Universidad Nacional de Rosario
Facultad de Ciencias Exactas, Ingeniería y Agrimensura

TESIS DOCTORAL

Generación automática de paisajes sonoros realistas con espectro, distribución de duraciones y categorías semánticas especificados

Ernesto Accolti

Director: Ing. Federico Miyara
Co-Director: Dr. Ing. Ernesto Kofman
Miembros del Jurado: Dr. Ing. Victor Cortinez
Dr. Ing. Leonardo Molisani
Dr. Ing. Fabián Tommasini

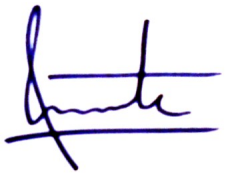
Tesis presentada en la Facultad de Ciencias Exactas, Ingeniería y Agrimensura, en cumplimiento parcial de los requisitos para optar al título de

Doctor en Ingeniería

Trabajo realizado en el Laboratoro de Acústica y Electroacústica (FCEIA-UNR)
y en el Laboratorio para el Control del Ambiente Construido (RiAS)
de la Segunda Universidad del Estudio de Nápoles (SUN)

Rosario, Argentina, 2015

Certifico que el trabajo incluído en esta tesis es el resultado de tareas de investigación originales y que no ha sido presentado para optar a un título de postgrado en ninguna otra Universidad o Institución.



Ernesto Accolti

Resumen

El aporte principal de esta tesis es el desarrollo de una herramienta para componer estímulos sonoros realistas con parámetros definidos por el usuario, mediante la combinación automática de sonidos pregrabados. El usuario controla el espectro por bandas del estímulo sonoro, la distribución estadística de las duraciones de los de eventos y sus categorías semánticas. Esto permite experimentar sobre los efectos del ruido ambiental en el ser humano en función de los parámetros que controla la herramienta.

Se estudian diversas estrategias posibles para diseñar la herramienta y solucionar problemas de implementación. Finalmente se desarrolla en detalle una estrategia basada en formular el problema como uno de programación lineal con soluciones mixtas en los enteros y los reales. También se implementan herramientas auxiliares como las de análisis de espectro, análisis de duración de eventos sonoros, sistema de auralización y una base de archivos de audio con sus respectivos metadatos.

Con el fin de validar la herramienta se diseña un estudio experimental piloto utilizando materiales especialmente creados e implementados para exponer a sujetos a sonidos ambientales realistas generados con esta herramienta. Usando estos estímulos se realizan encuestas proponiendo 8 escenarios a un grupo de 18 participantes. Los escenarios sonoros logrados son aceptables respecto a los parámetros requeridos a la herramienta. El estudio en sí revela resultados esperables en cuanto a la sensación de molestia y además es favorable respecto a la sensación de realismo reportada por los participantes.

Abstract

The main contribution of this thesis is the development of a tool that mixes pre-recorded sounds to compose realistic sound stimuli with parameters set by the user. The user can set the band spectrum of the sound stimulus, the statistical distribution of the durations of the events and their semantic categories. This tool can be used on experimental research on the effects of environmental noise on humans as a function of the controlled parameters.

Several strategies for developing and troubleshooting implementation tasks are studied. Finally, a strategy based on formulating the problem as a mixed integer linear programming one is full developed. Auxiliary tools such as spectrum analysis, duration analysis of sound events, an auralization system and a database of audio files with their metadata are also implemented.

In order to validate the tool an experimental pilot study is designed. The materials for the experiment were specially created and implemented to expose subjects to realistic ambient sounds generated using this tool.

Surveys were conducted within a group of 18 participants using 8 aural scenarios. The scenarios are acceptable regarding the parameters required to the tool. The study reveals the expected responses on annoyance and the sense of realism reported by participants is also favorable.

Agradecimientos

Esta tesis fue financiada por las siguientes cinco instituciones. Por la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) a través de una Beca de Iniciación Doctoral (en el marco del proyecto PICT Nº 38109), por la Segunda Universidad del Estudio de Nápoles a través de una Beca para el desarrollo parcial de tesis doctorales, por la Comisión Nacional de Investigación Científica y Tecnológica (CONICET) a través de una Beca tipo II para la finalización de tesis doctorales, por la Sociedad Americana de Acústica (ASA) a través del Comité de Educación e Investigación Internacional (CIRE) mediante la beca para estudiantes internacionales y por la Facultad de Ciencias Exactas, Ingeniería y Agrimensura (FCEIA) de la Universidad Nacional de Rosario (UNR) a través de una beca de exención de costos de cursos doctorales. Se reconoce con gratitud el apoyo de estas instituciones. El autor también agradece a otras instituciones como el Instituto Internacional de Control de Ruidos (I-INCE) que lo reconoció como Joven Científico y le otorgó la beca YSG 2010 para asistir al congreso INTERNOISE 2010 en Lisboa, el Laboratorio de Acústica y Electroacústica FCEIA-UNR que financió los costos de dos cursos internacionales y de asistencia a varios congresos nacionales, la Federación Iberoamericana de Acústica (FIA) por la beca FIA Grant 2010, la sociedad Europea de Acústica (EAA) por la beca EAA Grant para Jóvenes Investigadores 2010 para asistir al congreso EUROREGIO 2010 y la escuela de verano EAA Summer School 2010 en Ljubjana y, finalmente pero igual de importante, la Sociedad Española de Acústica (SEA) y la Comisión Internacional de Acústica (ICA) por la beca ICA-SEA para asistir al congreso Tecniacústica 2011 en Cáceres, España.

Gracias al Profesor Federico Miyara, director de esta tesis, director del Laboratorio de Acústica y Electroacústica donde se realizó la mayor parte de esta tesis, director de la beca de ANPCyT y codirector de la beca de CONICET. Por toda la guía, el soporte

y la paciencia dedicada al proyecto y por darle soporte a cada proyecto, idea y sueños relacionados con esta tesis y con la acústica en forma general.

Muchas personas aportaron de alguna manera con esta tesis. Les agradezco profundamente y a continuación los listo brevemente. Prof. Elizabeth Gonzales directora de la beca de CONICET, Prof. Luigi Maffei y Massimo Masullo director y codirector respectivamente durante mi estadía en Nápoles, Prof. Ernesto Kofman codirector de tesis, por su guía y soporte. Prof. Sergio Junco por sus aportes en charlas de pasillo en la Facultad. Susana Cabanellas, Marta Yanitelli, Vivian Pasch y Pablo Miechi del Grupo Ruido de la Facultad de Arquitectura, Planeamiento y Diseño UNR por su ayuda en diversos temas de esta tesis. Prof. Samir Gerges por su amistosa guía y soporte sobre las sociedades de acústica del mundo, su estructura y oportunidades para los acústicos jóvenes. Prof. Graciela Nasini y Daniel Severín por su guía sobre tópicos de Programación Lineal con soluciones en los Enteros usados en esta tesis. Varios amigos y colegas conectados con trabajos académicos relacionados a esta tesis son Matías Nacusse, Fernando Marengo-Rodriguez, Ezequiel Mignini, Federico Berguero, Gabriela Santiago, Nicolás Urquiza, Seçkîn Bastürk, Germán Miretti e Ignacio Roggio, muchas gracias amigos.

Finalmente pero en realidad primero que nada gracias a mi familia: mis padres, mi hermano, mi familia entera y especialmente a Gloria Mafalda Lucero.

A Gloria Mafalda Lucero.

Índice general

Capítulo 1

Introducción

La influencia del contenido espectral y el perfil temporal en los efectos del ruido en el ser humano ha sido ampliamente estudiada. Sin embargo el conocimiento actual aún no permite estimar de manera ajustada cuánto ni en que forma influyen estos factores. Las estimaciones actuales se basan en aproximaciones estadísticas gruesas que son usualmente utilizadas en normas y normativas que regulan el contaminante ruido en diversos ámbitos. Algunas de estas aproximaciones gruesas son correcciones por tipo de fuentes sonoras, audibilidad de tonos puros y contenido de baja frecuencia [ISO 1996-1, 2003; IRAM 4113-1, 2009; IRAM 4062, 2001] (notar que la norma IRAM 4062 es normativa en varias jurisdicciones de la República Argentina, por ejemplo en la Provincia de Santa Fe según lo establece la Resolución 201 de la Secretaría de Estado de Medio Ambiente y Desarrollo Sustentable de la Provincia de Santa Fe [Res 201, SEMADS, 2004]).

Esta tesis aporta al futuro desarrollo del conocimiento de la influencia de estos factores. En si la tesis no pretende desarrollar por completo un modelo predictivo de estos efectos sino que se propone y consigue implementar una herramienta que apoyará al desarrollo de este conocimiento en investigaciones futuras, con la posibilidad de avanzar en la búsqueda de un modelo predictivo más ajustado e incorporar otros parámetros.

Más abajo, en este mismo capítulo se describe el estado actual del conocimiento sobre los efectos del ruido en el ser humano. También en este mismo capítulo se formula el problema general en el cual se enmarca esta tesis definiendo el alcance y los objetivos.

En el capítulo 2 se describe la estructura general de la herramienta. En los capítulos 3 a 6 se describen principalmente las tareas de compilación y análisis de la base de datos introduciendo partes que son reutilizadas en las secciones siguientes (por ejemplo el capítulo 4 introduce el análisis espectral que también es utilizado en el análisis de los resultados de combinación (capítulo 9). En el capítulo 7 se describe el sistema de Auralización también introduciendo ciertas partes necesarias para la composición controlada.

La composición controlada, que es la parte principal de esta tesis, se describe en el capítulo 8. En el capítulo 9 se muestra como el usuario de la herramienta visualiza las discrepancias entre los valores de los parámetros solicitados a la herramienta y los valores que la herramienta puede lograr en el archivo de audio de salida.

Finalmente, mediante una prueba piloto, en el capítulo 10, se valida la herramienta mediante un estudio empírico preliminar. En las conclusiones se reflexiona sobre los alcances del trabajo y posibilidades de mejora y estudios a futuro.

1.1 Estado actual del conocimiento

Estudios epidemiológicos a gran escala sugieren un gran rango de efectos nocivos del ruido en el ser humano que van desde la molestia por ruidos hasta el incremento del riesgo de trastornos cardiovasculares, pasando por interferencia en la comunicación oral, disrupción del sueño, cambios (en algunos casos irreversibles) en la distribución de las etapas del sueño, fatiga y efectos en memoria de corto plazo debida a efectos del ruido durante el sueño, estrés, efectos endocrinos e inmunológicos, irritabilidad, aceleración de algunos desórdenes mentales latentes, efectos en la sensación de equilibrio, disrupción de tareas cognitivas incluyendo lectura y actividades que requieren atención y demandan memoria de trabajo, quejas generalizadas, protestas, mudanzas, efectos sensoriales como dolor auditivo y tinitus e hipoacusias inducidas por exposición a ruido [Berglund y Lindvall, 1995; Fritschi et al., 2011].

Un gran número de investigaciones sobre el efecto del ruido en el ser humano se han realizado en laboratorio utilizado sonidos sintetizados, material pregrabado o combinación de ambos como estímulo sonoro. Es de esperar que los sonidos grabados sean más similares a los sonidos de la realidad en comparación con los sonidos sintetizados a menos que se trate de síntesis muy realista lo cual tiene grandes costos

asociados en tanto para cada tipo de sonido se debe recurrir a un modelo de síntesis particular. Mientras los estímulos son más similares a los estímulos de la realidad, mayor será la validez ecológica de los experimentos a expensas de mayores dificultades para controlar los parámetros de los estímulos. El problema que resuelve esta tesis es el desarrollo de un método para mezclar archivos de audio preexistentes (ya sea de grabaciones previas o síntesis realista) de modo tal que el sonido resultante tenga parámetros que fueron previamente determinados.

En estudios actuales sobre efectos del ruido en el ser humano se utilizan estímulos controlando parámetros tales como el espectro, la duración de los eventos sonoros, la frecuencia de aparición de eventos y la categoría semántica de los eventos sonoros que aparecen. Por ejemplo Moorhouse et al. [2005] investigan sobre la influencia del espectro en el umbral de aceptabilidad de los sonidos de baja frecuencia, Vos y Houben [2013] investigan la influencia de la frecuencia de aparición de eventos sonoros impulsivos (de corta duración) en la probabilidad de despertar y, Hong y Jeon [2013] investigan sobre la influencia en la preferencia y la calidad ambiental de la categoría semántica de los eventos que conforman un ambiente sonoro.

Sin embargo, en estos estudios previos, cada estímulo se compone partiendo de una base de archivos de audio pequeña y ad-hoc sin proponer un control automático de diversos parámetros en simultáneo sino que solo grupos de unos pocos parámetros. Por ejemplo, Kim et al. [2010] controlan únicamente el nivel de banda ancha de un único evento sonoro, Vos y Houben [2013] generan los estímulos mediante el control del nivel de banda ancha y la frecuencia de aparición de eventos de una única fuente de ruido en dos niveles (una o cuatro repeticiones), Alayrac et al. [2011] mezclan solo dos archivos de audio para cada estímulo controlando el nivel de emergencia (*emergence level*) y, Hong y Jeon [2013] mezclan sólo tres o menos archivos controlando únicamente la categoría semántica de los eventos presentes en los archivos que serán incluidos en la mezcla final.

1.1.1 Paradigmas de investigación actuales

1.1.1.1 Paisaje Sonoro

El paradigma de paisaje sonoro se enmarca dentro de los paradigmas de investigación-acción en tanto la base del conocimiento se centra en los aspectos subjetivos y

sociológicos. Con esos principios trata de abordar holísticamente el contexto patrimonial y cultural que acompaña a los sonidos y sus referencias a las fuentes sonoras.

Este paradigma se diferencia de los paradigmas clásicos en varios aspectos y particularmente en el objetivo de preservar aquellos sonidos con connotación favorable para quienes los oyen. Ambos paradigmas confluyen en el objetivo de controlar aquellos sonidos con connotaciones negativas pero en el paradigma clásico la connotación de la importancia de los sonidos suele referirse a la inteligibilidad de un mensaje que se desea escuchar, normalmente la palabra y la música. El paradigma paisaje sonoro además de estos sonidos incorpora aquellos otros sonidos que, quizás no entran en la categoría de deseables, pero son juzgados con connotaciones positivas por sus oyentes habituales. Es decir, no incorpora nuevas categorías de sonidos con connotación positiva en forma directa sino que estudia si los habitantes y usuarios de una pequeña zona le dan tal connotación a sonidos como los producidos por fuentes de agua, aves y otros animales, sonidos naturales e incluso algunos sonidos generados artificialmente para emular sonidos naturales. Estos sonidos con connotación positiva se estudian además contextualizados por la presencia de los sonidos con connotación negativa que puedan estar presentes en esa región [Maffei, 2008; Raimbault y Dubois, 2005]. Este paradigma está actualmente en desarrollo y la herramienta de este trabajo ofrece importantes aplicaciones para tal fin. Aunque aún esté en estudio la posible relación entre la valoración de los sonidos y el contenido espectral y patrón temporal de los mismos, es sabido que la segregación auditiva está relacionada con el concepto de timbre que a su vez se relaciona con el contenido espectral y el patrón temporal. Inclusive otros grupos de investigación, dedicados a la aproximación de Paisaje Sonoro, trabajan en algoritmos para separación ciega que puede ser utilizada para clasificar las fuentes sonoras involucradas en el ruido ambiental para luego desarrollar descriptores adecuados [Bunting et al., 2009]. Si se lograsen desarrollar descriptores basados en los eventos sonoros que componen un determinado paisaje sonoro, por ejemplo realizando experimentos sistematizados con la herramienta presentada en esta tesis, luego será necesario separar los eventos sonoros que componen el paisaje sonoro medido en una situación real para poder estimar el valor de esos descriptores.

1.1.1.2 Saliencia Auditiva

En años recientes, imitando modelos de detección visual de objetos en imágenes [Itti y Koch, 2001; Itti et al., 1998], se han desarrollado diferentes modelos de saliencia auditiva [De Coensel y Botteldooren, 2010; Kayser et al., 2005].

Este tipo de modelos, visuales y auditivos, se basan en observaciones experimentales de comportamientos biológicos pero aún existen algunos aspectos desconocidas del comportamiento, por ejemplo los aspectos que estos modelos describen mediante una retroalimentación debida a la atención, el interés en hallar alguna particularidad determinada o incluso a entrenamiento sobre alguna particularidad.

En cuanto al comportamiento temporal este tipo de modelos utiliza diversas escalas para analizar un mismo sonido o un mismo ambiente sonoro. Estas escalas distinguen distintos tipos de detalles en el dominio temporal, espectral y combinado de ambos. Por ejemplo De Coensel y Botteldooren [2010] desarrollan un interesante modelo dinámico de atención a fuentes sonoras y no sonoras que permite determinar eventos de diversas duraciones basándose en modelos de saliencia auditiva.

Si bien estos modelos son interesantes para describir el patrón temporal del sonido ambiental, sucede que no se prestan aún a una especificación a partir de las correspondientes a los eventos sonoros que componen ese ambiente.

1.1.1.3 Psicoacústica y duración de un evento sonoro

La duración percibida o subjetiva de un evento sonoro difiere de su duración física [Fastl et al., 2002; Fastl, 1977]. Ambas duraciones (física y subjetiva) pueden ser medidas o estimadas en diferentes escalas. Por ejemplo un grupo de personas pasando cerca de un oyente pueden ser escuchadas como el paso de un grupo en una macroescala o como pasos individuales en una microescala. Esto se relaciona con observaciones de la actividad neurológica de la corteza auditiva en animales (que tiene un comportamiento similar a la corteza visual) y ha sido la inspiración de modelos de detección de eventos sonoros salientes como el denominado Saliencia Auditiva [Kayser et al., 2005] (Ver 1.1.1.2). La relación entre duración física y subjetiva no es lineal, sobre todo dentro del rango de hasta los cientos de milisegundos.

Hugo Fastl propuso un modelo basado en la sonoridad para determinar la duración subjetiva en una única escala [Fastl et al., 2002]. Este modelo predice la duración

subjetiva como el periodo en que el valor del enmascaramiento temporal, entendido como el patrón temporal de sonoridad) es superior a 10 dB sobre el mínimo del propio enmascaramiento temporal. Si bien la aproximación explica muy bien los datos empíricos, este modelo depende del contexto pues el valor mínimo del enmascaramiento está fuertemente relacionado con el contexto, por ejemplo por el ruido de fondo. En caso que se evalúe un evento sonoro aislado, este mínimo sería el umbral de audición, pero cuando hay otros eventos en el contexto el mínimo estará relacionado con el perfil de enmascaramiento de esos eventos. Notar que para este modelo un mismo evento sonoro puede tener diferentes valores de duración subjetiva si se cambia el contexto. Por ejemplo, el paso de un auto se percibe con una duración mayor en un ambiente silencioso, como podría darse en una zona rural, en contraposición al mismo auto pasando en un ambiente ruidoso, como podría darse en una zona urbana altamente congestionada. Entonces, la duración subjetiva, no solo depende de cada sujeto, sino también del contexto. Esa dependencia del contexto también puede ser explicada desde la Saliencia Auditiva.

1.2 Formulación del problema

El marco teórico de esta tesis se nutre de los paradigmas Paisaje Sonoro, Psicoacústica Ecológica y Análisis de Escena Auditiva. Estos paradigmas son cercanos entre si y se complementan mutuamente en diversos aspectos.

El paradigma de esta tesis es del tipo holístico involucrando aspectos de la acústica física, la psicología y la psicoacústica, el audio, la electrónica, neurociencias, ciencias de la computación y semiótica. Las posibles aplicaciones futuras están orientadas a aspectos legislativos y de toma de decisión sobre el bienestar social y la salud auditiva o, más allá de cómo definir los criterios legales, a aspectos de mejora y preservación del paisaje sonoro.

Un aspecto de los paradigmas Paisaje Sonoro y Análisis de Escena Auditiva incorporados en esta tesis es el estudio del ruido ambiental como una combinación de sonidos provenientes de diversas fuentes sonoras. Los parámetros psicoacústicos como sonoridad, fuerza de fluctuación, rugosidad y aspereza han sido validados para algunas fuentes individuales. Existen además modelos computacionales [Chalupper y Fastl, 2002] que permiten estimar estos parámetros psicoacústicos en función del

tiempo. Sin embargo algunos autores concluyen que una escena completa, con más de una fuente, no puede ser explicada simplemente mediante el promedio de los valores de estos parámetros [Accolti y Miyara, 2009, Ver bibliografía del artículo referido].

El mayor aporte de esta tesis a los paradigmas que la enmarcan es el desarrollo de una herramienta para componer señales de prueba con parámetros definidos por el usuario, mediante la combinación automática de sonidos pregrabados. Se estudian varias estrategias posibles para solucionar el problema de implementación y finalmente se desarrolla en detalle una de estas estrategias. También se implementan otras herramientas auxiliares conocidas (por ejemplo el sistema de auralización o de análisis espectral) pero adaptadas para ser compatibles con los requisitos del problema principal de esta tesis.

El problema general es investigar los efectos del contenido espectral y el patrón temporal del ruido ambiente en la percepción y valoración humana. Estos efectos se estudian en un ambiente controlado de laboratorio. Este problema principal se separa en dos partes, la primera es el desarrollo de las herramientas para controlar las señales de prueba y la segunda es una prueba piloto de las herramientas. El primer problema es el problema principal abarcado en esta tesis. El segundo problema es parte de la puesta a punto de la herramienta y aporta un antecedente para trabajos futuros que realicen una experimentación exhaustiva valiéndose de la herramienta. Si bien las conclusiones de la prueba piloto son en parte esperables, es también de esperar que la investigación sucesiva mediante el uso de esta herramienta logre arrojar más datos sobre la influencia del contenido espectral y el patrón temporal del ruido ambiental en el ser humano.

El problema del desarrollo de la herramienta consiste, principalmente, en un algoritmo capaz de combinar algunos archivos de audio de naturaleza espectral y temporal variada de manera controlada de forma tal que el sonido de salida, ya combinado, presente valores predefinidos por el usuario para los parámetros del contenido espectral y del patrón temporal. Este problema asocia otros problemas tales como la recolección de archivos para una base de sonidos, el análisis de los mismos y el sistema de auralización.

En la prueba piloto, el sistema de auralización se implementa de manera sencilla mediante un parlante. El mismo no es visible para los sujetos pues se encuentra detrás

de un vidrio y una persiana de madera con el objetivo de simular que el ruido viene del exterior. Se reflexiona sobre la extensión a otros sistemas de reproducción.

La prueba piloto consiste en estudiar los efectos de ciertos factores del ruido en la apreciación y percepción de ese ruido como ambiente. Este estudio preliminar se llevó a cabo mediante encuestas en condiciones controladas de laboratorio. Los factores estudiados son la forma del espectro, la forma del histograma de duración de eventos, el nivel sonoro total y la frecuencia de aparición del total de eventos. Las posibilidades de control, según se justifica en el capítulo 8, dependen del material, por ejemplo de la cantidad de eventos que tiene la base de sonidos o del sistema de auralización.

1.3 Objetivos

El objetivo principal es desarrollar una herramienta capaz de generar diferentes sonidos ambientales realistas, con ciertos parámetros controlados, mediante la combinación automática de sonidos pregrabados. Los parámetros que tendrán los sonidos ambientales deben ser controlados por el usuario permitiéndole generar escenarios de prueba para la investigación sobre la influencia de esos parámetros del ambiente en el ser humano. Los parámetros a controlar son el espectro sonoro por bandas y la distribución de duración de los eventos sonoros que compondrán cada sonido ambiental. Adicionalmente se dejará constancia de otras posibilidades de control, como ser respecto al tipo de fuentes sonoras intervinientes.

1.3.1 Objetivos Específicos

Para lograr dicha herramienta se deben desarrollar diversos objetivos previos de alcance más acotado. A continuación se citan los mismos.

1. Proponer los criterios para generar una base de archivos de audio, en la cual cada archivo contenga los símbolos sonoros de menor escala posible y que pertenezcan a una misma categoría de fuentes sonoras, y aplicarlos para obtener una base apta para las pruebas de la herramienta.

2. Diseñar un esquema de clasificación de eventos sonoros que permita manipular el tipo de ambiente sonoro en forma realista. Por ejemplo excluir/incluir grupos

eventos sonoros que habitualmente se encuentran en un tipo de ambientes pero no en otros.

3. Desarrollar e implementar un protocolo y software para analizar de manera semiautomática los archivos de audio que contiene la base.

4. Analizar auditivamente los archivos de audio identificando a que clasificación de eventos pertenecen, editar el archivo de audio si fuese necesario, calcular automáticamente el espectro sonoro, la duración total del archivo, la duración de los eventos dentro del archivo y otros parámetros que sean de utilidad para la combinación controlada.

5. Formular matemáticamente el problema de combinación controlada. La formulación debe permitir la solución de diversas instancias del problema, cada instancia corresponde a la composición de una señal de salida.

6. Desarrollar un algoritmo que combine automáticamente los sonidos de la base de datos, según la solución de cada instancia del problema matemático de combinación controlada, generando el archivo de audio de salida.

7. Realizar un estudio piloto sobre la influencia de los parámetros controlados en la molestia causada por ruido.

El objetivo específico 2 permite agregar otro parámetro de control por parte del usuario de la herramienta. Es decir, permite controlar la cantidad de eventos según el tipo de fuentes sonoras, lo cual es de gran utilidad en estudios del contenido semántico del ruido.

Capítulo 2

Estructura de la herramienta de combinación controlada de sonidos

En este capítulo se describe de manera general la estructura de la herramienta. Luego, en los capítulos 3 a 7 se describen en detalle ciertas partes relevantes de la herramienta. No todas las partes corresponden a algoritmos sino que algunos corresponden a actividades semiautomáticas o incluso a fenómenos físicos acústicos. En algunos capítulos, como en el capítulo 4, se describen partes de código que son reutilizadas en más de un bloque de la herramienta general.

2.1 Estructura general

Con el objeto de establecer un orden para facilitar la presentación de la herramienta, se definen niveles para las partes que componen la estructura de la misma. Las partes más globales se definen como módulos. En la figura 2.1 se muestra un esquema general de la herramienta utilizando módulos. Cada uno de estos módulos se analizará en el siguiente nivel de detalle que corresponde a los bloques y subbloques (figuras 2.2 a 2.5). Las flechas representan el flujo de datos o información de diversa índole.

El módulo principal de la herramienta es el denominado *Composición* que se encuentra en el centro de la figura 2.1. Este módulo corresponde a un algoritmo implementado en software de cálculo matricial que permite combinar de manera controlada los archivos de audio de una *base de datos* de tal modo que el archivo de audio de salida cumpla ciertas características especificadas mediante el *Control de Usuario*. El grado de concordancia entre el archivo de audio de salida y las características espe-

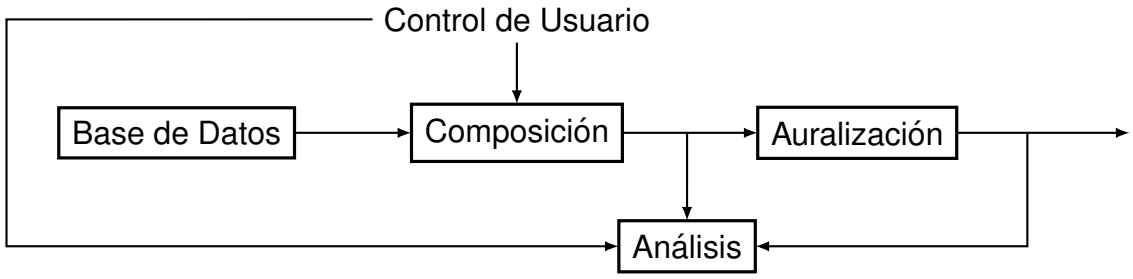


Figura 2.1: Estructura general

cificadas por el usuario son analizadas por esta herramienta en el módulo *Análisis*. El módulo *Análisis* muestra las diferencias entre los datos requeridos por el usuario y los que finalmente alcanza el archivo de salida. En realidad el módulo se ha incluido porque la adecuación depende de la base de datos y de los requerimientos del usuario. El último módulo corresponde a la *Auralización* que consiste en la presentación auditiva, de la prueba especificada por el usuario, a los posibles sujetos de un estudio experimental. El módulo *Auralización* comprende todo el recorrido de la información desde el archivo de audio de salida hasta que es percibido por el sujeto.

2.2 Módulo Base de Datos

Los bloques que componen el módulo base de datos se muestran en el esquema de la figura 2.2.

Los archivos de audio que compondrán la base de datos pueden ser tomados de dirvesas fuentes. En la figura 2.2 se sugiere dos categorías que corresponden a las bases de datos de otros autores, ya sean gratuitas o comercializadas, y grabación propia de quien implemente esta herramienta. En las bases de otros autores existen dificultades para encontrar cierta información pertinente como la sensibilidad del sistema de grabación (es decir, qué nivel sonoro representa el nivel de fondo de escala de la grabación o algún nivel relativo), la distancia entre el micrófono y la fuente sonora durante la grabación, si hubo algún postprocesamiento o incluso si se debe a una simulación. Sin embargo la ventaja es que estas bases contienen una gran cantidad y variedad de sonidos. Mediante la grabación propia es posible establecer un protocolo de registro que considere los datos de sensibilidad, distancia y postprocesamiento pero la desventaja es que se trata de un procedimiento muy costoso si la base de datos debe ser de gran tamaño.

