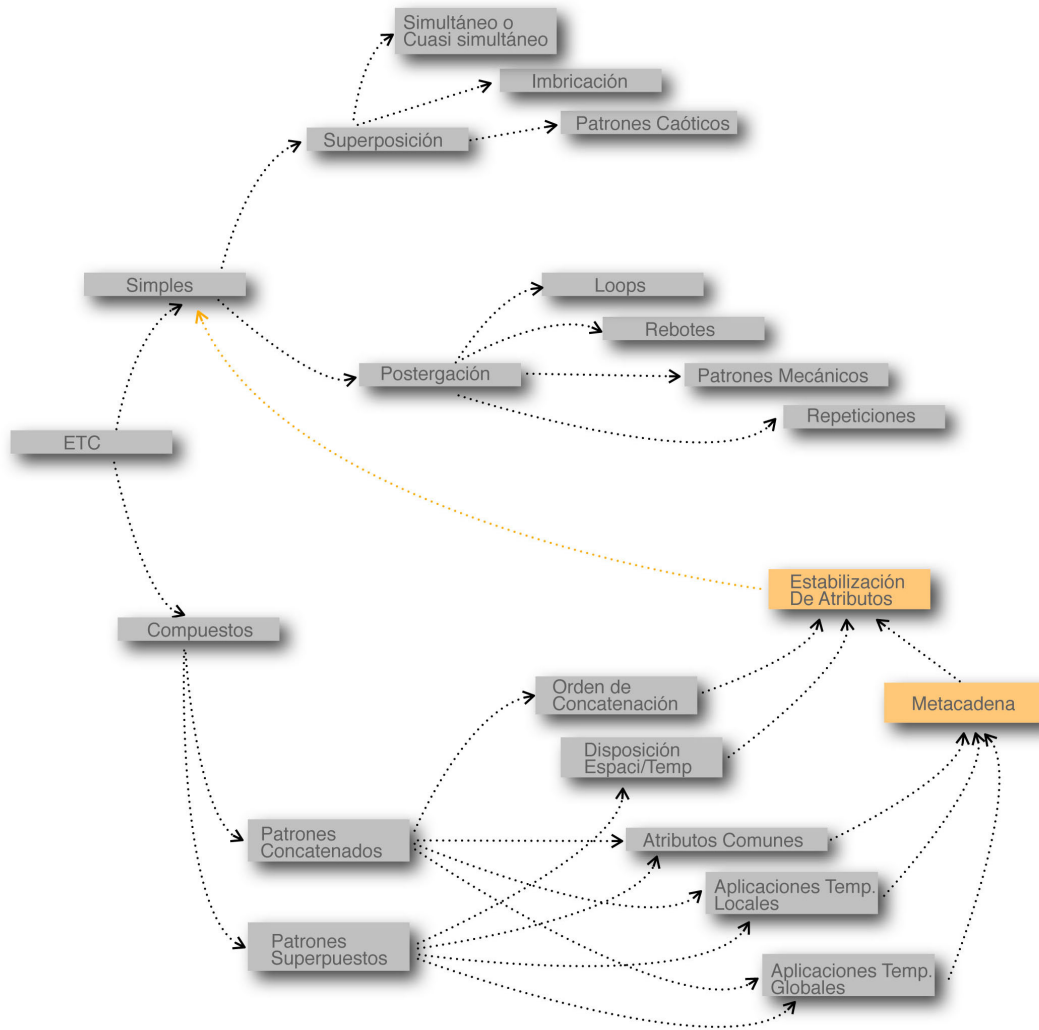


Figura 6.10 Taxonomía propuesta para ETC.

temporales automáticos que se adapten a las necesidades de la imagen, sin la necesidad de una costosa edición de audio.

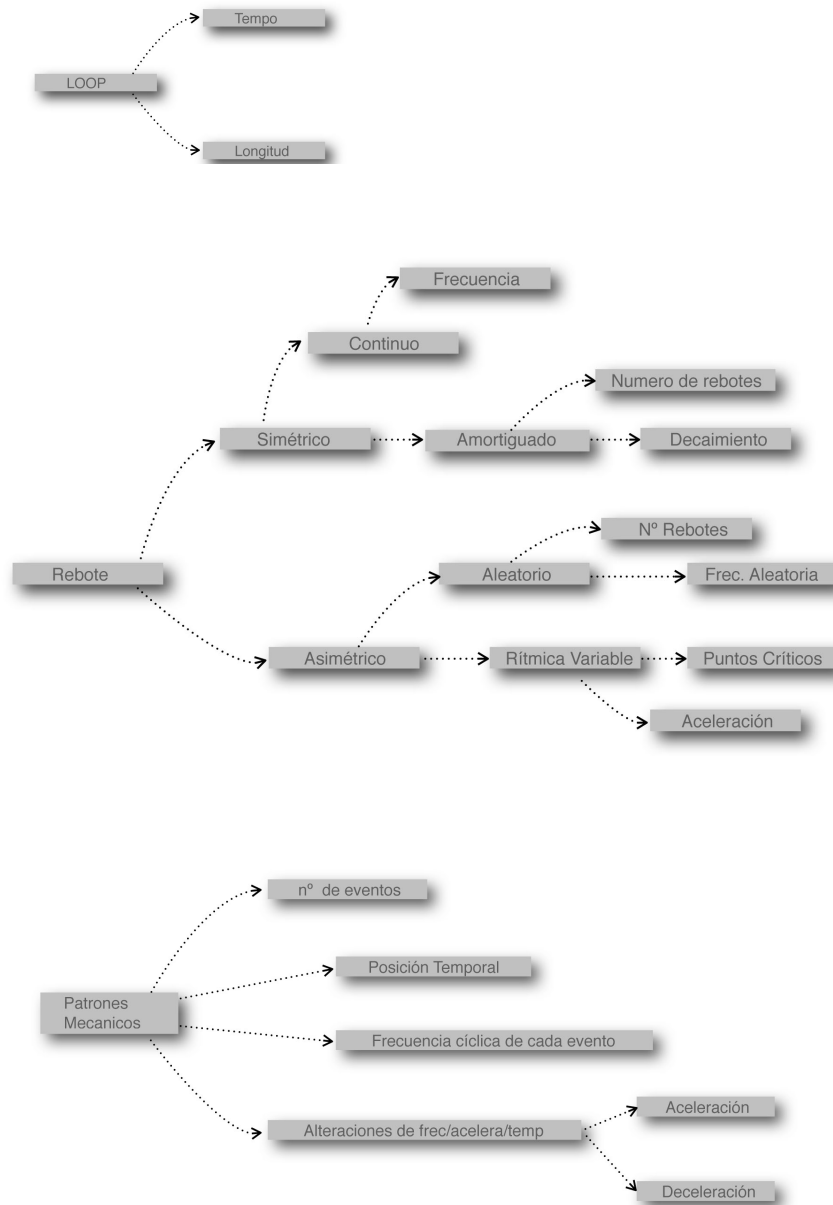


Figura 6.11 Taxonomía propuesta para patrones por postergación.

6.3.6 Análisis ontológico

Se ha propuesto un análisis ontológico derivado de las aproximaciones propuestas en este trabajo para el análisis de la escena. Ha sido desarrollado un mapa conceptual de análisis expuesto en la figura 6.12 (esta figura seta situada al final de la memoria). La estructura del mapa ha sido creada bajo tres conceptos:

1. Concepto ontología: La ontología o tipología de un sistema estudia qué entidades, relaciones y clases de entidades y relaciones hay en su dominio. Una ontología adecuada es capaz de expresar todo lo relevante de un dominio de forma práctica. Entidades y relaciones complejas pueden construirse a partir de las entidades más básicas y primitivas. Los objetos pueden cambiar en el tiempo, por lo que son capaces de adquirir propiedades tales como el crecimiento o su propia generación. Los sistemas pueden influirse o relacionarse de muchas maneras distintas, originando todo tipo de procesos.
2. Concepto análisis: El análisis del sistema se divide en partes más elementales y que se estudiarán localmente, a bajo nivel, con el máximo detalle posible. El diseño mostrará cómo el comportamiento global del sistema será el resultado de las interacciones básicas entre sus elementos. Analizar un sistema es descomponerlo y estudiar sus elementos y relaciones. Diseñar un sistema es construirlo como un todo unitario a partir de los elementos básicos, sus partes distinguibles. El resultado de las operaciones intelectuales complementarias de análisis y síntesis es un modelo conceptual del sistema. Por lo que el sistema estará caracterizado, no solamente en función de sus partes, sino también por el patrón que las conecta (relaciones, comunicación, organización), revelando las propiedades globales del sistema.
3. Concepto de niveles de organización: Es un sistema complejo, estructurado a múltiples niveles de integración, organización y abstracción de forma heterárquica, como una amplia red con elementos y relaciones, en la cual pueden distinguirse jerarquías

locales parciales. Los sistemas a niveles inferiores son más básicos, primitivos, concretos, fundamentales, y a partir de ellos, tomándolos como elementos disponibles y combinándolos de diversas maneras, se construyen los sistemas compuestos, derivados, de niveles superiores. Las distintas posibilidades de combinación de entidades y relaciones básicas pueden producir comportamientos globales complejos. Las reglas locales pueden producir orden global. Por ejemplo, el ajedrez tiene reglas básicas simples pero las estrategias de juego son muy complejas.

El mapa propuesto en este proyecto, se ha diseñado guardando una estructura de datos abierta y fácilmente ampliable, dotando al sistema la posibilidad de crecer y de mejorar. Los accesos a las ramas y caminos proporcionan datos de análisis que son tomados por el sistema para proponer los parámetros de diseño convenientes.

6.3.7 Banco de muestras

Las colecciones de muestras son la herramienta más común para agilizar el proceso del diseño de sonido en cine, RV, videojuegos... Dos son los principales problemas que presenta esta técnica. El primero forma parte del proceso de construcción de las muestras, estas se crean mediante la grabación y posterior procesado de señales de audio captadas de la naturaleza. Esto implica conocimientos de un experto diseñador de sonidos y la utilización de microfónica y tecnología asociada a la grabación de muestras de audio de alta calidad. El segundo es la escasa versatilidad que pueden llegar a ofrecer estas muestras. Con esto, me refiero a que la mayoría de estas muestras son grabaciones de audio de eventos sonoros que suelen estar compuestos por la acción, tanto simultánea como concatenada, de varios sonidos. Un ejemplo sencillo de entender es el sonido de un coche en movimiento. Partimos del punto como observadores, entonces un coche pasa por una carretera a cierta velocidad. Nosotros oiremos, o mejor dicho, nuestro cerebro entiende que el sonido asociado a lo que estamos viendo, es el de un coche en movimiento. En realidad ese evento sonoro que entendemos como algo único e indivisible, no lo es, es la suma del ruido del motor, la fricción de las ruedas contra el asfalto, y sonido del aire contra la superficie del coche. Además, hay que añadir el

efecto doppler, que se puede llegar a dar dependiendo de la velocidad y de la disposición del observador con respecto al coche y los posibles rebotes del sonido en paredes colindantes o edificios. Por lo que, si grabamos el sonido del evento que se ha comentado en líneas anteriores, nos encontraremos con que no podremos realizar muchas transformaciones sobre la muestra con el fin de cambiar algunas de las características físicas del coche. Este ejemplo nos ayuda a definir el concepto de estructura del evento sonoro. Un objeto sonoro está compuesto por unidades sónicas primitivas que, en su conjunto, forman la estructura del evento sonoro.

Nuestro banco de muestras debe presentar la ventajas que ofrece la tecnología basada en el muestreo y además intentar mejorar los problemas que esta presenta. La muestras deben cumplir con ciertos requerimientos, ya que no se trata de una colección clásica. Estos son:

- Flexibilidad: Es necesario que cada muestra genere el mayor número de muestras asociadas a ella como sea posible.
- Control: Deben presentar la capacidad de ser modificadas mediante las técnicas de análisis/resíntesis que propone este trabajo.
- Limpieza: Las señales contaminadas con ruido, o reverberaciones no pueden formar parte de nuestra colección.
- Rendimiento espectral: Las muestras pertenecientes a los mismos géneros descriptivos deberán tener un alto rendimiento sin que les afecten problemas asociados a las transformaciones espectrales propuestas.

6.3.7.1 Estructura

El tamaño de nuestro banco de datos vendrá determinado por el tamaño de la ontología desarrollada. Esta proporcionará el número de muestras primitivas o “muestras madre” que necesitamos para cubrir cualquier resultado propuesto por el análisis ontológico que realiza el usuario. La estructuración del banco de muestras se basa en la agrupación de estas por géneros asociados a los

descriptores utilizados. Cada genero se subdivide en clases, que dan lugar a muestras madre, estas muestras madre se unen formando una hermandad y entre ellas guardan una relación de hermanas, ya que no se pueden considerar como la misma muestra, pero si están creadas a partir del mismo objeto y en condiciones similares. En la siguiente figura 6.13 podemos ver un ejemplo de esta estructuración.

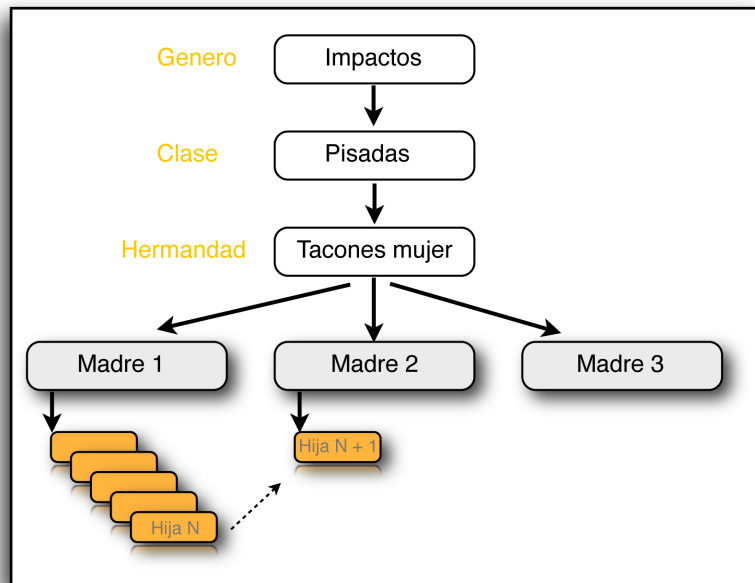


Figura 6.13 Estructura de las muestras dentro del banco de sonidos

Las muestras madre están limitadas a unas ciertas transformaciones, por ejemplo, el impacto producido por una barra de hierro presenta un espectro con una relación de formantes y frecuencias que están ligadas a la longitud de la barra, cuanto más larga más vibra y más parciales inarmónicos genera, por lo que deberemos disponer de muestras de la misma barra pero grabadas con diferentes longitudes, evitando problemas de distorsión a la hora de resintetizar. Otro ejemplo de limitación es el del estrechamiento temporal, el cual está limitado debido a los problemas de distorsión de la señal que conlleva. Las transformaciones capaces de hacer cada muestra madre se denominan hijas y son el resultado sonoro que estamos buscando.

En la figura 6.14 podemos observar cómo la muestra madre 1 (M1) puede llegar a alcanzar mediante transformaciones a las madres M2 y M3 sin embargo para llegar a M4 es necesaria M3.

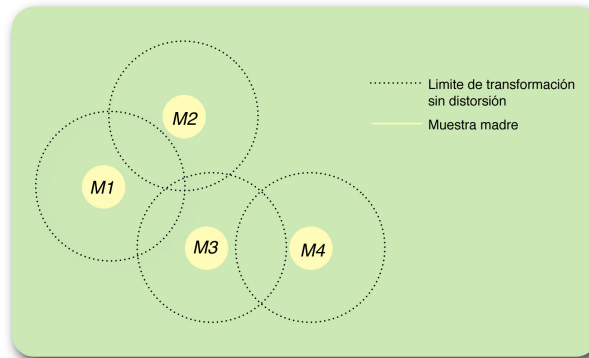


Figura 6.14 Alcance de transformación de cada muestra.

6.3.7.2 Metadatos asociados a cada muestra

La correcta localización de las muestras que más se ajusten a nuestro objetivo estará ligada a la cantidad de información que las acompañe. Estos datos proporcionarán:

- información temporal
- descripciones de la fuente
- características del sonido
- relaciones con otras muestras
- posibles transformaciones
- transformaciones propuestas por otros usuarios

Cuantos más datos adjuntemos a cada muestra mayor precisión tendrá nuestro análisis ontológico. Este tipo de datos que describen las muestras de audio es comúnmente utilizado en las librerías comerciales de sonidos para que el usuario identifique con rapidez y sin necesidad de reproducir una por una las muestras que cree necesitar. En la siguiente figura ejemplificamos una muestra y la cadena de metadatos asociada a ella. (ver figura 6.15).

La versatilidad de este método de diseño reside en la capacidad de reutilización y ampliación de la colección de muestras. Esta ampliación se puede realizar siempre y cuando se sigan los requerimientos de almacenaje y se practique la descripción de las muestras siguiendo el mismo análisis ontológico que realizamos con la escena, con el fin de ir clasificando y etiquetando cada muestra. Se trata de un proceso inverso. De esta manera conseguimos ampliar la librería de muestras.



Figura 6.15 Estructura metadatos asociados a cada muestra

6.4 Módulo de síntesis y procesado

Este módulo representa el corazón de la generación sonora. Está compuesto por varios procesos que configurarán el resultado final:

1. Análisis
2. Transformaciones
3. Resíntesis
4. Síntesis Adicional
5. Filtrado
6. Procesamiento patrones temporales

El estado inicial de este módulo pasa por un análisis de las muestras proporcionadas por el banco de sonidos. Una vez analizado y extraídos los parámetros acústicos que nos interesen se pasa a un proceso de transformación

que modela la muestra bajo condiciones impuestas en el mapeo. Estas transformaciones son ajustadas por el usuario. El siguiente paso es la resíntesis de la muestra. Paralelamente está el módulo complementario de síntesis, que forma una herramienta de apoyo para diseño de elementos menos formales. El contenido de estos dos procesos pasa a través de un banco de filtros encargados de modelar ciertos aspectos correspondientes a características físicas, propias de un sistema resonador fuente. Es aquí donde acaba el proceso de síntesis del objeto-fuente y empieza la etapa de procesamiento temporal y espacial. El procesamiento espacial se realizará mediante técnicas de auralización, reverberación y dinámica. Por ultimo propondré la aplicación de patrones temporales como parte de un completo proceso de diseño intuitivo. En la figura 6.16 se expone el diagrama del modulo de síntesis y procesado.

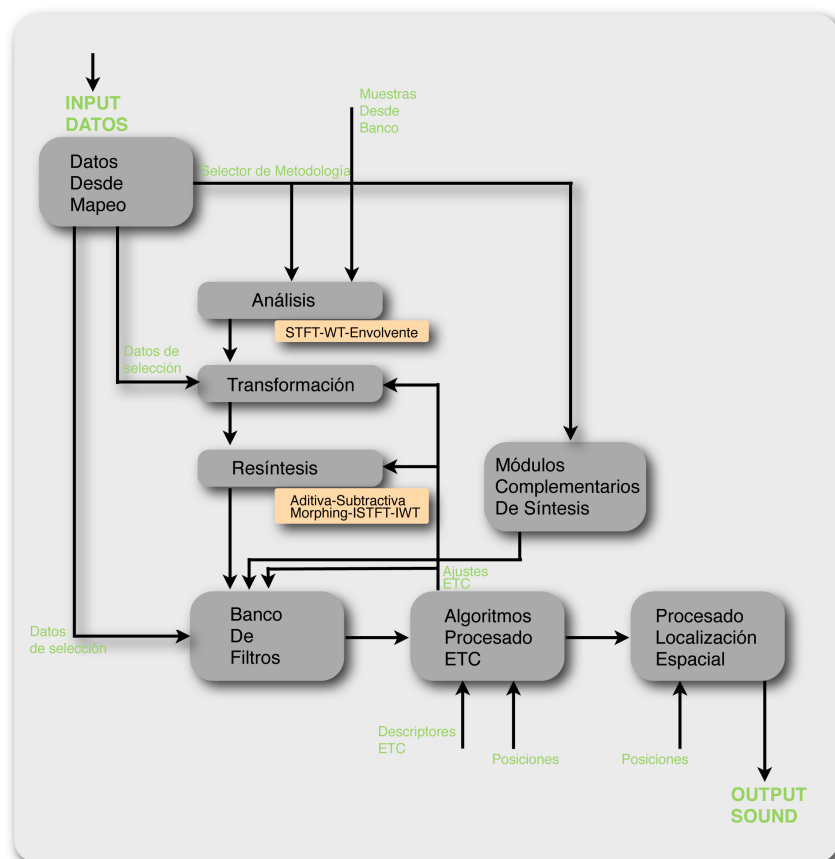


Figura 6.16 Diagrama del módulo de síntesis y su flujo de datos

6.4.1 Etapa de Análisis

Esta etapa es la primera en el camino de la señal en el modulo de síntesis. Nos permite recibir las muestras provenientes del banco de muestras. A su entrada se disponen dos tipos de datos:

- Muestras de audio provenientes del banco de muestras.
- Datos de selección de metodología que debe ser seleccionar el sistema.

Como hemos indicado en el capítulo 5, se plantean tres técnicas de análisis espectral:

- Transformada de Fourier de tiempo reducido STFT
- Transformada Wavelet
- Análisis por Autoregresión y Cepstral

El diagrama de funcionamiento de esta etapa es el que se muestra muestra en la figura 6.17.

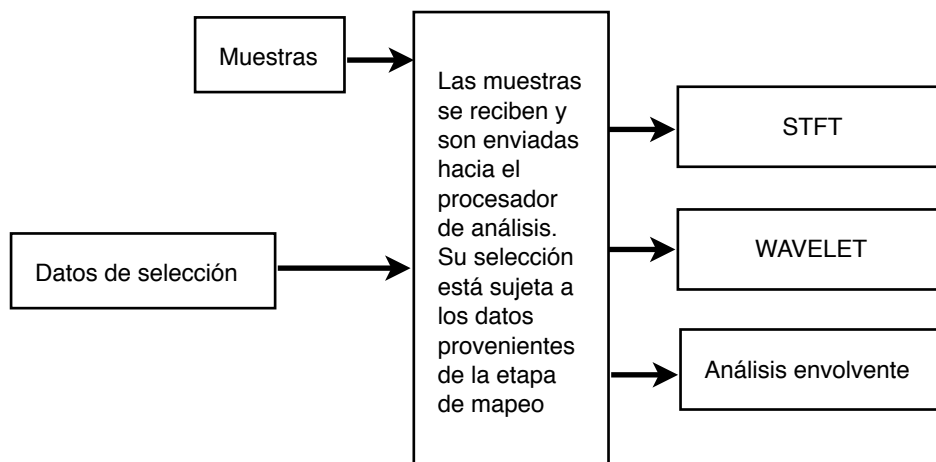


Figura 6.17 Diagrama de funcionamiento la etapa de análisis.

La salida de esta etapa devuelve los datos extraídos del análisis espectral. Estos son dirigidos hacia la etapa de transformación.

6.4.2 Etapa de transformaciones

6.4.2.1 Aplicaciones y técnicas

Tras el análisis de la señal y después de haber extraído las componentes de frecuencia, fase y amplitud, podemos realizar ciertas transformaciones en la señal, que nos permitirán llevar a cabo ajustes en el timbre, tiempo y tono. Estos nos permiten adaptarnos a las condiciones impuestas por la escena. En la figura 6.19 se especifican algunos de los efectos que permiten estas técnicas.

<i>Efecto</i>	<i>Técnica</i>
<i>Variación de muestras</i>	<i>Cambio de envolventes de amplitud y frecuencia mediante la edición multiplicación por funciones.</i>
<i>Escalado espectral</i>	<i>Multiplicado de la frecuencia amplitud y frecuencia mediante la edición o multiplicación por funciones.</i>
<i>Desplazamiento espectral</i>	<i>Suma de un valor n a todos los parciales.</i>
<i>Inversión espectral</i>	<i>Inversión de cada uno de los componentes frecuenciales antes de la síntesis.</i>
<i>Hibridación del timbre</i>	<i>Asignación de la envolvente de un sonido a la del otro.</i>
<i>Expansión y compresión temporal sin modificación del tono</i>	<i>Extensión de la duración de la envolvente de amplitud y frecuencia, o cambiando el ajuste del hop Size durante la reconstrucción.</i>
<i>Estiramiento de timbres percusivos en pasajes sintéticos prolongados</i>	<i>Retardo del tiempo inicial de cada parcial y suavizado de sus envolventes.</i>
<i>Realzado de regiones espectrales</i>	<i>Incremento de la amplitud de los parciales.</i>
<i>Síntesis cruzada</i>	<i>Escalado de la envolvente de un sonido por la de otro (convolución). Aplicación de la envolvente de amplitud de un sonido a la frecuencia o fase de otro. Aplicación del ruido residual de un sonido a la parte cuasi armónica de otro [36].</i>

figura 6.19 Transformaciones aplicadas sobre un proceso de análisis.

Entre estas técnicas destacaremos las técnicas denominadas de estrechamiento temporal (*time stretching*), que actúa modificando la duración de la señal sin que se modifique la altura y al desplazamiento de altura (*pitch shifting*) que transpone el tono de la señal sin que varía la duración temporal. Esta técnica son aplicadas en la síntesis por muestreo, pero aquí alteraciones de duración afectan la altura y viceversa. Por otro lado hay que tener en cuenta que en el diseño de sonido esta técnica ha sido muy utilizada con un gran éxito.

6.4.2.2 Warping

Las primeras técnicas de warping de tiempo discreto fueron en 1965 [40]. Sin embargo su interés en aplicaciones de sonido y de música es relativamente reciente. Este efecto actúa deformando el eje de tiempo o de frecuencia de la señal. El warping de tiempo deforma la forma de onda o la envolvente. Por otro lado el warping de frecuencia actúa deformando sus componentes frecuenciales. Algunos ejemplos y aplicaciones de esta técnica [37][38] aplicada al diseño de sonido o aplicaciones musicales son:

- Transformación de una señal armónica a otra inarmónica y viceversa.
- Incremento de la riqueza espectral de la señal mediante fluctuaciones o desafinaciones.
- Modelado de propagación de ondas de ciertos tipos de comportamientos acústicos.
- Efectos sobre la dinámica de la señal.
- Realzado de las características naturales del sonido.

A partir de ahora nos referiremos al warping de tiempo como TW (Time Warping) y al de frecuencia como FW (Frequency Warping).

El TW se puede explicar de una manera sencilla imaginando que el eje horizontal de la señal pueda ser comprimido o expandido entre dos puntos de la señal que deseemos, siendo solo este segmento de la señal el que varíe su resolución temporal. En la figura 6.20 podemos ver un ejemplo básico de TW sobre una onda sinusoidal. La representación es obtenida mediante la

adjudicación de los instantes de tiempo de la señal original hacia las nuevas características de la señal deformada $\theta(t)$, mediante líneas horizontales.

El papel de esta técnica en el diseño de sonido aplicado a entornos de interacción audio-imagen es esencial, sobre todo para corregir problemas de sincronización sin que las condiciones tímbricas de la señal se ven afectadas. El FW produce un efecto es comparable al escalado frecuencial que se realiza en métodos como el Phase Vocoder, pero la flexibilidad y calidad que ofrece el FW es mucho mayor.

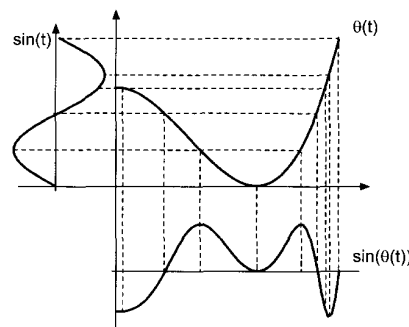


Figura 6.20 Efecto TW aplicado a una señal sinusoidal.

6.4.3 Técnicas de resíntesis

La idea bajo la definición de resíntesis es la de la reconstrucción de la señal al dominio del tiempo después de que haya sido analizada. La señal reconstruida será más o menos similar a la original dependiendo de las transformaciones que se hayan producido en el punto intermedio del proceso. Se han inventado una gran variedad de técnicas de resíntesis. Algunas son más precisas, mientras que otras presentan una mejor robustez.

- Método Overlap-Add (OA): Este método consiste en aplicar la transformada discreta inversa del método de transformación aplicado a la señal, en cada

uno de los segmentos de análisis. Se toman las magnitudes correspondientes de amplitud y fase de cada componente frecuencial y se genera correspondiente señal en el dominio del tiempo. Mediante un proceso de suma y solapado de los segmentos, con una atenuación de -3dB . Este método idealmente debe ser capaz de analizar y resintetizar numerosas veces, pero la pérdida de información en cada iteración provoca pérdidas. En la figura 6.21 se muestra el procedimiento.

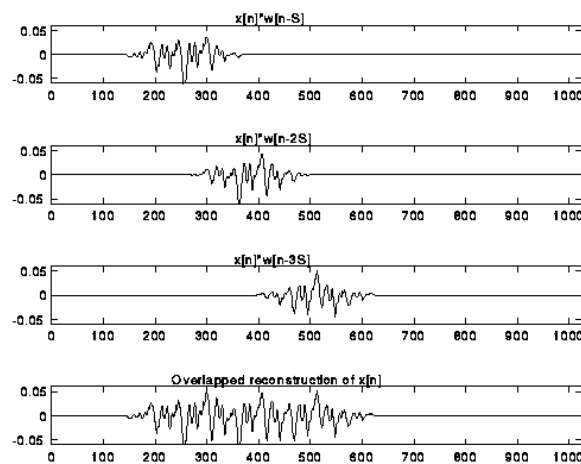


Figura 6.21 Señal reconstruida mediante el Método *Overlap-add*.

- Método de banco de osciladores o síntesis aditiva. La diferencia entre este método y el de OA es que, en vez de sumar cada una de las señales senoidales a través de la transformada inversa, este método convierte los datos del análisis en envolventes de frecuencia y amplitud, mediante osciladores. La principal ventaja que presenta este método es que es mucho más robusto a la hora de aplicar transformaciones a cada cuadro. Es el modelo de resíntesis que mejor aproximaciones devuelve, sin embargo, el coste computacional es muy elevado. Esta técnica de resíntesis es la empleada en el poderoso sistema de análisis/resíntesis Phase Vocoder (vocoder de fase) [42]. Utiliza un seguimiento de picos para localizar los parciales dominantes, así como su variación temporal,

después, estas componentes son reconstruidas mediante oscilares. La implementación de un oscilador por cada una de las sinusoides que necesitemos presenta un coste muy elevado. Los modelos de resíntesis basados en la combinación de síntesis aditiva y estocástica [45] son una solución mediante el ahorro de osciladores. Incorporan a la señal la componente de ruido presente en los transitorios, y que no se tiene en cuenta en los modelos de síntesis aditiva (figura 6.22).

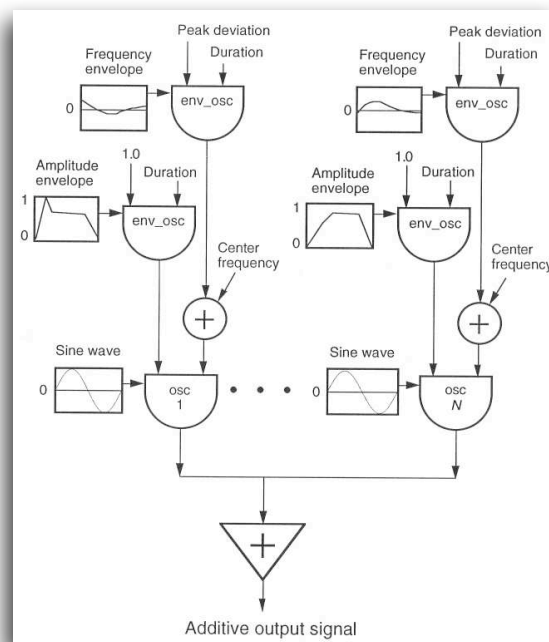


Figura 6.22 Banco de osciladores para la reconstrucción de la señal mediante síntesis aditiva.