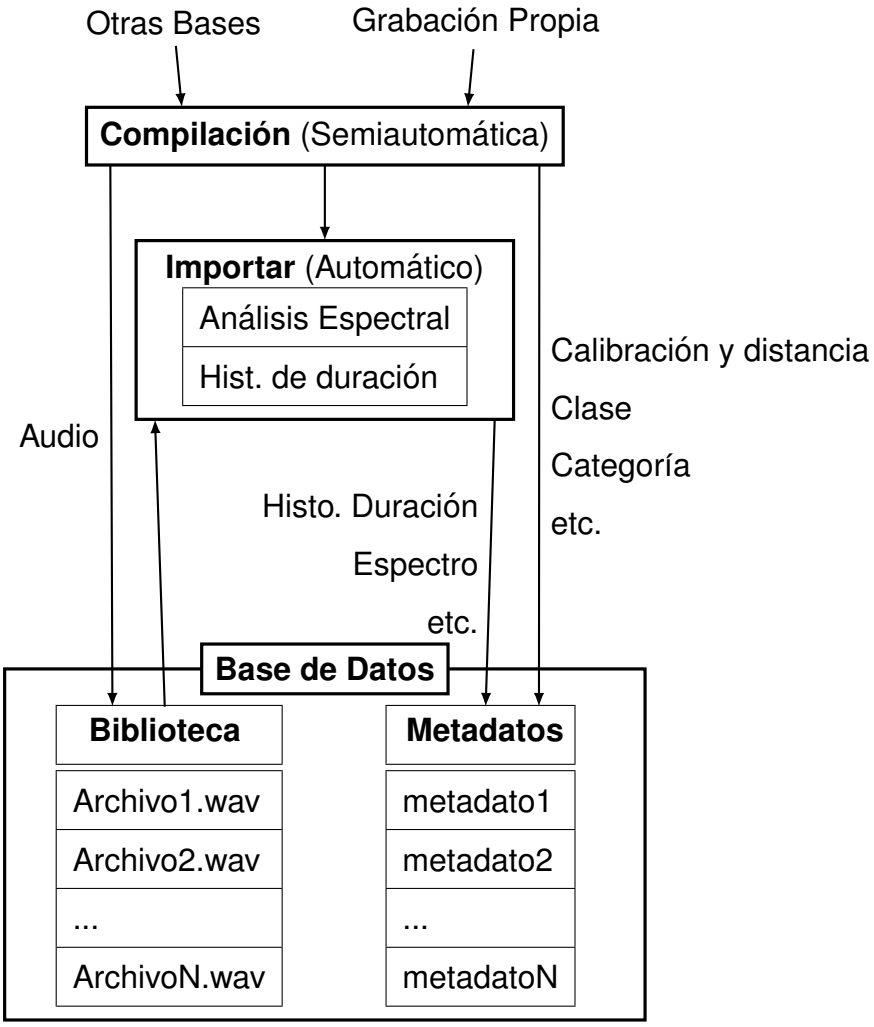


Figura 2.2: Estructura del módulo Base de Datos

A continuación de la selección de archivos de audio se implementa el bloque *Compilación* que corresponde a un procedimiento semiautomático mediante el cual se editan los archivos de audio y se les asignan ciertas etiquetas de utilidad para la herramienta y otras de utilidad para mejoras a futuro. En el capítulo 3 se especifican más detalles de este bloque. Los archivos de audio se envían directamente a la *Base de Datos* luego de la edición de audio realizada en el bloque *Compilación*. Para ser importados totalmente se genera, en el bloque *Importar*, una serie de metadatos para cada archivo de audio.

El bloque *Importar* consiste en la importación (ver capítulo 6) de los archivos a la base, la cual se realiza de manera automática una vez que un archivo, o un grupo de archivos, fue compilado. Dentro de este bloque existen subbloques que de manera automática analizan los archivos de audio generando nuevas etiquetas que se agregan

a las generadas durante la compilación para completar el archivo de metadatos. En el esquema se muestran los dos subbloques más importantes: *Análisis Espectral* (Ver Capitulo 4) e *Histograma de Duración* (Ver Capitulo 5)

Finalmente la base de datos se compone de dos subbloques, una *Biblioteca* de N archivos de audio y un archivo de *Metadatos*. Se debe notar que a cada archivo de audio le corresponde un grupo de metadatos dentro del archivo *Metadatos*. Es decir, al *archivo1.wav* le corresponde el grupo de metadatos denominado *metadato1* que se encuentra dentro del archivo *Metadatos* al *archivo2.wav* le corresponde *metadato2* y así sucesivamente para los N archivos. Cada grupo de *metadatos*, entonces, se compone de unas etiquetas generadas semiautomáticamente durante la *Compilación* (por ejemplo la categoría de evento sonoro y la clase de distribución temporal de evento sonoro, ver 3) y otras etiquetas generadas en el bloque *Importar* (ver 6) que son calculadas mediante algoritmos de análisis de señales de audio y corresponden por ejemplo al espectro sonoro por bandas de 1/1 octava (4) y el histograma de duración (5) de cada archivo.

2.3 Módulo Combinador

En la figura 2.3 se muestra el esquema de bloques del módulo *Combinador*, el módulo principal de este trabajo. Este módulo se describe de manera más detallada en el capítulo 8 pero en esta sección se da un panorama general que es de utilidad para comprender el resto del texto.

Este módulo tiene básicamente dos entradas de datos y su salida es el archivo de audio, logrado mediante la combinación controlada de los archivos de audio de la *Base de Datos*. Las entradas de datos corresponden a los bloques *Base de Datos* y *Usuario*.

El bloque *Usuario* contiene dos subbloques; el de *Configuración*, que contiene aspectos generales de configuración que, en general, no serán variados en un estudio dado, y el de *Prueba*, que se compone de los datos requeridos para cada prueba particular dentro de un estudio. Por ejemplo si el estudio fuese sobre el efecto del nivel sonoro de una banda en particular, a realizarse en dos pruebas, ambas pruebas contendrían los mismos valores en todos los parámetros salvo en el nivel de la banda en estudio, parámetro en el cual cada prueba tendrá un valor distinto (fijando o

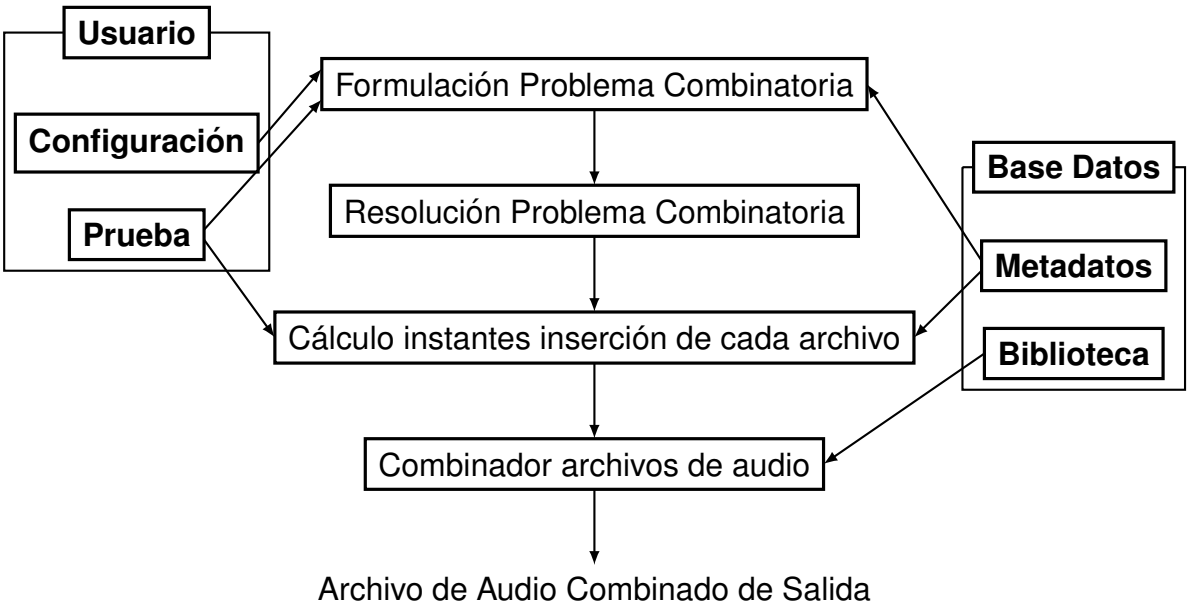


Figura 2.3: Estructura del módulo Combinador

controlando así todos estos parámetros).

Los datos de mayor interés en los subbloques *Prueba* y *Metadatos* son los del espectro y los del histograma de duración de eventos. La diferencia entre estos dos subbloques es que en *Prueba* el usuario define los valores requeridos para el archivo de salida mientras que en *Metadatos* se encuentran los valores de los metadatos correspondientes a cada archivo de audio en la *Biblioteca*.

El primer bloque de la combinación es la formulación del problema. Esta formulación puede definirse de diversas maneras que se exponen en el capítulo 8. Todas esas formulaciones utilizan tanto los datos de cada *Prueba* como los de *Configuración* del bloque *Usuario* y solo los *Metadatos* del bloque *Base de Datos*. La formulación corresponde a la formulación matemática y computacional del problema, y debe notarse que no incluye a los archivos de audio en si, sino a los metadatos de cada archivo.

El bloque siguiente corresponde a la *Resolución del Problema de Combinatoria* y es básicamente un algoritmo que resuelve el problema de optimización formulado, para cada instancia, en el bloque anterior (conocido como Solver, del inglés). En este trabajo se utiliza el algoritmo CPLEX [ILOG, 2011] para resolver el problema. Este bloque corresponde no sólo al algoritmo CPLEX sino también su configuración y adaptación de los datos de entrada y salida. La salida de este bloque corresponde a la ganancia que tendrá cada archivo de la base y cuántas veces se repetirá, cuando se combine en

el archivo de salida. Esa ganancia será nula para la mayoría de los archivos indicando que esos archivos no son incluidos en el archivo de salida. (Ver 8.3).

Una vez definidos cuáles son los archivos que estarán presentes en la combinación, y la ganancia que debe aplicarse a cada uno, se debe definir el instante de inserción de cada archivo. Incluso algunos archivos pueden insertarse en más de un instante. Estos instantes de inserción dependen de la clase de evento (que es un *metadato* en la *Base de Datos*, ver 3.2.1) y de la cantidad máxima (y mínima) potencial de reinserción de archivos que es un dato de cada *Prueba*. La forma en que se definen estos instantes es analizada con más detalles en las secciones 3.2.1 y como se usan en la combinación se describe en 8.9.1.

Finalmente, una vez que se han determinado los instantes y amplificaciones en cada uno de los archivos que serán incluidos, se combinan los archivos de audio correspondientes que se encuentran en la *Biblioteca*. Nótese que es recién en este bloque donde, mediante un algoritmo de procesamiento de audio, se accede a los archivos de audio de la *Biblioteca*, el mapeo de ganancias e instantes de inserción se genera solo con los *Metadatos*.

El resultado de este bloque es el archivo de audio listo para ser enviado al sistema de reproducción sonora (incluida en el módulo *Auralización*) como señal de excitación para la prueba correspondiente.

2.4 Módulo Auralización

En la figura 2.4 se muestra la estructura del módulo *Auralización*. En este módulo además se incluye un bloque de caracterización del sistema de reproducción que es necesaria para el subbloque de *Configuración* del bloque *Usuario* en el módulo *Combinador* (Ver figura 2.3).

El archivo de audio de salida pasa por diversos bloques antes de alcanzar al sujeto. En el esquema de la figura 2.4 se muestran los bloques que serán usados en la prueba piloto del capítulo 10. Sin embargo la herramienta permite el uso de otras configuraciones. El bloque *Interfaz de Audio* es siempre necesario debido a que la señal digital del archivo de audio debe ser convertido a una señal analógica, tanto para sistemas de reproducción sonora por altavoces como por auriculares. En este caso, por simplicidad, se prefirió utilizar un sistema con un altavoz potenciado (bloque *Altavoz*).

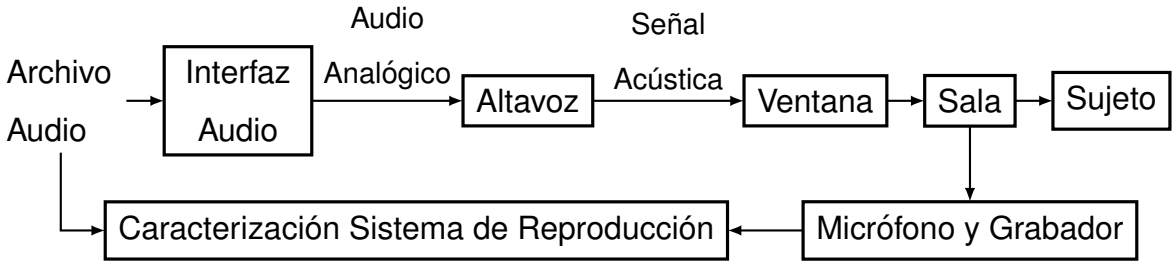


Figura 2.4: Estructura del módulo Auralización

Luego del *Altavoz* la señal eléctrica es transformada en una señal acústica que en este caso es transformada primero por una *ventana* y luego llega al sujeto a través de la *sala* donde el sujeto se encuentra, lo cuál implica otra transformación de la señal.

En caso de utilizar auriculares es posible simular, mediante un filtro digital, una ventana u otra superficie a través de la cuál el sonido ingresa a la sala. También en caso de utilizar auriculares se debe simular digitalmente la respuesta de la sala y la respuesta de la cabeza y torso del sujeto. Estas simulaciones se deben incluir de manera digital por lo cuál en ese caso los bloques estarían antes de la *Interfaz de Sonido*. Vale aclarar que en la base se prefieren los sonidos grabados con mínima intermediación de elementos que alterarían sustancialmente su distribución espectral, particularmente al aire libre o en su ámbito de generación, de preferencia, anecoico.

La señal que es captada por el sujeto también es captada por el bloque *micrófono y grabador* que registra la señal (sin la influencia del sujeto) para poder comparar y caracterizar al sistema de reproducción. Si bien la configuración es la misma, durante la configuración del sistema, se emplea una señal de audio determinística y rica armónicamente lo cual permite identificar el sistema.

2.5 Módulo Análisis

En la figura 2.5 se muestra el diagrama de bloques del módulo *Análisis*. El objetivo de este módulo es brindar al usuario información sobre la exactitud con la que los parámetros requeridos son alcanzados por el estímulo sonoro generado. En esta sección se introduce brevemente la estructura global del módulo y en el capitulo 9 se describe de manera más detallada. El módulo de análisis comprende la visualización de dos tipos de datos, los espectrales y los del histograma de duración identificados en las flechas y en el bloque *visualización* de la figura 2.5 con las leyendas “espectro”

e “histograma”.

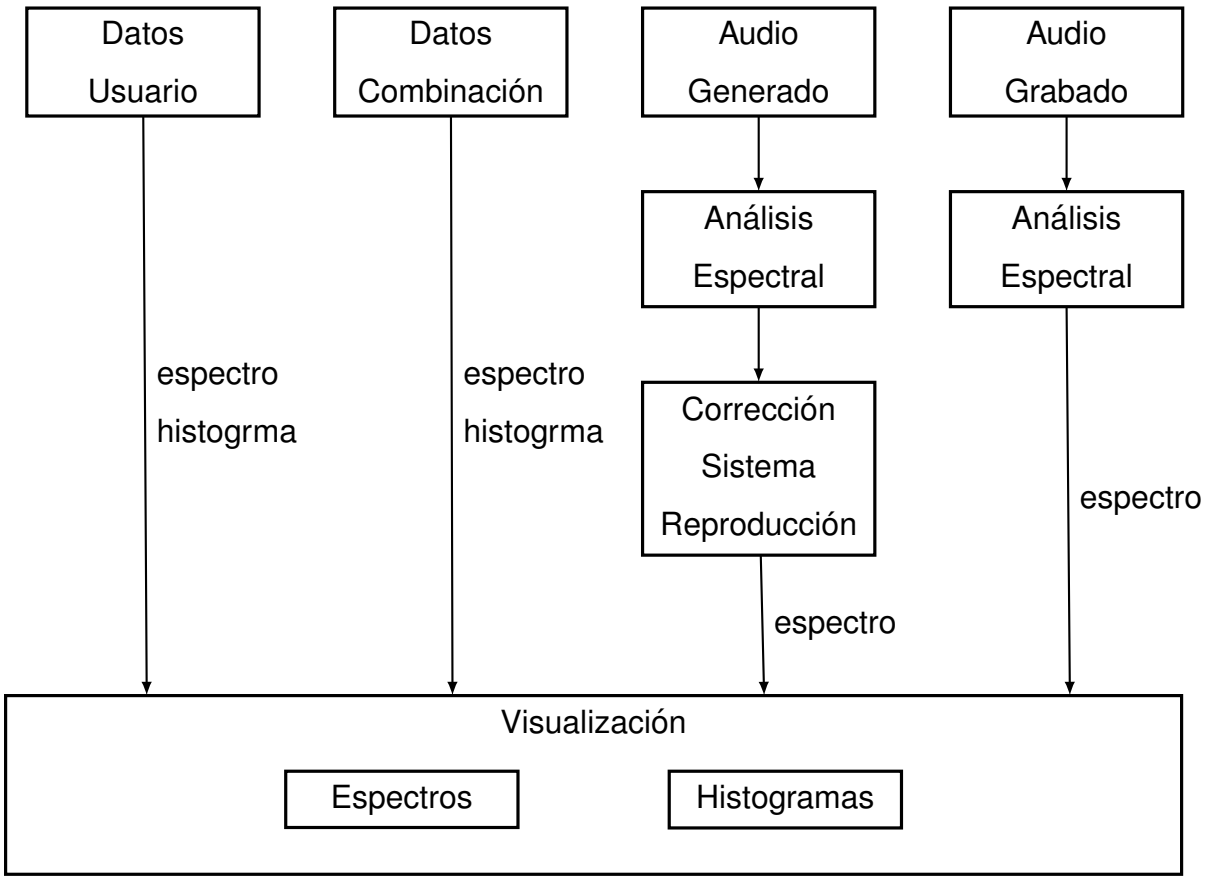


Figura 2.5: Estructura del módulo Análisis

La parte del histograma corresponde al análisis de los histogramas de duración y se implementa en el subbloque *histogramas*. Este bloque recibe los datos especificados por el usuario (ver el módulo *Combinador*) y los muestra junto a los datos alcanzados en el bloque *Resolución del Problema de Combinatoria* de la figura 2.3 permitiendo al usuario compararlos numéricamente para decidir si el audio generado es adecuado en este parámetro.

La parte del espectro corresponde al análisis de los niveles sonoros espectrales. En este caso se comparan los niveles espectrales en cuatro versiones diferentes. La primera versión es el espectro requerido por el usuario, la segunda es el espectro tomado del bloque *Resolución Problema Combinatoria* del módulo *Combinador*, la tercera es el archivo de salida y la cuarta es el sonido grabado en la sala en ausencia del sujeto.

Es decir, el espectro del archivo generado se toma de dos bloques internos del

módulo *Combinador*; por un lado directamente del módulo *Resolución Problema Combinatoria* y, por otro lado, el archivo de audio de salida es enviado a un bloque de análisis de espectro. Para que sean comparables se modifica el espectro del módulo *Resolución Problema Combinatoria* mediante una corrección que modela el sistema de reproducción sonora, incluyendo la sala (ver capítulo 7).

Capítulo 3

Compilación y etiquetado de sonidos

El bloque *compilación* del módulo *Base de Datos* (ver 2.2) corresponde a una serie de actividades a desarrollar por personal técnico a cargo de la compilación de la base. Estas tareas comprenden la recolección de archivos de audio, edición de los mismos, y generación de una serie de etiquetas para cada uno de esos archivos.

Debido a que la edición de los archivos de audio depende del etiquetado y viceversa, este capítulo no sigue el orden secuencial de las actividades. Para simplificar la descripción se introducen algunos aspectos de la edición dentro de la sección de etiquetado.

3.1 Recolección de archivos de audio

La tarea de recolección de archivos se puede realizar a partir de tres fuentes. Una de estas fuentes, descartada en este trabajo para simplificar el problema, es la simulación o síntesis de eventos sonoros. Un ejemplo de simulación realista de fuentes sonoras móviles se puede consultar en el artículo de Marengo et al. [Marengo-Rodriguez et al., 2011]. Las dos fuentes de archivos de audio utilizadas en esta tesis son las bases de otros autores y la grabación propia.

La grabación propia ofrece la ventaja de poder recabar otra información importante además del archivo de audio en si. La grabación de audio se realiza en este trabajo utilizando el micrófono de un sonómetro clase 1 [IEC 61672-1, 2002] después del preamplificador. Para la grabación se utiliza un grabador portable cuyas características han sido validadas para este fin [Miyara et al., 2010].

Mediante un protocolo interno se establece una planilla con los datos a recolectar

para cada evento grabado. La planilla contiene los campos de distancia estimada a la fuente sonora (r) y nivel de fondo de escala ($L_{\text{p,FS}}$), el cual puede introducirse numéricamente o mediante un archivo de audio llamado de calibración. El valor de fondo de escala corresponde al nivel sonoro, en decibeles, que habría sido registrado por el micrófono si el nivel de la señal hubiese sido el fondo de escala digital, es decir 0 dB. El nivel de presión de fondo de escala se obtiene mediante

$$L_{\text{p,FS}} = L_{\text{p,cal}} - L_{\text{wav,FS}} \tag{3.1}$$

donde $L_{\text{p,cal}}$ es el nivel sonoro que corresponde a un nivel $L_{\text{wav,FS}}$ relativo al fondo de escala digital en el archivo de audio de calibración o el dato ingresado numéricamente.

El protocolo se puede adaptar para contener también una planificación para la recolección a modo de guía sobre que tipo, clase y cantidad de eventos se debe registrar en cada campaña. Además se genera un croquis identificando posibles efectos de reflexiones sonoras y otras fuentes de ruido.

La recolección desde bases de otros autores ofrece la ventaja de evitar el trabajo de campo economizando recursos. Sin embargo, generalmente, presenta la desventaja de no contar con información importante sobre cómo fueron grabados, si fueron sintetizados o editados, a qué distancia se encontraba la fuente sonora o cuál es el nivel sonoro correspondiente al fondo de escala u otra información necesaria para la calibración. Otro problema es que la calidad no es uniforme, en muchos casos son archivos comprimidos con pérdidas de información. En esta tesis se prefirió usar archivos sin pérdidas con una tasa de muestreo mínima de $44\ 100$ Hz y una resolución de al menos 16 bits. En caso de utilizar archivos comprimidos, la calidad mínima tolerable fue de 320 kbps y se convierten a formato no comprimido durante la edición. En los casos de archivos con pérdidas recuperados además debían ser auralizados críticamente por el personal encargado de la recopilación quien siempre podía descartar archivos por razones de calidad no aceptable.

En este caso de bases de otros autores, el personal encargado de recolectar los archivos de audio debe contar con experiencia en la escucha de fuentes sonoras, cálculo de propagación sonora, análisis espectral y niveles sonoros habituales de fuentes sonoras. Los niveles habituales se obtienen en muchos casos desde la bibliografía. Utilizando los niveles del espectro relativos al nivel de una banda, y comparando con

niveles relativos reportados en la bibliografía para el mismo tipo de fuentes es posible determinar, en una aproximación, la distancia a la cual se registró la fuente sonora. Luego los niveles relativos son escalados de modo de alcanzar el nivel sonoro absoluto de los datos de la bibliografía para estimar el nivel correspondiente al fondo de escala. Es decir, el nivel absoluto de presión sonora, tomado de la bibliografía, reemplaza a $L_{\text{p,cal}}$ y el nivel absoluto de la señal de audio reemplaza a $L_{\text{wav,FS}}$ en la ecuación (3.1).

3.2 Etiquetas

En las actividades de etiquetado se debe llenar una planilla de algún software de hojas de cálculo. En esta tesis se utiliza el programa Calc del paquete OpenOffice[1]. Un ejemplo de las etiquetas generadas para 4 archivos se muestra en la Tabla 3.1.

Tabla 3.1: Etiquetas: Ejemplo para 4 archivos

Nombre	a1.wav	a2.wav	a3.wav	a4.wav
Categoría	Antropogénico Transporte Terrestre ... pasaje auto	Naturales Biológicos Domésticos ... ladridos	Naturales Geofísicos Lluvia ... lluvia piso duro	Antropogénico No mecánico No vocal ... campanas
Clase	Simple	Nube	Bucle	Nube
Instante Inserción	max	1	0	mid
$L_{wav,FS}(dB)$	c1.wav	c2.wav	c2.wav	-20
$L_{p,cal}$ (dB)	94	94	114	94
r (m)	10	20	1	100

Estas etiquetas conforman un grupo de metadatos asociadas al n-ésimo archivo de audio de la base de datos (ver figura 2.2 en 2.2). Es decir, para cada columna de la Tabla 3.1 debe haber un archivo de audio en la *Biblioteca*.

El campo nombre corresponde al nombre del archivo. En caso que el archivo se encuentre en la ruta de archivo predeterminada basta con escribir el nombre del ar-

[1] Calc en Web de OpenOffice: `http://www.openoffice.org/product/calc.html`

chivo; caso contrario se puede especificar la ruta del archivo, por ejemplo C:\combinador\base\otra ruta\a1.wav.

En las secciones siguientes se describen los demás metadatos. Cada metadato, específicamente los de *Categoría*, *Clase* e *Instantes de Inserción*, son necesarios para resolver una parte del problema de realismo y congruencia del estímulo sonoro de salida. La congruencia está asociada a las etiquetas de *Categoría* y se refiere a la congruencia con el ambiente que se pretende modelar y las fuentes sonoras que serán utilizadas. El usuario puede definir la cantidad de eventos de alguna categoría que aparecerán en el archivo de salida en función de la probabilidad de que dichos sonidos se encuentren en el ambiente sonoro que pretende generar. El realismo está relacionado con la congruencia y a su vez con la distribución temporal de los eventos sonoros. Las etiquetas de *Clase* son utilizadas para determinar si el intervalo temporal entre eventos sonoros responde a una distribución de Poisson o una tomada de la realidad. Las etiquetas de *Instantes de Inserción* son utilizadas para asegurar que partes importantes del evento sonoro no queden afuera del archivo de salida y de ese modo se pierda el sentido de haberlos incluido (ver 3.2.4).

3.2.1 Distribución temporal de eventos (Clase de eventos)

Muchos procesos que describen una serie de sucesos tienen una distribución aproximada de Poisson, que es además simple de generar. No obstante los procesos de Poisson no describen adecuadamente la distribución de todas las categorías de eventos sonoros. Es sabido que este tipo de procesos es adecuado cuando no existe relación entre los intervalos existentes entre eventos. En el caso del problema de este trabajo, el intervalo temporal que existe entre la aparición de eventos, generalmente, depende de los eventos anteriores y posteriores de la misma categoría de eventos (por ejemplo el canto de pájaros) y también depende de eventos sonoros de otras categorías (por ejemplo de paso de aeronaves). Esa dependencia ha sido observada en mediciones de campo de ruido de aeropuertos donde el canto de pájaros precedía al ruido del sobrevuelo o aterrizaje de aviones [Miyara et al., 2012]. Esta dificultad de modelar la dependencia entre eventos de diferentes categorías es simplificada en esta tesis bajo la suposición de que solo puede ser percibida por oyentes muy entrenados o especializados en la percepción de estas características del sonido ambiente.

El intervalo de tiempo entre eventos de la misma categoría depende de procesos como restricciones de velocidad, formación y disolución de embotellamientos en eventos de tránsito terrestre o de contexto de comportamientos en el caso de eventos sonoros de origen biológico (ver 11.1). En esta tesis se utilizarán unas simples reglas para manejar la dependencia del intervalo temporal entre eventos sonoros en tres grupos distintos de archivos de audio de la base.

Durante la compilación de archivos (ver módulo *Compilación* en la figura 2.3) se debe asignar un valor a la etiqueta clase de eventos (ver Tabla 3.1) según el grupo al que corresponda cada archivo.

El primer grupo se denomina *eventos simples* y contiene un solo evento por archivo. Pertenecen a este grupo aquellos eventos que pueden ser modelados de manera realista mediante procesos de Poisson. El segundo grupo se denomina *eventos largos* y contiene un solo evento de larga duración por archivo. Pertenecen a este grupo aquellos eventos que deben ser extendidos de modo de estar presentes durante toda la duración del sonido ambiente de salida. Por último el tercer grupo se denomina *nube de eventos* y contiene en cada archivo de sonido un grupo de eventos del mismo tipo de fuente y que están distribuidos temporalmente como fueron grabados de la realidad, como por ejemplo la frase completa del canto de una pájaro que puede contener más de un evento.

El intervalo temporal entre eventos simples se calcula mediante un algoritmo de generación de procesos de Poisson [Accolti et al., 2010a].

Los *eventos largos* deben tener al menos una gran proporción de muestras de audio respecto a las que tendrá el sonido ambiente de salida. Además estos sonidos son estacionarios y pueden ser repetidos en forma de bucle si sus secciones iniciales y finales tienen una amplitud y espectro similares y de banda ancha. Esto permite generar bucles cuando los archivos no alcanzan la duración que debe tener el archivo de salida o bien recortar el archivo cuando su duración es superior a la del archivo de salida. Ejemplos de estos eventos corresponden a eventos climáticos (lluvia, viento sobre una superficie, etc.), ruidos mecánicos (computadoras, sistemas de aire acondicionado), etc.

El intervalo de tiempo entre eventos en un archivo de audio del grupo *nube de eventos* se relaciona fuertemente con el intervalo de tiempo entre los eventos ante-

riores del mismo archivo. Por ejemplo el ladrido de perros, los gemidos, gruñidos, etc. se relacionan en una forma compleja que depende de infinidad de factores. Incluso los ladridos de un perro se relacionan con los de algún perro que se encuentra cercano al primero. Las características de esa dependencia están fuera de los alcances de esta tesis pero su estudio es de interés para los sistemas de realidad virtual (ver 11.1). Los eventos de los archivos de la clase *nube de eventos* no son separados de manera individual, obteniendo de esta manera la representación de una versión de la dependencia entre los intervalos temporales de los eventos.

3.2.2 Categorías de eventos

Dentro de las actividades de etiquetado de los archivos de audio de la base de datos, se encuentran las etiquetas de *Categoría* (ver Tabla 3.1). Estas etiquetas ofrecen control al usuario sobre diversos aspectos asociados al tipo de sonidos. A fin de ofrecer una vasta gama de opciones de control se utiliza una estructura de categorización jerárquica para describir el tipo de fuentes sonoras y el tipo de eventos sonoros. Esta estructura se implementó en trabajos previos [Accolti et al., 2010a,b].

En la figura 3.1 se muestra un esquema de las categorías jerárquicas utilizadas.

Los diferentes niveles de las categorías no buscan una descripción etimológica ni una división que refleje totalmente los principios de funcionamiento de las fuentes sonoras sino que se busca una división que permita estudiar aquellas fuentes sonoras que están más presentes en la mayoría de las ciudades y que permita separar aquellas fuentes que ya se conoce de antemano que son percibidas de alguna manera distinta al resto. El esquema está inspirado en el esquema de Stammets y Chesmore [Stammers y Chesmore, 2008], con algunas modificaciones basadas en trabajos de distintos autores [Schafer, 1977; Matsinos et al., 2008; Bunting et al., 2009] y otros criterios. Las modificaciones se deben a criterios del grupo de trabajo (autor, directores y colegas) respecto al interés para el estudio de los efectos del ruido en el ser humano en el contexto regional. Sin embargo no existe actualmente un acuerdo uniforme para categorización de eventos sonoros.

En el primer nivel, siguiendo la figura 3.1 para cada archivo, se debe indicar si el sonido es natural o antropogénico, es decir, si es generado por humanos o no.