- Método de resíntesis sustractiva o estocástica. Este método actúa de forma diferente a los dos anteriores, mientras que la resíntesis aditiva y el método OA reconstruyen la señal a partir de sus componentes frecuenciales, la resíntesis sustractiva lo hace a partir de las formantes. Utiliza la información de la envolventes espectral para la reconstrucción. Este método actúa bajo una estructura de fuente-resonador. Los filtros son el resonador y como fuente se puede utilizar un generador de ondas,

generadores de ruido o incluso otra señal(morphing). Las limitaciones de este método están en la imprecisión a la hora de identificar componentes espectrales extraídas del análisis con las generadas por la fuente. Sin embargo presenta unas condiciones excelentes, entre otras, para representar señales con un alto componente estocástico.

6.4.3.1 Estructura de la etapa de resíntesis

El funcionamiento de la etapa de resíntesis se expone en la figura 6.23.

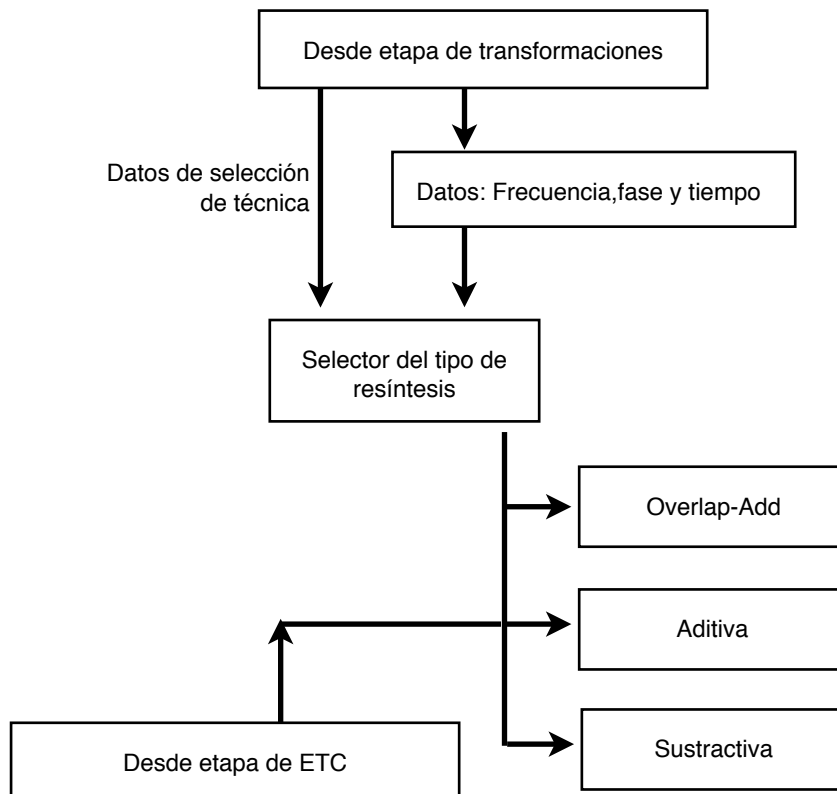


Figura 6.23 Banco de osciladores para la reconstrucción de la señal mediante síntesis aditiva.

6.4.5 El banco de filtros

Siguiendo las bases propuestas en este trabajo, en él, se toman las características físicas del objeto que es mostrado en la imagen, para realizar el correspondiente diseño sonoro del objeto. Una técnica a tener en cuenta es que una gran cantidad de estos eventos sónicos pueden ser descritos mediante un modelo de fuente-resonador.

Por lo tanto, partimos de que el resonador es modelado mediante un banco de filtros [46][47], al igual que se hace en síntesis de codificación lineal predictiva, LPC. Sin embargo en nuestro caso trataremos de modelar la respuesta que presentan los materiales en ciertos eventos sonoros, como los impactos, en los que intervienen dos objetos, uno de ellos es la fuente y otro el resonador. Este soporta la vibración impuesta por la excitación de la fuente, ligada a la potencia del impacto. El resonador guarda relación con el tipo de material del que está compuesto objeto y con su estructura interna. Finalmente, resaltaremos el factor de amortiguamiento, al estar relacionado con el tipo de material.[48][53].

La técnica propuesta en este proyecto para modelar la respuesta del resonador es un filtrado por convolución de la señal ante impulsos que respondan a:

- Tipos de materiales
- Tamaño de los objetos
- Diferentes fuerzas e intensidades
- Dispositivos electrónicos de megafonía, radiodifusión o amplificación

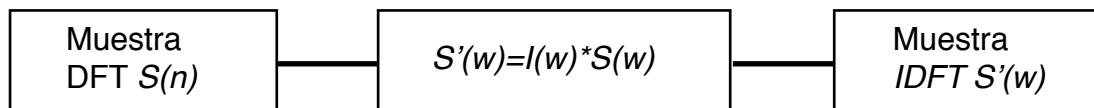


Figura 6.24 Diagrama de bloques del filtrado por convolución.

A parte de actuar como un módulo-resonador, esta técnica se puede utilizar para recrear efectos en los que la señal pasa a través de entornos que ya modulan la señal, como es el caso de un amplificador de guitarra, teléfonos

móviles, megafonía, radios, y un largo etcétera. La figura 6.24 es el diagrama del filtro.

6.4.6 Arquitectura del procesamiento para ETC

En esta etapa, las muestras de audio se procesan para crear patrones temporales. Consisten en repeticiones de la señal, dispuestas en determinados patrones temporales y de las consecuentes transformaciones en el timbre de la señal que conllevan estos patrones.

Algunos ejemplos que estos patrones y transformaciones proporcionarían son:

- flujo de fluidos continuado (olas del mar, ríos, cascadas)
- objetos cayendo al suelo
- sonidos de maquinaria industrial
- sonidos derivados de movimientos humanos (caminando, corriendo, etc...)
- sonido de animales (canto de pájaros)
- sonidos de movimientos provenientes de la naturaleza
- Efectos especiales

Según se describe en el apartado 6.3.5 de este capítulo, los ETC se dividen en simples y complejos. A su vez, estos describen sucesivos tipos de patrones. El procesado de la señal se hará mediante el uso de retardadores. Debemos tener en cuenta los diferentes patrones que debemos conseguir a la hora de diseñar los retardos sobre las muestras de audio. En base a la taxonomía propuesta en este proyecto la estructura y tipo de retardos son:

- Retardos simples en el tiempo
- Retardos complejos en el tiempo
- Retardos en el dominio frecuencial

El diagrama de bloques propuesto se presenta en la figura 6.25.

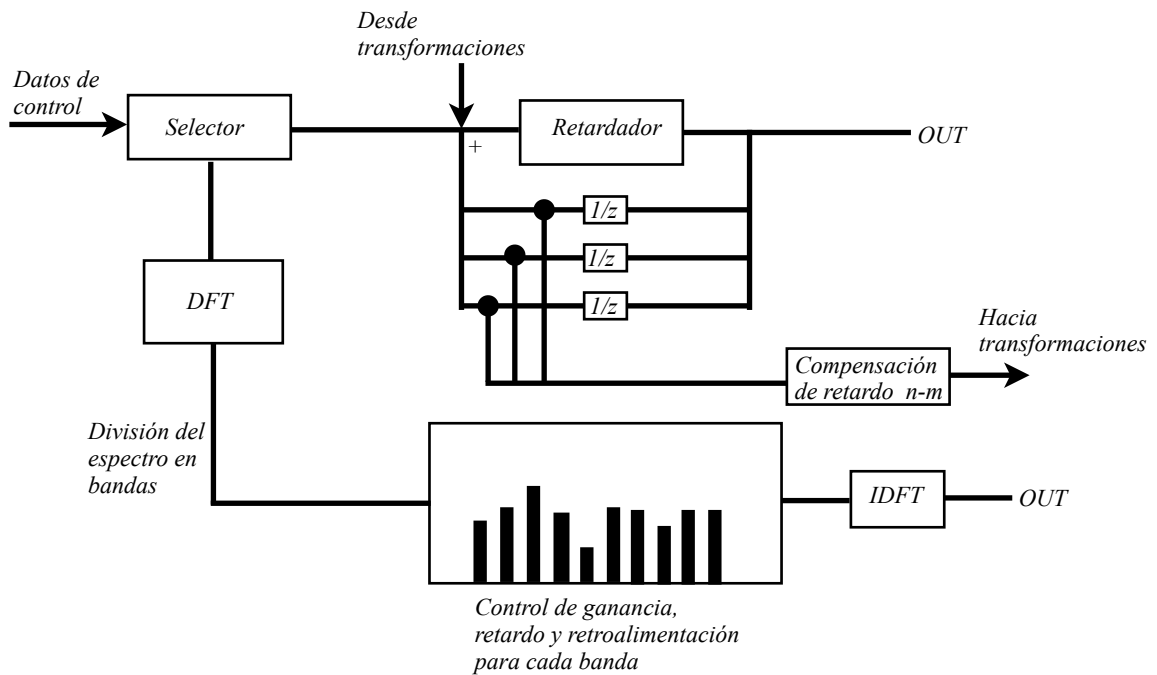


Figura 6.25 Diagrama de flujo del sistema de retardo y construcción de patrones.

Las transformaciones que se realizan sobre las muestras retardadas dependen de nuevo de las características físicas de la fuente. Estas varían respecto a los patrones temporales que presente cada evento. La mayoría de estos procesos son sistemas caóticos, mientras que otros deben de guardar una relación precisa con la fuente.

6.4.7 Procesado de localización espacial

En este módulo la señal se procesa por última vez con el objetivo de diseñar las condiciones espaciales que caracterizan la escena. Esto significa que habrá que dotar a la señal de un entorno físico en el que ubicarla, así como de su colocación en el plano aural. No debemos olvidar para qué tipo de entorno estamos trabajando, ya que si se trata de un entorno de RV o un videojuego, hay que permitir que ciertos parámetros de los efectos de procesado sean

modificables en el lenguaje de programación sobre el que se ejecutan. Un ejemplo de este tipo de plataforma a nivel comercial lo encontramos en software Wwise®.

Este módulo debe ser capaz de realizar las siguientes técnicas sobre los objetos sonoros diseñados:

- Posicionamiento de los objetos tridimensionalmente en el plano aural.
- Descripción de las alteraciones dinámicas provocadas por los objetos en movimiento.
- Identificación de la acústica ambiental y de reverberación de la escena.
- Descripción de las condiciones acústicas de objetos sonoros respecto al punto de escucha.
- Procesamiento de dinámica para controlar las condiciones de proximidad frente a los movimientos del personaje con respecto al entorno o a una fuente.

Estas premisas nos llevan a asumir y elegir las siguientes técnicas para realizar el procesamiento espacial de la señal de la manera más eficiente. No olvidemos que no se trata de hacer una mezcla en formato 5.1 ya que esta técnica pertenece a otro proceso dentro de la producción de bandas sonoras en el que no está centrado este proyecto.

6.4.7.1 Aspectos teóricos sobre espacialización y reverberación a tener en cuenta para el diseño.

Reflexión y refracción de las ondas

Cuando una onda alcanza la superficie de separación de dos medios de distinta naturaleza se producen, en general, dos nuevas ondas, una que retrocede hacia el medio de partida y otra que atraviesa la superficie límite y se propaga en el segundo medio. El primer fenómeno se denomina reflexión y el segundo recibe el nombre de refracción.

En las ondas monodimensionales como las producidas por la compresión de un muelle, la reflexión lleva consigo una inversión del sentido del movimiento ondulatorio. En las ondas bi o tridimensionales la inversión total se produce únicamente cuando la incidencia es normal, es decir, cuando la dirección en la que avanza la perturbación es perpendicular a la superficie reflectante. Si la incidencia es oblicua se produce una especie de rebote, de modo que el movimiento ondulatorio reflejado cambia de dirección, pero conservando el valor del ángulo que forma con la superficie límite.

En el caso de las ondas sonoras, la reflexión en una pared explica el fenómeno del eco. Si la distancia a la pared es suficiente, es posible oír la propia voz reflejada porque el tiempo que emplea el sonido en ir y volver permite separar la percepción de la onda incidente de la reflejada. El oído humano sólo es capaz de percibir dos sonidos como separados si distan uno respecto del otro más de 0,1 segundos, de ahí que para que pueda percibiéndose el eco la superficie reflectiva debe estar separada del observador 17 metros por lo menos, cantidad que corresponde a la mitad de la distancia que recorre el sonido en el aire en ese intervalo de tiempo ($17 \text{ m} = 340 \text{ m/s} \cdot 0,1 \text{ s}/2$).

En los espacios cerrados, como las salas, el sonido una vez generado se refleja sucesivas veces en las paredes, dando lugar a una prolongación por algunos instantes del sonido original. Este fenómeno se denomina reverberación y empeora las condiciones acústicas de una sala, puesto que hace que los sonidos anteriores se entremezclen con los posteriores. Su eliminación se logra recubriendo las paredes de materiales, como corcho o moqueta, que absorben las ondas sonoras e impiden la reflexión.

El fenómeno de la refracción supone un cambio en la velocidad de propagación de la onda, cambio asociado al paso de un medio a otro de diferente naturaleza o de diferentes propiedades. Este cambio de velocidad da lugar a un cambio en la dirección del movimiento ondulatorio. Como consecuencia, la onda refractada se desvía un cierto ángulo respecto de la incidente.

La refracción se presenta con cierta frecuencia debido a que los medios no son perfectamente homogéneos, sino que sus propiedades y, por lo tanto, la

velocidad de propagación de las ondas en ellos, cambian de un punto a otro. La propagación del sonido en el aire sufre refracciones, dado que su temperatura no es uniforme. En un día soleado las capas de aire próximas a la superficie terrestre están más calientes que las altas y la velocidad del sonido, que aumenta con la temperatura, es mayor en las capas bajas que en las altas. Ello da lugar a que el sonido, como consecuencia de la refracción, se desvía hacia arriba. En esta situación la comunicación entre dos personas suficientemente separadas se vería dificultada. El fenómeno contrario ocurre durante las noches, ya que la Tierra se enfría más rápidamente que el aire. Este efecto de proximidad es muy útil para diseñar entornos de intimidad y proximidad durante pasajes de la escena nocturnos y pasajes más dinámicos durante el día.

La difracción

Las ondas son capaces de traspasar orificios y bordear obstáculos interpuestos en su camino. Esta propiedad característica del comportamiento ondulatorio puede ser explicada como consecuencia del principio de Huygens y del fenómeno de interferencias.

Así, cuando un frente de ondas alcanza una placa con un orificio o rendija central, cada punto de la porción del frente de ondas limitado por la rendija se convierte en foco emisor de ondas secundarias todas de idéntica frecuencia. Los focos secundarios que corresponden a los extremos de la abertura generan ondas que son las responsables de que el haz se abra tras la rendija y bordee sus esquinas. En los puntos intermedios se producen superposiciones de las ondas secundarias que dan lugar a zonas de intensidad máxima y de intensidad mínima típicas de los fenómenos de interferencias.

Ambos fenómenos que caracterizan la difracción de las ondas dependen de la relación existente entre el tamaño de la rendija o del obstáculo y la longitud de onda. Así, una rendija cuya anchura sea del orden de la longitud de la onda considerada, será completamente bordeada por la onda incidente y, además, el patrón de interferencias se reducirá a una zona de máxima amplitud idéntica a un foco. Es como si mediante este procedimiento se

hubiera seleccionado uno de los focos secundarios descritos por Huygens en el principio que lleva su nombre.

Los principales parámetros de reverberación son :

- **Tiempo de decaimiento:** se define como el tiempo que tarda el sonido reverberado en disminuir 60 dB (a menudo se denomina TR60). Las salas grandes tienen tiempos largos (un segundo o más), mientras que las habitaciones de una casa tienen tiempos muy cortos (menos de medio segundo).
- **Retardo de las primeras reflexiones:** en salas grandes las primeras reflexiones tardan en llegar más tiempo que en salas pequeñas, pudiendo sonar incluso como una especie de eco.
- **Intensidad de las primeras reflexiones:** está determinada por la distancia del oyente y de la fuente sonora respecto a las superficies reflectantes y el coeficiente de absorción de estas. Si el oyente o la fuente sonora están junto a ellas las primeras reflexiones sonarán con mucha intensidad.

Manipulando los 3 parámetros anteriores podemos crear la sensación de tamaño del recinto, y de posicionamiento de fuente y oyente dentro de él. Pero además, podemos crear diferentes sensaciones relacionadas con los materiales de las paredes, suelo y techo con parámetros tales como:

- **Tipo de reverberación:** una reverberación tipo *hall* nos proporciona una coloración diferente que una de tipo *plate*, o de tipo *room*. Otros tipos de reverberación como las *gate-reverbs* o las *reverbs* no lineales (en las que la intensidad de las reflexiones no se van atenuando a medida que pasa el tiempo) pueden alterar poco la coloración, pero en cambio provocar sensaciones extrañas (ya que son "anti-naturales").
- **Densidad de las reflexiones:** aumenta en función de la cantidad de trayectorias reflejadas que lleguen al oyente (debido a que hay muchas superficies reflectantes (paredes con angulaciones cambiantes,

objetos interpuestos en la trayectoria del sonido, paredes de materiales poco absorbentes...).

- Absorción selectiva de determinadas frecuencias: puede simularse aplicando una determinada ecualización; la absorción está directamente relacionada con los materiales de las superficies reflectantes (una pared de hormigón reflejará muchas más altas frecuencias que una cortina gruesa, por ejemplo).

Es importante remarcar que cuando uno empieza a utilizar reverberaciones suele hacer un abuso de ellas. La mejor manera de evaluar su efectividad consiste en ajustar el equilibrio entre sonido seco y sonido reverberado (dry/wet) según creamos apropiado, y a continuación eliminar la reverberación; si "aparecen" detalles o instrumentos que en la mezcla no se oían quiere decir que seguramente estábamos a punto de sobre-reverberar. Para hacer este tipo de escucha es importante configurar el mezclador de manera que en dos canales tengamos la señal seca, en los dos contiguos la señal reverberada, y que el procesador nos entregue 0% de señal original y 100% de procesada.

En los últimos años han aparecido equipos e incluso programas que permiten simular el posicionamiento de una fuente sonora no sólo en un espacio acústico y en un eje horizontal, sino también en el eje vertical, así como simular con credibilidad trayectorias de la fuente dentro de ese espacio.

6.4.7.2 Procesado de Reverberación por Convolución

Los procesadores digitales de reverberación son y han sido muy útiles a lo largo de los años, pero muchos encuentran limitadas sus prestaciones debido a que emulan entornos realizando un gran número de cálculos y basándose en complejos algoritmos internos. Esta limitación "electrónica" repercute en que los efectos que generan resultan fríos, metálicos, demasiado artificiales, etc. Esto depende de la calidad del procesador, ya que algunas unidades han llegado a unas cotas de evolución realmente impresionantes, creando efectos "de la nada" que suenan bastante realistas, e incluso ofrecen la posibilidad de

crear efectos "inhumanos" y exageradamente extraños, a los que siempre podemos acudir.

Estas limitaciones son las que propiciaron el nacimiento de la reverberación por convolución, basada en impulsos. Estos son extraídos de entornos acústicos reales, o sea, se reproduce una señal impulsiva en un recinto que deseemos muestrear, y se graba, con lo que contamos con un muestra de audio de la reverberación del recinto o entorno que posteriormente podremos procesar para aplicar dicho efecto a nuestras señales.

Las ventajas de este tipo de reverberación es la posibilidad de recrear entornos totalmente realistas de espacios acústicos con características varias, así como de casi cualquier sitio o lugar que creamos conveniente.

La convolución es un proceso matemático entre dos señales para obtener una tercera, donde las dos primeras representan, por ejemplo, la repuesta al impulso de un sistema y la señal de entrada a este, y la tercera, es decir el resultado de la operación, es la señal de salida del sistema.

La convolución FFT se basa en la propiedad que establece que la multiplicación en el dominio de frecuencia corresponde a la convolución en el dominio de tiempo. Así, en este proceso se obtiene el espectro de la señal de entrada mediante transformada discreta de Fourier (DFT), luego se multiplica con la respuesta al impulso en el dominio de la frecuencia, la respuesta en frecuencia, y se retorna el resultado al dominio de tiempo mediante la IDFT (inversa DFT). Ver figura 6.26.

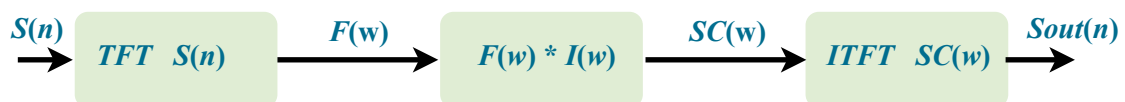


Figura 6.26 Filtrado por convolución.

Teniendo en cuenta que el comportamiento acústico de un recinto -para un par de puntos de fuente y receptor dados - puede ser descrito por medio de la respuesta al impulso, las alteraciones que este realiza a una señal acústica desde que es generada por una fuente hasta que llega a los oídos de un receptor, pueden ser simuladas computando la operación de convolución entre la señal de entrada (señal generada por la fuente) y la respuesta al impulso medida en la misma ubicación del receptor. Usando este mismo razonamiento, es posible desarrollar un sistema de procesamiento de audio digital que calcule la convolución entre una señal de entrada y una respuesta al impulso previamente medida, dando como salida la señal de entrada “reverberada” , así la señal de entrada sonará como si hubiese sido grabada, donde fue tomada la respuesta al impulso. Donde el teorema de la convolución confirma que :

$$SC(w) = S(w) * I(w)$$

Siendo $S(w)$ la señal en el dominio de la frecuencia, $I(w)$ es el impulso, $SC(w)$ es la señal convolucionada.

Este trabajo propone que, junto la colección de muestras, se complete un banco de impulsos que abarque una gran cantidad de entornos.

6.4.7.3 Procesado sobre la dinámica y la intensidad del sonido

Todos los efectos aquí descritos realizan diversas multiplicaciones a cada muestra de un fragmento.

Efectos simples sobre la amplitud

- Modificar ganancia: consiste en multiplicar cada una de las muestras por un valor real. Si el valor está comprendido entre 0 y 1 el nivel sonoro disminuye, mientras que a partir de 1 aumenta. Se puede aplicar para potenciar sonidos que se han grabado con un nivel excesivamente bajo.
- Silenciar: consiste simplemente en multiplicar por cero la zona seleccionada.

- Puerta de ruido (noise gate): silencia las muestras cuya envolvente de amplitud este por debajo de determinado valor umbral, introducido como parámetro. Este efecto permite eliminar el ruido de fondo, aunque existen formas más sofisticadas de reducción de ruido, basadas en el análisis de la señal, ya que este sistema sólo puede eliminar el ruido en los fragmentos en los que hay señal de baja amplitud. También se utiliza para procesar señales impulsivas, con lo que se consiguen ataques y decaimientos más bruscos.
- Aplicar envolventes: Una envolvente es una curva que fuerza determinada evolución temporal de la amplitud. Fade in y fade out son dos envolventes particulares. La primera (fundido de entrada) es una envolvente con valor inicial cero y valor final uno, que se aplica al inicio de un fragmento, mientras que la segunda (fundido de salida) se inicia con uno y termina con cero, y se aplica normalmente al final de un fragmento. Pueden ser curvas lineales o logarítmicas.

Procesadores de rango dinámico

Recordemos que el rango dinámico de un fragmento sonoro viene dado por la diferencia (en dB) entre la intensidad más fuerte y la más débil. Los efectos incluidos en este apartado, operan directamente sobre la amplitud de las muestras, aunque de formas más sofisticadas que los descritos anteriormente, modificando de diversos modos el rango dinámico de un determinado fragmento. Los más importantes son los compresores, los expansores, los limitadores y las puertas de ruido (comentadas ya).

Una buena forma de caracterizar estos procesos es a través de su función de transferencia, que establece una correspondencia entre las amplitudes de entrada y las amplitudes de salida. En la figura 6.27 se muestran varias funciones de este tipo, junto con los efectos que producen en un fragmento sonoro. La función 6.27.a es una recta con una inclinación de 45°, lo que significa que, en este caso, no se produce ninguna modificación del rango dinámico (a cualquier valor de amplitud de entrada, le corresponde el mismo valor en la salida).

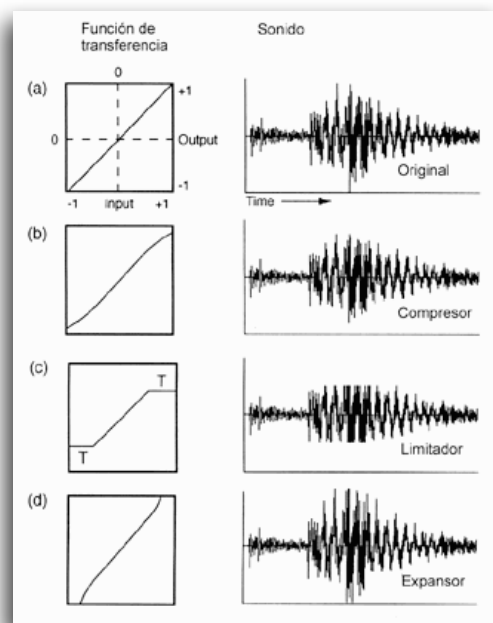


Figura 6.27 Aplicación de procesadores de dinámica sobre una señal de audio.

- Los compresores y los limitadores

Los compresores se utilizan para reducir el rango dinámico de una señal. Su uso es muy frecuente en la grabación de partes vocales, ya que en muchos casos la señal emitida por la voz y recogida por un micrófono presenta mínimos muy débiles que se confundirían con el ruido de fondo. Con una compresión más exagerada se consigue un efecto intimista, ya que se acentúan los sonidos de la respiración y de los movimientos bucales (lengua, saliva, etc.).

Los limitadores son un caso extremo de compresores, que limitan la amplitud máxima posible a un valor umbral. Su función de transferencia presenta pendientes horizontales en los extremos.

Este tipo de procesado puede ser usado también para dar intensidad y vigorosidad al sonido, así como para crear efectos

de proximidad, tanto para voces, como para cualquier tipo de sonido. Aplicado a sonidos de tipo aerodinámico es esencial para recrear eventos sonoros que simulan explosiones o impactos muy contundente

- Los expansores

Un expansor es el opuesto de un compresor. Acentúa los cambios, disminuyendo los niveles débiles y aumentando los fuertes. Se utiliza frecuentemente en combinación con puertas de ruido y para realzar grabaciones antiguas que presentan un rango dinámico estrecho .

- La distorsión

Este término tiene connotaciones peyorativas ya que normalmente define la pérdida o la degradación inevitables en una señal, ocasionadas por los diferentes dispositivos o procesos (micrófono, grabación, amplificador, altavoces) a los que se ve sometida. Sin embargo, todos los efectos descritos anteriormente son, en realidad, casos particulares de distorsión, que, por su uso frecuente, reciben un nombre propio.

6.4.8 Módulos complementarios de síntesis

Debido a que las necesidades del diseño no siempre nos piden que diseñemos sonidos naturales provenientes de la naturaleza, sino que en muchos casos los sonidos que necesitemos diseñar serán sonidos electrónicos, producidos por un aparato electrónico que utiliza síntesis para crear su sonoridad. Además los SFX se basan en una técnica de hibridación de sonidos naturales con sonidos sintetizados.

Este módulo puede admitir cualquier tipo de síntesis que hemos visto en el capítulo 3, sin embargo debe asumirse que los parámetros de control de estos sintetizadores deben ser mapeados con relación a lo impuesto en este trabajo.

6.5 Sistema de mapeo

Son muchos los estudios que demuestran la habilidad del cerebro humano para destacar características del sonido asociadas a las características de la fuente sonora [49][50][51]. El oyente presenta una gran capacidad para categorizar la forma, estructura y material del objeto, incluso algunas propiedades mucho más complejas como el tipo de interacción entre objetos o la diferenciación del género de una persona al andar. Esta habilidad para enlazar características físicas de la fuente con otras a un nivel perceptivo establece uno de los principales enfoques de este trabajo. A partir de aquí, se han creado las taxonomías de diferentes tipos de eventos sonoros, como son los sonidos de impactos, sonidos aerodinámicos o sonidos de líquidos. Bajo este conjunto de descriptores de la fuente podemos relacionar los descriptores de audio que se identifican con cada una de las propiedades de la fuente, tanto a nivel físico como perceptivo.

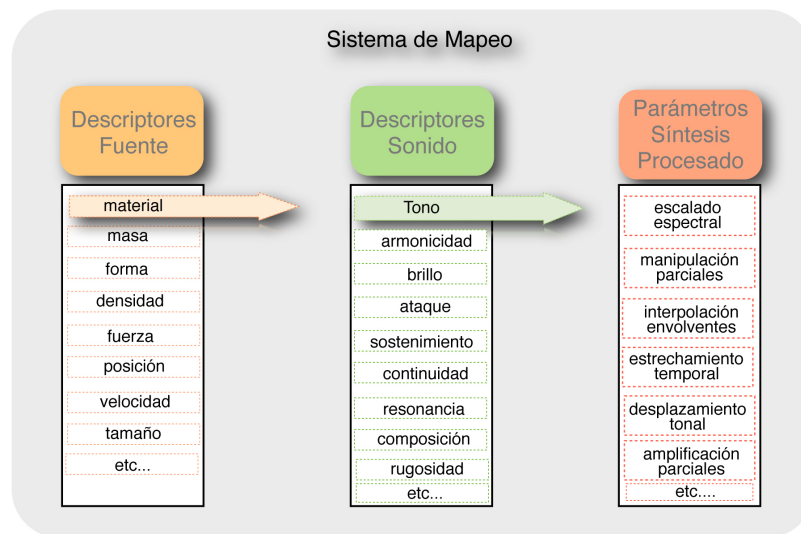


Figura 6.28 Sistema de mapeo .

Podemos definir el mapeo en dos fases (ver figura 6.28):

1. Identificar los descriptores de la fuente con sus correspondientes descriptores de sonido.
2. Relacionar los descriptores de audio con los parámetros de síntesis y procesamiento que se corresponden entre sí.

Este sistema de mapeo realizará una función automática de selección, mostrando al usuario el camino más adecuado para llegar a su objetivo.

Se han realizado varios estudios en los que encontramos la relación que existe entre las características de la fuente con características del sonido. Por lo que es posible relacionar la fuerza de impacto de un objeto con el brillo y este con un descriptor de audio denominado centroide espectral, que es la medida de la frecuencia media ponderada del espectro. A su vez este puede ser modificado con un filtrado de la señal de audio, capaz de alterar las altas frecuencias de la señal. La relación de descriptores, fuente y sonido, como de las transformaciones asociadas a ellos, se relacionan bajo el desarrollo de una ontología.

6.5.1 Descriptores de sonido

El hecho de disponer de una buena información de las características espectrales y temporales del sonido, nos permitirá tener un mayor control y mejor precisión a la hora de definir que parámetros serán los más indicados para realizar nuestros ajustes.

Es esta sección se revisan un conjunto de descriptores de audio desarrollados y usados en el proyecto CUIDADO [52], del IRCAM, así como de otros propuestos en el estándar de descripción de contenido multimedia MPEG-7. Han sido clasificados por categorías según sus características.

1_Características temporales globales:

- Log-Attack Time, Tiempo Log-Ataque [LAT]: Es el logaritmo en base decimal de la duración temporal entre, el tiempo de inicio de la

señal, hasta que esta alcanza su parte estable. Es uno de los descriptores perceptuales más importantes:

$$\text{lat} = \log_{10}(\text{stop_attack} - \text{start_attack}) \quad (6.1)$$

- Temporal Increase, Incremento Temporal [Incr]: El incremento temporal se define como la media de la pendiente temporal de la energía durante el tiempo de ataque. Se computan las pendientes locales de la energía correspondiente a cada esfuerzo w_i .
- Temporal Decrease, Decremento Temporal [decr]: Mide el aumento y el decremento de la energía de la señal. Nos permite distinguir entre sonidos sostenidos y no sostenidos, como los sonidos percusivos. Su cálculo está basado en el siguiente modelo de envolvente temporal, el cual comienza en el punto máximo de energía de la envolvente (t_{max}):

$$S(t) = A \cdot \exp(-\alpha(t - t_{max})) \quad t > t_{max} \quad (6.2)$$

α se calcula por regresión lineal en el logaritmo de la envolvente de energía de la señal.