En el segundo nivel, los sonidos de la categoría antropogénicos, se dividen en

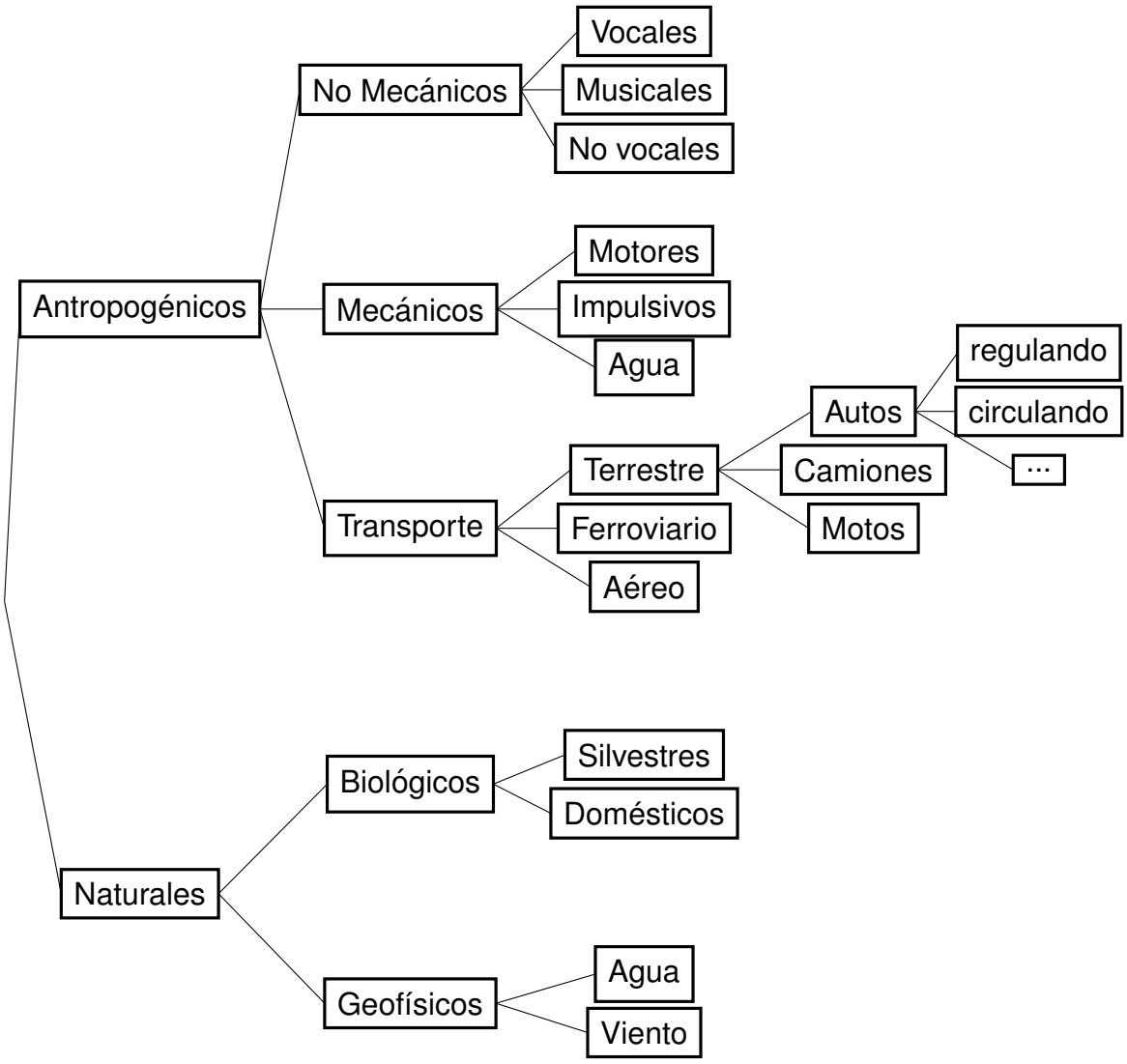


Figura 3.1: Estructura clasificación jerárquica de tipo de evento

mecánicos, no mecánicos y de transporte; mientras que los naturales se dividen en geofísicos y biológicos. Si bien los sonidos de transporte en general se deben a principios mecánicos, resulta de interés poder separar los ruidos de transporte debido a que de esa manera se han separado habitualmente en investigaciones previas, dado que este tipo de fuentes son, en general, los de mayor presencia en la mayoría de las ciudades, y a otras características como su regulación legislativa o la necesidad de transporte.

El tercer nivel, para los sonidos no mecánicos, se divide en vocales, musicales y no vocales. Por supuesto que los sonidos no vocales podrían a su vez dividirse en otras categorías que serían más representativas, por ejemplo calzados sobre suelo, roce de

ropa, aplausos, etc. pero, esto obligaría a complicar el esquema siendo que la base de datos no contiene actualmente una gran cantidad de sonidos en estas categorías. La base no contiene gran cantidad de sonidos en estas categorías porque no son categorías con un gran interés, en comparación con las que si están presentes, en cuanto a efectos del ruido en el ser humano.

El tercer nivel, para los sonidos mecánicos, se divide en motores, impulsivos y agua. Nuevamente estas categorías responden a un compromiso entre criterios relacionados con dos tipos de sonidos, por un lado los estacionarios, los de motores habitualmente asociados a sonidos con connotaciones negativas y los de agua asociados habitualmente a sonidos con connotaciones positivas y por otro los impulsivos que en general son sonidos que captan la atención de quien los escucha generando una interferencia con la actividad que realiza ese oyente.

El tercer nivel, para los sonidos de transporte, se divide en aéreo, terrestre y ferroviario. Esta división responde a cómo han sido divididos en investigaciones anteriores [Miedema y H., 1998] donde se ha demostrado empíricamente que, en promedio, a un mismo nivel sonoro, los ruidos de transporte aéreo son más molestos para la sociedad que los de tránsito automotor y estos más molestos que los de ruido ferroviario.

El tercer nivel para los sonidos geofísicos se divide en agua y viento. Se trata de sonidos de lluvia y viento que interactúan con otros objetos. Es decir, el agua de la lluvia por si sola o el viento por si solo difícilmente generan ruido por el roce con el aire (en el caso del viento sucede esto cuando las ráfagas son turbulentas) pero si lo hacen cuando interactúan con algún otro objeto como el suelo, bordes de construcciones, hojas secas, etc.

El tercer nivel para los sonidos biológicos se divide en domésticos y silvestres. Esta división trata de separar sonidos de modo tal que permita la simulación de ambientes rurales o urbanos pasando por ambientes intermedios si el usuario final de la herramienta especifica porcentajes de fuentes sonoras utilizando algún criterio.

A su vez, el esquema permite agregar más niveles hasta llegar al individuo y/o la acción. Por ejemplo en la figura 3.1 se ha detallado un cuarto y quinto nivel para transporte terrestre. El sonido que en el cuarto nivel corresponde a auto tiene un quinto nivel con características como circulando, regulando, etc. Además podría continuar identificando marca, modelo, revoluciones por minuto del motor, marcha, etc. En ge-

neral la estrategia usada es llegar hasta el nivel de detalle con que se cuente según la información disponible desde donde se incorporó el archivo.

3.2.3 Calibración de archivos de la base

Los archivos de audio, si fueron registrados con un sistema que no distorsione la señal acústica, contienen una señal digitalizada con la forma de onda temporal de la señal acústica. La señal digitalizada tiene una amplitud relativa a la presión sonora pero es de interés conocer la presión sonora absoluta o el factor de escala que permita calcular cuál fue la presión sonora que captó el micrófono.

Para poder reproducir nuevamente ese sonido al mismo nivel sonoro será necesario caracterizar también el sistema de reproducción sonora. Este tema se ampliará en el capítulo 7. Por el momento nos enfocaremos en representar matemáticamente la señal acústica a partir de la señal digital registrada.

Existen diversas formas de obtener la constante de calibración para poder escalar el archivo de audio de modo de obtener los valores de presión sonora instantánea que captó el micrófono. En general los métodos propuestos requieren un mínimo conocimiento de los eslabones que componen la cadena electroacústica del sistema de registro. Lamentablemente esta información no está disponible en todas las bases de sonido existentes.

Un sistema habitual de registro de audio está compuesto de los siguientes elementos conectados secuencialmente y en el siguiente orden: i) un micrófono, ii) un preamplificador con ganancia regulable, iii) un conversor analógico-digital y iv) un sistema de almacenamiento digital. Los eslabones i a iii tienen una determinada sensibilidad. En el eslabón i la sensibilidad corresponde a la relación entre la tensión de salida del micrófono y la presión sonora, se expresa en unidades de tensión por unidad de presión sonora. El eslabón ii, independientemente de su función que es la de adaptar impedancias, tiene una sensibilidad que corresponde a la relación entre la tensión de salida y la de entrada, es decir, se trata de una ganancia adimensional. La sensibilidad del eslabón iii corresponde a la relación entre la palabra binaria y la tensión que representa. Para simplificar la explicación, y teniendo en cuenta cómo representan las señales la mayoría de los softwares de cálculo matricial, se supondrá que el valor máximo admisible para la palabra digital es el que corresponde a la unidad en

un sistema decimal, y el mínimo a la unidad negativa. Con esta definición, la sensibilidad será la tensión máxima (o mínima) representable digitalmente por el sistema iii. El último bloque simplemente archivará estas palabras digitales que representan los valores de las muestras de audio.

El eslabón con mayores incertidumbres es habitualmente el micrófono. Esto se debe a que la transducción de señales acústicas en eléctricas involucra la vibración de una superficie, habitualmente una membrana. Asegurar la estabilidad de las condiciones mecánicas de esa membrana, por ejemplo una tensión uniforme, presenta ciertas dificultades tecnológicas. Es por esta razón que la recomendación para registrar sonidos, en el contexto de este trabajo y similares, consiste en utilizar el micrófono de un sonómetro calibrado en laboratorio, como mínimo cada dos años, asegurando así su trazabilidad a patrones internacionalmente aceptados y luego comprobar, al momento de iniciar la medición, el estado de calibración mediante un calibrador acústico a su vez verificado también en laboratorio.

El eslabón de ganancia regulable suele ser un potenciómetro con pocos puntos fijos o ninguno. Por esta razón no es sencillo saber cuál es la sensibilidad de este bloque simplemente observando la posición del potenciómetro. Sin embargo este eslabón es mucho más estable que el micrófono en cuanto a su sensibilidad. El eslabón de conversión digital también es estable, en cuanto a su sensibilidad, comparado con el micrófono.

Los sistemas de registro de audio actuales cuentan con buenas características en cuanto a relación señal ruido y distorsión armónica. Suponiendo que esas distorsiones no son importantes y que no se utiliza ninguna ganancia autoajustable, aun la respuesta en frecuencia de cada eslabón puede no ser plana introduciendo otro tipo de distorsión a la señal original. La respuesta en frecuencia se puede corregir mediante filtros digitales, si esta respuesta es conocida.

Uno de los métodos para obtener la constante de calibración requiere conocer la sensibilidad de cada eslabón de la cadena electroacústica que compone el sistema de registro. Este método presenta la dificultad de conocer el valor de ganancia ajustable y de corregir pequeñas diferencias en la sensibilidad del micrófono debidas a las condiciones de campo. En ese caso se puede establecer el nivel a fondo de escala para un sonido hipotético de 1 Pa multiplicando la sensibilidad de cada eslabón. Es decir,

en la tabla 3.1 se utiliza $L_{p,cal} = 94$ dB que corresponden a 1 Pa y se calcula $L_{wav,FS}$ de la siguiente manera:

$$L_{wav,FS} = 10 \log \left(G_m G_p G_D \right) \tag{3.2}$$

siendo G_m, G_p y G_D la sensibilidad del micrófono, del preamplificador y del conversor analógico-digital, respectivamente. Cualquier otro eslabón que pudiese afectar la cadena electroacústica se puede caracterizar por su sensibilidad e incluir la misma multiplicando dentro del logaritmo.

Este método se usa con fines de diseño, apelando a constantes nominales más que reales. Esto se verá reflejado seguramente en una desviación respecto al valor de presión sonora que habría captado un sonómetro. Sin embargo es de esperar que esta desviación no sea muy elevada y que no cause una pérdida de realismo desde el punto de vista de la percepción.

Un segundo método es utilizar un calibrador de campo. En este caso, si se utiliza un tono de 1 Pa o, en términos de nivel sonoro, de 94 dB, el nivel de la sensibilidad del total de la cadena electroacústica de registro se puede calcular a partir del tono que queda grabado en el archivo de audio. Luego, $L_{p,cal} = 94$ dB, u otro nivel si se utiliza un tono con otro nivel sonoro conocido, y $L_{wav,FS}$ se calcula usando los datos de audio registrado del tono y como referencia el fondo de escala.

El protocolo de etiquetado admite los dos métodos. Es decir, se puede simplemente dar un valor numérico a $L_{wav,FS}$ o una ruta al archivo de audio que contiene el tono de calibración correspondiente (ver tabla 3.1).

La etiqueta de distancia debe ser completada con la distancia aproximada desde la fuente sonora, o grupo de fuentes sonoras, hasta el micrófono. Esto permitirá que el usuario final tome decisiones respecto al nivel sonoro máximo y mínimo que tendrán los sonidos de entrada, presentes en el sonido de salida, basado en la distancia máxima y mínima que puede haber entre la fuente sonora virtual y el oyente.

3.2.4 Instantes de inserción

Dado que la combinación final es una mezcla de archivos de audio de distintas duraciones puede suceder que algunos archivos de entrada se incluyan parcialmente. Esto hace necesario definir una estrategia para que las partes que queden fuera del archivo de salida influyan lo menos posible en el resultado final, es decir, en cuanto

se adecuan los parámetros del archivo de salida respecto a los especificados por el usuario.

La estrategia es de crucial importancia para eventos dentro del grupo eventos simples (ver 3.2.1). Por ejemplo el pasaje de un auto puede durar unos 30 segundos pero si se debe insertar unos 10 segundos antes que finalice el archivo de salida, lo más probable es que éste no se perciba debido a que su mayor energía no se concentrará en los primeros 10 segundos de dicho archivo de entrada y será excluida. La estrategia es no utilizar la primera muestra del archivo como el punto de referencia temporal para ser insertado en la mezcla final sino otro instante dependiendo de cada archivo y el criterio del personal a cargo de la compilación.

En las etiquetas, el personal encargado de la compilación, debe especificar el punto de referencia temporal de cada archivo. Puede utilizar directamente el número de muestra $j_{\mathrm{b},n}$, el valor nulo o las palabras *max* o *mid*. Un valor muy usado es la unidad. Este valor indica al bloque *Importar* que el instante de inserción es la primera muestra. Este valor es útil para eventos de tipo impulsivo, de impacto o cuando la parte más importante del sonido está en el inicio del archivo.

El valor nulo para la etiqueta instante de inserción es el preferido para eventos de clase *eventos largos* y, si bien es redundante, indica al bloque *Importar* que el archivo es candidato a ser repetido en bucle si su duración es menor a la duración del archivo de salida.

El valor *max* es ideal para pasajes de medios de transporte y eventos de la clase *eventos simples*. Este valor indica al bloque *Importar* que debe reemplazar esta etiqueta por el valor de la muestra $j_{\mathrm{b},n}$ donde el nivel sonoro instantáneo es máximo.

El valor *mid* es ideal eventos de la clase *nube* y eventos en los cuales la parte más importante del archivo, bajo algún criterio, se encuentre en la mitad temporal del archivo.

3.2.5 Otras etiquetas

El protocolo permite además incorporar otras etiquetas con información de interés. Por ejemplo información respecto a la biblioteca original, rutas a archivos de imagen conteniendo fotos, esquemas, etc., detalles del sistema completo de registro y detalles de la edición entre otros datos útiles.

El protocolo indica que se debe recabar toda la información posible con vistas al aprovechamiento de esta base para futuros trabajos.

3.3 Edición

Las tareas de edición de audio sirven para adecuar los archivos de audio para respetar las características necesarias de los archivos de audio para conformar parte de la base de datos. Es recomendable que la persona a cargo de esta tarea sea la misma encargada de la recolección de archivos y del etiquetado pues gran cantidad de la información necesaria para estas tareas es de rápido acceso al momento de abrir cada archivo en un software editor de audio.

Para las tareas de edición de audio de esta tesis se utiliza el software Audacity®[2] en diversas versiones. La edición comprende unas tareas básicas para todos los archivos. La primera es cortar silencios anteriores y posteriores a los eventos sonoros que estarán presentes en ese archivo. Esta tarea involucra la selección de clase de evento que luego será introducida en la etiqueta correspondiente a ese archivo.

En el caso de *eventos largos*, que luego se pueden presentar en forma de bucle, el corte inicial y final debe ser cuidadosamente ubicado. Para ello se utilizan las herramientas de zoom y reproducción sonora de Audacity y se busca cuidadosamente las muestras más cercana al valor nulo del inicio y del fin. En el caso del inicio la muestra debe ser seguida por una muestra con valor positivo mientras que en el caso del fin la muestra debe ser posterior a una muestra con valor negativo. De esta forma, si el contenido espectral es de banda ancha y aproximadamente similar, no se escuchará un clic al hacer el bucle. Se debe utilizar la herramienta de reproducción en bucle de Audacity para corroborar auditivamente que no se escuchan clics.

Una vez editado el archivo, se guarda en la ruta de archivos donde estará la biblioteca de archivos de audio y se escribe dicha ruta en la etiqueta correspondiente. Notar que de un mismo archivo proveniente de la recolección es posible generar más de un archivo para introducir en la biblioteca.

En caso que el método de calibración, para el archivo que se está editando, haya

[2] El software Audacity® está registrado © 1999-2014 Audacity Team. [Web site: `http://audacity.sourceforge.net/`. Es un software libre distribuido bajo los términos de la licencia GNU (General Public License)] El nombre Audacity® es marca registrada de Dominic Mazzoni.

sido el de un calibrador acústico de campo se suma la tarea de editar el archivo de audio de calibración (o en caso que ya haya sido editado para otro archivo en la biblioteca simplemente se debe poner la etiqueta correspondiente). En algunas ocasiones el registro de audio contiene los momentos en los cuales el calibrador se colocó y se retiró del micrófono. En esos instantes el tono estará presente pero la amplitud no será representativa. La edición de estos archivos corresponde a dejar solo la porción de audio que tiene una forma de onda de amplitud constante evitando los momentos de colocación y retiro del calibrador. La estrategia para el inicio y fin es idéntica a la usada para *eventos largos*, en este caso con la finalidad de que el algoritmo que calculará el nivel sea más preciso, contemplando periodos enteros del tono.

Notar que un mismo archivo de calibración puede corresponder a más de un archivo de eventos sonoros en la biblioteca. Normalmente esto ocurre cuando en una misma campaña de campo se registran varios eventos. Sin embargo es recomendable que las campañas no sean muy extensas o, en caso de serlo, registrar un nuevo tono de calibración cada 2 horas y cada vez que las condiciones climáticas varíen considerablemente.

Capítulo 4

Análisis espectral

El módulo de análisis espectral se utiliza para caracterizar la composición espectral de los sonidos de la Base de Datos, pero también para analizar el sonido ambiente de salida. Este bloque está presente, como subbloque, dentro del bloque *Importar* en el módulo *Base de Datos* y en dos bloques del módulo *Análisis*.

4.1 Espectro de líneas

El análisis espectral es implementado mediante la suma energética de las lineas espectrales de la Transformada Rápida de Fourier (FFT) promediadas energéticamente en el tiempo. La FFT se calcula usando $N_{\mathsf{FFT}} = 4\ 096$ muestras. La n_{FFT}-ésima frecuencia de la transformación corresponde a

$$f(n_{\mathsf{FFT}}) = n_{\mathsf{FFT}} \times F_{\mathsf{S}}/N_{\mathsf{FFT}} \tag{4.1}$$

para una señal con tasa de muestreo F_{S}, siendo $n_{\mathsf{FFT}} = 0, 1, \cdots, N_{\mathsf{FFT}} - 1$.

La resolución en frecuencia, dada por la diferencia entre dos lineas contiguas, es de $F_{\mathsf{S}}/N_{\mathsf{FFT}}$. Los sonidos de la base de datos están muestreados a una tasa de $F_{\mathsf{S}} = 44\ 100$ Hz alcanzando una resolución de aproximadamente 10 Hz. Para obtener una mejor resolución en frecuencias bajo los 688 Hz, sin aumentar la cantidad de muestras que debe procesar la FFT, los archivos son diezmados por un factor de 32 y se calcula una nueva versión de la FFT. Esta nueva versión de la FFT ($FFTD$) alcanza una resolución de $0{,}3$ Hz para sonidos originalmente muestreados a $F_{\mathsf{S}} = 44\ 100$ Hz.

En la figura 4.1 se muestra un ejemplo del análisis espectral realizado a un sonido de la base de datos que corresponde al pasaje de una moto. Las ordenadas expresan

el nivel de presión sonora (ver Apéndice A, ecuación (A.3)).

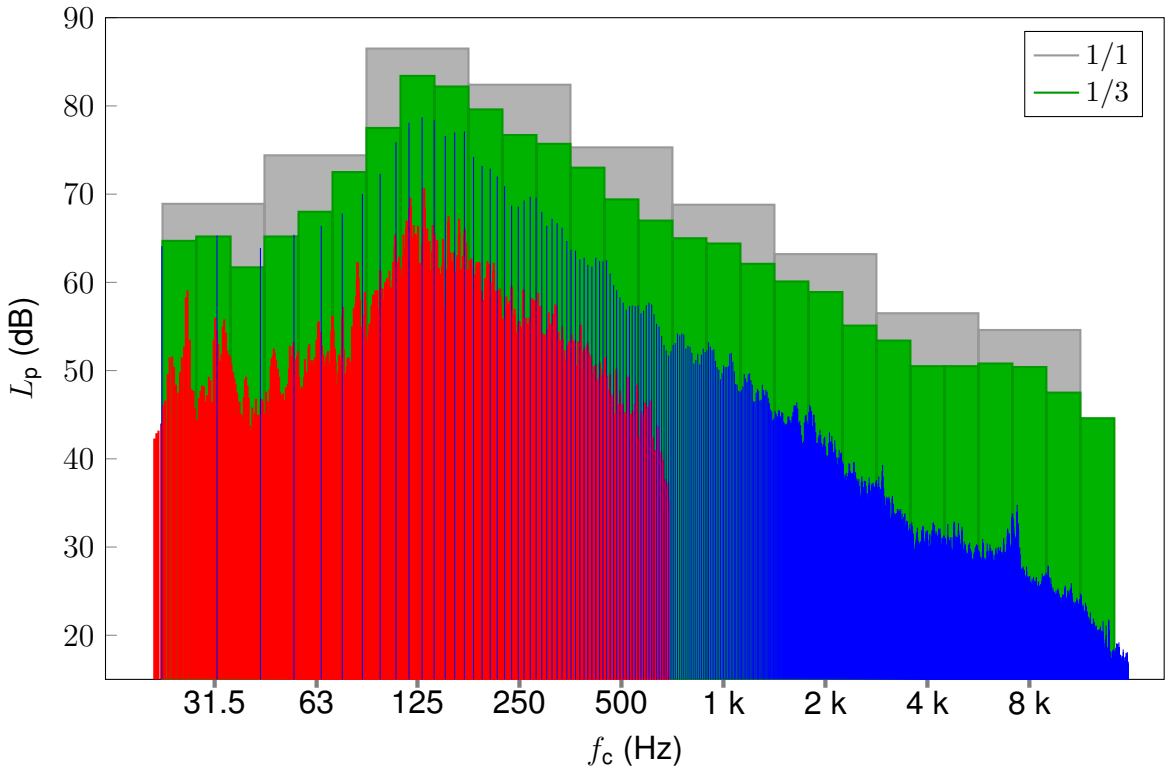


Figura 4.1: Análisis espectral del pasaje de una moto. Azul: FFT ($F_S = 44\ 100$ Hz). Rojo: $FFTD$ ($F_S \approx 1\ 400$ Hz). Barras grises: Bandas de $1/1$ octava. Barras verdes: Bandas de $1/3$ octava.

En líneas azules se muestra el promedio temporal energético de la FFT aplicada a la señal con la tasa de muestreo original (FFT). En líneas rojas se muestra la misma transformación pero aplicada sobre la versión diezmada de la señal ($FFTD$). Se visualiza rápidamente que en bandas de bajas frecuencias la versión $FFTD$ contiene más líneas que la FFT pero esta última alcanza frecuencias más altas respecto a la $FFTD$.

4.2 Espectro de bandas

Una vez obtenidos estos dos espectros de líneas, la energía de cada banda se obtiene mediante la superposición energética de todas las líneas espectrales que caen dentro de esa banda, donde la frecuencia de cada línea se calcula según la ecuación (4.1).

La superposición de las líneas espectrales se justifica por la identidad de Parseval. Partiendo de la definición de presión sonora eficaz o rms (ver p_{rms} en A.2), que se

usará más adelante en 8.3,

$$p_{\text{rms}} = \sqrt{\frac{1}{T}\int_0^T p(t)^2 \mathrm{d}t} \tag{4.2}$$

es posible demostrar la identidad

$$\int_{-\infty}^{+\infty} s(t)^2 \mathrm{d}t = \int_{-\infty}^{+\infty} S(f)^2 \mathrm{d}f \tag{4.3}$$

La identidad de Parseval establece que la energía total de la señal s sumada en todo el tiempo t es igual a la energía de su transformada de Fourier $S(f)$ sumada en todo el rango de frecuencias f. Si el filtro aplicado para cada banda de frecuencias fuese un filtro ideal, entonces, los valores de la transformada de Fourier serían nulos fuera de las frecuencias de corte inferior y superior.

4.2.1 Espectro de bandas de fracción de octava

Para el caso de las bandas de octava, o fracción de octava, basta con calcular las frecuencia de corte e integrar las líneas que caen dentro de cada banda según el termino del lado derecho de la ecuación (4.3), es decir el valor cuadrático de cada línea espectral dentro de la banda. Las frecuencias de corte inferior (f_{inf}) y superior (f_{sup}) [IEC 61260, 1995; IRAM 4081, 1977] vienen dadas por

$$f_{\text{inf}} = f_{\text{c}} 2^{-\frac{1}{2b}} \qquad\qquad f_{\text{sup}} = f_{\text{c}} 2^{\frac{1}{2b}} \tag{4.4}$$

donde $b = 1$ para bandas de octava, $b = 2$ para bandas de media octava, $b = 3$ para bandas de tercio de octava, etc.

Las frecuencias centrales nominales, las que identifican a cada banda, son las definidas en la norma IRAM 4081 [1977] en concordancia con IEC 61260 [1995]. Las frecuencias centrales exactas, usadas para calcular las frecuencias de corte, son calculadas según

$$f_{\text{c}} = 1\,000 \times 2^{\frac{m}{b}} \tag{4.5}$$

donde m es un entero que simboliza el índice de cada banda.

En la figura 4.1 se muestra las bandas de 1/1 octava, calculadas con este método, en color gris detrás de los espectros FFT y $FFTD$ correspondientes a las líneas espectrales. Entre las líneas espectrales y las bandas de 1/1 octava se muestran las bandas de 1/3 de octava con color verde. En el eje de las abscisas se marcan las

frecuencias centrales nominales de las bandas de octava entre los índices m en el intervalo $[-5, 3]$. Notar que la misma ecuación de suma energética, que corresponde a la versión discreta de la integral energética de la identidad de Parseval, se puede aplicar cada 3 bandas de 1/3 de octava para obtener las bandas de 1/1 octava. Este método de sumar energéticamente las bandas de 1/3 es el método que utiliza el algoritmo implementado en esta tesis para calcular las bandas de 1/1 octava.

Hasta la banda de octava centrada en $f_c = 250$ Hz, y hasta la banda de 1/3 de octava centrada en $f_c = 500$ Hz, se utiliza la suma energética de las líneas espectrales $FFTD$. Para el resto de las bandas se utiliza FFT. En la figura 4.1 se visualiza rápidamente que en las bandas de 1/3 de octava centradas entre $f_c = 31{,}5$ Hz hasta $f_c = 63$ Hz solo hay una línea espectral FFT por cada banda. Esta es la razón por la cuál se utiliza una mayor resolución en bajas frecuencias, mediante la versión $FFTD$, lo cual permite una mayor cantidad de líneas espectrales y por lo tanto reduce el error debido al derrame espectral (para mayores detalles sobre el derrame espectral ver [Oppenheim y Schafer, 1975]). Además del derrame espectral, notar que la banda de 1/3 centrada en $f_c = 25$ Hz no contiene ninguna línea espectral, lo cual hace imposible calcular su energía. El error por derrame espectral es evidente en la banda de 1/3 centrada en $f_c = 40$ Hz, donde la única línea espectral de la versión FFT tiene un nivel sonoro de $L_p = 63{,}9$ dB comparado con $L_p = 61{,}5$ dB que se obtiene de la suma energética de la versión $FFTD$ para la cual se cuenta con 27 líneas.

El bloque *importar* del módulo *base de datos*, analiza y registra no sólo los espectros en bandas de 1/1 octava y 1/3 de octava sino también los espectros FFT y $FFTD$, además del espectro en bandas críticas que se describe a continuación. El registro de los espectros FFT y $FFTD$ permite al usuario configurar, en el bloque *Configuración* del módulo *Combinador*, otro tipo de bandas en caso de ser necesario. Por defecto se utilizan las bandas de 1/1 octava en la configuración.

4.2.2 Espectro de bandas críticas

Las bandas críticas son de especial interés cuando se analizan respuestas del ser humano a estímulos sonoros. Estas bandas tienen un ancho equivalente al ancho en el cual, aproximadamente, el sistema auditivo integra la energía antes de ser percibida como sonoridad. Entre bandas separadas por una o más bandas críticas el sistema

auditivo no integra al percibir sino que percibe la sonoridad de aquella banda que tenga mayor energía (teniendo en cuenta no solo la suma energética sino también el enmascaramiento). Para una descripción más detallada se recomienda el libro *Psychoacoustics* [Fastl y Zwicker, 2005].

Para el caso de las bandas críticas, los extremos de cada banda, $f\left(z_{\text{inf}}\right)$ y $f\left(z_{\text{sup}}\right)$ en unidades de Hz, se calculan mediante la inversa de la ecuación (4.6) [Fastl y Zwicker, 2005].

$$z = 13 \times \arctan\left(\frac{f}{1316}\right) + 3{,}5 \times \arctan\left(\frac{f}{7500}\right)^2 \tag{4.6}$$

La ecuación (4.6) relaciona la variable psicoacústica razón de banda crítica (z), en unidades de barks, con la variable física frecuencia (f). La función inversa se calcula mediante el método de Newton-Raphson. Finalmente, de manera similar a las bandas de fracción de octava, se suman energéticamente las líneas espectrales que están entre $f\left(z_{\text{inf}}\right)$ y $f\left(z_{\text{sup}}\right)$ para cada banda crítica.

En la figura 4.2 se muestra un ejemplo del análisis espectral realizado al mismo pasaje de una moto de la figura 4.1 pero en este caso por bandas críticas. El eje inferior de abscisas muestra la frecuencia en unidades de Hz y el eje superior muestra la razón de banda crítica en unidades de barks.

La frecuencia central de cada banda crítica corresponde a la inversa de la ecuación (4.6) para $z_{\text{c}} = 0{,}5; 1{,}5; ...; 23{,}5$, la frecuencia de corte inferior corresponde a la inversa para $z_{\text{inf}} = z_{\text{c}} - 0{,}5$ y la frecuencia de corte superior corresponde a la inversa para $z_{\text{sup}} = z_{\text{c}} + 0{,}5$.

En la figura 4.2 se muestra el espectro en bandas críticas detrás de los espectros FFT y $FFTD$. Hasta la banda centrada en $z_{\text{c}} = 5{,}5$ barks, se utiliza la suma energética de las líneas espectrales $FFTD$ y para el resto de las bandas se utiliza FFT.

Un ejemplo de posibilidades brindadas al usuario, en la configuración del análisis espectral, es la de definir bandas de fracciones de bandas críticas. Por ejemplo tomando bandas de ancho inferior a 1 barks como podría ser $0{,}1$ barks.

4.3 Espectro en instante de nivel máximo

Es habitual que sonómetros y analizadores de espectro tengan una función que calcula el nivel máximo en cada banda de frecuencias. Pero en este caso el interés está en encontrar el espectro en el instante en que el nivel sonoro es máximo. En general, para

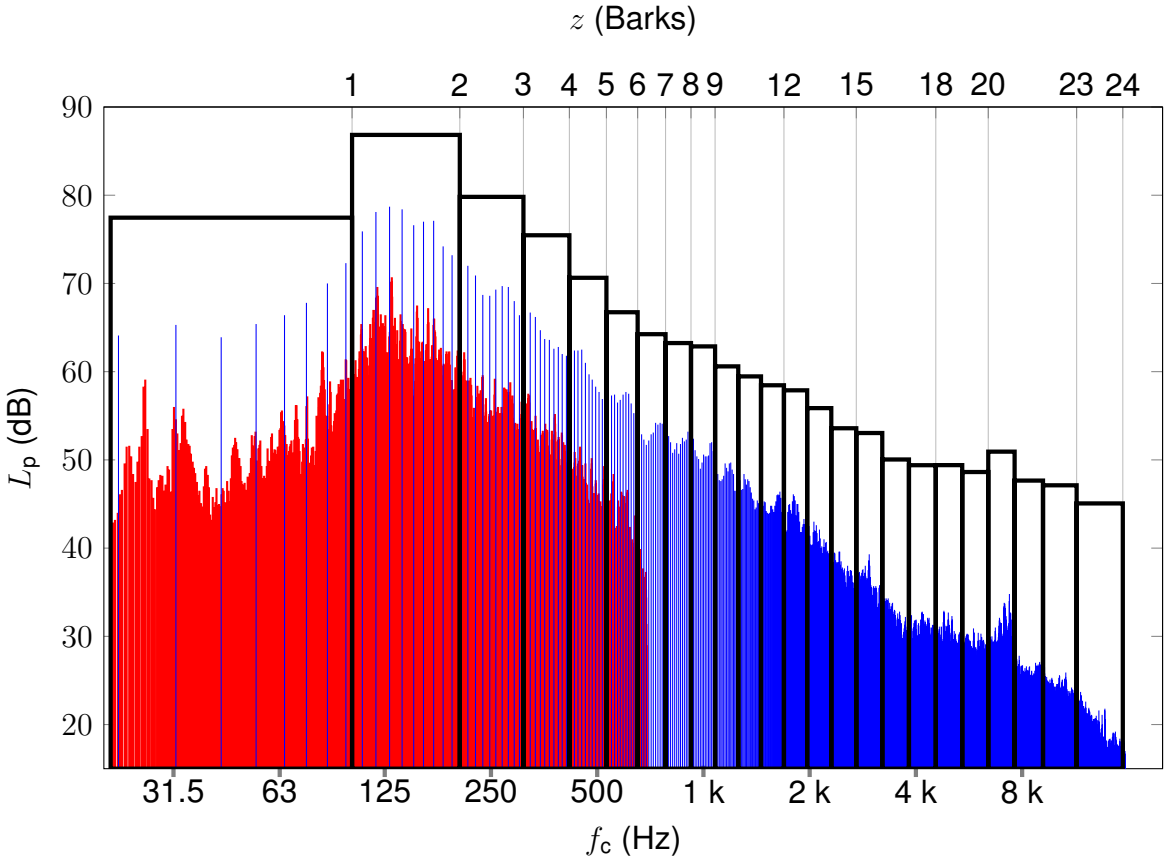


Figura 4.2: Análisis espectral del pasaje de una moto. Azul: FFT ($F_S = 44\ 100$ Hz). Rojo: $FFTD$ ($F_S \approx 1\ 400$ Hz). Barras transparentes (borde negro): Bandas Críticas.

un solo evento sonoro el instante donde ocurre el máximo de nivel sonoro en banda ancha es similar al máximo en cada una de las bandas pero no siempre es así. Puede existir un archivo de audio en el cual un evento sonoro contenga más energía sonora en una banda que otro evento que contiene más energía en otra banda. Por esta razón se introduce el espectro en el instante de nivel máximo. Este nuevo descriptor es mucho más representativo que los máximos de cada banda del espectro que pueden no darse nunca todos de manera simultánea.

Este indicador es de gran utilidad no solo para los propósitos de esta tesis y trabajos similares sino también para cálculos de propagación sonora de eventos impulsivos o de impacto. En el módulo de análisis espectral, especialmente cuando es usado para generar los metadatos de la base de sonidos (ver 2.2 y 6.2), se calcula este indicador.

Se calcula el nivel de presión sonora $L_{p,10\text{ms}}$ con un algoritmo que simula un circuito de promedio móvil con constante temporal $\tau = 10$ ms (ver Apéndice A) utilizando la

señal de audio s. Luego se identifica el instante $t_x = \arg\max[L_{p,10\text{ms}}(t)]$.

Por otra parte se aplica, a la misma señal s, un banco de filtros de la fracción de octava correspondiente obteniendo $L_{p,10\text{ms},m}$ para la m-ésima banda. Finalmente el espectro del nivel máximo corresponde a los M valores de $L_{p,10\text{ms},m}$ en el instante t_x, es decir $L_{p,10\text{ms},m}(t_x)$. Utilizar por un lado el espectro de líneas permite que el usuario defina bandas de otro tipo en caso que no utilice el espectro del nivel máximo. Utilizar el banco de filtros permite trabajar con constantes temporales para el promedio móvil lo cual es deseable para comparar con las normativas actualmente vigentes.

Los filtros de fracción de octava pueden aplicarse mediante diversos tipos de filtros digitales. En Miyara et al. [2009a,b] se contrasta, contra un sonómetro con análisis por bandas de octava y tercio de octava, algoritmos basados en técnicas de filtros de respuesta al impulso finita (FIR), de respuesta al impulso infinita (IIR) e implementación de FIR mediante transformada de Fourier (FFT). En esta tesis se utiliza el método de filtrado mediante FIR implementado mediante la técnica de FFT [Miyara et al., 2009b; Miyara, 2013a].

Finalmente se podría definir el espectro en instante de nivel máximo $L_{p,10\text{ms},m}(t_x)$ como el nivel de presión sonora, con constante temporal $\tau = 10$ ms, para cada banda m en el instante t_x donde el nivel sonoro de banda ancha, con constante temporal $\tau = 10$ ms, es máximo.

Capítulo 5

Histograma de duración

El histograma de duración de eventos es de especial interés por ser un indicador relacionado con las variaciones del patrón temporal. Para poder calcular el histograma de duraciones se debe definir la variable duración de un evento sonoro. Esta definición asocia implícitamente la definición de un evento sonoro.

Si bien existen tres clases diferentes de archivos (ver eventos simples, eventos largos y nubes de eventos en 3.2.1), según la manera en que se decida manejar la distribución temporal de eventos para cada archivo de audio, el histograma de duración para todas las clases se calcula de la misma manera. La diferencia entre estas tres clases se da en el instante y la forma en que se insertaran estos archivos de la base en el archivo de salida, según se describe en 8.9.1.

5.1 Modelo de duración

Según se introduce en 1.1.1.3, un modelo bastante adecuado para calcular la duración subjetiva de un evento sonoro es a través de la sonoridad en función del tiempo o bien del patrón de enmascaramiento. El modelo de Fastl et al. [2002] se basa en umbrales definidos de manera relativa a los mínimos locales. Si bien el modelo podría ser de gran utilidad para los objetivos de esta tesis, en cuanto se ajusta a cómo los sujetos perciben la duración, resulta que esto no es compatible con la posibilidad de predecir como serán estas duraciones cuando el evento sea combinado con otros eventos. La predicción de la duración subjetiva, al ser definida desde mínimos locales, depende fuertemente del contexto. El contexto en este caso corresponde a la presencia de otros eventos que puedan elevar esos mínimos locales. Por esta razón, y teniendo en

cuenta que el contenido espectral también se está evaluando en un sentido físico y no psicofísico, se decide utilizar una definición física sugerida en las normas ISO 1996-1 [2003] e IRAM 4113-1 [2009] con algunas adaptaciones para adecuar la definición a eventos cuyo nivel no es del todo uniforme entre eventos dentro del mismo archivo.

Según ISO 1996-1 [2003] e IRAM 4113-1 [2009], la duración de un suceso se debe especificar en función de algunas características del ruido tales como el número de veces que se sobrepasó un nivel umbral fijado. Es decir, no se trata de una definición cerrada sino que deja abierta la posibilidad de definir la duración a partir de otras características, dependiendo del análisis que se deba realizar.

Para los eventos de la clase Simple, es decir, si el archivo de audio contiene un solo evento, sería sencillo definir un criterio de duración en función del nivel máximo de su envolvente. Por ejemplo, en la figura 5.1, se muestra un criterio posible definido como el tiempo durante el cual el nivel de presión sonora es superior al de un umbral de -10 dB relativo al máximo.

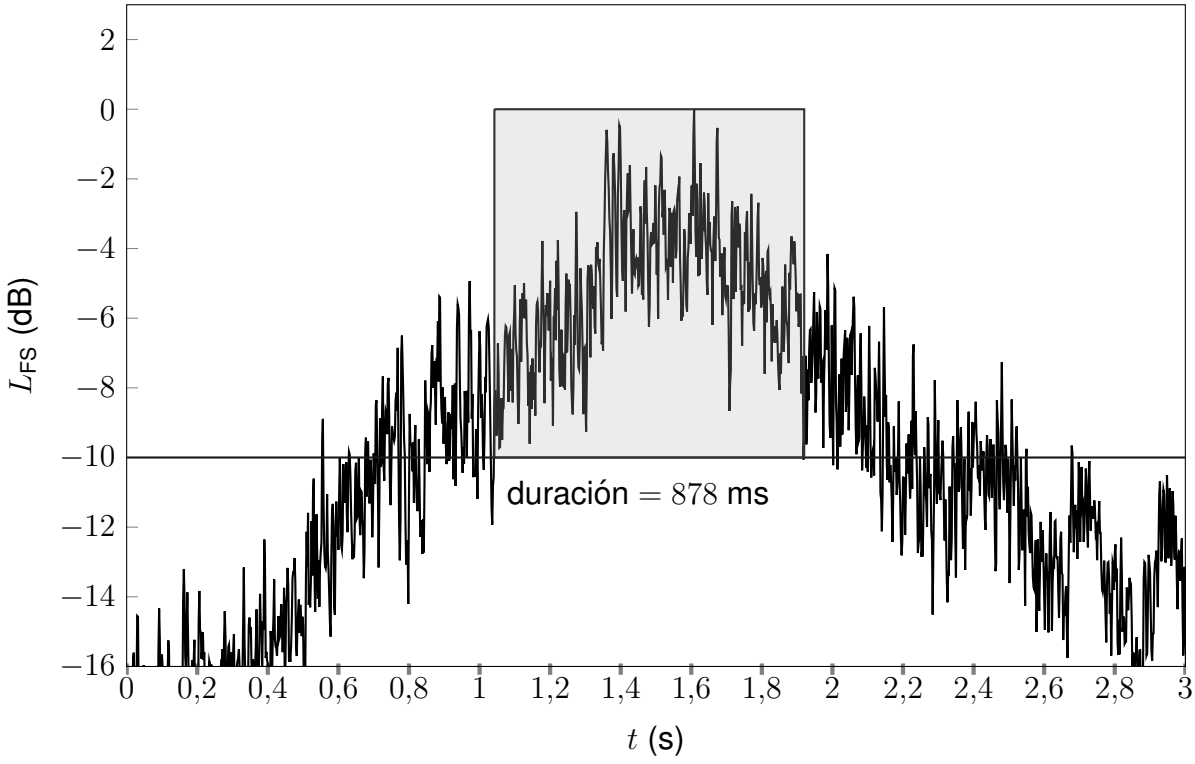


Figura 5.1: Criterio de duración con umbral en -10 dB relativo al máximo. Señal acústica correspondiente al paso de un automóvil

En la figura se muestra el nivel de presión sonora, calculada como promedio energéti-

co móvil con un filtro de primer orden aplicado a la señal cuadrática, y luego submuestreada a una tasa de muestreo de 500 Hz (ver Apéndice A). Este nivel está referido al fondo de escala, por lo tanto 0 dB corresponde al nivel máximo digital. Con un cuadro sombreado en gris se muestra el intervalo temporal en que el nivel supera el umbral de -10 dB relativo al máximo. Esta definición es congruente con las notas informativas de las normas ISO 1996-1 [2003] e IRAM 4113-1 [2009]. El máximo en este caso corresponde a 0 dB debido a que la señal está normalizada. La señal corresponde al pasaje de un automóvil.

Notar que en la figura 5.1 existen otras diferencias de nivel del orden de 10 dB, pero no son relativas al máximo total sino a máximos locales. Este hecho dificulta que la definición sea extensible a casos con más de un evento por archivo, como los son los de la clase *Nube*. Sucede que, los niveles mínimos locales con constante de tiempo $\tau = 2$ ms, no son percibidos pues quedan enmascarados. Y los niveles máximos locales, con dicha constante de tiempo, pueden ocurrir a niveles por sobre los 10 dB de esos mínimos locales sin ser percibidos como eventos separados. Esto redunda en la detección de eventos que no son percibidos.

La estrategia empleada en esta tesis para definir la duración de eventos, es a través de un umbral móvil que depende de niveles estadísticos locales (ver definición de niveles estadísticos en Apéndice A). La búsqueda de niveles estadísticos locales, alrededor de un máximo local, implica en sí la determinación de un intervalo de tiempo que contiene al evento. Es decir, la definición de duración está involucrando dos intervalos temporales, el intervalo en el cual se busca el nivel estadístico L_N y el intervalo de duración del evento según la definición misma.

A los efectos de esta tesis, se define la duración de un evento del siguiente modo: el tiempo durante el cual el nivel de presión sonora, con constante de tiempo $\tau = 2$ ms, se encuentra sobre un nivel umbral de 10 dB por debajo del nivel estadístico L_5 del intervalo de tiempo local de incremento de nivel. Se utiliza L_5 porque es un valor más representativo del pico comparado con el máximo que puede ser muy corto debido a un aleatorio irrelevante. La estimación del intervalo de tiempo local también es móvil. Depende de dónde terminó el evento anterior o si se trata del primer evento del archivo, de la posibilidad de encontrar un máximo posterior asociado a un instante anterior y uno posterior que se encuentren bajo el umbral móvil, y de que no se solapen con otro

evento.

5.1.1 Algoritmo de búsqueda de eventos

Bajo las circunstancias antes descritas, es necesario construir un algoritmo para poder calcular la duración. La definición de la duración de los eventos sonoros quedará entonces asociada a este algoritmo. En esta sección se introduce dicho algoritmo mediante el diagrama de flujo de la figura 5.2. En la siguiente sección (sección 5.2) se detalla cómo se define el evento sonoro y su duración en el algoritmo y el bloque *calcular histograma* se detalla en la sección 5.3.

En los primeros pasos del algoritmo (ver figura 5.2) se calculan los máximos locales $M(i)$ del nivel de presión sonora. El nivel de presión sonora se calcula con un filtro de promedio móvil de constante de tiempo $\tau = 2$ ms (ver Apéndice A) y se submuestrea a una tasa de $F_{\mathrm{s},2} = 500$ Hz[1] obteniendo así una estimación de nivel $L(j)$ para cada muestra $j \in \mathbb{N}$. Notar que el periodo de muestreo es de 2 ms en la señal submuestreada por lo tanto un incremento de una muestra en j corresponde a un incremento de 2 ms en el sentido temporal. Los índices $i \in \mathbb{N}$ corresponden al orden de aparición de cada máximo local y tienen asociado el valor de la muestra $j_M(i)$ donde ocurre el máximo local $M(i)$, es decir, $j_M(i)$ es la muestra donde ocurre $M(i)$.

No todos los máximos locales corresponden al máximo de un evento por lo cual se utiliza un índice i_{e} para indicar el orden secuencial de cada evento en el archivo. El máximo de cada evento no se puede estimar sin haber definido el inicio y final del evento por lo cual simplemente se retendrá el valor de la muestra j_c donde sucede el máximo más grande dentro de los límites del evento. El algoritmo fija tentativamente un umbral $u = 10$ dB por debajo del nivel máximo de cada evento para luego buscar la muestra j_0 que se encuentre debajo del umbral y más cercana al inicio del archivo o el final del evento anterior según se trate del primer evento o uno subsiguiente distinto al primero, respectivamente. Por esa razón se introduce j_1', para poder identificar el inicio de archivo o final del evento anterior según corresponda. El algoritmo inicializa $i = 0$, $i_{\mathsf{e}} = 0$, $j_c = 1$ y $j_1' = 1$ (ver tercer cuadro después del bloque inicio en figura 5.2).

En la primera iteración se inicia con el primer máximo local y se incrementa al

[1] Previo al submuestreo se aplica un filtro anti-alias mediante un filtro Chebyshev tipo I de orden 8 implementado como filtro de respuesta infinita

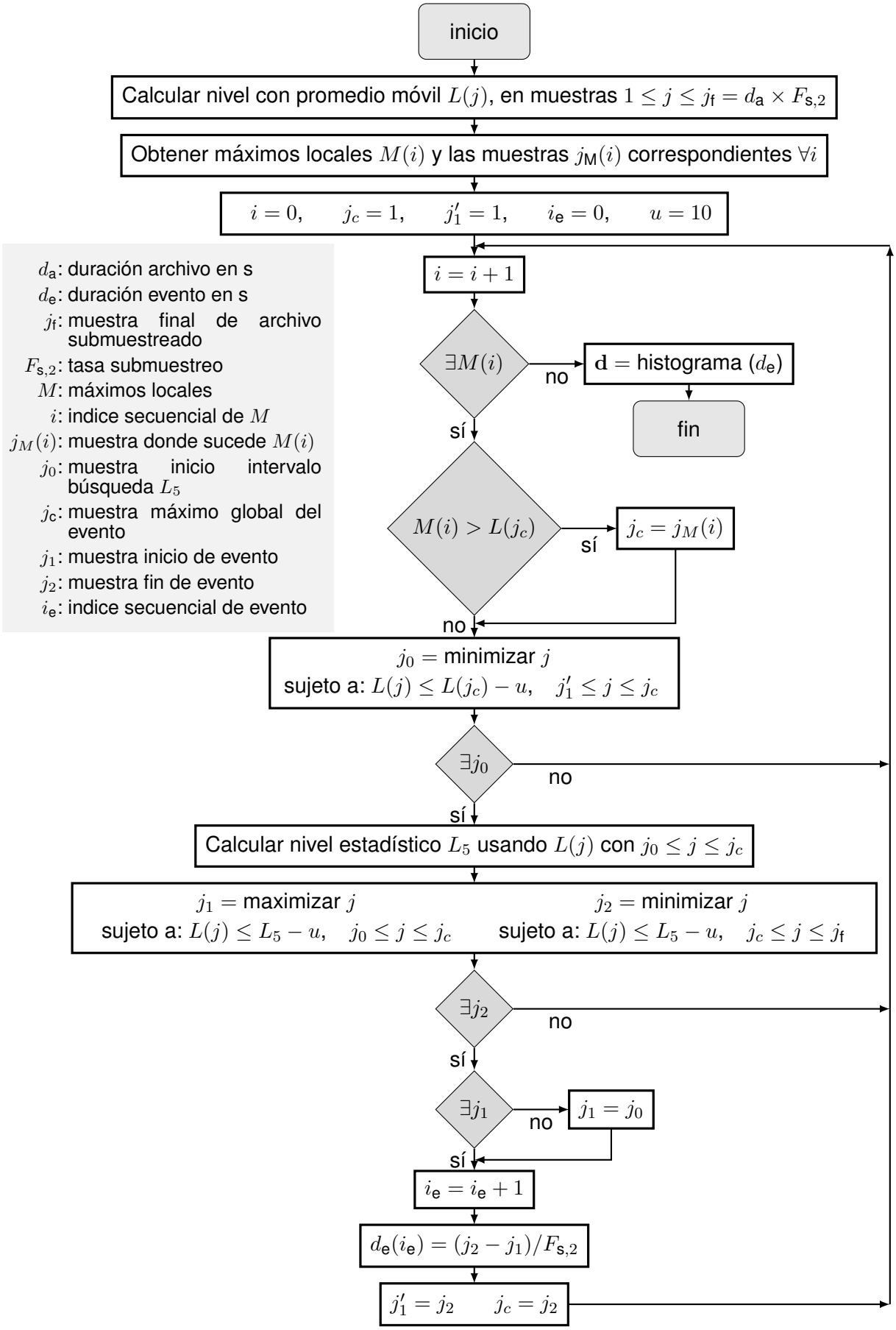


Figura 5.2: Diagrama de flujo de algoritmo de histograma de duraciones

siguiente en cada iteración (cuarto cuadro). El primer rombo (figura 5.2) es un bloque de decisión que identifica si se ha alcanzado el último máximo y en ese caso termina el algoritmo previo paso por el bloque de cálculo del histograma (ver sección 5.3). Caso contrario continúa.

A continuación, en el segundo bloque de decisión (ver segundo rombo en figura 5.2) se identifica si el nuevo máximo $M(i)$ es mayor que el valor $L(j_c)$ retenido en la iteración anterior. En caso de que se cumpla tal condición se actualiza j_c tomando el nuevo valor $j_c = j_M(i)$, con lo cual se actualiza j_c al valor de la muestra donde ocurre el mayor de los máximos locales analizados para ese evento. Caso contrario, si el nuevo máximo $M(i)$ es menor al anterior $L(j_c)$, se retiene el valor del máximo anterior al no modificar el valor de j_c.

En el siguiente bloque (primer bloque de minimización) se busca el índice j_0 correspondiente al primer valor de nivel sonoro del evento, anterior al índice j_c y posterior al índice j_1', que está debajo del umbral referido al valor $L(j_c)$, es decir el nivel máximo tentativo del evento. Si al evaluar el tercer bloque de decisión no existe un valor j_0 que satisfaga tal condición, entonces se inicia una nueva iteración. Es decir, en la nueva iteración se postula al siguiente máximo local como máximo global del evento. En caso de que exista un valor tentativo de j_0 se obtiene el nivel estadístico L_5 usando los valores de nivel de presión sonora $L(j)$ entre j_0 y j_c.

En el bloque siguiente (bloque de maximización y minimización) es donde se determinan las muestras inicial j_1 y final j_2 de cada evento sonoro y de allí se obtiene su duración. Esta minimización busca los valores más cercanos al nivel máximo que estén apenas bajo el umbral respecto a L_5. El valor de $L(j)$ por debajo del umbral correspondiente al inicio j_1 del evento se busca entre j_0 y j_c mientras que el valor correspondiente al final j_2 se busca entre j_c y j_f.

En caso de que no exista un valor j_2 que satisfaga la condición, lo cual se determina en el cuarto bloque de decisión, se inicia una nueva iteración. Si existe j_2 y el valor obtenido para j_1 no existiese, se valida para j_1 al valor tentativo de j_0 obtenido en el primer bloque de minimización.

Al existir j_2 existirá un valor de j_1 o se le asignará uno (el de j_0) por lo tanto se define un nuevo evento entre j_1 y j_2 cuyo índice se obtiene incrementando i_e en el bloque que sigue al último bloque de decisión. La duración de ese evento $d_\mathsf{e}(i_\mathsf{e})$ se

calcula en el bloque siguiente según

$$d_{\mathsf{e}}(i_{\mathsf{e}}) = \frac{j_2 - j_1}{F_{\mathsf{s},2}} \tag{5.1}$$

Luego se reinicia $j_1' = j_2$ y $j_c = j_2$ para comenzar una nueva iteración en búsqueda del siguiente evento o bien de determinar si se alcanzó el último evento y se debe calcular el histograma de duración.

5.2 Definición de evento sonoro y duración

Cada vez que una iteración satisface la condición del cuarto bloque de decisión, de la figura 5.2, se define un nuevo evento que es guardado momentáneamente en una tabla en memoria. Se utilizará un archivo que contiene ladridos y gruñidos de un perro para ejemplificar la obtención de la duración de cada evento.

En la figura 5.3 se muestra el nivel de presión sonora para el ejemplo. Con cuadros sombreados se muestran los eventos definidos por el algoritmo, descrito en la sección anterior, y con elipses sombreadas se muestran unas porciones de audio que hubiesen sido identificadas como eventos en caso de haber usado la definición con un umbral móvil dependiente solo del nivel máximo y no de un nivel estadístico.

En la primera fila de la tabla 5.1 se muestra el número de evento, es decir el valor de i_{e}, para el archivo del ejemplo. Cada columna de la tabla 5.1 es generada cada vez que se define un nuevo evento, es decir, cada vez que i_{e} se incrementa en el algoritmo de la figura 5.2.

Tabla 5.1: Eventos y duraciones de archivo en figura 5.3

evento	1	2	3	4	5	6	7
muestra inicial	235	433	615	1150	1918	2073	2243
muestra final	325	520	705	1253	2018	2160	2335
duración (ms)	180	175	180	205	200	175	185

En la segunda y tercera fila se muestran los valores de las muestras inicial j_1 y final j_2 de cada evento, respectivamente. En la cuarta fila se muestra la duración estimada para cada evento. La duración para cada evento será $(j_2 - j_1)/500$, dado que el nivel de presión sonora, estimado a partir de la señal de cada archivo, se submuestrea a una tasa de muestreo $F_{\mathsf{s},2} = 500$ Hz.

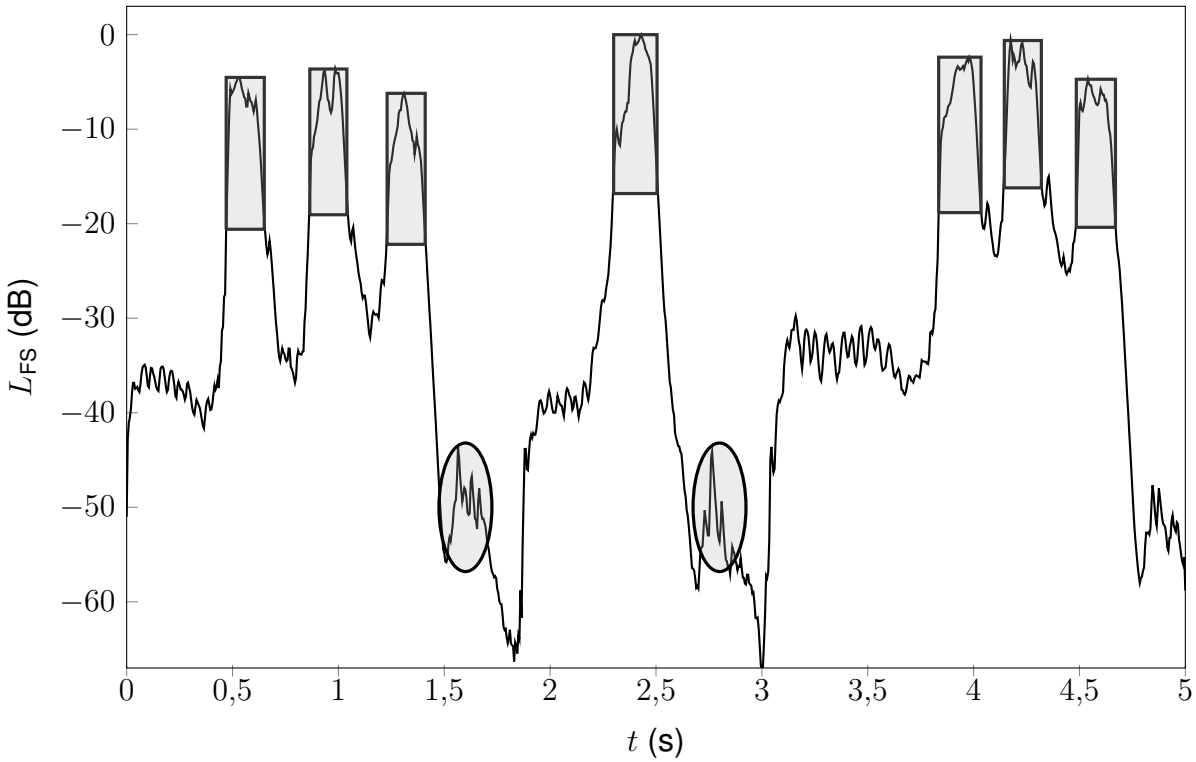


Figura 5.3: Criterio de duración con umbral móvil -10 dB relativo al nivel superado el 5 % del evento. Señal ladridos y gruñidos de perro

En el ejemplo de la figura 5.3 la duración promedio de los 7 eventos es de 186 ms siendo 205 ms la duración del evento más largo y 175 ms la duración del evento más corto (la duración de cada evento se muestra en la tabla 5.1).

5.3 Cálculo del histograma de duración

Una vez que se ha realizado la última iteración se calcula el histograma de duración de eventos sonoros para el archivo que está siendo analizado. Visto en el diagrama de la figura 5.2 después de alcanzado el último máximo local se inicia el bloque de cálculo del histograma. Se calculan dos histogramas, uno genérico y otro con intervalos predefinidos.

El histograma con intervalos genéricos ofrece al usuario la posibilidad de definir nuevos intervalos según alguna estrategia que sea adecuada para algún estudio en particular. Este histograma se obtiene en intervalos de ancho constante de aproximadamente 2 ms entre 0 s y 120 s. Es decir, se definen 65536 intervalos siendo $[0; 2]$ el primer intervalo en milisegundos y $[120; \infty]$ el último intervalo en segundos.

La figura 5.4 muestra un ejemplo del histograma de duración con intervalos genéri-

cos.

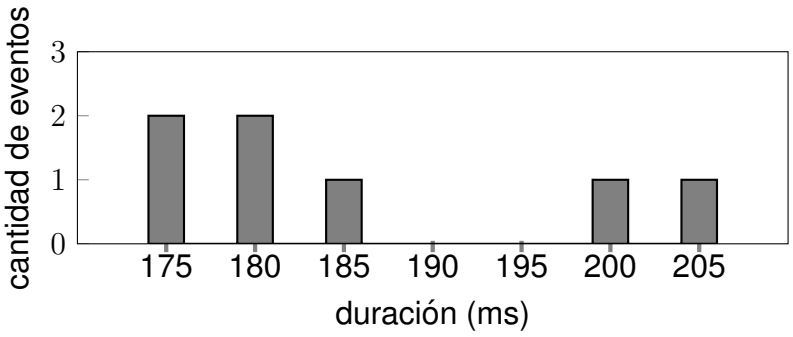


Figura 5.4: Intervalos genéricos. Histograma de duración archivo ladridos y gruñidos de perro

El histograma con intervalos predefinidos se calcula en los intervalos $]0{,}0; 1{,}0[$; $[1{,}0; 4{,}9[$; $[4{,}9; 12{,}4[$; $[12{,}4; 23{,}5[$; $[23{,}5; \infty]$ en segundos. Los intervalos de los extremos se han elegido para incluir posibles eventos demasiado largos o demasiado cortos. La figura 5.5 muestra un ejemplo del histograma de duración con estos intervalos predefinidos.