- Temporal Centroid, Centroide Temporal [TC]: Es el promedio de tiempo de la energía de la envolvente de la la señal:

$$tc = \frac{\sum_t e(t) \cdot t}{\sum_t e(t)} \quad (6.3)$$

2_Características temporales instantáneas:

- Auto-Correlation, Auto-Correlación [XCORR]: La correlación cruzada representa la distribución espectral de la señal en el dominio del tiempo:

$$xcorr(k) = \frac{1}{x(0)^2} \sum_{n=0}^{N-k-1} x(n) \cdot x(n+k) \quad (6.4)$$

- Zero-Crossing rate, Cruce por Cero [ZCR]: Mide el número de veces que la señal cruza por cero. Los sonidos periódicos tienden a tener un valor pequeño de este, mientras que los sonidos muy ruidosos tienden a tener un valor alto.

3_Características de energía:

- Total Energy, Energía Total [TOT]: Estima la potencia de la señal en un tiempo dado.
- Harmonic Part Energy, Energía de la parte Armónica [HARMO]: Estima la potencia de la parte armónica de la señal en un tiempo dado. Es estimado desde la amplitud de la parte armónica en un tiempo dado.
- Noise Part Energy, Energía del ruido [NOISE]: Estima la potencia de la parte ruidosa de la señal en un tiempo dado. Es estimado del resultado obtenido de la sustracción de la parte armónica a la señal en un tiempo dado.

4_Características espectrales:

- Spectral Centroid, Centroide Espectral [SC]: El centroide espectral es el baricentro del espectro. Indica la media ponderada de la potencia espectral. Está relacionado con la percepción del brillo del sonido. Este se computa considerando el espectro como una distribución, cuyos valores son las frecuencias y las probabilidades de observarlas son la amplitud de estas, normalizadas:

$$\text{Donde: } \mu = \int x \cdot p(x) \, dx$$

x es el dato observado: $x = freq_v(x)$

$p(x)$ es la probabilidad de observar x : $p(x) = \frac{ampl_v(x)}{\sum_x ampl_v(x)}$

- Spectral Spread, Propagación Espectral [SS]: Se define como la propagación del espectro alrededor de su valor medio.
- Spectral Skewness, Asimetría Espectral [skew]: Nos da la medida de la asimetría de una distribución al rededor de su valor medio. Este se computa del momento de 3^o orden:

$$m_3 = \int (x - \mu)^3 \cdot p(x) \delta x$$

Por lo que el valor de SK es: $\gamma_1 = \frac{m_3}{\sigma^3}$ (6.5)

Este descriptor nos proporciona el grado de asimetría de la distribución, siendo:

SK = 0 ; indica una distribución simétrica.
SK < 0 ; indica mas energía a la derecha.
SK > 0 ; indica mas energía a la izquierda.

- Spectral Slope, Pendiente Espectral [Slope]: Representa el aumento y decremento de la amplitud espectral. Se computa mediante la regresión lineal de la amplitud espectral:

$$\hat{a}(f) = slope \cdot f + const$$

Donde:

$$slope = \frac{1}{\sum_k a(k)} \frac{N \sum_k f(k) * a(k) - \sum_k f(k) * \sum_k a(k)}{N \sum_k f^2(k) - \left(\sum_k f(k) \right)^2} \quad (6.6)$$

- Spectral Decrease, Decremento Espectral [decr]: Este descriptor también representa el aumento y decremento de la amplitud espectral. Su formulación viene derivada de estudios perceptuales:

$$decrease = \frac{1}{\sum_{k=2:K} a} \sum_{k=2:K} \frac{a(k) - a(1)}{k - 1} \quad (6.7)$$

- Spectral Roll-off, Roll-off Espectral [rolloff]: Es la frecuencia para la que, el 95% de la energía de la señal, es contenida por debajo de esta:

$$\sum_0^{fc} a^2(f) = 0.95 \sum_0^{sr/2} a^2(f) \quad (6.8)$$

Donde: fc es la frecuencia del factor Roll-off, $sr/2$ es la frecuencia de Nyquist.

- Spectral Variation o Spectral Flux, Variación Espectral o Flujo Espectral[Var]: Representa el aumento de variaciones del espectro a lo largo del tiempo. Se computa mediante la correlación cruzada normalizada entre dos espectros de amplitud sucesivos. Si el valor es cercano a cero los espectros son similares, si es cercano a 1 son muy distintos:

$$variation = 1 - \frac{\sum_k a(t-1, k) \cdot a(t, k)}{\sqrt{\sum_k a(t-1, k)^2} \sqrt{\sum_k a(t, k)^2}} \quad (6.9)$$

- Mel Frequency Cepstral Coefficients, Coeficientes Cepstrales de Frecuencia MEL [MFCC]: Representan la forma del espectro con muy pocos coeficientes. Es derivado de la representación cepstral pero, esta basado en la escala de frecuencias MEL, la cual aproxima la respuesta del sistema auditivo humano.

5_Características armónicas:

- Noisiness, Ruidosidad [noisiness]: Es la relación entre la energía del ruido y la energía total:

$$\text{noisiness} = \frac{\text{ener_noise}}{\text{ener_tot}} \quad (6.10)$$

- Inharmonicity, Inarmonicidad [Inharmo]: Representa la divergencia de los componentes espectrales de una señal puramente armónica.
- Harmonic Spectral Deviation, Desviación Espectral Armónica [Desv]: Es la desviación de la amplitud de los picos armónicos de la envolvente espectral global:

$$HDEV = \frac{1}{H} \sum_h (a(h) - SE(h)) \quad (6.11)$$

Donde: H es el número total de armónicos que se han considerado, $a(h)$ la amplitud del h -ésimo armónico, $SE(h)$ la amplitud de la envolvente espectral evaluada a la frecuencia $f(h)$.

- Odd to Even Harmonic Energy Ratio, Ratio de Energía de Armónicos Pares a Impares [oeratio]: Nos permite distinguir sonidos con armónicos impares de energía predominante, de otros con energía armónica igualmente importante:

$$OER = \frac{\sum_{h=1:2:H} a^2(h)}{\sum_{h=2:2:H} a^2(h)} \quad (6.12)$$

- Tristimulus [tri]: Este indicador es un equivalente en el timbre, como los atributos del color lo son en la visión. Este nos proporciona tres tipos diferentes de relaciones de energía, que nos permite una descripción bien ajustada del primer armónico del espectro:

$$\begin{aligned}T1 &= \frac{a(1)}{\sum_h a(h)} \\T2 &= \frac{a(2) + a(3) + a(4)}{\sum_h a(h)} \\T3 &= \frac{\sum_{h=5:H} a(h)}{\sum_h a(h)}\end{aligned}\tag{6.13}$$

6.5.2 Diseño basado en la correlación de atributos

Debemos tener en cuenta que los sonidos presentan atributos que están correlacionados entre sí [54]. Esto supone que si cambiamos uno de ellos, los otros atributos relacionados, también cambiarán. Las propiedades y atributos de un sonido son consecuencia de las propiedades de la forma del objeto o de las características de su excitación. Por lo tanto, esta correlación entre atributos nos ayuda a construir un mapa de características del sonido, asociadas entre sí, como por ejemplo: un sonido con un volumen alto está relacionado con más cantidad de brillo. Sabiendo que el brillo está relacionado con el centroide espectral podemos categorizar y mapear este hacia las características de la fuente. Los atributos correlacionados entre sí deben de ser tratados de manera conjunta, formando conjuntos de atributos.

Esta característica del sonido y de la fuente sonora, derivan en un tipo de diseño que se puede afrontar de dos maneras:

- Como un diseño constructivo
- Como un diseño destructivo

Mediante un diseño constructivo, los atributos correlacionados entre sí, deben ser construidos guardando las proporciones entre atributos, con el fin de preservar el carácter de un sonido determinado. Por el contrario, rompiendo esa relación entre atributos podemos conseguir interesantes sonidos y diseñar objetos sonoros que no estén sujetos a un carácter fijo o predeterminado, por ejemplo sería el caso de efectos de sonidos.

6.5.2 Estructura de la etapa de mapeo

En la siguiente figura 6.29 esta representada la estructura de la etapa de mapeo.

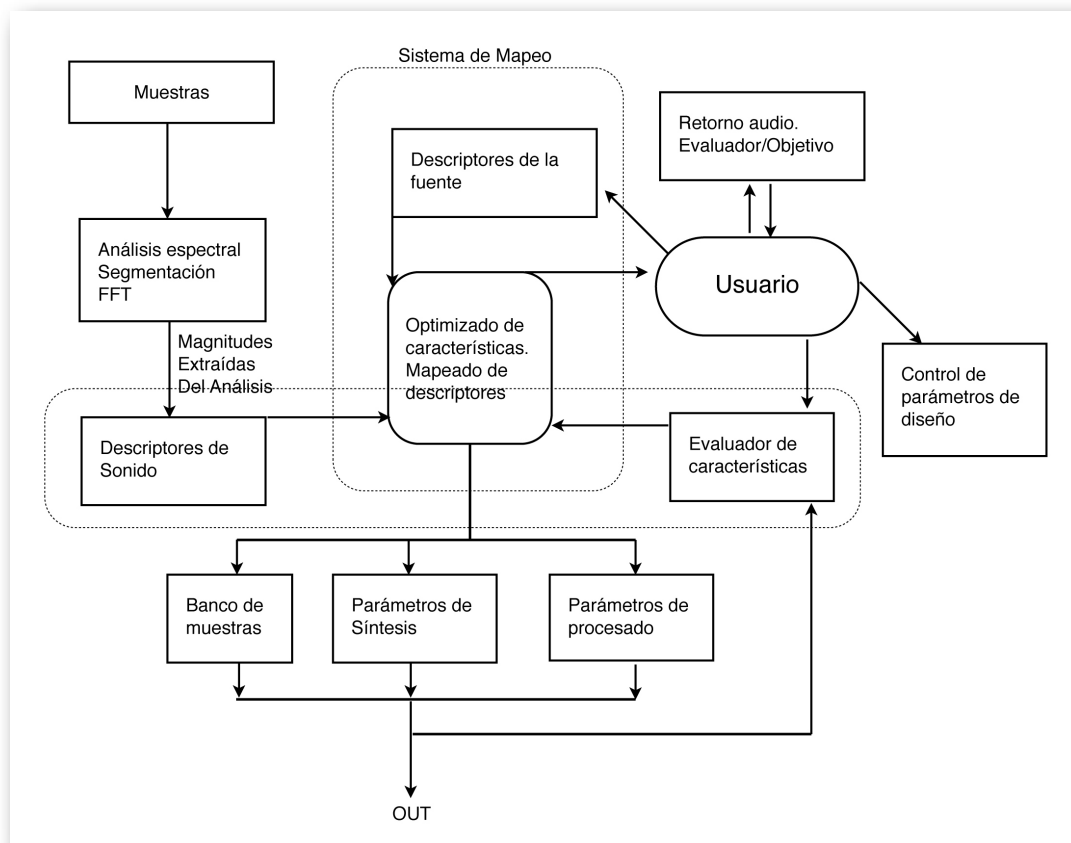


Figura 6.29. Diagrama de la etapa de mapeo.

En esta etapa el mapeo de descriptores esta condicionado a los siguientes procesos:

- El usuario selecciona los descriptores correspondientes a la fuente u objeto sonoro que se pretende diseñar.
- Las muestras del banco de muestras son analizadas y descritas mediante los descriptores de audio propuestos.
- Los descriptores de la fuente proporcionan una o varias muestras.
- Los datos y metadatos asociados a cada muestra proporcionan información de la muestra al sistema y al usuario. Esta ha sido analizada mediante el uso de descriptores.
- Esta información es evaluada y optimizada.
- La salida de la etapa de optimización nos proporciona parámetros de control sobre la síntesis, procesado y el banco de muestras.
- El usuario evalúa los parámetros mediante el control del modulo de síntesis y procesado.
- La salida de esta etapa es retornada al usuario y se utiliza la muestra para determinar y corregir el resultado.
- El usuario compara los resultados. Si no es el resultado adecuado, el sistema se retroalimenta, proporcionando nuevos parámetros de control.

7

Sonorización de varios clips de audio. Experimentación y resultados

En este capítulo se expondrá el procedimiento y técnicas empleadas para llevar a cabo la sonorización de varios clips de video. Mediante esta sonorización se pretende poner en práctica el modelo de análisis/diseño que se ha desarrollado en el capítulo 6, y que es uno de los objetivos de este proyecto. El punto de partida son tres clips de video sin sonorizar (ver contenido en el DVD adjunto

al proyecto), escogidos por presentar una gran cantidad de objetos sonoros, a la vez de servirnos para ejemplificar las conclusiones expuestas en el modelo.

7.1 Equipamiento

Para el desarrollo de la sonorización se estudió la instrumentación necesaria para alcanzar nuestro objetivo bajo las mejores condiciones. En primer lugar se expondrá un diagrama (figura 7.1) del sistema y equipo utilizado.

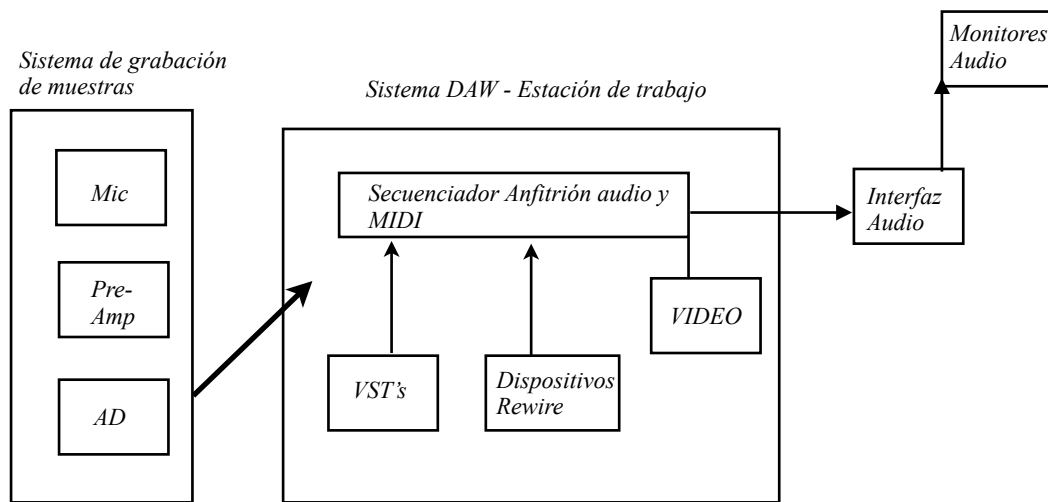


figura 7.1 . Diagrama del equipamiento utilizado

A continuación se describen cada uno de los bloques del diagrama:

- Sistema de grabación de muestras: el sistema está formado por un micrófono, un previo y un convertor AD/DA.
- Microfonía: Micrófono de válvulas con patrón cardioide.
- Previo de micrófono.
- Convertor AD/DA.
- Sistema DAW: El sistema compuesto por un programa anfitrión para secuenciar audio y video de forma no lineal, un conjunto de efectos y

sintetizadores de audio bajo la tecnología VST, un programa de síntesis que se sincroniza con el anfitrión mediante tecnología Rewire y un dispositivo de monitorización de video sincronizado con el programa anfitrión.

- Interface de audio.
- Sistema de monitores estéreo.