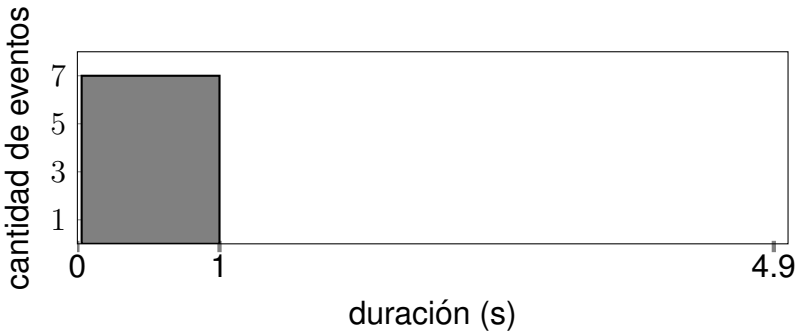


Figura 5.5: Intervalos predefinidos. Histograma de duración archivo ladridos y gruñidos de perro

Es de especial interés notar que al integrar el histograma de duración de un archivo, en cualquiera de sus dos versiones, se obtiene la cantidad total de eventos de ese archivo, dato que es utilizado en el capítulo 10.

5.4 Duración de eventos en el archivo de salida

Se ha observado que para, ciertos sonidos y contextos, la duración percibida de un eventos sonoro se asocia a una componente de la categoría semántica y no necesa-

riamente su duración en el sentido físico. Por ejemplo un sonido aislado normalmente tiene una duración mayor que la duración que tendrá inmerso en un contexto de otros sonidos que pueden enmascarar parcialmente al primer sonido. Sin embargo en estas situaciones el oyente puede reconstruir particularidades que fueron enmascaradas en el contexto como ser la duración del sonido aislado.

Sean N archivos de audio con histogramas de duración $\mathbf{d_1}, \cdots, \mathbf{d_n}, \cdots, \mathbf{d_N}$. Cada uno de los L elementos $d_{l,n}$ de los vectores $\mathbf{d_n}$ corresponde a la cantidad de eventos en el l-ésimo rango de duración. El histograma de duraciones del archivo de audio generado al combinar estos N archivos se define como la suma de estos histogramas, es decir, para cada rango de duraciones, la suma de los aportes a ese intervalo de cada archivo. Notar que si uno de los N archivos se repite $y \in \mathbb{N}_0$ veces la suma es equivalente a multiplicar cada elemento del histograma por y.

Para el l-ésimo rango de duración el valor del histograma del archivo de audio de salida viene dado por

$$d_{l,\mathbf{s}} = y_1 d_{l,1} + y_2 d_{l,2} + \cdots + y_N d_{l,N} \tag{5.2}$$

En el capítulo 8 se describe la metodología para controlar el histograma de duración del archivo de salida $d_{l,\mathbf{s}}$ mediante la optimización matemática de las variables y_n entre otras.

Capítulo 6

Importar sonidos a la base

Este capítulo describe el bloque *Importar* del módulo Base de Datos (ver 2.2). La función de este bloque es extraer y adecuar todos los metadatos que serán empleados en el módulo *Combinador* (ver 2.3).

Durante la importación se utilizan las etiquetas (ver 3.2) y los archivos de audio que han sido cargados a la biblioteca durante la *Compilación* (ver figura 2.2 y capítulo 3). Este capitulo se divide en dos secciones que tratan la lectura de las etiquetas (6.1) y la extracción de datos desde los archivos de audio (6.2), respectivamente.

La importación no implica ninguna edición de los archivos de audio, las ediciones correspondientes se hacen previamente de manera no automática en el bloque *Compilación* (ver 3.3). El bloque de importación puede ser ejecutado cada vez que se deban incorporar uno o más archivos a la base. No es necesario aplicar este bloque a toda la base cada vez que se amplia la base sino simplemente a los nuevos archivos que serán importados.

El archivo de metadatos es un archivo que contiene diferentes tipos de variables: vectores, matrices y arreglos de celdas. A cada archivo de audio en la *biblioteca* (ver 2.2) le corresponde un valor en una posición determinada de cada una de las variables. En los vectores cada elemento corresponde a un archivo, en las matrices cada columna corresponde a un archivo y lo mismo sucede con los arreglos de celdas. Es decir, cada metadato, que corresponde a un archivo de audio, se compone por un conjunto de datos con una posición común.

6.1 Lectura de etiquetas

Las etiquetas han sido previamente guardadas en una hoja de algún software de planilla de cálculo durante la compilación (ver tabla 3.1). La planilla puede ser la misma que se haya utilizado previamente para importar otro grupo de archivos o una nueva planilla. El algoritmo agrega una nueva etiqueta en una celda de la planilla de cálculo en la columna correspondiente a cada archivo de audio que ya ha sido importado. Esa etiqueta permite evitar que se genere más de un metadato para un mismo archivo de audio.

La primera tarea del bloque *importador* es leer la planilla de cálculo y guardar algunas de estas etiquetas en las variables que compondrán los metadatos. A su vez todas las etiquetas (ver tabla 3.1) quedan en memoria para ser utilizadas en la segunda secuencia de tareas que implican cálculos asociados a archivos de audio.

La tabla 6.1 muestra las variables que son guardadas en esta secuencia de tareas. La primera columna indica el nombre de la variable, la segunda el tipo y la tercera el tamaño de la misma. El valor de N_g es la cantidad de archivos de audio que se importan simultáneamente.

Tabla 6.1: Metadatos generados por lectura de etiquetas (ver detalles en el texto)

Nombre de Variable	tipo	tamaño
Nombre de archivo	arreglo de celdas	$1 \times N_\mathrm{g}$
Categoría	arreglo de celdas	$N_\mathrm{t} \times N_\mathrm{g}$
Clase	arreglo de celdas	$1 \times N_\mathrm{g}$
r (m)	vector	$1 \times N_\mathrm{g}$
Otros datos	arreglo de celdas	$N_\mathrm{o} \times N_\mathrm{g}$

El *Nombre de Archivo*, que puede incluir la ruta en caso de no ser la predefinida, es una cadena de caracteres ubicada en cada celda del arreglo correspondiente a cada metadato.

La *Categoría* (ver 3.2.2) contiene N_t cadenas de caracteres por cada archivo de audio. Cada metadato n_t se introduce en una fila diferente N_g dentro de la columna correspondiente al archivo n_g. Pueden quedar algunas celdas vacías de alguno de estos arreglos de celdas. Por ejemplo algún archivo de audio puede ser clasificado

hasta un nivel de detalle más profundo que otro (comparar rama de eventos *Autos* en 4^{to} nivel de la figura 3.1 con rama de eventos *Viento* en 3^{er} nivel de la misma figura).

La *Clase* (ver 3.2.1) es un arreglo de celdas cuyas celdas tienen un valor de las tres opciones *s*, *n* o *b*. La opción *s* indica que el archivo es de la clase *eventos simples*, la opción *n* corresponde a la clase *nube de eventos* y la opción *b* a *eventos largos*.

La distancia r es un vector en el cual cada elemento corresponde a una estimación de la distancia entre la fuente sonora (o fuentes sonoras) y el micrófono utilizado durante la grabación del archivo correspondiente (ver pg. 31 en 3.2.3).

Finalmente en *Otros datos* se guardan otras etiquetas de interés que hayan sido recopiladas durante la compilación (ver 3.2.5). Esta variable tiene una estructura similar a la de la variable *Categoría* pero contiene N_{o} celdas por cada archivo. Estas celdas también corresponden a cadenas de caracteres y admite celdas vacías.

6.2 Análisis de los archivos de audio

La segunda secuencia de tareas del bloque *importador* es extraer y calcular el tipo de metadatos que se desprenden de la propia señal de audio. Estas tareas se realizan mediante un bucle que accede a cada archivo utilizando la ruta de cada archivo que dejó en memoria la secuencia de tareas anterior.

La tabla 6.2 muestra las variables que son guardadas en esta secuencia de tareas. Las columnas siguen la misma estructura de la tabla 6.1, es decir, cada columna corresponde a un archivo.

Las primeras dos variables simplemente agrupan datos extraídos directamente del correspondiente archivo de audio. La variable tasa de muestreo contiene en cada elemento el valor de la tasa de muestreo de cada archivo. (Los archivos utilizados en esta tesis están muestreados a una tasa de $44\ 100$ Hz, es decir que esta variable es en todos los casos un vector con todos los elementos iguales, pero contar con este dato permitirá futuras actualizaciones con remuestreo de algunos archivos). La variable *cantidad de muestras* contiene en cada elemento la longitud de cada archivo en muestras. Con estos dos datos es sencillo calcular la duración total, en unidades de tiempo, del i-ésimo archivo

$$d_{\mathsf{a},i} = N_{\mathsf{m},i}/F_{\mathsf{S},i} \tag{6.1}$$

siendo $N_{\mathsf{m},i}$ la cantidad de muestras y $F_{\mathsf{S},i}$ la tasa de muestreo.

Tabla 6.2: Metadatos generados calculando parámetros o extrayéndolos directamente de los archivos de audio (ver detalles en el texto)

Nombre de Variable	tipo	tamaño
Tasa de muestreo (Hz)	vector	$1 \times N_{\text{g}}$
Cantidad de muestras	vector	$1 \times N_{\text{g}}$
$L_{\text{p,FS}}$ (dB)	vector	$1 \times N_{\text{g}}$
p^2_{rms} líneas FFT	matriz	$4\,096 \times N_{\text{g}}$
p^2_{rms} líneas $FFTD$	matriz	$4\,096 \times N_{\text{g}}$
p^2_{rms} bandas de 1/3	matriz	$31 \times N_{\text{g}}$
p^2_{rms} bandas de 1/1	matriz	$10 \times N_{\text{g}}$
Histograma de duración genérico	matriz	$65\,536 \times N_{\text{g}}$
Histograma de duración predefinido	matriz	$5 \times N_{\text{g}}$
Inicio y duración de eventos	arreglo de celdas	$1 \times N_{\text{g}}$
Muestra de Inserción	vector	$1 \times N_{\text{g}}$
$p^2_{\text{10ms}}(t_x)$ bandas de 1/3	matriz	$31 \times N_{\text{g}}$
$p^2_{\text{10ms}}(t_x)$ bandas de 1/1	matriz	$10 \times N_{\text{g}}$

La variable $L_{\text{p,FS}}$, contiene en cada columna el nivel de presión sonora que generaría una señal sinusoidal a fondo de escala para cada archivo. Cada elemento se calcula según la ecuación (3.1)

$$L_{\text{p,FS}} = L_{\text{p,cal}} - L_{\text{wav,FS}}$$

$L_{\text{p,cal}}$ fue cargado en memoria en las tareas anteriores y viene de la etapa de compilación. El valor de $L_{\text{wav,FS}}$ se calcula como el valor eficaz de un archivo de audio de calibración (distinto al archivo que se importará a la base) si el valor en memoria correspondiente es una cadena de caracteres correspondiente al nombre del archivo o directamente leído de memoria si el valor correspondiente es numérico.

Las cuatro variables p^2_{rms} se calculan según se detalla en el capítulo 4 previa multiplicación de los valores de la señal digital por la constante de calibración $10^{L_{\text{p,FS}}/20}$ (ver 8.1). Los datos quedan registrados en cuatro matrices cuyas columnas corresponden a cada archivo y cada una de cuyas filas corresponde a una línea o banda de frecuencia según sea el caso.

Las dos variables *histograma de duración* se calculan según se detalla en el capítulo 5, particularmente en 5.3. Estos datos conforman una matriz donde cada columna corresponde a un archivo y cada fila corresponde a un intervalo de duración. En esta tesis se utiliza el histograma de duración predefinido pero el histograma genérico permite definir nuevos intervalos en estudios futuros.

La variable *inicio y duración de eventos* es un arreglo de celdas en el cual cada celda corresponde a un archivo. Cada celda del arreglo es una matriz de dos filas y tantas columnas como eventos contenga el archivo de audio correspondiente. La primera fila indica el instante de inicio de cada evento (los valores de tiempo correspondiente a j_1 en 5.1.1). La segunda fila indica la duración de cada uno de esos eventos (por ejemplo ver la segunda fila de la tabla 5.1).

La variable *muestra de inserción* se determina según el valor de la etiqueta instante de inserción asignada durante la compilación y retenido en memoria en la primera secuencia de tareas de la importación. Esta variable es un vector cuyos elementos indican la muestra del archivo correspondiente que será utilizada durante la composición para ubicar temporalmente al archivo (ver 3.2.4).

Las dos variables $p^2_{\mathsf{10ms}}(t_x)$ se calculan según se detalla en 4.3 previa multiplicación de los valores de la señal digital por la constante de calibración $10^{L_{\mathsf{p,FS}}/20}$ (ver 8.1). Los datos quedan registrados en dos matrices cuyas columnas corresponden a cada archivo y cuyas filas corresponden a una banda de frecuencia, en el primer caso bandas de 1/1 octava y en el segundo de 1/3 de octava.

Capítulo 7

Auralización

En 2.4 se introdujo de manera general el sistema de auralización implementado en esta tesis. En esta sección se aborda específicamente el problema de la auralización, particularizando en el sistema empleado.

Existen dos grandes enfoques en cuanto a sistemas de auralización. Una de ellas es a través de auriculares y la otra a través de altavoces. Los sistemas con auriculares necesitan de estrategias para tener en cuenta la función de transferencia de la cabeza y el torso además de la función de transferencia del camino desde la fuente al receptor influido por la sala. Los sistemas con altavoces pueden requerir esas mismas estrategias pero en este caso pueden ser resueltas sin apelar a modelos digitales sino ubicándolos a una distancia adecuada del sujeto de modo que las funciones de transferencia reales hagan el trabajo por si solas.

Un esquema más general que el de la figura 2.4, que tiene en cuenta tanto auriculares como parlantes, se muestra en la figura 7.1.

En la figura 7.1 los módulos *parlante*, *ventana* y *sala* de la figura 2.4 han sido reemplazados por el módulo *sección real* cuya respuesta al impulso se designa h_r. También se ha agregado con línea de puntos la posibilidad de simular una porción del sistema en el módulo *sección simulada* con respuesta al impulso h_s. Si todos los bloques son lineales e invariantes en el tiempo (LTI) es posible cambiar la posición de alguno de los bloques en la sección real ubicándolo en la posición de las secciones simuladas. Esto permite complementar la porción real del sistema mediante un complemento simulado que permita obtener una respuesta global diferente. En estos casos será necesario modelar dicha porción del sistema y evitar que la misma esté presente en la realidad.

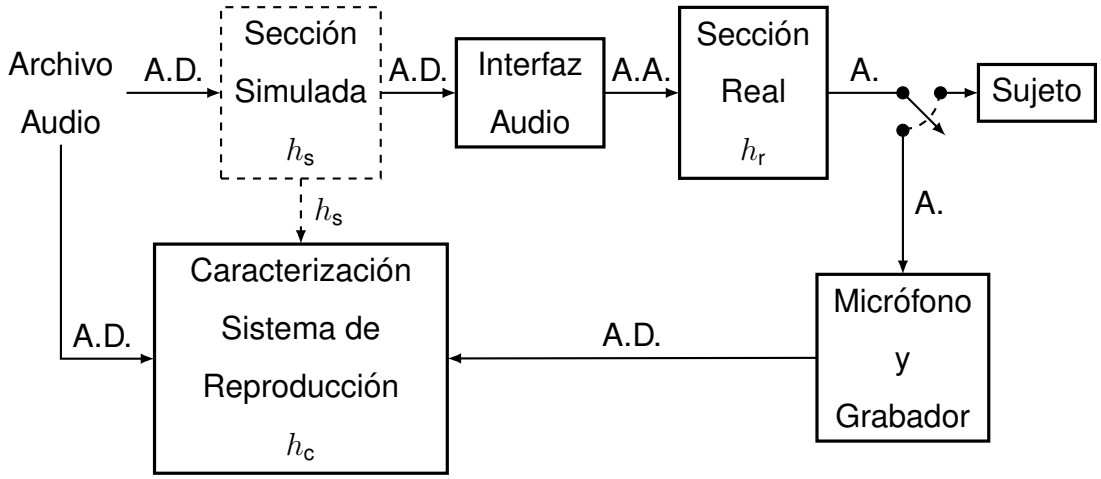


Figura 7.1: Diagrama General de Auralización (A.D.: señal de audio digital, A.A.: señal de audio analógico y A.: señal acústica)

Es decir, si se simula una ventana se debería evitar que el sonido que emite el parlante se propague a través de una ventana en la realidad, antes de llegar al sujeto, a menos que se desee recrear una situación donde el sonido se propaga a través de dos ventanas o una ventana diferente.

En las siguientes secciones se da una breve reseña sobre consideraciones a tener en cuenta en el diseño del sistema de auralización. En 7.1 se describen las consideraciones de sistemas con parlantes y en 7.2 las de sistemas con auriculares. El sistema particular utilizado en esta tesis para la prueba piloto corresponde a sistemas con parlantes y sus características se detallan en 10.1.1. En 7.3 se describe cómo caracterizar los sistemas detallando particularmente la metodología empleada en esta tesis. Finalmente en 7.4 se introducen brevemente los sistemas de realidad virtual inmersiva.

7.1 Auralización con parlantes

En principio los sistemas con parlantes ofrecen razonables ventajas respecto a los sistemas con auriculares para simular escenarios virtuales o mixtos para experimentación de respuestas del ser humano a estímulos sonoros. Estas ventajas tienen que ver con que no es necesario simular la función de transferencia de la cabeza (ver 7.2) y que además el sujeto no estará influenciado por la sensación, y en algunos casos incomodidad, que pueda causar tener puestos los auriculares.

Como punto de partida, la forma más simple de recrear algún escenario acústico, sería utilizar un parlante por cada fuente sonora virtual. Otros sistemas un poco más sofisticados pero basados en principios simples como los del sonido estéreo o multicanal han sido utilizados, por ejemplo para estudios de acústica de salas [Ando, 1998]. Si bien este tipo de sistemas en principio parece simple, con una prueba configurada adecuadamente puede ser suficiente. En la prueba piloto de esta tesis (ver capítulo 10) se implementa un sistema monofónico muy simple. Se trata de un parlante ubicado detrás de una ventana. Es decir, el sonido que proviene del parlante no llega directamente al sujeto sino a través de la ventana como es habitual en situaciones reales dado que las ventanas suelen ser los elementos de menor atenuación en la mayoría de las construcciones. A su vez, si no estuviese la ventana, el sonido sería radiado desde el centro aparente del parlante. Contrariamente, al estar presente la ventana, el sonido que esta recibe desde el parlante es nuevamente radiado por la ventana pero además por los bordes y toda su superficie dando una sensación más envolvente y realista comparando con un sistema monofónico sin interponer ningún objeto en el camino de propagación (camino parlante-oyente).

Existen también sistemas de audio multicanal que comprenden la grabación y la reproducción multicanal [Okamoto et al., 2014; Alvarsson et al., 2014]. Estos sistemas ofrecen una auralización realista desde el punto de vista perceptivo y existen diversas metodologías para controlar la reproducción dependiendo de cómo fueron grabados los sonidos.

Posiblemente los sistemas más completos son los sistemas de síntesis de campo sonoro (en inglés: wave field synthesis) [Berkhout et al., 1993; Ahrens et al., 2008; Oldfield, 2013; Wierstorf et al., 2013]. Sin embargo estos sistemas presentan un elevado costo en cuanto requieren una gran cantidad de canales de audio y altavoces además del procesamiento de audio. Estos sistemas modelan el campo sonoro mediante la contribución controlada de un gran número de altavoces. El concepto general detrás de este modelo es el principio de Huyghens que establece que un frente de onda puede ser reconstruido mediante una serie de fuentes puntuales. Este tipo de sistemas de auralización permite lograr, con gran realismo, un fuerte control de la ubicación de la fuente sonora. Incluso se pueden simular fuentes sonoras dentro de la sala real [Wierstorf et al., 2013], lo cual se suele referir como holografía acústica.

7.2 Auralización con auriculares

En el caso de sistemas con auriculares es necesario modelar la función de transferencia de la cabeza y torso (HRTF). Esto puede ser una ventaja en cuanto al control de la dirección de inmisión desde cada fuente e incluso desde cada reflexión. Por otra parte puede significar una desventaja en cuanto a las posibilidades de implementación y medición o modelado de las HRTF.

Existen, en la actualidad, bases de datos abiertas de funciones de transferencia de la cabeza y el torso. Algunas de estas bases son las de CIPIC [Algazi et al., 2001b] y la del Proyecto LISTEN del IRCAM [IRCAM, 2003]. Estas bases contienen mediciones de la HRTF de varios participantes asociadas a distintos tipos de datos, por ejemplo las de CIPIC se asocian las HRTF a determinados datos antropométricos de los sujetos.

Uno de los grandes desafíos, que es un tópico abierto de investigación, es la personalización de la HRTF. Sucede que cada persona tiene una HRTF particular y al usar la HRTF medida en otra persona no se puede asegurar que obtendrá una experiencia realista. Por ejemplo puede haber confusiones entre las direcciones de inmisión. Algunos de los trabajos más recientes en personalizar la HRTF [Iida et al., 2014; Katz y Parseihian, 2012] y algunos anteriores utilizan estrategias basadas en los datos antropométricos o en respuestas de percepción [Algazi et al., 2001a, 2002].

Además existen modelos comerciales y desarrollos académicos con sistemas de seguimiento de la posición y orientación de la cabeza que permiten aplicar en tiempo real la HRTF de tal manera que el sujeto perciba que las fuentes sonoras no siguen los movimientos de su cabeza. Un ejemplo de sistemas comerciales son los de la marca Sony modelo VIP-1000 y serie MDR o de la marca Beyerdynamic modelos Headzone, entre otros. Dos ejemplos de sistemas de desarrollo académico en nuestro país se pueden consultar en la bibliografía [Riva, 2008; Tommasini, 2012].

7.3 Caracterización del sistema de auralización

Es habitual caracterizar los sistemas LTI mediante su su respuesta al impulso o su función de transferencia. Ambas funciones se relacionan entre sí mediante la transformada de Fourier y su inversa. Existen diversos métodos para identificar sistemas en un sentido general. Particularmente en sistemas acústicos y de audio, incluyen-

do respuestas impulsivas de salas, es habitual el uso de técnicas de deconvolución o simplemente convolución de la señal medida con una función especial detallada a continuación.

Sea h el sistema que se desea identificar y x la señal de excitación, entonces es de esperar que la señal tomada por el micrófono, en ausencia de otras señales parásitas, sea

$$y = x * h \tag{7.1}$$

La función especial usada para convolucionar, apoyándose en la propiedad conmutativa de la convolución, es un filtro inverso respecto al sistema al cuál podría asociarse la señal x al conmutar la convolución. Es decir, se puede imaginar que la señal es el sistema y viceversa. Al convolucionar el filtro inverso de x con la señal y se obtiene una aproximación de h [Farina, 2000]. Esta última técnica es la utilizada en esta tesis.

La respuesta al impulso global h_t se puede caracterizar por los subsistemas simulado h_s y real h_r o bien de manera completa. El subsistema simulado generalmente se caracteriza mediante modelado de algún fenómeno acústico con determinadas simplificaciones. Por ejemplo mediante modelos de rayos acústicos que simulan la respuesta al impulso de una sala o modelos ondulatorios que simulan un divisorio mediante técnicas de filtrado digital. El subsistema real generalmente debe ser medida aunque también podría ser caracterizada por un modelo a partir de otras mediciones. Por ejemplo a partir de mediciones de los parámetros Thielle-Small [Thiele, 1971a,b; Small, 1971] es posible caracterizar la respuesta al impulso de un parlante. Sin embargo lo más recomendable es medir para evitar diferencias entre el modelo y la realidad.

Las funciones de transferencia del subsistema real H_r, la del subsistema simulado H_s y la respuesta del sistema total H_t se pueden obtener aplicando la transformada rápida de Fourier a las respuestas al impulso h_r, h_s y h_t, respectivamente. O bien, en caso de contar por separado con H_r y H_s, se puede obtener la función resultante mediante

$$H_\mathrm{t} = H_\mathrm{r} H_\mathrm{s} \tag{7.2}$$

Una opción para tener en cuenta estas respuestas en el problema general de combinatoria del capítulo 8 sería modificar los metadatos espectrales de la base de sonidos (ver figura 2.2) importando nuevamente estos datos con los mismos archivos de

la base original pero en este caso modificados por H_t. Esta opción se descartó en esta tesis por el tiempo de cómputo que esto requiere. Se prefirió un enfoque más simple que, admitiendo diferencias mayores de los resultados alcanzados respecto a los requeridos por el usuario, permita un trabajo más fluido durante el desarrollo de la herramienta en esta tesis.

El enfoque consiste en caracterizar por bandas de frecuencia la respuesta H_t. La forma empleada es a través de la suma de líneas espectrales detallada en el capítulo 4. De esa manera se obtiene un vector $\mathbf{h}_\mathrm{o}$ de M elementos $h_{\mathrm{o},m}$, siendo M el total de bandas que se utilizarán. Se define entonces $\mathbf{h}_\mathrm{o}$ como la respuesta total, por bandas de frecuencia, del sistema de auralización completo que tiene en cuenta tanto respuestas simuladas como reales.

Estos datos quedan guardados en el bloque *configuración* del módulo *combinador* evitando modificar los metadatos de manera permanente. Luego, durante la *formulación del problema*, del módulo *combinador*, el primer paso consiste en aplicar esta corrección a los metadatos (ver 2.3).

En esta tesis no se implementó ningún sistema simulado. En caso de implementar un sistema de este tipo, se deberá aplicar el sistema simulado a los archivos de audio a combinar según una estrategia adecuada. Esta modificación se puede implementar inmediatamente antes o después del bloque *combinador de audio* del módulo *combinador* (ver 2.3).

Si el sistema simulado tiene en cuenta múltiples direcciones de inmisión para las fuentes sonoras, se abre la posibilidad de contar con un camino de propagación sonora distinto para cada fuente. En este caso es posible multiplicar la cantidad de metadatos, correspondientes a cada archivo de la base, por la cantidad de caminos diferentes que se pueden simular para cada uno de ellos. Se debe considerar cuál es el grupo de metadatos para cada archivo al momento de definir la cantidad máxima de repeticiones de un mismo archivo durante la *formulación del problema*, a diferencia de la estrategia usada en esta tesis que sólo admite un metadato por archivo.

7.4 Sistemas de realidad virtual inmersiva

En general los efectos del ruido en el ser humano no dependen solamente del estímulo sonoro sino también del contexto. Dentro del contexto se pueden identificar factores

como la atención a algún estímulo en particular (ya sea auditivo o de otra índole), el tipo de actividad que realiza el sujeto, la responsabilidad del mismo en esa actividad, los estímulos recibidos mediante otros sentidos como el visual, entre otros. Este tipo de relaciones son tenidas en cuenta en estudios sobre los efectos del ruido y la percepción sonora [Ruotolo et al., 2012; Maffei et al., 2014].

Los sistemas de realidad virtual inmersiva ofrecen la posibilidad de completar el contexto llegando a una situación de mayor validez ecológica cuando no se puede implementar una configuración real del problema. Es decir, sin duda tendría mayor validez ecológica la realidad por si sola pero, ante la imposibilidad de controlar algún factor de la realidad, estos sistemas virtuales ofrecen una interesante alternativa.

Este tipo de sistemas tienen sensores de las acciones humanas, por ejemplo sistemas de seguimiento de la posición del sujeto, de la orientación de la cabeza, de la posición de alguna extremidad, de la fuerza que realiza una mano, etc. En función de los datos censados el sistema se adapta para dar la sensación de que el sujeto se encuentra en un ambiente determinado. Normalmente el punto de partida es un modelo espacial tridimensional que puede ser *renderizado*[1] en tiempo real y presentado por visores tipo gafas o grandes pantallas que revisten una habitación. Simultáneamente el audio es renderizado en tiempo real y reproducido por parlantes o auriculares. Otros estímulos o acciones también pueden ser variadas en tiempo real como por ejemplo la fuerza que ejerce un guante sobre la yema de los dedos o el desplazamiento de una cinta sobre la cual está parado el sujeto, por supuesto en ambos casos el objetivo es que estos estímulos sean congruentes con la acción de haber tomado un objeto virtual o haber avanzado en alguna dirección del escenario virtual.

La medición de la calidad de la experiencia en estos sistemas, desde el punto de vista del sujeto, es un tema abierto de investigación [Larsson et al., 2003]. Se debate sobre cómo se relacionan y cuál es la validez de la sensación de presencia, envolvimiento, inmersión y realismo, entre otros factores [Witmer y Singer, 1998].

[1] Del inglés render. Término habitual en la jerga informática para referirse al proceso de generar una imagen o un video con audio partiendo de modelos computacionales en tres dimensiones. En este caso solo se hace alusión al renderizado de audio.

Capítulo 8

Composición Controlada

Este capítulo describe la parte principal de este trabajo que corresponde al módulo *Combinador* (ver figura 2.1). En la sección 2.3 se mostró una descripción introductoria del módulo.

Las primeras secciones (sección 8.1 a sección 8.3) de este capítulo describen cómo se interpreta la información de los archivos de audio de entrada y salida, así como sus metadatos, en términos de presión sonora y sus descriptores acústicos relacionados.

Las secciones 8.4 a 8.6 describen alternativas de implementación del bloque *formulación del problema* del modulo *Combinador* (ver figura 2.3). En este trabajo se implementa sólo la última alternativa, descrita en la sección 8.6. Las primeras alternativas se incluyen, no sólo como alternativas en si, sino también, a efectos de describir de manera paulatina los aspectos considerados en la alternativa implementada.

La sección 8.8 describe el bloque *Resolución Problema Combinatoria* y la sección 8.9 describe los bloques *Cálculo instantes inserción de cada archivo* y *Combinador archivos de audio* (ver figura 2.3).

8.1 Calibración de datos de entrada y salida

8.1.1 Audio

Los archivos de audio, como cualquier señal digital, en un instante cualquiera pueden tomar una cantidad finita de valores de amplitud determinada por la cantidad de bits empleados para su codificación. Es habitual, en procesamiento de señales, interpre-

tar la señal digital $s(t)$ con una escala de base decimal en la cual el máximo valor posible corresponde aproximadamente a la unidad y el mínimo a la unidad negativa, es decir, los posibles valores de amplitud quedan acotados en el intervalo $[-1, +1]$ (el valor $+1$ es aproximado de $1 - 2^{-N_{\text{bit}}}$, siendo N_{bit} el número de bits empleados para la codificación de la señal).

La conversión entre la señal de presión sonora $p(t)$ y la señal digital $s(t_k)$ en un sistema de grabación se introdujo en 3.2.3. En la sección 6.2 se introdujo brevemente la calibración de los datos espectrales de presión sonora cuadrática. Cada vez que se carga un archivo de audio, en cualquiera de los módulos de esta herramienta, la señal $s(t)$ se expresa en términos de presión sonora $p(t)$ multiplicando por el valor correspondiente a la amplitud del fondo de escala que se obtiene del nivel de fondo de escala $L_{\text{p,FS}}$, según

$$p(t) = 10^{L_{\text{p,FS}}/20} s(t) \tag{8.1}$$

En el caso de la mezcla de audio, (sección 8.9), se debe guardar la información en el archivo de audio de salida. Luego de haber mezclado, la señal resultante queda expresada en términos de presión sonora y debe ser escalada para poder ser registrada en un archivo de audio. El factor de escala en este caso es el inverso de $\max(|p_{\text{s}}(t)|)$ y la señal escalada que se registra en el archivo de audio se calcula según

$$s_{\text{s}}(t) = \frac{1}{\max(|p_{\text{s}}|)} p_{\text{s}}(t) \tag{8.2}$$

El nivel máximo que puede alcanzar una señal en esta escala es 0 dBFS o una amplitud máxima de ± 1. Para que ese nivel máximo de la señal digital corresponda a un nivel de presión sonora determinado se deben tener en cuenta diversos factores. Estos factores como mínimo involucran las siguientes etapas: conversor digital analógico, preamplificador, amplificador de potencia, altavoz, sala y camino altavoz-oyente. Además, cada una de esas etapas, especialmente la sala, tienen una respuesta en frecuencia que puede no ser plana.

La estrategia empleada en esta tesis para calibrar la señal de salida es generar un tono puro de 1 kHz a nivel de fondo de escala (es decir, cuya amplitud máxima es la unidad), y ajustar la ganancia del amplificador de potencia hasta medir, en la posición del receptor, un nivel sonoro $L_{\text{p}_{\text{cal,s}}}$. Por otra parte se mide la respuesta en frecuencia del camino completo entre la señal digital y el receptor (ver sección 7.3), que incluye

todas las etapas intermedias nombradas anteriormente, y se aplica esta respuesta a los metadatos de espectro de exposición antes de proceder a la optimización.

8.1.2 Metadatos

En el caso de los metadatos espectrales se debe aplicar, como paso inicial de la *formulación del problema* del módulo *combinador* (ver figura 2.3), la respuesta del sistema de auralización. Todos los valores de presión cuadrática de la tabla 6.2 que vayan a ser utilizados para formular el problema se afectan por $h_{\mathrm{o},m}$ (ver sección 7.3) según

$$p_m^2 = h_{\mathrm{o},m}^2 \times p_{m,\mathrm{importado}}^2 \tag{8.3}$$

8.2 El problema de optimización

Sea $W = \{w_1, w_2, \cdots, w_N\}$ el conjunto de los N archivos de audio que conforman la base de datos. El problema de la combinación controlada es encontrar un subconjunto de archivos de audio $I \subseteq \{1, \cdots, N\}$ y, para cada $i \in I$, una ganancia $g_i \in \mathbb{R}^+$ y una cantidad de repeticiones $y_i \in \mathbb{N}$, tal que, al mezclar las señales de los archivos en un nuevo archivo de audio y reproducirlo, genere un sonido ambiente realista cuyas características fueron previamente especificadas por el usuario. El usuario especifica el espectro de exposición $\mathbf{e}_{\mathrm{u}}$ y el histograma de duración de eventos $\mathbf{d}_{\mathrm{u}}$. Por la mezcla se entiende que se suman las señales de los archivos aplicando previamente la ganancia g_i y repitiendo y_i veces la señal de cada archivo.

Se deben agregar ciertas restricciones para evitar que el sonido ambiente generado resulte poco realista. Una de estas restricciones busca evitar demasiadas repeticiones de un mismo archivo de audio w_i que puede ser percibido como una situación forzada y poco realista, esto da una condición de borde superior para y_i. Una segunda restricción busca evitar que algún archivo sea demasiado amplificado pudiendo causar la sensación de que la fuente sonora está demasiado cercana en comparación con la distancia a la cual se encuentra habitualmente, es decir una restricción de borde superior para g_i. Finalmente, para simplificar el problema, se permite que y_i y g_i tomen simultáneamente un valor nulo y en cuyo caso (es decir $y_i = 0$ y $g_i = 0$) el archivo w_i es excluido del subconjunto I.

Este problema es de tipo indeterminado en tanto se espera que la cantidad de

archivos de audio sea mayor que la cantidad de ecuaciones que se pueden escribir. Esto significa que el problema es encontrar valores g_i e y_i a partir de unas pocas ecuaciones dadas por la cantidad de bandas de frecuencia M que involucran a todas las variables g_i y por otras pocas ecuaciones dadas por la cantidad de rangos del histograma de duración de eventos sonoros L siendo $N >> M + L$.

Una forma de tratar de transformar el problema en uno del tipo sobredeterminado sería ampliar la cantidad de bandas e intervalos pero para que exista una solución, al menos cercana al óptimo, será necesario también ampliar la cantidad de archivos de sonido en la base para poder asegurar que el espacio vectorial generado por la base que conforman los metadatos de espectro e histograma de duración contenga puntos cercanos a las posibles especificaciones del usuario. Un problema, en el caso de utilizar bandas de menor ancho, es que disminuirá la medida en que se cumple el requisito de no correlación entre las señales de dos o más archivos de sonido por bandas. Este requisito se introduce más adelante para obtener la ecuación (8.9). El caso extremo sería un ancho de banda en el cual solo se da un tono puro y sucede que dos o más tonos puros de una misma frecuencia tienen cierta correlación dependiente de la fase entre estos. Se puede afirmar que la cantidad de archivos de audio necesarios crece con la cantidad de bandas de frecuencia e intervalos del histograma de duración y que además, si se desea ampliar el rango de soluciones posibles, también se deberá ampliar la cantidad de archivos. Por estas razones el problema no puede ser de tipo sobredeterminado sino que es de tipo indeterminado.

Los problemas indeterminados no tienen una solución exacta para todos los casos sino que pueden tener una solución exacta, varias soluciones exactas o ninguna solución exacta. Para los casos que no hay una solución exacta es posible aproximar la solución más cercana. Definiendo una función de distancia (que se conoce como función de costo o de utilidad) y manejando el problema para darle forma de un problema de tipo convexo es posible encontrar la solución más cercana (en términos de la distancia previamente definida) utilizando técnicas conocidas de optimización. La función de costo puede entenderse como una medida de la falta de exactitud o error entre los valores de los parámetros requeridos por el usuario y los valores posibles para estos parámetros en el sonido ambiental de salida. Un nombre más general para esta función de costo o distancia es *función objetivo* , adoptado de la terminología de

tópicos de optimización matemática. También se utilizara el término *variable objetivo* para referirse a las variables que se busca optimizar a modo de minimizar la función de costo.

Existen varios métodos que pueden ser aplicados para obtener una solución al problema general de combinación controlada de archivos de audio. La base de todos estos métodos es la optimización pero difieren en el tipo de optimización y uso de técnicas complementarias para el tratamiento de los datos de audio que permiten formular el problema. En las secciones 8.4 a 8.6 se analizan algunas de las posibles técnicas analizando sus ventajas y desventajas. Además, el orden en que se presentan estas cuatro secciones permite abordar paulatinamente la técnica finalmente empleada (a través de la formulación de un problema de combinación lineal con solución mixta en los enteros y los reales).

8.3 Exposición sonora

Evidentemente mientras mayor sea la ganancia de uno de los archivos de la base de datos más influirá su espectro en el espectro del archivo de salida. Para poder identificar esta relación de manera exacta es útil introducir el descriptor de ruido denominado exposición sonora. La exposición sonora es la integral temporal del cuadrado de la presión sonora instantánea sobre un intervalo de tiempo determinado, que en este caso, equivale a la duración total de un archivo de audio d_{a} (ver Apéndice A) [IRAM 4113-1, 2009; ISO 1996-1, 2003], según

$$E(p) = \int_0^{d_{\mathsf{a}}} p(t)^2 \mathrm{d}t \tag{8.4}$$

Es decir, si se tiene un archivo que genera una presión sonora $g \cdot p(t)$, cuando es afectado por una ganancia g, su exposición será

$$E(g \cdot p) = E(p) g^2 \tag{8.5}$$

Por simplicidad denotaremos $x_i = g_i^2$ en las secciones siguientes.

De la la definición de la ecuación (8.4) y la definición de presión sonora eficaz, dada por la ecuación (4.2), es sencillo verificar que

$$E = d_{\mathsf{a}} p_{rms}^2, \tag{8.6}$$

es decir, la exposición sonora, durante la duración d_{a} de un archivo de audio, es igual a la duración del audio por el valor cuadrático de presión sonora eficaz extendida al intervalo.

8.3.1 Linealidad de la suma de exposición sonora

Una gran ventaja de utilizar la exposición sonora es la posibilidad de describir el problema como una serie de ecuaciones lineales. Esto permite utilizar algoritmos ampliamente desarrollados de optimización, pero se deben respetar ciertos requisitos en las señales de audio para que esta suma sea lineal.

La exposición es proporcional a la energía sonora según se observa en 8.4.

Es posible demostrar que el valor eficaz total que se obtiene al sumar dos señales estacionarias de presión sonora $p_1(t)$ y $p_2(t)$ viene dado por

$$p^2_{\text{rms,total}} = p^2_{\text{rms},1} + p^2_{\text{rms},2} + 2 \times p_1(t) \star p_2(t), \tag{8.7}$$

donde $p_1(t) \star p_2(t)$ es la correlación cruzada entre las señales p_1 y p_2, siendo $p_{\text{rms},1}$ el valor eficaz de p_1, $p_{\text{rms},2}$ el valor eficaz de p_2 y $p_{\text{rms,total}}$ el valor eficaz de la señal resultante $p_{\text{total}}(t) = p_1(t) + p_2(t)$.

La correlación cruzada entre dos señales, $p_1(t) \star p_2(t)$, tiene un valor nulo si las dos señales $p_1(t)$ y $p_2(t)$ son incoherentes entre si. Es decir, si las señales no tienen la misma frecuencia o son de un ancho de banda amplio sin ser similares entre si, el tercer término de la ecuación (8.8) se hace nulo y es posible obtener el valor cuadrático eficaz total como la suma de los valores cuadráticos de ambas señales según

$$p^2_{\text{rms,total}} = p^2_{\text{rms},1} + p^2_{\text{rms},2} \tag{8.8}$$

A continuación, salvo que se indique lo contrario, se supondrá que las señales de los archivos de audio de la *Base de Datos* no son coherentes entre si. Dado que los archivos de la base no son estacionarios, los valores eficaces de las señales (por ejemplo de $p_1(t)$ y $p_2(t)$ en ecuación (8.8)) deben ser calculados utilizando el mismo periodo T utilizado para la señal resultante, que será el archivo de salida. Sin embargo, durante la importación de los sonidos de la base, el valor de la duración del archivo de salida es desconocido. Por esa razón, durante la importación, el cuadrado del valor eficaz de cada archivo se calcula, según su definición, integrando su valor cuadrático a lo largo de todo el archivo y dividiendo por su duración d_{a}. Generalizando la ecuación

(8.8), y teniendo en cuenta la consideración de duración, es posible demostrar que la suma de N archivos de audio incoherentes resultará en un valor eficaz según

$$p^2_{\text{rms,total}} \times T = p^2_{\text{rms},1} \times d_{\text{a},1} + p^2_{\text{rms},2} \times d_{\text{a},2} + \cdots + p^2_{\text{rms},N} \times d_{\text{a},N} \tag{8.9}$$

Reemplazando la ecuación (8.6) en la ecuación (8.9) se obtiene la exposición total

$$E_{\text{total}} = E_1 + E_2 + \cdots + E_N \tag{8.10}$$

donde E_{total} es la exposición que causa un archivo de audio que combina los archivos de audio $w_1, w_2, \cdots, w_N$ cuyas exposiciones sonoras corresponden a $E_1, E_2, \cdots, E_N$, cada cual calculada con un periodo de integración igual a la duración de cada archivo $d_{\text{a},1}, d_{\text{a},2}, \cdots, d_{\text{a},N}$ respectivamente.

Si los archivos de audio fuesen previamente afectados por una ganancia $g_i = \sqrt{x_i}$, la combinación de exposiciones se obtiene reemplazando la ecuación (8.5) en la ecuación (8.10) según

$$E_{\text{total},\mathbf{g}} = x_1 E_1 + x_2 E_2 + \cdots + x_N E_N \tag{8.11}$$

Luego por cada repetición y_i de cada archivo w_i, la ganancia que realmente se aplique al archivo de audio deberá multiplicarse por el recíproco de y_i. Es decir, la ganancia que realmente se aplicará a cada archivo es

$$g'_i = g_i / y_i \tag{8.12}$$

Otra posibilidad sería aplicar ganancias aleatorias dentro de ciertos márgenes que sumen g_i. De ese modo se evitaría que sean idénticas.

Notar que las ecuaciones 8.4 a 8.11 se definieron sin identificar un ancho de banda particular. Se utilizará $\mathbf{e_i} = (e_{1,i}, e_{2,i}, \cdots, e_{m,i}, \cdots, e_{M,i})^T$ para denotar el espectro de exposición para M bandas de frecuencia de un archivo w_i de la base. A su vez la exposición, obtenida a partir de la especificación espectral y la duración del archivo de salida definidos por el usuario, se denota $\mathbf{e}_{\text{u}} = (e_{1,\text{u}}, e_{2,\text{u}}, \cdots, e_{m,\text{u}}, \cdots, e_{M,\text{u}})^T$, donde $e_{m,\text{u}} = p_{\text{rms},m,\text{u}} \times T$, siendo $p_{\text{rms},m,\text{u}}$ y T los parámetros especificados por el usuario.

8.3.2 Problemas de combinatoria

La parte espectral del problema general se puede expresar como un problema de combinatoria en el cual para cada banda m se escribe una ecuación de la forma mostrada

en la ecuación (8.11). Es decir, se trata de un sistema lineal de M ecuaciones con N incógnitas.

$$\begin{array}{ccccccccccccc}
x_1e_{1,1} & + & x_2e_{1,2} & + & \cdots & + & x_ne_{1,n} & + & \cdots & + & x_Ne_{1,N} & = & e_{1,\mathsf{u}} \\
x_1e_{2,1} & + & x_2e_{2,2} & + & \cdots & + & x_ne_{2,n} & + & \cdots & + & x_Ne_{2,N} & = & e_{2,\mathsf{u}} \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
x_1e_{m,1} & + & x_2e_{m,2} & + & \cdots & + & x_ne_{m,n} & + & \cdots & + & x_Ne_{m,N} & = & e_{m,\mathsf{u}} \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
x_1e_{M,1} & + & x_2e_{M,2} & + & \cdots & + & x_ne_{M,n} & + & \cdots & + & x_Ne_{M,N} & = & e_{M,\mathsf{u}}
\end{array} \tag{8.13}$$

o bien, en notación matricial

$$\mathbf{A}\mathbf{x} = \mathbf{e}_{\mathsf{u}} \tag{8.14}$$

donde $\mathbf{A}$ es una matriz de $M \times N$ cuya columna n corresponde al vector $\mathbf{e_n}$ para $w_1, w_2, \cdots, w_N$ y $\mathbf{x}$ es un vector cuyos elementos representan la ganancia cuadrática para cada archivo w_n.

La solución exacta para cualquier espectro de exposición no se puede asegurar a menos que los vectores que representan el espectro de exposición de cada archivo de audio contengan los M posibles vectores ortogonales en el hiperoctante $\Re^M_{\geq 0}$. No es posible en general obtener M vectores ortogonales en un hiperoctante salvo que coincidan con los ejes, es decir, por ejemplo, que correspondan a sonidos con contenido en una sola banda cada uno. Si este fuese el caso, se podría generar cualquier punto en $\Re^M_{\geq 0}$, mediante una combinación lineal con coeficientes $x_i \geq 0$. En la figura 8.1 se muestra un ejemplo con los espectros de exposición en dos bandas, tanto los requeridos por el usuario $\mathbf{e}_{\mathsf{u}} = (e_{\mathsf{u},1}, e_{\mathsf{u},2})^T$ cómo los correspondientes a los archivos de audio de la base de datos. En este ejemplo la base de datos solo cuenta con dos archivos, cuyos espectros de exposición son $\mathbf{e_1} = (e_{1,1}, e_{1,2})^T$ y $\mathbf{e_2} = (e_{2,1}, e_{2,2})^T$, respectivamente. El subconjunto en $\Re^2_{\geq 0}$ que puede ser generado por $\mathbf{e_1}$ y $\mathbf{e_2}$ se muestra sombreado. Con flechas se muestran los vectores $\mathbf{e_1}$ y $\mathbf{e_2}$ y con un círculo se muestra $\mathbf{e}_{\mathsf{u}}$.

La figura 8.1 muestra tres criterios de solución más cercana ($\mathbf{e}^*_{\mathsf{o}}$) al espectro de exposición requerido por el usuario $\mathbf{e}_{\mathsf{u}}$. Los criterios de distancia son las normas ℓ_1, ℓ_2 y ℓ_∞, siendo las ganancias cuadráticas óptimas (o las que proveen la solución más cercana) $\mathbf{x}^*(\ell_1)$, $\mathbf{x}^*(\ell_2)$ y $\mathbf{x}^*(\ell_\infty)$, respectivamente para cada criterio. El superíndice $*$

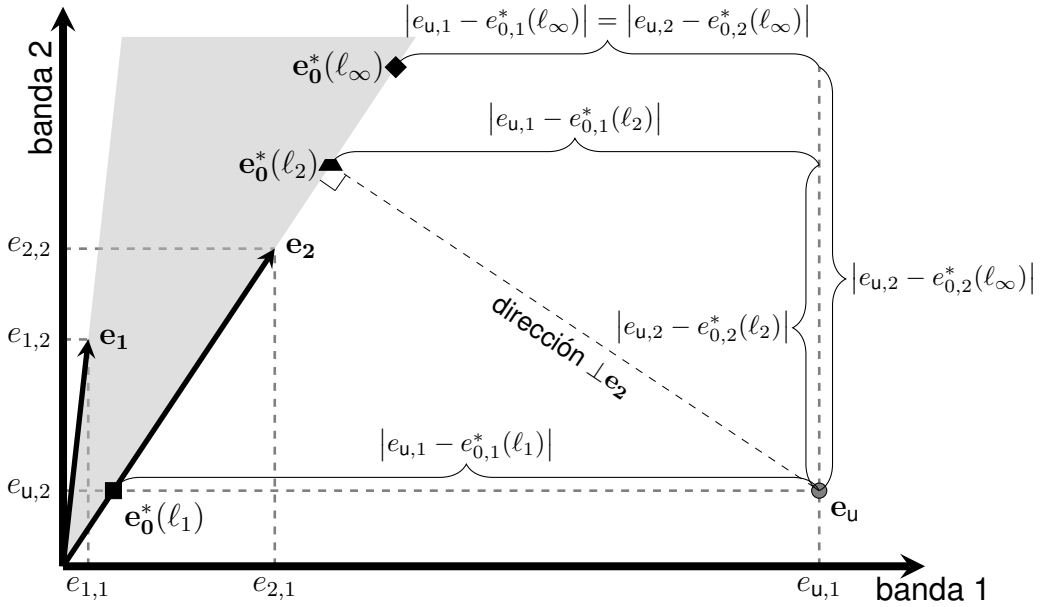


Figura 8.1: Subconjunto generado por sonidos w_1 y w_2 con espectros $\mathbf{e_1}$ y $\mathbf{e_2}$ respectivamente, y soluciones aproximadas para instancia $\mathbf{e}_u$ requerida por usuario. Aproximación por norma ■ ℓ_1, ▲ ℓ_2 y ◆ ℓ_∞

se ha empleado para indicar que la solución no es exacta. Para el caso que el espectro de exposición requerido por el usuario $\mathbf{e}_u$ esté dentro del subconjunto generado por los espectros de exposición $\mathbf{e_1}, \mathbf{e_2}, \cdots, \mathbf{e_N}$, sucede que los tres criterios coinciden en hallar al menos una solución exacta (es decir, $\mathbf{e}_o[\mathbf{x_e}(\ell_1)] = \mathbf{e}_o[\mathbf{x_e}(\ell_2)] = \cdots = \mathbf{e}_o[\mathbf{x_e}(\ell_\infty)] = \mathbf{e}_u$ (donde $\mathbf{x_e}$ corresponde a puntos óptimos de ganancia cuadrática que producen la solución exacta).

En el caso que $\mathbf{e}_u$ no esté dentro del subconjunto, es posible hallar un punto que esté dentro del subconjunto y a su vez se encuentre a la menor distancia de $\mathbf{e}_u$, según algún criterio de distancia. Es habitual utilizar el criterio de la norma euclídea ℓ_2 en casos donde la distancia representa justamente unidades de longitud, por ejemplo una distancia recorrida sin obstáculos. El criterio dado por la distancia Manhattan o ciudad (City-Block) ℓ_1 puede ser adecuado para distancias recorridas en una ciudad cuyas calles son equidistantes.

En el caso de la combinación de sonidos estas distancias deben interpretarse de otra manera. Una norma ℓ_1 significa que la suma de las diferencias absolutas entre la exposición sonora en las bandas del espectro especificado por el usuario y en las bandas del espectro del sonido ambiente de salida será mínima. Una norma ℓ_∞ signi-

fica que la máxima de las diferencias absolutas de nivel entre las bandas del espectro especificado por el usuario y las bandas del espectro del sonido ambiente de salida será mínima. Esto significa que para el criterio ℓ_1 el error en una banda puede ser mayor que el error para otra de ellas (en el ejemplo de la figura 8.1 el error más grande se da en la banda 1 mientras que en la banda 2 el error es nulo). En cambio con el criterio ℓ_∞ esto significa que el error tiende a distribuirse homogéneamente entre las bandas.

En el ejemplo de la figura 8.1 el error con el criterio ℓ_∞ es idéntico para ambas bandas y menor que el mayor de los errores con los otros dos criterios, es decir $\left|e_{\mathsf{u},1} - e^*_{0,1}(\ell_\infty)\right| = \left|e_{\mathsf{u},2} - e^*_{0,2}(\ell_\infty)\right| < \left|e_{\mathsf{u},1} - e^*_{0,1}(\ell_2)\right| < \left|e_{\mathsf{u},1} - e^*_{0,1}(\ell_1)\right|$. El error con el criterio ℓ_∞ es mayor que el menor de los errores con los otros criterios, es decir $\left|e_{\mathsf{u},2} - e^*_{0,1}(\ell_\infty)\right| > \left|e_{\mathsf{u},2} - e^*_{0,2}(\ell_2)\right| > \left|e_{\mathsf{u},2} - e^*_{0,2}(\ell_1)\right| = 0$. Si bien es deseable que el error sea lo menor posible, resulta de mayor relevancia que el error de la banda con mayor error sea lo menor posible lo cuál en general redunda en una distribución más homogénea del error entre las bandas.

Para normas intermedias entre ℓ_1 y ℓ_∞ el error tiende a una distribución mas homogénea cuanto mayor sea el orden de la norma. Este análisis puede extenderse a más de dos bandas, observando la misma conclusión. Para casos con más de dos bandas la norma ℓ_∞ no logrará siempre que los errores de todas las bandas sean idénticos pero si que estén distribuidos de manera más homogénea.

Este análisis sugiere que es más adecuado el criterio ℓ_∞ pues un error grande en una banda es mucho más perceptible que si ese mismo error se distribuye homogéneamente entre las demás bandas involucradas. Un error grande en una banda es mucho más perceptible, pues se aleja más de las diferencias apenas perceptibles, en comparación con un error menor aunque distribuido en varias bandas. Por ejemplo si esa diferencia es por exceso podría causar una sensación de tonalidad no buscada.

8.4 Programación lineal con solución en los reales

Este método consiste en separar el problema en dos pasos. El primer paso consiste en ajustar el histograma de duración y el segundo en ajustar el contenido espectral mediante métodos de optimización denominados de programación lineal con solución en los reales.

Antes de encarar la solución del problema, la base de datos puede ser editada para contener solamente un evento sonoro en cada archivo de audio que compone la base. Mediante un análisis previo se genera una variable que asigna a cada archivo un grupo según la duración del evento que ese archivo contiene. De esta manera los coeficientes y_i pueden ser interpretados como la cantidad necesaria de eventos, dentro de cada grupo de duración, que deben estar presentes en el sonido ambiental de salida. Este paso no necesita un algoritmo de solución sino uno de selección aleatoria de archivos w_i que conformaran la solución de la parte del histograma. Por ejemplo si el usuario especifica 10 eventos en el intervalo de duración $]0; 1]$ s, este algoritmo simplemente identifica 10 archivos del grupo correspondiente a ese intervalo. De esta manera se halla el subconjunto de archivos I (aquellos que estarán presentes en el archivo de salida) y los valores de y_i.

En el segundo paso, se obtienen los coeficientes x_i solucionando el problema espectral de la ecuación (8.14) como un problema de optimización matemática. La solución se puede hallar por ejemplo mediante optimización de mínimos cuadrados con soluciones no negativas [Accolti y Miyara, 2010, 2008] o bien mediante un problema de programación lineal usando una aproximación de norma infinito de manera similar a la usada en 8.6.

El ejemplo de la figura 8.2 muestra el dominio para un problema en una sola banda de frecuencias y solo dos sonidos en la base bajo la restricción de que la máxima ganancia es g_1^{max} para el sonido 1 de la base y g_2^{max} para el sonido 2 de la base, dando las cotas máximas x_1^{max} y x_2^{max} respectivamente para los coeficientes de combinación de exposición.

Además de las restricciones de borde superior también se deben tener en cuenta unas restricciones inferiores. Esas restricciones equivalen a fijar un mínimo de ganancia g_n^{min}, con su correspondiente coeficiente x_n^{min}, para cada archivo w_n, pues si el nivel de presión sonora del sonido no supera cierto umbral no será percibido por el oyente.

Una alternativa es evitar las restricciones x_n^{min}. Pero en ese caso la solución puede, y es lo que ocurre generalmente, dar un gran número de ganancias nulas. Esto implica que varios eventos i, que se esperaba estuviesen presentes debido al primer paso, no estarán presentes en el sonido ambiental de salida. Pero esto puede ser corregido mediante diferentes técnicas. Una de estas técnicas es iterar la solución del segundo

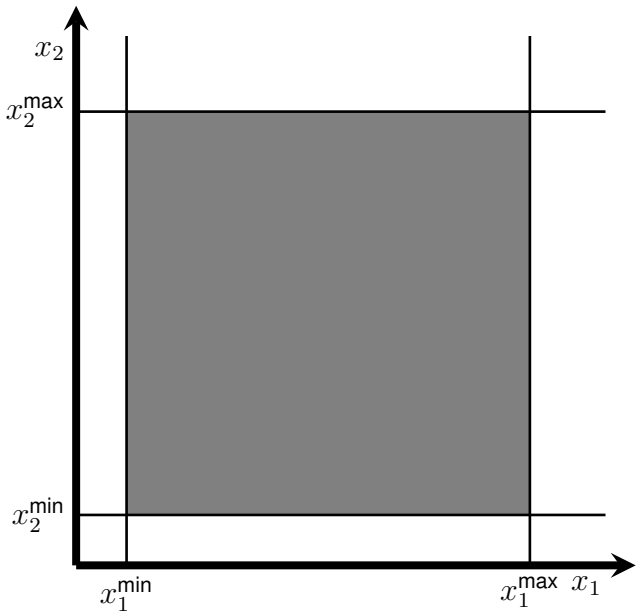


Figura 8.2: Dominio de un problema de combinación controlada de archivos de audio con solución en los reales no negativos

paso usando subconjuntos de la base de sonidos completa, por ejemplo mediante el proceso VARISON utilizado en trabajos previos [Accolti y Miyara, 2010, 2008].

Sin embargo, la gran desventaja de este método está en la dificultad de encontrar los instantes en los cuales cada evento sonoro debería aparecer de modo de obtener un sonido ambiental realista. Esos instantes dependen de múltiples factores en el mundo real. Una posibilidad es suponer que no existe ninguna relación entre los instantes de aparición de los eventos (lo cual no es real para todos los eventos) y distribuirlos según un proceso estocástico de Poisson. Aunque en la realidad los procesos estocásticos de Poisson no describen exactamente la distribución de eventos sonoros, para algunos eventos permiten una simulación bastante realista.

8.5 Programación lineal con solución en los enteros

Una vez introducidos los grupos de archivos de la sección 3.2.1, el problema completo (sin separar en pasos como en la sección anterior) se puede solucionar usando técnicas de optimización matemática. Incluso, si la función objetivo se expresa como una ecuación lineal, o el problema se plantea como un problema equivalente con función objetivo lineal, entonces se pueden utilizar algoritmos conocidos de programación lineal (por ejemplo mediante solución de subproblemas con relajaciones resueltas me-

diante algoritmo Simplex o algoritmo de Punto Interior).

El problema completo se puede expresar como un problema típico de optimización convexa [Boyd y Vandenberghe, 2004] y obtener una solución mediante la solución a un problema equivalente de programación lineal con solución en los enteros (IP) [Wolsey, 1998].

Este problema de optimización convexa tiene la forma

$$\begin{aligned} &\underset{\mathbf{y}}{\text{minimizar}} \quad f_0(\mathbf{y}) = \|A\mathbf{y} - \mathbf{u}\| \\ &\text{sujeto a} \qquad\qquad y_n \leq y_n^{max}, \quad n \in \{1, 2, \cdots, N\}, \qquad \text{(a)} \\ &\qquad\qquad\qquad\quad y_n \geq y_n^{min}, \quad n \in \{1, 2, \cdots, N\}, \qquad \text{(b)} \\ &\qquad\qquad\qquad\quad y_n \in \mathbb{N}_0, \quad n \in \{1, 2, \cdots, N\}, \qquad \text{(c)} \end{aligned} \tag{8.15}$$

donde la variable objetivo $\mathbf{y} = (y_1, y_2, \cdots, y_N)^T$ es un vector cuyos elementos son enteros no negativos ($y_n \in \mathbb{N}_0$). La matriz $\mathbf{A}$ tiene la forma

$$\mathbf{A} = \begin{pmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ e_{M,1} & e_{M,2} & \cdots & e_{M,N} \\ d_{1,1} & d_{1,2} & \cdots & d_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{L,1} & d_{L,2} & \cdots & d_{L,N} \end{pmatrix} \tag{8.16}$$

Cada una de las M primeras filas de $\mathbf{A}$ representan la m-ésima banda del espectro y cada una de las L últimas filas representan el l-ésimo rango del histograma de duraciones. Cada columna representa el n-ésimo archivo de los N archivos de audio que componen la base.