Los programas y equipos han sido escogidos por presentar las características de procesamiento y edición que se exponen como herramientas en el modelo. A continuación se expone el listado del equipo utilizado:

- Secuenciador Logic Pro 8
- Instrumentos virtuales incluidos en Reason 3: Sampler, sintetizador sustractivo, vocoder, secuenciador por pasos para generar secuencias.
- Sintetizador FM8, Native Instruments
- Sintetizador Massive, Native Instruments
- Sintetizador Reaktor 4, Native Instruments
- Sampler Kontakt 4, Native Instruments
- Procesador de impulsos de reverberación por convolución Altiverb 6, AudioEase
- Modulo de Mastering Tracks IK Multimedia: Ecu, Dinámica y visor de espectro y fase.
- Delay espectral, Spectral Delay, Native Instruments
- Ecuadorizador Sonalksis
- Compresor Sonalksis
- DAW MAC Book Pro
- Interfaz de audio, Motu Ultralite MKIII
- Monitores de respuesta plana Monitor 8, DAS
- Micrófono de válvulas Peluso CK12, patrón cardioide
- Previo de micrófono John Hardy M1
- Conversor AD/DA Aurora 16, Lynx
- Sistema monirotaje 2.1 y 2.0 para mastering Earthquake Tigris
- 2 Etapas de amplificación Lynn

7.2 Metodología y técnicas

En esta sección se describe la metodología empleada en el proceso de sonorización. Tendremos en cuenta principalmente tres:

1.Fase de análisis: En esta primera fase se estudia cada clip de audio. Primero se determinan las escenas, a continuación, se estudia la escena siguiendo tres premisas:

- Análisis del entorno
- Localización de los objetos sonoros
- Análisis detallado de cada evento sonoro

Para esta fase de localización y análisis se han tenido en cuenta las consideraciones propuestas en el modelo. Los eventos sonoros han sido estudiados siguiendo un estudio de su estructura y clasificados según la taxonomía planteada en este trabajo.

2.Fase de síntesis: Después de concluir con el análisis disponemos de la suficiente información para describir y diseñar los objetos requeridos por la escena. Esta fase de síntesis, está basada en una técnica híbrida de síntesis de sonido, grabación y edición de muestras. Por lo tanto, se lleva a cabo el diseño de los eventos mediante la creación de muestras primitivas de audio. Estas se componen y disponen teniendo en cuenta en todo momento la imagen como referente de sincronía.

3.Fase de Procesado: Este proceso se aplica en dos partes: en la primera se aplica una vez se ha completado cada una de las partes que forman la escena, con el fin de unificar elementos comunes. La segunda se consigue, aplicando procesos, de forma individual a cada muestra o conjunto de ellas.

7.3 Procedimiento y aplicación práctica del modelo mediante la sonorización de varios clips de video

En este apartado se expone el procedimiento práctico que se ha llevado a cabo en la sonorización de los vídeos. Para ello se han elegido dos secciones de una

duración limitada, pero que nos ayudaran a ejemplificar como se ha llevado a cabo todo el proceso.

7.3.1 Sonorización de la sección “Peripetics” [2:30:20 / 3:01:12]¹

La sección elegida forma parte del cortometraje Peripetics, el cual, ha sido sonorizado al completo en este proyecto. La duración de la sección escogida es de aproximadamente 29 segundos. El proceso de sonorización se ha producido mediante la puesta en práctica del modelo de análisis/diseño propuesto en este trabajo.

Paso 1_ Análisis de la escena

El primer paso conlleva un análisis de la escena. Este pasa por la descripción de cinco fases de análisis: espacial, de movimiento, del objeto, del entorno y psicoacústico.

- Análisis espacial: La escena se localiza dentro de un espacio cerrado. Su geometría responde a la de un cubo, por lo que, el espacio se entiende como una habitación cerrada. Durante toda la duración del clip, solo se localiza un objeto dentro de espacio indicado. Este se identifica como un objeto sonoro. No se localizan ni eventos ni más objetos sonoros dentro de la escena.
- Análisis de los Objetos: En el caso que nos procede, solo hemos identificado un objeto, por lo que, pasaremos a realizar su análisis. El objeto es identificado como un objeto abstracto y poco definido. Como estamos trabajando con un cortometraje realizado mediante infografía y animación, debemos caracterizar el objeto basándonos en la información visual y el contexto de la escena. En este caso, deducimos que el objeto guarda cierta relación con el sistema respiratorio humano, debido al color de este y textura, así como, al propio contexto del cortometraje. Sin embargo, el movimiento y la

¹ Tiempo de inicio y fin de la sección de video expresada en minutos, segundos y centésimas

formación-deformación del objeto, son los marcadores que nos llevan a describir con mayor facilidad a este. El objeto se expande y contrae por lo que, identificamos el objeto sonoro como un sonido de aspiración y expiración de aire. Definido el objeto, pasamos a la descripción taxonómica. Se trata de un evento sonoro compuesto, no simple. ya que el objeto se puede descomponer en varias unidades más simples. Estas unidades son sonidos aerodinámicos continuos a ráfagas. Su duración es extraída a partir de la imagen. La sincronización con los movimientos del objeto debe de ser cuasi-precisa y sincronía. La intensidad de las ráfagas varían en función a la intensidad de expansión-contracción que refleja el objeto en la escena.

- Análisis de los movimientos: El movimiento del objeto se identifica como un movimiento natural acelerado y decelerado. Se repite formando un bucle.
- Análisis del entorno: El paisaje sonoro de esta escena viene determinado por un solo objeto y la reverberación característica de la habitación en la que se sitúa.
- Análisis psicoacústico: El desarrollo de las imágenes de la escena transmiten la sensación de continuidad, debido a la repetición de las contracciones en forma de bucle. Los cambios de posición del punto de vista de la cámara, dotan una perspectiva del objeto en la que las expansiones de este cada vez son más exageradas. Esto provoca una sensación de increchendo e intensidad a lo largo de la secuencia.

En la Imagen figura 7.2 se expone una secuencia de imágenes extraídas del clip de video, en las que podemos ver el objeto y diferentes fases de sus movimientos.

Paso 2 Diseño del objeto sonoro

El sonido del objeto está compuesto por 4 pistas de audio. Cada pista contiene una secuencia de muestras de audio que han sido grabadas y procesadas. A continuación pasaremos a la descripción de cada una de ellas.

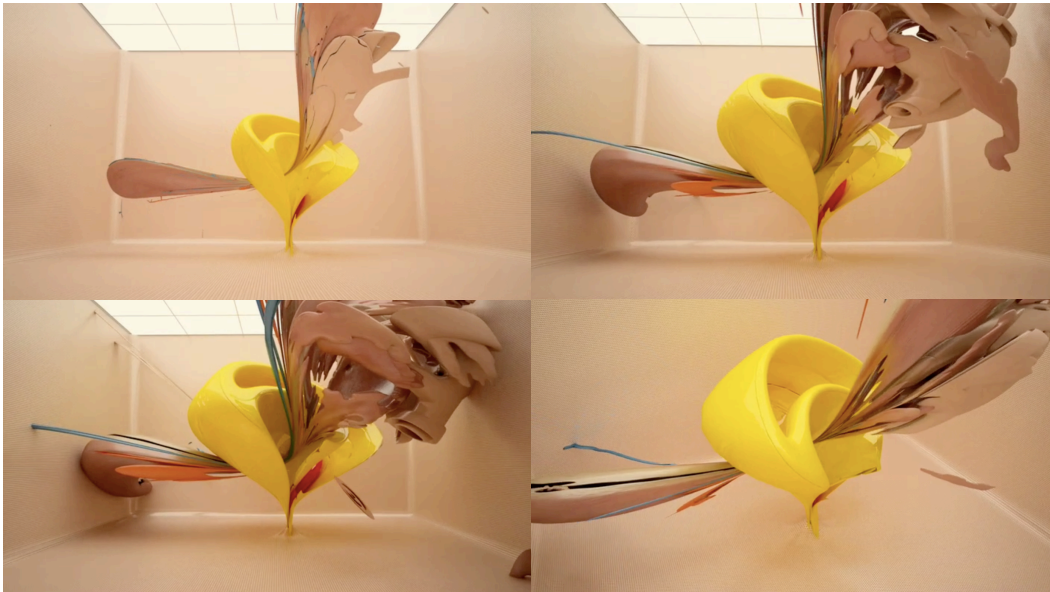


Figura 7.2. Secuencia de imágenes de los diferentes movimientos del objeto.

- Pista 1: Contiene una muestra de audio de un sonido de aspiraciones y espiraciones de aire. La muestra de audio se ha grabado mediante el sistema indicado en la figura 7.1 y con la instrumentación indicada en el apartado 7.1. Esta instrumentación fue escogida debido a que se trata de un equipo de alta calidad, que ofrece una muy buena relación S/N y que presenta una respuesta en frecuencia muy plana. La muestra ha sido editada para eliminar los ruidos y partes no deseadas. Para ello se han utilizado las herramientas de edición del secuenciador Logic 8. Después de esta edición, nuestra muestra de audio se ha convertido en muestras mas pequeñas, que pertenecen a una secuencia de aspiraciones y exhalaciones, como vemos en la figura 7.3. El siguiente paso es corregir los posibles errores de sincronización que puedan haberse producido en la grabación. El paso final, es la etapa de procesado. Se le han aplicado tres procesos a la señal: una ecualización, una compresión y una

reverberación. El ecualizador utilizado es de tipo paramétrico. Con él filtramos paso alto, para recortar frecuencias no deseadas en los sub-graves y realzamos las frecuencias medio-altas para realzar el brillo y dar mas presencia a la muestra, figura 7.4.

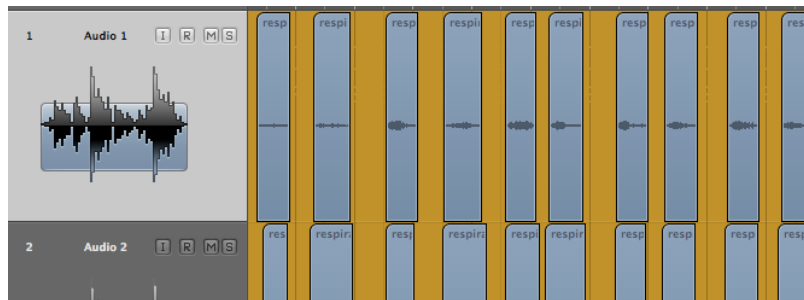


Figura 7.3. Muestras de la pista 1.

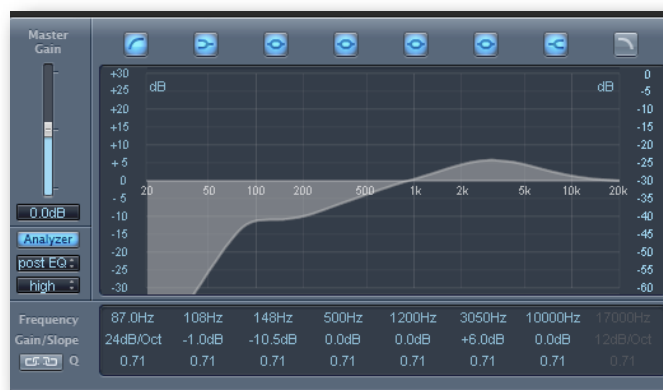


Figura 7.4. Ecuación aplicada a la muestras de la pista 1.

Con el compresor ajustamos la velocidad de los transitorios de la muestra. Esto nos permite dotar al objeto de cierta aceleración y deceleración, así como, aumentar la intensidad de la fuente, en este caso, la ráfaga de aire. Con este efecto, nos da la sensación de que el objeto aumenta de tamaño o esta más cerca a nosotros. Los parámetros del compresor son los que se aprecian en la figura 7.5:



Figura 7.5. Compresión aplicada a la muestras de la pista 1.

Por ultimo, se procesa la señal con un procesador de impulsos de convolución. El impulso elegido es de una reverberación tipo *hall*, con un tiempo de caída de 4 seg.

- Pista 2: Esta pista contiene las muestras de la pista 1, con la diferencia de que estas muestras han sido resintetizadas utilizando un vocoder y un sintetizador sustrativo, con un generador de ruido como excitador. El resultado es una señal con unas características tímbricas algo diferentes a la muestra de la pista 1. Con esto conseguimos que al superponer la dos señales se produzca un efecto de coro que otorga al objeto de más aletoriedad y dinamismo. Para ello, la disposición de las muestras deben no ser siempre síncronas entre si. A esta pista se le aplica los mismos procesos que se le aplicaron a la pista 1: ecualización, compresión y reverberación.

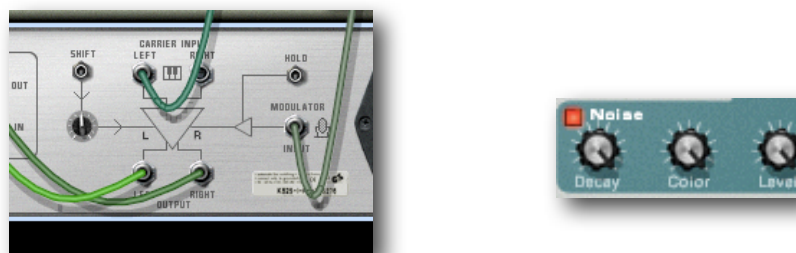


Figura 7.6. Generador de ruido del sintetizador y conexiones del vocoder.

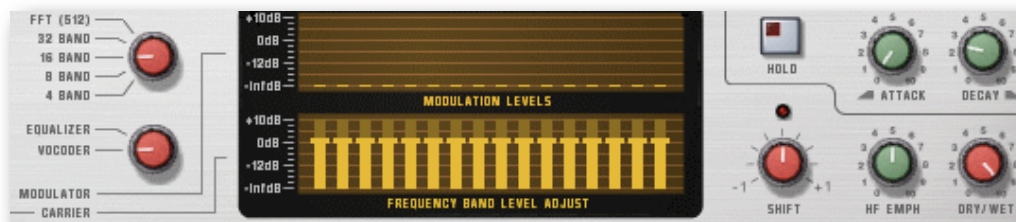


Figura 7.7. Visualizador de análisis del vocoder

Las figuras 7.6 y 7.7 nos muestran las herramientas utilizadas para la resínteis de la muestra.

- Pista 3: Para la creación de las muestras de esta pista se ha utilizado un sintetizador sustrativo y un secuenciador midi, para disparar la secuencia. Con el sintetizador hemos generado un ruido rosa, que ha sido filtrado paso bajo y paso banda. La intención de esto, es generar sintéticamente una exhalación y una inhalación. Los parámetros y ajustes del sintetizador han sido escogidos en base a que necesitamos generar un sonido con solo componente estocastica y que sea filtrado por un resonador. Estos ajustes nos proporcionan un sonido del tipo aerodinámico. Lo siguiente es ajustar la envolvente temporal para que comience con un ataque lento para la inhalación(figura 7.8) y una relajación alta y un ataque medio-rápido para la exhalación.



Figura 7.7. Parámetros de síntesis escogidos para el sonido de la inhalación.

Una vez sintetizados los dos sonidos, la muestra correspondiente a la inhalación pasa a la pista 3 y la de exhalación a la pista 4. En esta pista nos encargaremos de diseñar la parte del objeto sonoro que choca sobre la paredes de la habitación. Para diseñar estos rebotes o choques contra las paredes, utilizamos un retardador espectral. con el fin de crear micro rebotes relacionados con las bandas de frecuencias medias altas. Este efecto evoca la sensación de que las paredes son metálicas o de una superficie muy pulida, y resuenan debido al choque del objeto con las paredes, provocando un eco que rebota en la habitación. Hay que destacar que este efecto es aplicado a esta pista porque se trata de la muestra correspondiente a la inhalación. Es aquí donde el objeto se expande y choca contra las paredes. De nuevo la pista pasa por un efecto de reverberación. Elegimos un impulso de una habitación real, con un tamaño que imita a la que tenemos.

- Pista 4: Como hemos explicado en el paso anterior, la otra muestra procedente del sintetizador se lleva a la pista 4. A esta se le aplican los mismos procesos que a las pistas 1 y 2. Las muestras se sincronizan con las otras muestras de las pistas 1, 2 y 3. Esta pista imita fugas de aire, diseñando así, la descomposición estructural de cada expansión y contracción.

Paso 3 Diseño de los patrones temporales

Como es fácil de observar, el patrón que se da en la imagen corresponde a un evento compuesto que se repite a modo de bucle y que guarda una relación asimétrica entre muestras continuas, es decir, no forman un patrón rítmico que pueda ser medido. Al igual que cuando oímos los pasos de una persona caminando nuestro cerebro lo identifica gracias a la construcción del patrón, la respiración humana, también esta marcada por un patrón fácilmente identificable. Cada una de las muestras de cada pista que corresponden a bucles distintos, son desplazadas en relación a las muestras del bucle anterior. Esto provoca un efecto de dinamismo sobre la imagen, la cual también es dinámica y cambiante.

Paso 4 Mezcal y masterización

El paso final consiste en ajustar el volumen de cada una de las pistas a lo largo de todo el clip. Sobre la mezcla final se han realizado ajustes de masterización², con el fin de depurar la calidad de sonido de la secuencia.

7.3.2 Diseño de un objeto sonoro perteneciente a la secuencia Tierra compuesto por múltiples impactos.

En este apartado se describe el procedimiento llevado a cabo en el diseño de uno de los objetos sonoros del clip de video Tierra. El clip Tierra se ha sonorizado por completo, aunque en este apartado nos centraremos solo en uno de los objetos contenidos en él. Este está formado por la concatenación y superposición de impactos de piedras cayendo sobre otras piedras (figura 7.8).



Figura 7.8 Imagen que muestra piedras cayendo sobre un montículo.

² Procesado sobre la totalidad de una mezcla de audio. Su finalidad es la de corregir errores de ecualización, edición y amplitud máxima.

Al igual que en apartado 7.2.1 de este capítulo, el diseño pasa por una primera fase de análisis y una segunda de diseño. Partiendo de la fase de análisis nos centraremos en primer lugar en describir el objeto en cuestión.

- Descripción del objeto sonoro: Se trata de recrear el sonido de cientos de piedras y arenisca al caer al suelo. Estas son transportadas por una especie de cinta transportadora (la cinta transportadora no aparece en ningún momento en la secuencia de video, aun así, se sobrentiende por contextualización que se trata de este mecanismo) que las vierte a un montículo con más piedras y arenisca, las cuales son del mismo tipo y de formas similares que las que son arrojadas por la cinta. Según muestra la imagen, las piedras tienen un tamaño pequeño-medio, de entre 6 cm de diámetro, a otras muchas más pequeñas, como arenisca o granos.

- Estrategia desarrollada:
 1. Localización de muestras. Este proceso se ha llevado a cabo mediante una técnica de granulación de muertas. Esta técnica consiste en coger un archivo de audio, de una duración determinada, y estrechar su duración temporal, así conseguimos que su espectro sea mucho más rico y su frecuencia fundamental sea mayor. Otra técnica de granulación que hemos usado es la de recortar pequeños paquetes de muestras de muy corta duración, de otras muestras más largas, las cuales mostraban un espectro ruidoso y con alto contenido en altas frecuencias. Las muestras de más duración fueron grabadas utilizando el sistema de grabación que ya se indicó en este capítulo. Las muestras que se grabaron, pertenecen al sonido de una persona tecleando en un teclado de ordenador y la otra al sonido producido cuando arrugamos una bolsa. Una vez elegidas las muestras que más se adecuaban a nuestro objetivo, se procesaron modificando su envolvente temporal con el fin de recrear el mismo comportamiento temporal que presentan los sonidos de impactos. En la figura 7.9 vemos dos de las muestras conseguidas.

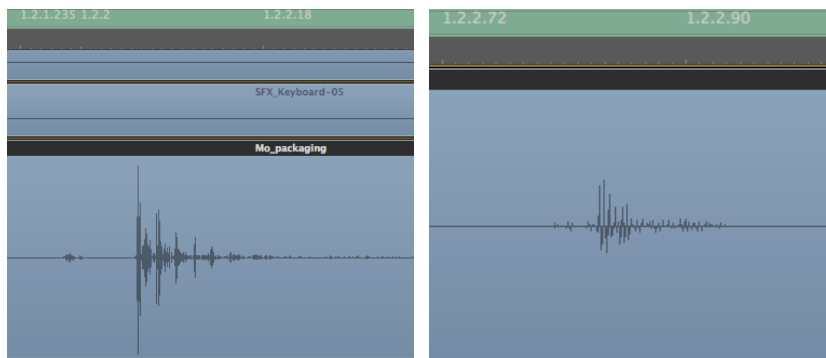


Figura 7.8 Imagen que muestra dos de las muestras finalmente escogidas.

2. Diseño de la evolución temporal. Una vez localizadas las muestras, debemos diseñar la evolución temporal del objeto. Para ello creamos un patrón de repetición con cada una de las muestras. Cada uno de estos patrones se asigna a una pista del secuenciador. Analizando la evolución temporal del objeto sonoro y de la estructura que lo conforma, podemos describirlo como un patrón por postergación espacial, asíncrono y aleatorio. Cada una de las muestras se corresponden con un sonido de impacto, producido por las piedras. Este se supone producido por el mismo tipo de objeto, la piedra, pero con tamaños y fuerzas de impacto distintos. La solución de diseño escogida pasa por trasponer el tono de varias muestras de forma aleatoria (ya que la imagen, nos ofrece una disposición de la caída las piedras también aleatoria y difícil de identificar, consecuencia de la velocidad que estas llevan) y de controlar distintas intensidades, esto se consigue por medio de un filtro paso bajo. Destacaremos que debido a que no se aprecia en la imagen grandes fluctuaciones en la fuerza de los impactos de las piedras, los parámetros de ganancia del filtro fluctúan muy poco entre muestras. Si por el contrario, la diferencia entre la intensidad de los impactos hubiese sido mucho mayor, tendríamos que haber utilizado un procesador de compresión para cada muestras, con el fin de caracterizar el efecto que produce una mayor intensidad de excitación.

3. Superposición de las pistas de muestras de micro-impactos. Una vez realizado este proceso para cada una de las pistas, el siguiente paso es el de superponer cada una de estas y realizar un ajuste de los volúmenes de cada una de las pistas. El objetivo es empastar todos los sonidos de impactos, recreando el sonido de un vertido de piedras. En este momento las muestras más pequeñas (muestras primitivas) son reconstruidas, como si de una orquesta se tratase, formando en su conjunto, un nuevo sonido que es aceptado por nuestro cerebro como un objeto sonoro único y distinto.
4. Diseño de unidad complementaria 1. Este objeto sonoro no solo está formado por los impactos de las piedras al caer al suelo, sino que durante la caída, estas chocan entre sí y junto con la arenisca, producen rozamientos e impactos de poca intensidad, pero que serían notablemente audibles. Este sonido se ha diseñado a partir del generador de ruido de un sintetizador. La secuencia ha sido generada con un secuenciador Midi por pasos, reproducidos de forma aleatoria. La envolvente temporal ha sido ajustada con valores de sostenimiento y caída iguales a 0 y con tiempos de ataque rápidos y tiempos de relajación más largos. El resultado obtenido es una secuencia de ruido, con fluctuaciones rápidas en su envolvente. Por último ha sido ecualizando, realzando altas frecuencias y recortando frecuencias graves como vemos en la figura 7.9.

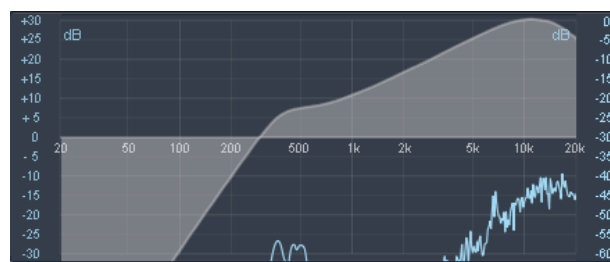


Figura 7.9 Imagen de la ecualización aplicada a la muestra.

5. Diseño de unidad complementaria 2. El último sonido que hemos diseñado para complementar el objeto se corresponde con los “aludes” que se aprecian en el montículo, provocando el

derrumbamiento de este. Este sonido, se describe como un sonido de rozamiento continuo. Para ello hemos utilizado un sampler y hemos estirado una de las muestras anteriores por transposición. Este efecto nos ha ayudado a bajar la frecuencia fundamental y a variar su envolvente. A partir de la muestra “madre” hemos creado otras muestras “hija” . Posteriormente filtramos paso bajo, para realzar las frecuencias bajas. Con esto conseguimos caracterizar los desplazamientos de rocas, que suelen provocar sonidos muy ruidosos y graves debido a su gran tamaño.

8

Conclusiones

Este capítulo tiene como finalidad realizar una evaluación crítica y constructiva de los análisis, resultados y aplicaciones prácticas que se han realizado en el presente proyecto. También serán planteadas las conclusiones tomadas por el autor a cerca del trabajo realizado y la evolución del mismo, así como de las futuras líneas de trabajo.

8.1 Objetivos del proyecto. Revisión y valoración

Comenzaremos replanteando cuales han sido los objetivos y propuestas de los que partía este proyecto:

- Análisis de los campos relacionados con el diseño de sonido en entornos de interacción audio-imagen.
- Estudio y análisis de las técnicas de síntesis de sonido.
- Estudio y análisis de técnicas de procesamiento de audio.
- Propuesta de un modelo de análisis /diseño para la producción de sonido en entornos audiovisuales.
- Desarrollo del modelo.
- Aplicación practica del modelo mediante la sonorización de varias secuencias de video.

Una vez planteados y revisados los objetivos preliminares del proyecto, comparamos estos con el trabajo y los resultados que se han obtenido a lo largo de esta memoria. La evaluación que podemos hacer del proyecto es que ha logrado desarrollar todos los objetivos propuestos.

No obstante, creemos conveniente evaluar más detalladamente cada uno de los planteamientos preliminares con los resultados obtenidos y descritos en la memoria de este proyecto.

- Análisis de los campos relacionados con el diseño de sonido en entornos de interacción audio-imagen. Este primer análisis pretende, mediante el estudio del campo en cuestión, extraer datos de interés, así como conocer en profundidad métodos, técnicas y procedimientos empleados en esta disciplina. Este análisis ha sido desarrollado y descrito en el capítulo 2 de esta memoria. En él se han estudiado los antecedentes históricos del

diseño de sonido de cuyo análisis hemos extraído una visión más completa de las técnicas y metodología que hoy en día se emplean en este campo. También se han estudiado tres grandes campos audiovisuales donde el diseño de sonido está más presente (cine, videojuegos y realidad virtual), las características propias de cada uno y las diferencias existentes entre ellos. Para concluir se ha llevado a cabo un estudio de la metodología, técnicas, herramientas y jerarquías empleadas para la producción y diseño de sonido en estos campos. Por tanto, podemos decir que de este capítulo se han extraído importantes resultados que se han utilizado como parámetros a la hora de proceder en el modelo y prácticas que este proyecto plantea.

- Estudio y análisis de las técnicas de generación de sonido, procesamiento de audio y de análisis espectral. A través de este estudio se ha conseguido realizar un buen análisis comparativo de las técnicas y herramientas de las que se dispone actualmente en el campo del diseño de sonido. En la memoria se ha descrito el funcionamiento de las técnicas de síntesis estudiadas así como de procesamiento de audio. Además, se ha realizado un buen estudio sobre el análisis espectral de sonido. Para ello se han comparado diversas técnicas de análisis, como son el análisis basado en la transformada de Fourier y el análisis basado en la transformada Wavelet. En ambos casos se han extraídos conclusiones derivadas de las características que ofrecen cada uno de ellos y sus posibles aplicaciones en este proyecto.
- Desarrollo de un modelo análisis/diseño. El modelado de un posible sistema para la creación y diseño de sonido de una manera más intuitiva se propuso como uno de los objetivos más importantes de este proyecto. Por lo tanto, podemos afirmar que uno de los objetivos principales de este proyecto ha sido cumplido. Se ha tenido en cuenta tanto los aspectos técnicos, más relacionados con

la síntesis de sonido, como los aspectos cognitivos del proceso de diseño. Se ha realizado una exhausta exploración taxonómica de los tipos de sonidos que podemos encontrar en la naturaleza. La solución que se ha planteado para poder automatizar el modelo de una manera intuitiva ha sido mediante la proposición de un mapa ontológico del proceso de análisis, basado en la información que nos proporciona la escena. Los datos extraídos en este análisis se mapean con atributos y parámetros característicos de diseño, proporcionando al usuario el tipo de diseño más adecuado en cada momento.

- Puesta en práctica del modelo. La evaluación del modelo propuesto se ha llevado a cabo mediante la sonorización de varios clips de audio, la cual se ha puesto en práctica mediante el uso de herramientas comerciales de tratamiento de audio y síntesis. Durante la sonorización se ha planteado la misma estrategia de análisis y diseño que el modelo propone. Podemos evaluar los resultados obtenidos como favorables, ya que se ha conseguido recrear con éxito todas las escenas de los clips de vídeo. También añadiremos que con la sonorización se ha pretendido ejemplificar un gran número de tipos de sonidos (impactos, ambientes, rozamientos, líquidos...) y creemos que esto ha sido conseguido con éxito. Otro de los retos que planteaba la puesta en práctica del modelo era el de proporcionar suficiente información al usuario para que el diseño fuese lo más intuitivo posible. Debido a que el modelo no ha sido implementado computacionalmente no ha podido ser probado por más usuarios. Esto plantea un problema a la hora de evaluar el grado de efectividad de este. Sin embargo, los resultados expuestos creo que son suficientemente convincentes como para poder evaluar favorablemente el sistema planteado.

8.1 Futuras líneas de trabajo

En referencia a la evolución de este trabajo se plantearán posibles mejoras y líneas de trabajo, que pueden formar parte de futuros proyectos para el desarrollo de aplicaciones relacionadas con este trabajo:

1. Implementación computacional del modelo propuesto en el proyecto.
2. Ampliación de la taxonomía de sonidos propuesta.
3. Ampliación de la ontología planteada a partir de nuevas líneas de análisis.
4. Implementación de un procesador de eventos temporales. Hoy en día solo es posible realizar esta tarea mediante el uso de secuenciadores y efectos de audio como el *delay*. Sin embargo, el planteamiento de este tipo de procesos en el modelo y el resultado obtenido en la sonorización, hacen de este una buena herramienta que debería ser estudiada con más detenimiento.
5. Creación de un banco de muestras de construcción acorde a las características representadas en la memoria.
6. Implementación de un programa de análisis de muestras de audio mediante el uso de la transformada Wavelet

B

Bibliografía

- [1] Gaver, William W.
Everyday Listening and Auditory Icons.
PhD. University of California, San Diego. 1988

- [2] Miner, N. A Thomas P. Caudell
Using Wavelets to Synthesize Stochastic-based Sounds for Immersive Virtual Environments.
Proceedings of the International Conference on Auditory Displays, Santa Clara, CA, Nov. 1997

- [3] Miner, N.
A Wavelet Approach to Synthesizing Perceptually Convincing Sounds for Virtual Environments and Multi-Media.
PhD dissertation. University of New Mexico. 1998

- [4] Misra ,Ananya. Cook, Perry R. Wang, Ge
A new paradigm for sound desing.
Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06),
Montreal, Canada, September 18-20, 2006
- [5] CLOSED Project, IRCAM, Univerona, HGKZ.
*Closing the Loop of Sound Evaluation and Design. Everyday sound
classification: Sound perception, interaction and synthesis.*
<http://closed.ircam.fr/> 2006
- [6] Van den Doel, Kees.
Physically-Based models for liquid sounds.
Proceedings of ICAD 04-Tenth Meeting of the International Conference
on Auditory Display, Sydney, Australia 2004
- [7] Cook, Perry R.
*Modeling Bill's gait: analysis and parametric synthesis of walking
sounds*
22nd International Conference on Virtual, Synthetic and Entertainment
Audio, Finlandia. 2002
- [8] Van den Doel, Kees. Kry, Paul G. Pai, Dinesh K.
*FoleyAutomatic: Physically-based Sound Effects for Interactive
Simulation and Animation.*
Computer Graphics (ACM SIGGRAPH 01 Conference Proceedings), pp.
537-544, 2001.
- [9] O'Brien, James F. Cook, Perry R. Essl, Georg.
Synthesizing Sounds from Physically Based Motion.
Computer graphics proceedings , annual conference series. 2001.
- [10] Aramaki, Mitsuko. Gondre, Charles. Kronland-Martinet, Richard.
Voinier, Thierry. Ystad, Sølvi.
Thinking the sounds: An intuitive control of an impact sound synthesizer.
Proceedings of the 15th International Conference on Auditory Display,
Copenhagen, Denmark. 2009.
- [11] Schaeffer, Pierre.
Tratado de los objetos musicales.
Alianza Editorial. 1977.