El máximo numero admisible de repeticiones del archivo w_n en el sonido ambiente es y_n^{max}, según la restricción de la ecuación (8.15.a) y el mínimo número de repeticiones es y_n^{min}, según la restricción de la ecuación (8.15.b). El mínimo debe tener en cuenta que todos los archivos para los cuales $y_n = 0$ no están incluidos en el subconjunto I y por consiguiente no estarán presentes en el archivo de audio de salida. O por el contrario, si para un determinado archivo w_{n_0} se especifica $y_{n_0}^{min} > 0$, resultará que el archivo w_{n_0} estará presente en el archivo de audio de salida. La función objetivo f_0 se define mediante $\|\bullet\|$ que es una norma (más adelante se introducen algunas normas

vectoriales). El vector $\mathbf{u}$ tiene dimensión $M + L$, siendo sus primeros elementos el espectro de exposición y los últimos el histograma de duraciones, ambos especificados por el usuario. En notación matemática esto es $\mathbf{u} = ({\mathbf{e}_\mathsf{u}}^T, {\mathbf{d}_\mathsf{u}}^T)^T$.

Reemplazando (8.16) en (8.15) resulta la función de costo

$$f_0 = \left\| \begin{array}{ccccccccc} y_1 e_{1,1} & + & y_2 e_{1,2} & + & \cdots & + & y_N e_{1,N} & - & e_{1,\mathsf{u}} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_1 e_{m,1} & + & y_2 e_{m,2} & + & \cdots & + & y_N e_{m,N} & - & e_{m,\mathsf{u}} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_1 e_{M,1} & + & y_2 e_{M,2} & + & \cdots & + & y_N e_{M,N} & - & e_{M,\mathsf{u}} \\ y_1 d_{1,1} & + & y_2 d_{1,2} & + & \cdots & + & y_N d_{1,N} & - & d_{1,\mathsf{u}} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_1 d_{l,1} & + & y_2 d_{l,2} & + & \cdots & + & y_N d_{l,N} & - & d_{l,\mathsf{u}} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_1 d_{L,1} & + & y_2 d_{L,2} & + & \cdots & + & y_N d_{L,N} & - & d_{L,\mathsf{u}} \end{array} \right\| \tag{8.17}$$

Los primeros M elementos dentro de la norma son equivalentes a la ecuación (8.13) salvo que en este caso los coeficientes son enteros y representan la cantidad de repeticiones. Notar que, al repetir los archivos en distintos instantes de tiempo, se puede aceptar la incoherencia entre las repeticiones y aplicar la suma energética según las ecuaciones (8.9) y (8.10) considerando cada repetición como una señal p_i.

Los últimos L elementos dentro de la norma son equivalentes a la ecuación (5.2). Es decir, cada uno de estos últimos L elementos representan la combinación de archivos de la base necesaria para obtener el valor del histograma de duración requerido por el usuario para cada uno de los L rangos.

Se puede mejorar esta técnica mediante una modificación simple de la matriz $\mathbf{A}$ introducida en la ecuación (8.16). Esta modificación consiste en repetir las columnas de $\mathbf{A}$ modificando solo los valores del espectro en cada repetición y manteniendo intactos los valores del histograma de duración de cada archivo para cada repetición. Cada repetición puede representar, en términos acústicos, los mismos eventos de la matriz original pero suponiendo que la fuente esta a una distancia distinta para cada repetición, por ejemplo teniendo en cuenta la atenuación por divergencia geométrica y la absorción del medio aire según algún modelo matemático como el de la norma ISO 9613-2 [1996]. En este caso se deberán agregar variables y_n. Por ejemplo si se repite

K veces la matriz original, N debe ser remplazado por $N' = K \times N$. A las restricciones de la inecuación (8.15.a) se agregan términos que contienen las variables $y_{k\times n}$, con $k \in \{1, 2, \cdots, K\}$ y $n \in \{1, 2, \cdots, N\}$. Es decir, la cantidad de restricciones siguen siendo N pero cada una de ellas tiene la forma $y_n + y_{2\times n} + \cdots + y_{K\times n} \leq y_n^{max}$. Por otra parte las N restricciones de la inecuación (8.15.b) pasarán a ser N' restricciones de tipo $y_{k\times n} \geq y_n^{min}$ con $k \in \{1, \cdots, K\}$ y $n \in \{1, \cdots, N\}$.

La función objetivo es convexa pero no necesariamente lineal salvo para la norma ℓ_1. Para la norma ℓ_∞, si bien la función objetivo no sería lineal, es posible definir un problema análogo cuya función objetivo sí sea lineal [Boyd y Vandenberghe, 2004].

En el ejemplo de la figura 8.3 se muestra el dominio para un problema con sólo dos archivos w_1 y w_2. Las abscisas representan la cantidad de repeticiones y_1 del primer archivo y las ordenadas y_2 lo propio del segundo archivo. Las cotas superiores y_1^{max} e y_2^{max}, en conjunto con las cotas inferiores para las cuales $y_n \geq 0$, determinan la cápsula convexa[1] sombreada en color gris claro. Dentro de esta cápsula convexa, en los puntos marcados con círculos gris oscuro, se encuentran las soluciones posibles, que, como se expresó anteriormente, son de naturaleza discreta.

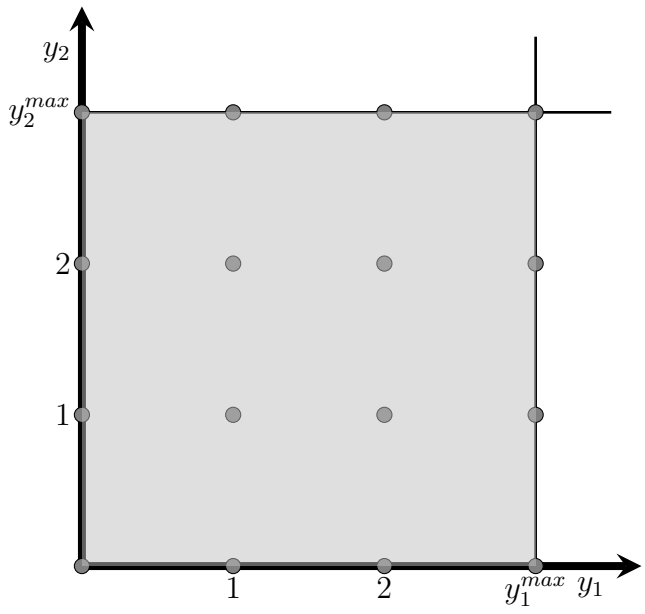


Figura 8.3: Dominio de un problema de combinación controlada de archivos de audio con solución en los enteros no negativos

[1] La cápsula convexa es el conjunto convexo más pequeño que contiene las posibles soluciones.

En general, es imprescindible una buena formulación de cualquier problema de combinación, enmarcado en técnicas de programación lineal con solución en los enteros (o enteros y reales), para asegurar que exista una solución al menos en un tiempo de cómputo razonable. La técnica principal consiste en formular el problema, y de no ser posible formular un problema equivalente, de tal modo que las restricciones demarquen la cápsula convexa de las soluciones posibles. Existen diversas técnicas orientadas a la formulación del problema, para una revisión detallada consultar Wolsey [1998].

La dificultad que surge al plantear todo el problema como uno equivalente IP es el gran tamaño y el costo computacional que esto significa. Sin embargo esta formulación es muy realista (ver página 116 en 11.1).

8.6 Programación lineal con solución mixta

Esta formulación, enmarcada en problemas de programación lineal con soluciones mixtas en enteros y reales (MILP), es la implementada en esta tesis. La formulación de este problema consiste en separar una parte con soluciones en los enteros y otra con soluciones en los reales, pero formular todo junto en un mismo problema [Accolti y Miyara, 2015]. Las restricciones relacionadas con el histograma de duración tienen solución en los enteros, pues para que un evento de una duración específica esté presente se deberá tener ese evento completo y no será posible conseguir un evento de una duración específica mediante partes de uno o más de un evento. Esto no sucede con la exposición sonora en una banda de frecuencias. Esta exposición sí se puede componer sumando la exposición de varios eventos según la ecuación (8.10). Accolti y Miyara [2015] incluyen el control de las categorías semánticas con soluciones en los enteros usando restricciones para las mismas variables enteras objetivo, es decir agregan restricciones para y_n.

La interpretación físico-acústica de aplicar una ganancia a un sonido tiene que ver con su sonoridad. Se postula en este caso que la sonoridad depende de la distancia del oyente a la fuente sonora virtual. Es decir, no se hace referencia al altavoz como fuente sonora sino a las fuentes que virtualmente representa el archivo de audio de salida cuando es reproducido relativo al oyente. El efecto de mayor relevancia al simular el efecto de distancia entre la fuente sonora y el oyente es la divergencia geométrica

[Kinsler et al., 2000; ISO 9613-2, 1996] que es independiente de la frecuencia. Sin embargo no es el único efecto de la distancia. Al formular el problema de esta manera se pierde la posibilidad de modelar individualmente para cada fuente los efectos dependientes de la frecuencia como la atenuación en el medio. Otros efectos globales si se pueden modelar con facilidad según se describe en el capítulo 7. Sin embargo el efecto de la atenuación del medio aire es pequeño (por ejemplo en $f = 500$ Hz a una temperatura de 30 ºC y 20 % de humedad relativa, la atenuación será de apenas 0,4 dB cada 100 m) y se puede descartar sin grandes consecuencias. Esto es así porque la mayoría de las fuentes distinguibles están ubicadas a distancias relativamente cercanas, y las fuentes más lejanas constituyen el ruido de fondo que en general ya está incluido en la mayoría de los sonidos grabados de la realidad.

La formulación en este caso utiliza las variables objetivo en el vector $\mathbf{q} = (\mathbf{x}^T, \mathbf{y}^T)^T$, siendo los elementos de $\mathbf{x}$ los coeficientes de combinación de exposición y los de $\mathbf{y}$ los coeficientes de repeticiones. Las restricciones son similares a los casos anteriores (secciones 8.4 y 8.5).

El problema se formula según

$$
\begin{aligned}
\underset{\mathbf{q}=(\mathbf{x}^T,\mathbf{y}^T)^T}{\text{minimizar}} \quad & f_0(\mathbf{q}) = \|A\mathbf{q} - \mathbf{u}\| \\
\text{sujeto a} \quad & y_n \leq ymax_n,\ n = 1, 2, \cdots, N. && \text{(a)} \\
& y_n - \frac{ymax_n}{x_n^{min}} x_n \leq 0,\ n = 1, 2, \cdots, N. && \text{(b)} \\
& -y_n + \frac{1}{xmax_n} x_n \leq 0,\ n = 1, 2, \cdots, N. && \text{(c)} \\
& x_n \in \mathbb{R}, \quad y_n \in \mathbb{N}_0 \quad n = 1, 2, \cdots, N. && \text{(d)}
\end{aligned}
\tag{8.18}
$$

donde la matriz $\mathbf{A}$ tiene la forma

$$
\mathbf{A} = \begin{pmatrix}
e_{1,1} & e_{1,2} & \cdots & e_{1,N} & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
e_{M,1} & e_{M,2} & \cdots & e_{M,N} & 0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 & d_{1,1} & d_{1,2} & \cdots & d_{1,N} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & d_{L,1} & d_{L,2} & \cdots & d_{L,N}
\end{pmatrix}
\tag{8.19}
$$

y el vector de requerimientos del usuario nuevamente tiene la forma empleada en la sección 8.5, esto es $\mathbf{u} = (\mathbf{e_u}^T, \mathbf{d_u}^T)^T$.

Al juntar ambas partes del problema se debe mantener cierta congruencia para que los valores de y_n y x_n sean nulos para todo n fuera del subconjunto $I \subset \{1, 2, \cdots, n, \cdots, N\}$. En otras palabras si un archivo de la base de datos no está presente por tener una ganancia nula no tendrá sentido repetir ese archivo, o un archivo cuya cantidad de repeticiones es nula no estará presente y por lo tanto no tendrá sentido aplicar una ganancia. Para mantener esta congruencia se aplicarán unas cotas inferiores que permitan modelar para todo n las condiciones: si $x_n = 0$, entonces $y_n = 0$ y viceversa. Se aprovecha esta restricción para restringir también los valores mínimos x_n^{min} y máximos x_n^{min} de x_n que están relacionados con el nivel sonoro mínimo que tendrá un archivo de la base que esté presente en el archivo de salida.

Las restricciones (8.18.a) corresponden a la máxima cantidad de repeticiones del archivo w_n y las restricciones (8.18.b) y (8.18.c) a la congruencia del subconjunto y la ganancia del n-ésimo archivo respectivamente.

En este caso la función objetivo f_0 es similar a la ecuación (8.17) salvo que los coeficientes de la parte espectral corresponden a ganancias cuadráticas x_n en lugar de repeticiones y_n. La ecuación (8.20) muestra el desarrollo de la función objetivo para el caso MILP.

$$f_0 = \left\| \begin{array}{ccccccccc} x_1 e_{1,1} & + & x_2 e_{1,2} & + & \cdots & + & x_N e_{1,N} & - & e_{1,\mathsf{u}} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ x_1 e_{m,1} & + & x_2 e_{m,2} & + & \cdots & + & x_N e_{m,N} & - & e_{m,\mathsf{u}} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ x_1 e_{M,1} & + & x_2 e_{M,2} & + & \cdots & + & x_N e_{M,N} & - & e_{M,\mathsf{u}} \\ y_1 d_{1,1} & + & y_2 d_{1,2} & + & \cdots & + & y_N d_{1,N} & - & d_{1,\mathsf{u}} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_1 d_{l,1} & + & y_2 d_{l,2} & + & \cdots & + & y_N d_{l,N} & - & d_{m,\mathsf{u}} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_1 d_{L,1} & + & y_2 d_{L,2} & + & \cdots & + & y_N d_{L,N} & - & d_{M,\mathsf{u}} \end{array} \right\| \tag{8.20}$$

La figura 8.4 muestra, a modo de ejemplo, el dominio para un problema MILP con un solo archivo. En las abscisas se muestra el valor del coeficiente de exposición y en las ordenadas el coeficiente de repeticiones para ese archivo.

El borde derecho (figura 8.4), de manera similar al caso de ambos bordes en la figura 8.2, representa el coeficiente de exposición máximo, por supuesto dependiendo

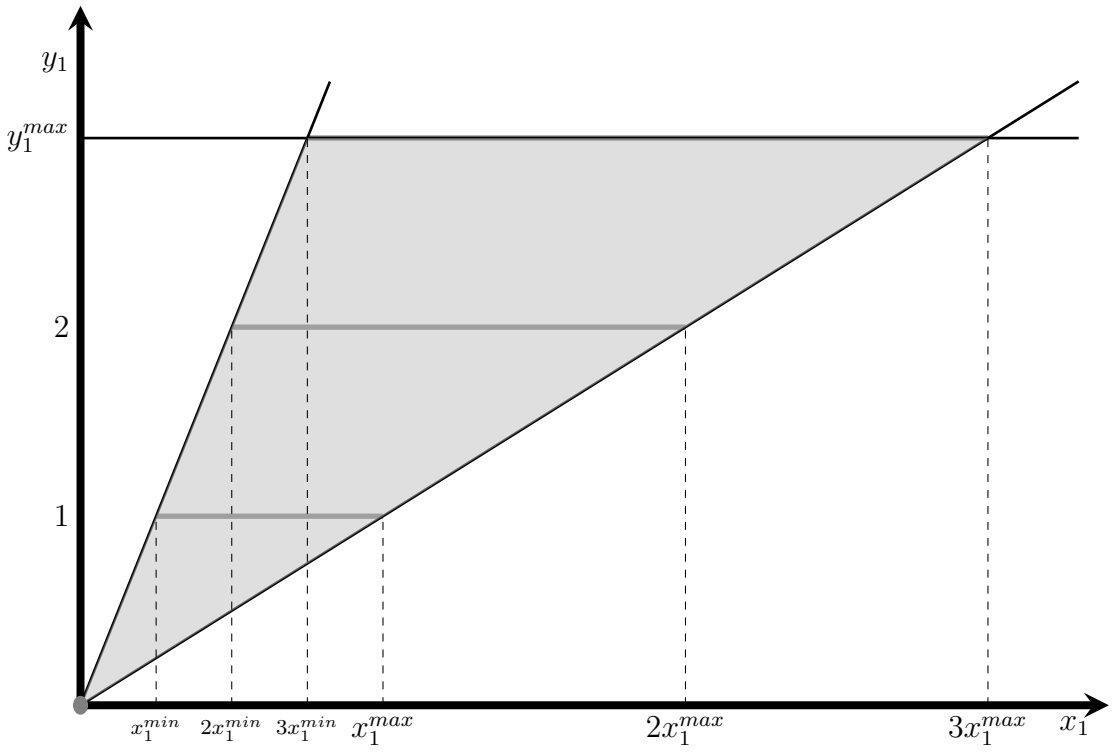


Figura 8.4: Dominio de un problema de combinación controlada de archivos de audio con solución mixta en los enteros no negativos y los reales no negativos. Se muestra la cápsula convexa sombreada en gris claro.

de la cantidad de repeticiones. El borde superior, de manera similar al caso de ambos bordes en la figura 8.3, representa el coeficiente de repetición máximo. El borde izquierdo corresponde a la restricción mínima para el coeficiente de exposición, nuevamente dependiente de la cantidad de repeticiones. Con líneas grises oscuros, y un círculo gris oscuro en el origen, se marcan las soluciones posibles. Notar que los bordes derecho e izquierdo se cierran hacia el valor nulo, es decir $x_1 = 0 \Leftrightarrow y_1 = 0$ lo cuál asegura la congruencia mencionada anteriormente.

Si el problema tuviese dos archivos, el plano representado por las variables y_1 e y_2 sería idéntico al de la figura 8.3, y el plano x_1 e x_2 sería idéntico al de la figura 8.2. En efecto, en el problema con N archivos, los planos correspondientes a dos variables objetivo serán similares a los de las figuras 8.2, 8.3 y 8.4 según el tipo de variable y modificados según los valores que el usuario requiera para cada una de las cotas o bordes.

8.7 Linealización de la función objetivo

Las funciones objetivo de los problemas, descritos en las secciones 8.4 a 8.6, involucran una norma de las variables objetivo. Es decir, no son lineales pues las normas involucran valores absolutos, potencias, raíces y, en el caso de la norma infinito, la función máximo. La técnica habitual de programación lineal, para resolver este tipo de problemas de combinatoria, es definir otro problema que tienda a una solución similar. En esta sección se analizará cómo definir el nuevo problema, para problemas de optimización convexa cuya función objetivo involucra la norma infinito y cuyas restricciones son lineales respecto a las variables objetivo, como es el empleado en esta tesis. Como definir el problema para otro tipo de normas se puede consultar en la bibliografía [Wolsey, 1998; Boyd y Vandenberghe, 2004].

El nuevo problema tendrá una función objetivo lineal muy simple dada por una nueva variable v, y una formulación similar al argumento de la norma pasará a formar parte de las restricciones de tipo inecuación. El nuevo problema es equivalente al problema de la ecuación (8.18) y se formula como

$$
\begin{aligned}
\underset{\mathbf{q}}{\text{minimizar}} \quad & f_0(\mathbf{q}) = v & & \\
\text{sujeto a} \quad & \mathbf{A}\mathbf{q} - \mathbf{u} \leq v\mathbf{1},\ n = 1, 2, \cdots, N. & \text{(a)} & \\
& -(\mathbf{A}\mathbf{q} - \mathbf{u}) \leq v\mathbf{1},\ n = 1, 2, \cdots, N. & \text{(b)} & \\
& y_n \leq ymax_i,\ n = 1, 2, \cdots, N. & \text{(c)} & \quad (8.21) \\
& y_n - \frac{ymax_n}{x_n^{min}} x_n \leq 0,\ n = 1, 2, \cdots, N. & \text{(d)} & \\
& -y_n + \frac{1}{xmax_n} x_n \leq 0,\ n = 1, 2, \cdots, N. & \text{(e)} & \\
& x_n \in \mathbb{R}, \quad y_n \in \mathbb{N}_0 \quad n = 1, 2, \cdots, N. & \text{(f)} &
\end{aligned}
$$

donde **1** es un vector de dimensión N y todos sus elementos son iguales a 1. El vector que contiene las variables objetivo es $\mathbf{q} = (\mathbf{x}^T, \mathbf{y}^T, v)^T$.

Las nuevas $2 \times N$ restricciones (8.21.a) y (8.21.b), en notación más compacta corresponden a $-v\mathbf{1} \leq \mathbf{A}\mathbf{q} - \mathbf{u} \leq v\mathbf{1}$, y desarrollando la notación para cada elemento se

obtiene

$$\begin{array}{ccccccccccccc}
-v & \leq & x_1e_{1,1} & + & x_2e_{1,2} & + & \cdots & + & x_Ne_{1,N} & - & e_{1,\mathsf{u}} & \leq & v \\
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \\
-v & \leq & x_1e_{m,1} & + & x_2e_{m,2} & + & \cdots & + & x_Ne_{m,N} & - & e_{m,\mathsf{u}} & \leq & v \\
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \\
-v & \leq & x_1e_{M,1} & + & x_2e_{M,2} & + & \cdots & + & x_Ne_{M,N} & - & e_{M,\mathsf{u}} & \leq & v \\
-v & \leq & y_1d_{1,1} & + & y_2d_{1,2} & + & \cdots & + & y_Nd_{1,N} & - & d_{1,\mathsf{u}} & \leq & v \\
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \\
-v & \leq & y_1d_{m,1} & + & y_2d_{m,2} & + & \cdots & + & y_Nd_{m,N} & - & d_{m,\mathsf{u}} & \leq & v \\
\vdots & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \\
-v & \leq & y_1d_{M,1} & + & y_2d_{M,2} & + & \cdots & + & y_Nd_{M,N} & - & d_{M,\mathsf{u}} & \leq & v
\end{array} \tag{8.22}$$

Es evidente que v es una cota para el máximo valor absoluto de los elementos de $\mathbf{Aq} - \mathbf{u}$. Es decir, $v \leq \max(|\mathbf{Aq} - \mathbf{u}|) = \|\mathbf{Aq} - \mathbf{u}\|_\infty$. Por esa razón el nuevo problema, que implica minimizar la nueva variable v, es un problema análogo al original de minimizar $\|A\mathbf{q} - \mathbf{u}\|_\infty$ pero lineal.

8.8 Resolución

La ecuación (8.22) formula 2 restricciones por cada fila. Habitualmente esta información se debe manejar algebraicamente de modo tal que respete el modo de entrada de datos del algoritmo de resolución. Típicamente las restricciones se escriben como ecuaciones (o inecuaciones) con cotas superiores constantes según se utiliza en esta tesis y se muestra en la ecuación (8.21). Normalmente, para enviar los datos al algoritmo de resolución, se compone un vector que contiene los coeficientes de la función objetivo, una matriz que contiene los coeficientes del lado izquierdo de las inecuaciones, un vector que contiene los coeficientes del lado derecho de las inecuaciones y un vector cuyos elementos, de manera binaria, indica cuáles variables objetivo son reales y cuáles son enteros.

El bloque *Resolución Problema Combinatoria* del módulo *Combinador* básicamente llama al algoritmo de resolución con los datos de entrada de la ecuación (8.21) y devuelve un vector de índices que identifica el subconjunto I de archivos w_i, un vector con los coeficientes x_i y otro vector con cantidades de repetición y_i. Para un conjunto de pruebas, a ser realizados con un mismo sistema de auralización, la formulación

del problema será la misma para cada prueba salvo la instancia del problema que cambiará los valores del vector $\mathbf{u}$.

El algoritmo de resolución empleado en esta tesis es el del software CPLEX [ILOG, 2011] con licencia de uso académico. CEPLEX implementa un algoritmo de propósito general para resolver problemas MILP entre otros. El algoritmo se basa en iteraciones que inician con una relajación del problema, continúan agregando cortes, aplicando estrategias de heurística, subdividiendo el problema, aplicando nuevos cortes a cada problema subdividido hasta que la relajación del problema ofrezca una solución tolerable bajo algún criterio [Wolsey, 1998].

La relajación del problema consiste en resolver el problema sin las restricciones de algunas variables en $\mathbb{Z}$. Estos problemas son resueltos por los algoritmos *Simplex* o de *Punto Interior*. Las técnicas de corte y estrategias heurísticas permiten una solución más rápida identificando rápidamente grupos de puntos que no pertenecen a las soluciones. Para mayores detalles sobre estas técnicas se sugiere el libro *Integer Programming* de Wolsey [1998].

8.9 Mezcla de audio

En esta sección se muestra cómo son interpretados los datos de salida del bloque *Resolución Problema Combinatoria* por los bloques *Cálculo de Instantes de Inserción* y *Combinador archivos de audio* del módulo *Combinador* para generar el archivo de audio de salida.

8.9.1 Instantes de inserción

El bloque *Cálculo de Instantes de Inserción* del módulo *Combinador* utiliza los valores de y_i del bloque anterior y los valores de muestra de inserción $j_{\mathsf{b},i}$ de cada archivo asignados según se detalla en 3.2.1, 3.2.4 y 6.2.

A partir de la muestra de inserción de cada archivo, y su tasa de muestreo F_{s}, se calcula el instante de inserción según

$$t_{\mathsf{b},i} = \frac{j_{\mathsf{b},i}}{F_{\mathsf{s}}} \tag{8.23}$$

Para los eventos de la clase *simple* y *nube* los instantes de inserción en el archivo de salida se calculan de la siguiente manera. Sea T_u la duración total especificada por el usuario para el archivo de salida y o un subíndice para cada repetición del archivo

w_i. El primer paso es seleccionar aleatoriamente y_i instantes $t_{\mathsf{rnd},i,o}$ del intervalo $[0; T_u]$, según una función de distribución uniforme, para cada archivo w_i. El segundo paso es calcular los instantes $t_{\mathsf{s},i,o}$ de la señal de salida en los cuales se insertará la muestra inicial de la señal s_i según

$$t_{\mathsf{s},i,o} = t_{\mathsf{rnd},i,o} - t_{\mathsf{b},i} \tag{8.24}$$

Para los archivos de la clase *bucle*, o bien aquellos cuyo instante de inserción corresponde a un valor nulo, se calculan los instantes de inserción en el archivo de salida según

$$t_{\mathsf{s},i,o} = d_{\mathsf{a},i} \times o \tag{8.25}$$

para todo o tal que $t_{\mathsf{s},i,o} \leq T_u$.

8.9.2 Combinación

Para simplificar las instrucciones para la mezcla se compone una matriz de mezcla $\mathbf{X}$ cuyas columnas corresponden a cuatro vectores. El primero contiene el nombre de cada archivo en I repitiendo y_i veces cada uno, el segundo vector contiene los valores en muestras correspondientes de $t_{\mathsf{s},i,o}$, el tercero los valores correspondientes de g_i y el cuarto contiene la duración, en muestras, de cada archivo $d_{\mathsf{a},i}$ (ver tabla 8.1).

Cada señal s_i es multiplicada por una ganancia $g_i = \sqrt{x_i/y_i}$. Se debe notar que si bien los archivos en la clase *bucle* pueden ser repetidos en el archivo de salida, en realidad el valor correspondiente del coeficiente de repeticiones es $y_i = 1$ y por lo tanto $g_i = \sqrt{x_i}$ independiente de la cantidad de veces que realmente se inserte.

Finalmente la combinación de audio se realiza mediante un algoritmo que recorre cada una de las filas de la matriz $\mathbf{X}$ y acumula en el archivo de salida las muestras de la señal correspondiente al archivo identificado en esa fila. Antes de sumar cada señal, esta se multiplica por g_i y se desplaza en muestras el equivalente a $t_{\mathsf{s},i,o}$. Se suman solo las muestras que, una vez desplazadas, están dentro del intervalo $[0; T_u]$ excluyendo así las muestras que no entran en dicho intervalo de tiempo.

En la figura 8.5 se muestran las señales de cada uno de los archivos que componen la mezcla, y de la mezcla en si, del ejemplo de la tabla 8.1, en unidades relativas de amplitud en función del tiempo.

En la figura 8.5 se muestra un ejemplo para un estimulo generado con la herramienta propuesta en esta tesis. Cada señal s_i de cada archivo w_i se muestra en las

Tabla 8.1: Matriz de mezcla. Ejemplo para un archivo de salida de 120 s ($5{,}29 \times 10^6$ muestras) de duración y tasa de muestreo de $44\ 100$ Hz

nombre archivo	instante (muestra)	ganancia (adim)	duración (muestras)
avion.wav	$3{,}31 \times 10^5$	$1{,}0 \times 10^0$	$2{,}48 \times 10^6$
camion.wav	$-2{,}25 \times 10^5$	$2{,}6 \times 10^0$	$1{,}59 \times 10^6$
camion.wav	$1{,}86 \times 10^6$	$2{,}6 \times 10^0$	$1{,}59 \times 10^6$
camion.wav	$4{,}11 \times 10^6$	$2{,}6 \times 10^0$	$1{,}59 \times 10^6$
moto125.wav	$4{,}09 \times 10^6$	$3{,}7 \times 10^0$	$5{,}54 \times 10^5$
moto650.wav	$-3{,}75 \times 10^5$	$2{,}8 \times 10^0$	$2{,}04 \times 10^6$
viento.wav	$1{,}00 \times 10^0$	$1{,}7 \times 10^1$	$5{,}02 \times 10^6$
viento.wav	$5{,}02 \times 10^6$	$1{,}7 \times 10^1$	$5{,}02 \times 10^6$
lluvia.wav	$1{,}00 \times 10^0$	$1{,}4 \times 10^1$	$3{,}81 \times 10^6$
lluvia.wav	$3{,}81 \times 10^6$	$1{,}4 \times 10^1$	$3{,}81 \times 10^6$
taladro.wav	$5{,}42 \times 10^5$	$1{,}1 \times 10^1$	$1{,}72 \times 10^6$
taladro.wav	$3{,}17 \times 10^6$	$1{,}1 \times 10^1$	$1{,}72 \times 10^6$
taladro.wav	$3{,}40 \times 10^6$	$1{,}1 \times 10^1$	$1{,}72 \times 10^6$

primeras 7 filas y en la última fila se muestra la señal s_o resultante para el archivo de salida w_o.

Cada una de las gráficas de la figura 8.5 muestra todas las repeticiones de cada una de las señales de los archivos incluidos. La última fila muestra la mezcla resultante al sumar todas las señales incluidas. En los sonidos camión, viento, lluvia y taladro hay más de una repetición (3; 2; 2 y 3 respectivamente). En caso del sonido de taladro las ultimas dos repeticiones están solapadas pero en la mezcla esas dos repeticiones han sido sumadas.