- [12] Chowning, J.
The synthesis of complex audio spectra by means of frequency modulation.
Journal of the audio engineering society. 1973
- [13] Roads, Curtis.
Microsound.
The MIT Press. 2001.
- [14] Roads, Curtis.
Granular synthesis of sound.
In C. Roads and J. Strawn, eds. 1985. Foundations of Computer Music.
Cambridge, Massachusetts: MIT Press. pp. 145±159. 1985
- [15] Strawn, John. McGill, James F.
Digital audio engineering: An anthology.
Computer music and digital audio series; [v. 3]. W. Kaufmann, Los
Altos, California. 1985.
- [16] Moore, F. Richard.
Elements of computer music.
Prentice Hall, Englewood Cliffs, N.J., 1990.
- [17] Serra, Xavier.
Current perspectives in the digital synthesis of musical sound.
Formats, 1, 1997.
- [18] Roads, Curtis.
Musical signal processing.
Studies on new music research 2. Swets & Zeitlinger, Lisse Netherlands;
Exton PA, 1997.
- [19] Roads, Curtis.
The computer music tutorial.
MIT Press, Cambridge, Mass., 4th edition, 1999
- [20] Dodge, Charles. Jerse, Thomas A.
Computer music: synthesis, composition, and performance.

- Schirmer Books; Prentice Hall International, New York London, 2nd edition, 1997.
- [21] Smith, Julius O.
Viewpoints on the history of digital synthesis.
In Proceedings of the 1991 International Computer Music Conference, San Francisco, CA, 1991.
- [22] Bank, Balázs. Márkus, János. Nagy, Attila. Sujbert, László.
Signal and physics-based sound synthesis of musical instruments.
Periodica Polytechnica Ser. El. Eng., 47(3-4):269-295, 2004
- [23] Schafer, R. Murray.
The Tuning of the World (The Soundscape).
Arcana Editions.1997.
- [24] Wishart, Trevor.
On Sonic Art.
Hardwood academic publishers. 1996.
- [25] Sarkar, Mihir. Lan, Cyril. Diaz, Joseph. Vercoe, Barry.
The effect of musical experience on describing sounds with everyday words.
157th Meeting of the Acoustical Society of America (ASA), Portland, OR, May 18-22, 2009.
- [26] Sarkar, Mihir.
Perceptual Synthesizer.
Project presentation, MAS.641, MIT, Cambridge, Mass., USA, Dec. 2005.
- [27] Sarkar, M. Vercoe, B. Yang, Y.
Words that Describe Timbre A Study of Auditory Perception Through Language.
Language and Music as Cognitive Systems Conference (LMCS-2007), Cambridge, UK, May 11-13, 2007.
- [28] Rocchesso, Davide. Fontana, Federico.
The Sounding Object

- Edizioni di Mondo Estremo “The Sounding Object Project,” Multi-lab EU project as part of “The Disappearing Computer,” <http://www.soundobject.org>. 2003.
- [29] Avanzini, Federico.
Interactive Sound.
Department of Information Engineering, University of Padua edited in Sound to Sense, Sense to Sound. Logos Verlag Berlin. 2008.
- [30] Donnadieu, Sophie.
Mental Representation of the Timbre of Complex Sounds.
Edited in Analysis, Synthesis, and Perception of Musical Sounds James W. Beauchamp. Editor: University of Illinois at Urbana, USA.
- [31] Avanzini, Federico. Rocchesso, Davide.
Controlling Material Properties in Physical Models of Sounding Objects.
In Proc. Int. Computer Music Conf., La Habana. 2001.
- [32] Klatzky, R. L. Pai, D. K. Krotov, E. P.
Perception of Material from Contact Sounds.
Presence 9(4), 399–410. Lutfi, R. A. and E. L. Oh (1997).
Auditory Discrimination. 2000.
- [33] Masri, P. Bateman, A. Canagarajah, N.
A review of time-frequency representations, with an application to sound/ music analysis-resynthesis.
Organised Sound 2(3): 193±205. 1997.
- [34] Chion , Michel.
Audio-Vision: Sound on Screen.
New York: Columbia University Press. 1994.
- [35] Gibson, Jame. J.
The Ecological Approach To Visual Perception.
Lawrnce Erlbaum Associates. 1986.
- [36] Serra, X. Smith, J.
Spectral modeling synthesis: a sound analysis/synthesis system based on a deterministic plus stochastic decomposition.
Computer Music Journal 14(4): 12±24. 1990.

- [37] Zölzer, Udo.
DAFX - Digital Audio Effects.
Edited by ISBN: 0-471-49078-4 John Wiley & Sons. 2002.
- [38] Gomes, Jonas.
Time Warping of Audio Signals.
Siome Goldenstein University of Pennsylvania
IMPA–Instituto de Matemática Pura e Aplicada. June 18, 1999.
- [39] Rodet, Xavier. Scharz, Diemo.
Spectral Envelopes and Additive + Residual Analysis/Synthesis.
Edited in Analysis, Synthesis, and Perception of Musical Sounds, The
Sound of Music. James W. Beauchamp.
Editor University of Illinois at Urbana, USA. Springer. 2007
- [40] Mallat, Stéphane.
A wavelet tour in signal processing.
Elsevier. 2009.
- [41] Visella, F. Fontanac, F. Giordanod, B. L. Nordahle, R. Serafine, S.
Bresinf, R.
Sound design and perception in walking interactions.
Elsevier. 2007.
- [42] Flanagan, J. L. Golden, R.
Phase vocoder.
Bell System Technical Journal 45: 1493±1509. 1966.
- [43] Klatzky, Roberta L. Pai, Dinesh K. Krotkov, Eric P.
Perception of Material from Contact Sounds.
The MIT Press, 2000.
- [44] Giordanoa, Bruno L. McAdamsa, Stephen.
*Material identification of real impact sounds: Effects of size variation in
steel, glass, wood, and plexiglass plates*.
Acoustical Society of America. 2006.

- [45] Serra, X.
A system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition.
Stanford: Center for Computer Research in Music and Acoustics, Department of Music, Stanford University. 1989.
- [46] Lee, Jung Suk. Depalle, Philippe. Scavone, Gary.
Analysis/synthesis of rolling sounds using a source-filter approach.
Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10), Graz, Austria, September 6-10, 2010.
- [47] Roads, C.
Sound transformation by convolution.
Musical Signal Processing, C. Roads, S.T. Pope, A. Piccialli, & G. De Poli (Eds.). Lisse: Swets & Zeitlinger, 411-438. 1997.
- [48] Aramaki, M. Besson, M. Kronland-Martinet, R. Ystad, S.
Timbre perception of sounds from impacted materials: behavioral, electrophysiological and acoustic approaches.
In Computer Music Modeling and Retrieval – Genesis of Meaning of Sound and Music, Springer Berlin, Heidelberg, 2009.
- [49] Walker, Bruce N. Kramer, Gregory. Lane, David M.
Psychophysical Scaling of Sonification Mappings.
Proceedings of the 6th International Conference on Auditory Display ICAD. 2000.
- [50] Neuhoff, John G. Kramer, Gregory. Wayand, Joseph.
Sonification and the Interaction of Perceptual Dimensions: Can the Data Get Lost in the Map?
Proceedings of the 6th International Conference on Auditory Display, ICAD. 2000.
- [51] Moody, Niall.
Motion as the Connection Between Audio and Visuals.
Centre for Music Technology, University of Glasgow UK. 2007.
- [52] Peeters, Geoffroy.
A large set of audio features for sound description (similarity and classification) in the CUIDADO project.

- Ircam, Analysis/Synthesis Team. 2004.
- [53] Aramaki, M. Besson, M. Kronland-Martinet, R. Ystad, S.
A Percussive Sound Synthesizer Based on Physical and Perceptual Attributes.
Computer Music Journal, 30:2, pp. 32–41, Summer 2006.
- [54] Serra, X.; Bonada, J.
Sound Transformations Based on the SMS High Level Attributes
International Conference on Digital Audio Effects. 1998.

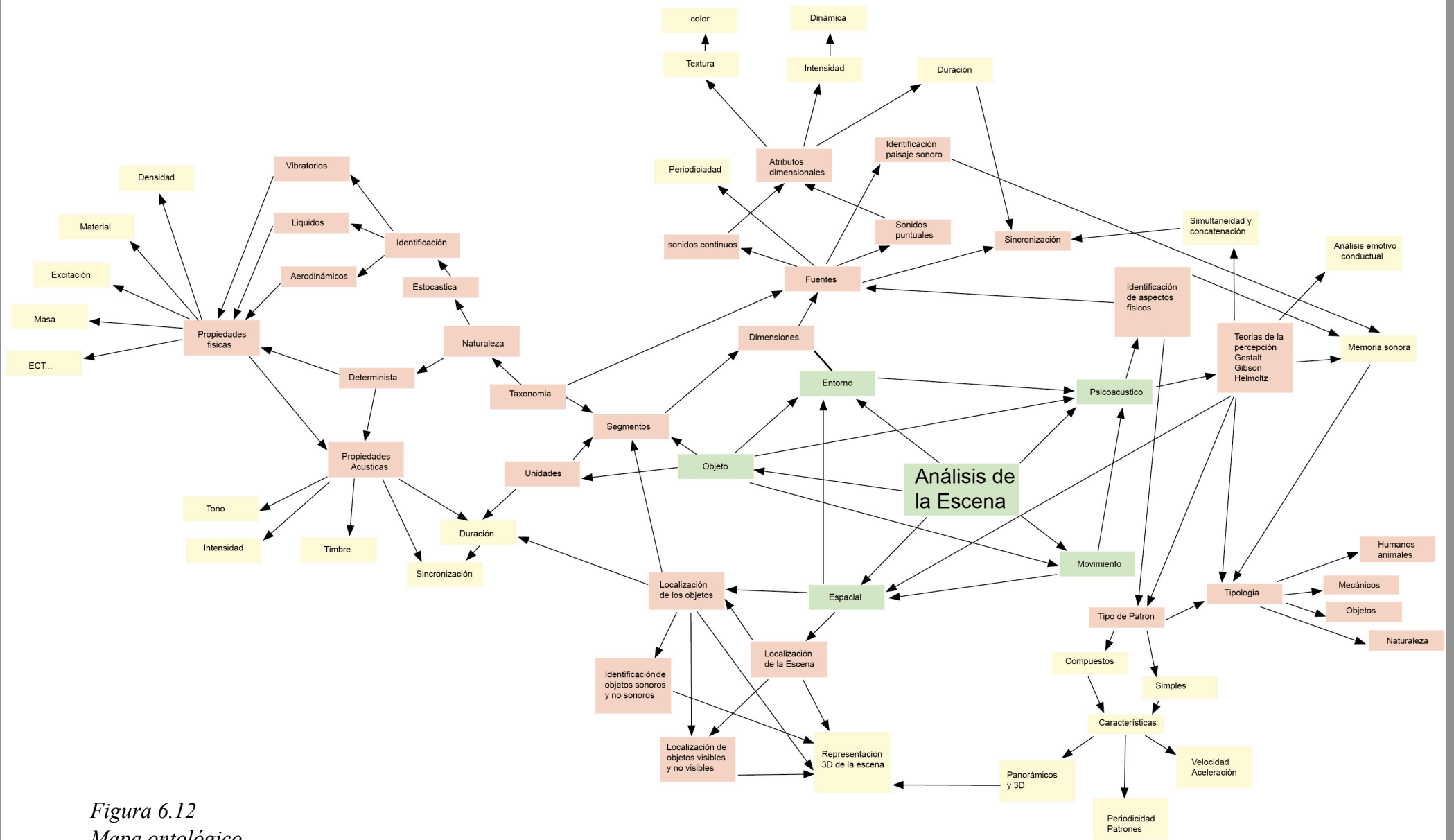


Figura 6.12
Mapa ontológico

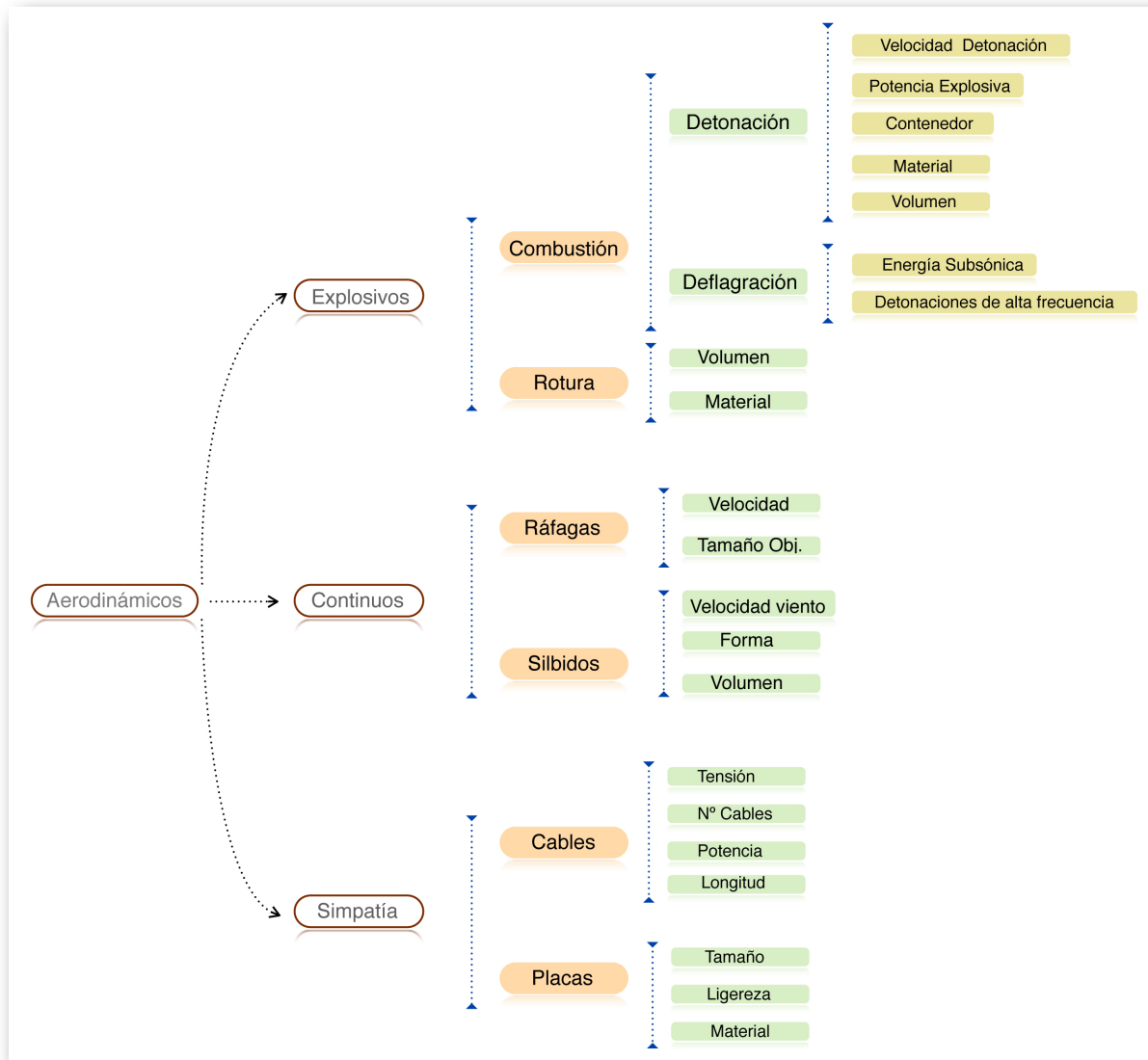


Figura ext. 2 Taxonomía de sonidos aerodinámicos