Para simplificar la gráfica no se ha hecho coincidir cada fila de la figura 8.5 con cada fila de la tabla 8.1, sino que se han agrupado en una fila por archivo. Se debe notar que el algoritmo de mezcla no opera por archivo para evitar cargar demasiados datos en memoria. El algoritmo de mezcla divide el archivo de salida en cuadros y repite el recorrido a través de las filas de la tabla 8.1 para cada cuadro. Esto permite cargar en memoria solo las muestras que, una vez desplazadas, entran en un cuadro de tiempo del archivo de salida.

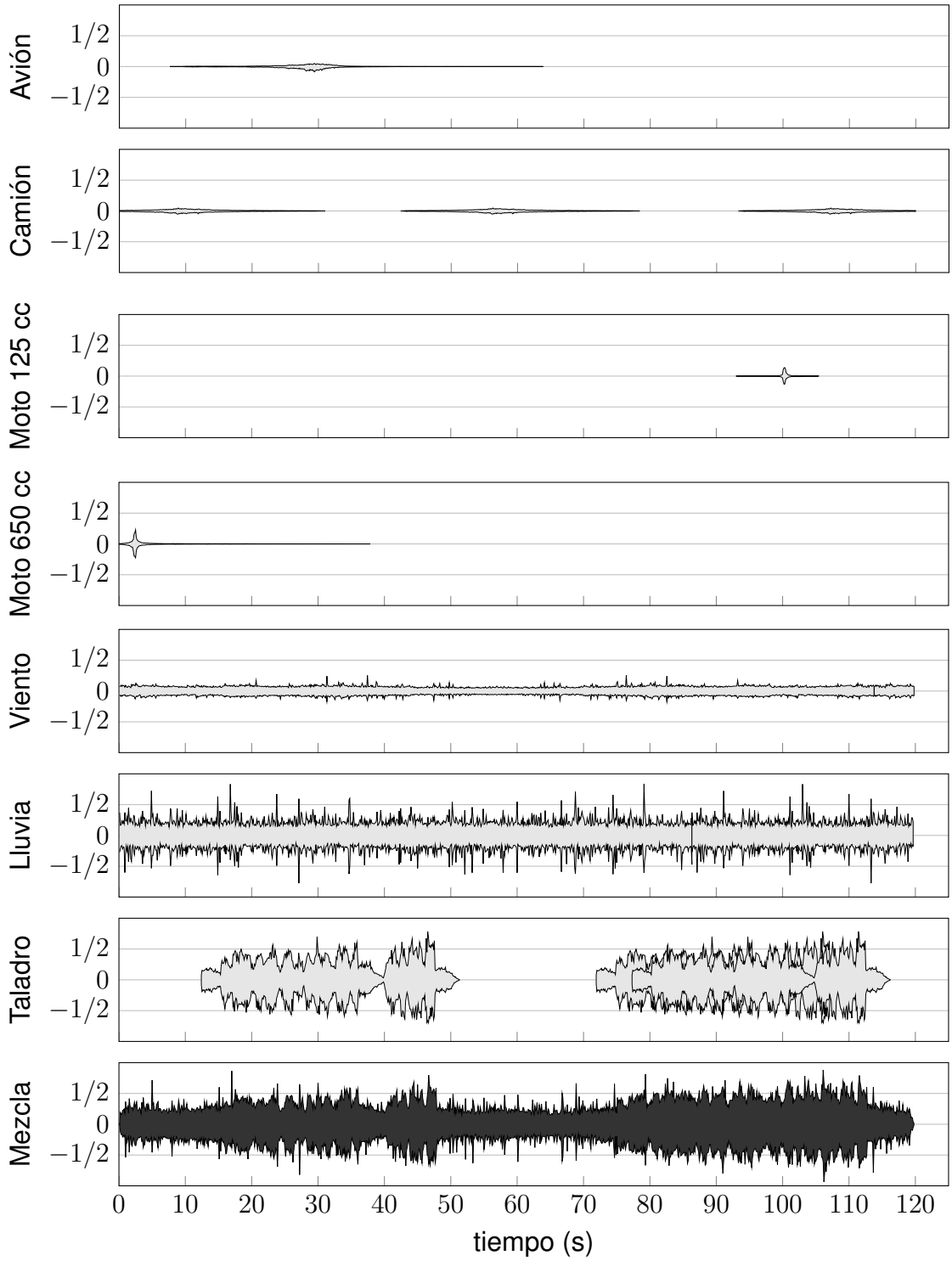


Figura 8.5: Mezcla de señales de audio $s_i(t)$ de la base y señal resultante $s_{\mathsf{o}}(t)$. Relleno gris claro: señales de la base afectadas por g_i y desplazados en $t_{\mathsf{s},i,o}$. Relleno gris oscuro: señal mezclada de salida

Capítulo 9

Análisis de discrepancias

Como es de esperar, la solución del problema de combinación no es necesariamente exacta. Sin embargo se pueden admitir pequeñas diferencias entre los valores requeridos por el usuario y los valores que se obtienen al aplicar la solución que minimiza justamente estas diferencias. Por otra parte, también se observarán ciertas diferencias entre la optimización, la señal del archivo de audio generado y la señal acústica medida en la posición de receptor debido a ciertas simplificaciones realizadas para poder formular el problema.

Si estas diferencias están acotadas por algún criterio, que dependerá del experimento para el cual se utilice la herramienta, el usuario puede admitirlas o en caso contrario, deberá tomar alguna medida. Un criterio habitual, en experimentos que involucran la percepción sonora, es el de las *diferencias apenas perceptibles* JNDs [Fastl y Zwicker, 2005]. El usuario podría admitir aquellas señales generadas cuyos parámetros difieran de los inicialmente requeridos si estas diferencias no superan las JNDs. En caso que las diferencias superen las JNDs se pueden tomar diversas decisiones que van desde incorporar nuevos archivos a la base hasta planificar nuevamente el experimento.

Por esta razón el módulo de Análisis (ver figura 2.5 y sección 2.5) ofrece al usuario la visualización de los valores de los parámetros alcanzados en las distintas etapas de la composición controlada en contraste con los datos requeridos. Es decir, la visualización de estos datos y el criterio adoptado por el usuario le permitirá evaluar si acepta la señal generada cómo válida o toma alguna medida para obtener una señal con mayor validez.

9.1 Diferencias en Espectro

En esta sección se detalla cómo se obtienen las cuatro versiones del espectro del archivo de salida introducidas brevemente en la sección 2.5.

La primera versión corresponde en realidad a los valores que el usuario especifica y que debería satisfacer el archivo de salida. Estos valores se simbolizan en este texto como $e_{m,\mathsf{u}}$, para la m-ésima banda con $m \in \{1, 2, \cdots, M\}$. Sin embargo es de esperar diferencias de distinto tipo entre la especificación del usuario y el valor que puede ser finalmente medido en la posición donde estaría el sujeto del experimento.

La segunda versión corresponde a los valores alcanzados al resolver el problema de optimización formulado en la ecuación (8.21). Según se introdujo en la sección 8.2, no es posible asegurar una solución exacta para cualquier instancia del problema. Sea $x_n^\star \in \mathbb{R}_0^+$ el valor óptimo del coeficiente cuadrático de ganancia del n-ésimo archivo hallado al resolver el problema formulado en la ecuación (8.21). Entonces el espectro resultante $\hat{e}_{m,\mathsf{u}_2}$, donde el subíndice u_2 corresponde a la segunda versión del espectro, se calcula según

$$\hat{e}_{m,\mathsf{u}_2} = x_1^\star e_{m,1} + x_2^\star e_{m,2} + \cdots + x_N^\star e_{m,N} \tag{9.1}$$

donde $e_{m,n}$ es la exposición sonora en la banda m del archivo n.

La tercera versión es el espectro calculado desde el archivo de audio de salida con las correcciones necesarias para que pueda ser comparable con los datos que especifica el usuario. Existen diferentes formas de implementar este cálculo. Una de ellas, la utilizada en esta tesis, es la mostrada previamente en la tercera columna de la figura 2.5. En esta versión se calcula $\hat{e}_{m,\mathsf{u}_3}$, donde el subíndice u_3 corresponde a la tercera versión del espectro, aplicando el análisis espectral detallado en el capítulo 4 al archivo de audio de salida obtenido del bloque *combinador archivo de audio* del módulo *combinador* (ver figura 2.3) y afectando los datos espectrales por h_m de manera similar a la ecuación (8.3). Sea $e_{m,\mathsf{s}}$ la exposición sonora para la m-ésima banda de frecuencias, obtenida mediante el análisis espectral detallado en el capítulo 4 y la ecuación (8.6) aplicados al archivo de audio de salida. Entonces se obtiene la tercera versión del espectro $\hat{e}_{m,\mathsf{u}_3}$ según

$$\hat{e}_{m,\mathsf{u}_3} = h_{\mathsf{o},m}^2 e_{m,\mathsf{s}} \tag{9.2}$$

Otra forma, similar a la tercera versión, es invertir el orden de los bloques *análisis espectral* y *corrección sistema reproducción* en la tercera columna del módulo *análisis* (ver figura 2.5). Otras versiones posibles surgen en los casos en los cuáles existe un sistema simulado con una respuesta determinada pero no se analizan en esta tesis. En esos casos, sobre todo si se tienen en cuenta características de la función de transferencia de la cabeza y el torso del sujeto, es recomendable generar un modelo para el sujeto y otro modelo para un micrófono de medición.

La cuarta versión es el espectro medido con un sonómetro en la posición donde estaría el sujeto pero en ausencia del mismo. Se debe utilizar un micrófono de campo libre o campo difuso según cuál sea el campo sonoro del experimento. En el caso de esta tesis se utiliza un sonómetro clase I con micrófono de campo sonoro difuso (en realidad el micrófono es de campo libre pero el mismo sonómetro corrige digitalmente su respuesta para ser empleado en campo difuso). Esta versión del espectro (subíndice u_4) corresponde entonces al espectro de exposición sonora medido $\hat{e}_{m,\mathrm{u}_4}$ con periodo de integración T_{u} correspondiente a la duración del archivo de audio.

Esta cuarta versión del espectro, medido en ausencia de sujeto, no será tan simple de medir en el caso que se utilicen auriculares. Una forma es utilizar técnicas para convertir datos medidos con sondas microfónicas insertadas en los oídos [ISO 11904-1, 2002] o datos medidos con un maniquí con micrófonos en el lugar del tímpano [ISO 11904-2, 2004] a datos en campo sonoro difuso o libre según sea el caso del experimento.

9.2 Diferencias en histograma de duraciones

En esta sección se detalla cómo se obtienen las dos versiones del histograma de duraciones del archivo de salida introducidas brevemente en la sección 2.5.

La primera versión corresponde nuevamente a los valores que el usuario especifica que debería cumplir el archivo de salida. Estos valores son simbolizados con $d_{l,\mathrm{u}}$, para el l-ésimo intervalo de duración con $l \in \{1, 2, \cdots, L\}$. También en este caso es de esperar diferencias de distinto tipo entre la especificación del usuario y el valor que puede obtenerse finalmente en la posición donde estaría el sujeto del experimento.

La segunda versión también es similar al caso del espectro en cuanto corresponde a los valores alcanzados al resolver el problema de optimización formulado en la

ecuación (8.21). Tampoco es posible asegurar que la solución sea exacta para estas variables. Sea $y_n^\star \in \mathbb{N}_0$ el valor óptimo del coeficiente repeticiones hallado al resolver el problema formulado en la ecuación (8.21). Entonces la cantidad de eventos $\hat{d}_{l,\mathrm{u}_2}$ que aparecerán en el archivo de salida y cuya duración cae dentro del intervalo l, donde el subíndice u_2 corresponde a la segunda versión del histograma, se calcula según

$$\hat{d}_{l,\mathrm{u}_2} = y_1^\star d_{l,1} + y_2^\star d_{l,2} + \cdots + y_N^\star d_{l,N} \tag{9.3}$$

Dada la forma en la que se definen los eventos sonoros al calcular su duración, no será posible identificar ni los eventos ni la duración con el mismo algoritmo detallado en 5.1.1. Tampoco sería posible utilizando la definición de duración subjetiva de Fastl [Fastl, 1977].

Un interesante punto de partida para estimar la duración de eventos, y para detectar eventos sonoros, en un archivo de audio que contiene una mezcla de eventos sonoros de diverso tipo es a través de los modelos de saliencia auditiva, y la aplicación de máscaras [De Coensel y Botteldooren, 2010]. Este enfoque es compatible en cuanto se debe contar, además de la mezcla de eventos, con archivos de audio de los eventos sonoros por separado. La dificultad en este caso viene dada por la imposibilidad de formular el problema de combinatoria en términos de programación lineal, razón por la cuál no se implementó en esta tesis.

Capítulo 10

Prueba piloto

Con la finalidad de poner a prueba la herramienta se realizó un estudio piloto incluyendo una breve encuesta sobre aspectos relacionados con el realismo de los escenarios propuestos en las pruebas. El estudio busca identificar el efecto de factores espectrales y temporales del ruido en la molestia causada por ruido durante actividades de tiempo libre, específicamente durante lectura como pasatiempo. El estudio sigue los lineamientos de las normas Nordtest 111 [2002] e ISO 15666 [2003]. El estudio se realiza utilizando técnicas de diseño estadístico de experimentos. Particularmente se utiliza un diseño factorial fraccionario de resolución III con 4 factores a 2 niveles [Box et al., 1988].

En la actualidad no existe una encuesta estandarizada para estudiar el realismo de escenarios virtuales. Existe, en el marco de escenarios de realidad virtual, un debate abierto sobre métricas del realismo, la relación con la sensación de presencia, las posibilidades de interacción con el ambiente y otras escalas incluyendo subescalas de las ya citadas [Witmer y Singer, 1998]. El cuestionario de realismo desarrollado en esta tesis representa en si un acercamiento inicial para escenarios mixtos, es decir escenarios virtuales con componentes del mundo real, en los cuales la parte virtual solo corresponde al ambiente sonoro (ver 11.1).

10.1 Método

En esta sección se describen los métodos empleados para la configuración de la prueba incluyendo diseño del experimento, estímulos, equipamiento, configuración, participantes y procedimientos. Estos métodos se han agrupado en cuatro subsecciones.

En 10.1.1 se describen las particularidades de la sala y demás materiales utilizados, en 10.1.2 se describe la metodología empleada en la selección de participantes y aspectos de distribución de grupos según diversas características, en 10.1.3 se describen los estímulos usados en función del diseño del experimento y en 10.1.4 se describe la forma en la cual se condujeron las encuestas incluyendo la introducción y la conclusión de cada sesión con los participantes.

10.1.1 Materiales

Las pruebas se condujeron en una habitación silenciosa ($L_{\text{A,eq}} = 28{,}2$ dB) cuyo nivel sonoro por bandas de frecuencia de 1/1 octava se reporta en la tabla 10.1. La habitación está dentro de un departamento en un edificio residencial, creando una situación con mayor validez ecológica en comparación con haber simulado la habitación en un verdadero laboratorio o en el ambiente del campus universitario.

Tabla 10.1: Ruido de fondo por bandas de 1/1 octava

f_c (Hz)	31,5	63	125	250	500	1 000	2 000	4 000	8 000	16 000
L_{eq} (dB)	42	47,1	34,5	29,7	28,4	18,4	13,6	12,6	12,8	12,5

El ruido de fondo, reportado en la tabla 10.1, corresponde al nivel sonoro continuo equivalente sin ponderación en frecuencia medido en dos puntos de la sala (ver figura 10.1) durante un periodo de 30 minutos en cada punto en días y horarios similares a los destinados a las pruebas con sujetos. Para esta medición se utilizó un sonómetro clase 1 calibrado en laboratorio en el mismo año y en campo antes y después de la medición.

El suelo de la habitación está revestido por una alfombra, en la misma pared de la puerta de entrada tiene un armario de aproximadamente 2 metros de altura por 2 metros de largo (ver figura 10.2), en la superficie opuesta al armario hay un ventanal y sus paredes y techo están terminadas en revoque fino y pintado. Aproximadamente en el centro de la habitación se colocaron cuatro sillas al rededor de una mesa sobre la cual se presenta material de lectura (ver figuras 10.2 y 10.3). En una esquina de la habitación, se ubicó el encuestador.

El material de lectura incluía periódicos del día, revistas del mes sobre temas de interés general tanto para hombres como para mujeres cómo así también para diversos

(a) Posición A

(b) Posición B

Figura 10.1: Medición

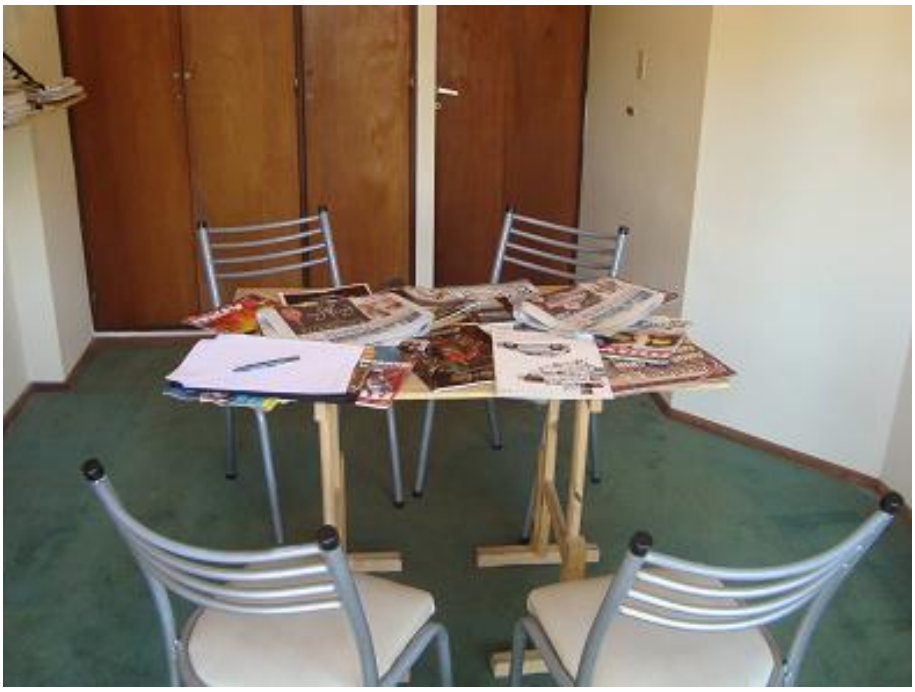


Figura 10.2: Sala

grupos etarios, libros de cuentos, novelas, etc. También se presentaron libros de pasatiempos como crucigramas, juegos de memoria y otros juegos similares pero ningún participante los eligió.

El sistema de reproducción sonora empleado es monofónico. Desde un ordenador portátil actual (Banghó doble núcleo de 2 GHz y 1 GB de memoria RAM con sistema operativo Windows XP service pack III), con una interfaz de sonido marca M-Audio modelo Mobile Pre se envía el audio a un parlante marca Alesis, modelo M1 Active MKII. El parlante está ubicado en el espacio contiguo, según se muestra en la foto de la figura 10.3, detrás del ventanal, a una distancia de un metro y orientado hacia el

Figura 10.3: Ubicación del parlante

ventanal. Durante las pruebas el parlante no es visible desde el lugar asignado a los participantes (ver figura 10.1).

Esta configuración permite que el ensayo sea realizado en grupos de hasta 4 personas. El tiempo de reverberación de la sala se estima en $t_{R,\text{mid}} = 0{,}67$ s (estimado a partir de T_{30} para la banda de octava de $f_c = 500$ Hz). Se midió el tiempo de reverberación según la norma IRAM 4109-2 [2011] equivalente a ISO 3382-2 [2008], según el método de la respuesta al impulso. La respuesta al impulso se midió según el método del barrido sinusoidal de la norma ISO 18233 [2006]. En la tabla 10.2 se reportan los resultados promedio.

Tabla 10.2: Estimación del Tiempo de Reverberación

f_c (Hz)	31,5	63	125	250	500	1 000	2 000	4 000
T_{10} (s)	1,04	1,07	1,31	0,75	0,72	0,47	0,59	0,55
$r_{T_{10}}$	0,97	0,94	0,98	0,99	0,99	0,99	1,00	1,00
T_{20} (s)	1,04	0,97	1,17	0,77	0,77	0,65	0,71	–
$r_{T_{20}}$	0,97	0,97	0,99	1,00	0,98	0,99	0,99	–
T_{30} (s)	–	–	1,17	0,86	0,67	0,67	–	–
$r_{T_{30}}$	–	–	1,00	0,99	0,99	0,99	–	–

La tabla muestra las estimaciones de tiempo de reverberación T_{10}, T_{20} y T_{30} [IRAM 4109-2, 2011] y los coeficientes de correlación $r_{T_{10}}$, $r_{T_{20}}$ y $r_{T_{30}}$ [ISO 18233, 2006], para cada banda de frecuencias. En la estimación del tiempo de reverberación se han aceptado sólo los datos cuyos coeficientes de correlación son superiores a $0{,}96$ salvo una excepción con $r_{T_{10}} = 0{,}94$ para T_{10} en la banda centrada en $f_c = 63$ Hz debido a que su valor es semejante a los valores de T_{20}.

La respuesta al impulso, para la posición de receptor A (ver figura 10.1.a) y de parlante según figura 10.3 pero con la ventana y la persiana cerradas, se muestra en la figura 10.4. La respuesta ha sido submuestreada por propósitos gráficos. El tiempo de reverberación se midió en las posiciones de las cuatro sillas y en las posiciones A y B para la posición de parlante de la figura 10.3.

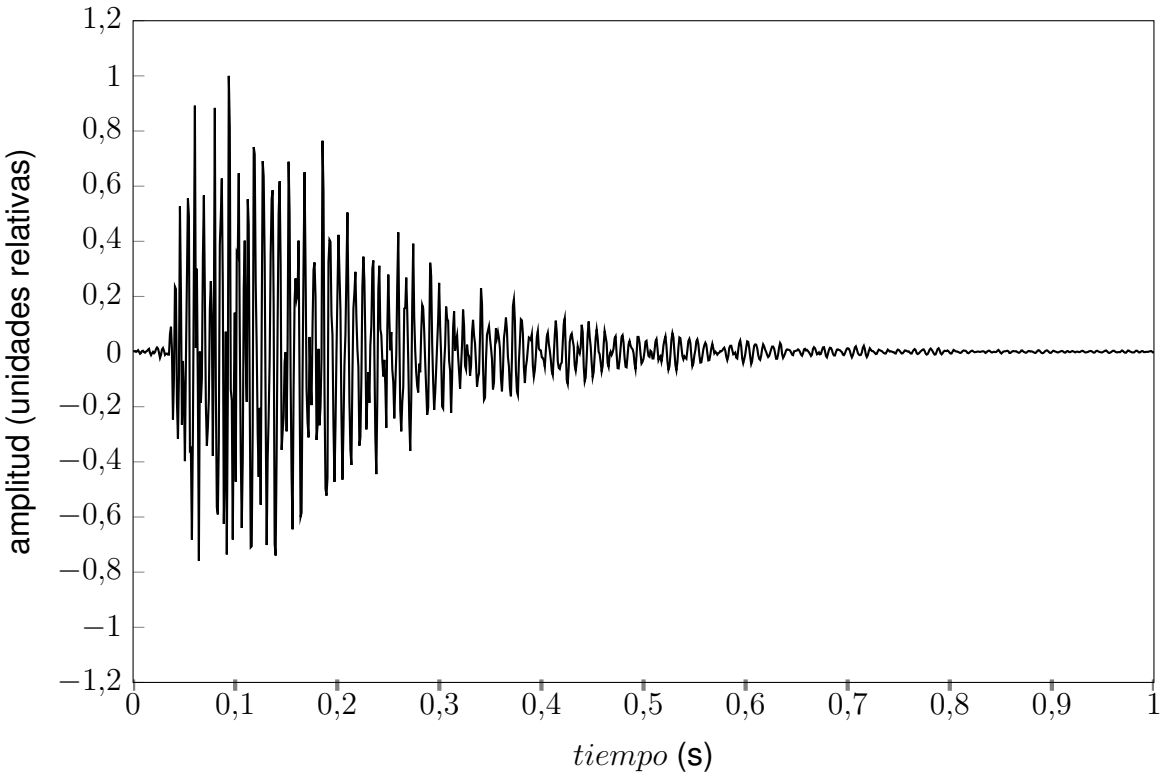


Figura 10.4: Respuesta al impulso entre parlante y posición A de micrófono según figura 10.1 (submuestreada a $F_{\mathrm{S},2} = 1\ 024$ Hz).

10.1.2 Participantes

Se seleccionaron 18 participantes de un grupo de 30 personas que respondieron al llamado inicial realizado a través de listas de correo electrónico. El criterio de selección

fue que los participantes cumplieran con ciertos requisitos, en base a sus respuestas escritas a un breve formulario adjunto al llamado. Los requisitos fueron, no tener dificultades para oír y no tener el hábito de exponerse durante más de 16 horas por semana a sonidos de nivel muy elevado. Respecto al nivel sonoro los interesados podían responder en los siguientes 5 segmentos: 1) sonidos moderados como la voz humana, 2) sonidos un poco más elevados que la voz humana, 3) sonidos elevados, 4) sonidos muy elevados y 5) sonidos de nivel extremadamente elevados (ejemplo: una industria ruidosa). Respecto al tiempo de exposición semanal los intervalos eran: 1) menos de 4 horas, 2) entre 4 horas y 16 horas, 3) entre 16 horas y 35 horas, 4) entre 35 horas y 50 horas y 5) más de 50 horas. El último criterio fue una selección al azar entre los interesados que cumplían los requisitos anteriores.

La muestra, compuesta por los participantes seleccionados, tiene una edad promedio de 30 años, entre 19 y 45 años salvo una participante de 64 años. El 67 % de los participantes eran de género masculino y el 33 % restante femenino. El 67 % de la muestra son estudiantes (de grado y posgrado), el 22 % son docentes y la ocupación del 11 % restante se distribuye entre trabajadores autónomos o en relación de dependencia en sus respectivas profesiones u oficios.

Las pruebas se realizaron en 6 grupos de participantes. Los horarios para cada grupo se acordaban con cada participante dentro de las opciones posibles que correspondían a uno de tres días consecutivos en dos franjas horarias posibles (de 15 hs a 16 hs o de 17 hs a 18 hs). Se intentó dividir en 6 grupos de 3 participantes cada grupo pero, debido a otros compromisos y obligaciones de los participantes, finalmente se conformaron 4 grupos de 3 participantes, un grupo de 2 participantes y un último grupo de 4 participantes.

Se ofreció una compensación económica por la participación que fue aceptada por todos excepto dos sujetos.

10.1.3 Estímulos

Este estudio piloto busca una primera aproximación a la respuesta *molestia por ruido en actividades de tiempo libre* frente a los factores *patrón temporal* y *contenido espectral* del ruido ambiente. A modo de estudio preliminar se estudian los siguientes factores:

1. Nivel sonoro equivalente ponderado A $L_{\text{A,eq}}$ (dBA)

2. Pendiente del contenido espectral $\Delta L_{\text{eq}}/\Delta m$ (dB/octava)

3. Total de eventos q_e (eventos)

4. Pendiente del histograma de duración de eventos $\Delta q_{\text{e}}/\Delta l$ (eventos/intervalo)

En este estudio se toman dos niveles para cada factor. Estos niveles se resumen en la tabla 10.3.

Tabla 10.3: Factores y niveles estudiados

Factor	niveles	
	-	+
$L_{\text{A,eq}}$ (dBA)	45	55
$\Delta L_{\text{eq}}/\Delta m$ (dB/octava)[1]	-5.5	-7.5
q_{e} (eventos)	15	30
$\Delta q_{\text{e}}/\Delta l$ (eventos/intervalo)	-1	1

En la tabla 10.4 se muestra el nivel al que se usa cada variable en cada prueba utilizando el símbolo "$-$" para el nivel bajo y el símbolo "$+$" para el nivel alto.

Tabla 10.4: Pruebas del experimento

Prueba	$\Delta L_{\text{eq}}/\Delta m$	$L_{\text{A,eq}}$	$\Delta q_e/\Delta l$	q_e
1	$-$	$-$	$-$	$-$
2	$-$	$-$	$+$	$+$
3	$-$	$+$	$-$	$+$
4	$-$	$+$	$+$	$-$
5	$+$	$-$	$-$	$+$
6	$+$	$-$	$+$	$-$
7	$+$	$+$	$-$	$-$
8	$+$	$+$	$+$	$+$

Este tipo de experimentos, factorial fraccionario de resolución III[2], permite estudiar las efectos principales de los factores sin confundir sus efectos entre si. Por otra parte no sucede lo mismo con las interacciones de dos o más factores. Confunde las interacciones dobles entre si. En particular se utilizan las tres primeras variables de la tabla 10.4 en un diseño factorial completo y se genera los niveles correspondientes de la cuarta variable multiplicando los niveles de las tres primeras. Esto implica que las interacciones triples del diseño estarán confundidas con las simples pero, es de esperar que el efecto de las interacciones triples no sea tan notorio como el de las interacciones simples[Box et al., 1988].

Luego, para expresar los niveles en los términos que son requeridos por la herramienta de combinación, se calcula el nivel de exposición $L_{\mathsf{e},m}$ para cada banda de frecuencia de octava para cada prueba según la ecuación

$$L_{\mathsf{e},m} = \frac{\Delta L_{\mathsf{eq}}}{\Delta m} m + L_{\mathsf{A,eq}} + 10\log\left(T_{\mathsf{u}}\right) - 10\log\left[\sum_{m}\left(10^{\frac{\frac{\Delta L_{\mathsf{eq}}}{\Delta m}m + A_m}{10}}\right)\right] \tag{10.1}$$

donde A_m es la corrección por ponderación A para la m-ésima banda (ver taba 10.5) y T_{u} es la duración del estímulo sonoro.

Tabla 10.5: Corrección por bandas de frecuencia 1/1 octava para obtener el nivel sonoro ponderado A

m	1	2	3	4	5	6
f_c (Hz)	125	250	500	1 000	2 000	4 000
A_m (db)	$-16{,}1$	$-8{,}6$	$-3{,}2$	$0{,}0$	$+1{,}2$	$+1{,}0$

La cantidad de eventos d_l del l-ésimo intervalo temporal según la ecuación

$$d_l = \frac{\Delta q_e}{\Delta l} l + \frac{q_e}{L} - \frac{1}{L}\sum_{l}\left(\frac{\Delta q_e}{\Delta l} l\right) \tag{10.2}$$

donde L es la cantidad total de intervalos temporales usados para describir el histograma de duración de eventos sonoros.

[2]La razón por la cuál se utilizó un diseño fraccionario y no uno completo fue reducir la cantidad de pruebas para evitar sesiones demasiado largas donde el cansancio de los participantes influya sus respuestas o múltiples sesiones para un mismo sujeto.

Para evitar clics al inicio y al final de cada una de las señales de prueba, se aplica una envolvente lineal de 100 ms de duración al inicio y al final de cada una de las señales.

En la figura 10.5 se muestra el espectro y el histograma de duración de cada prueba. Las figuras 10.5.a-10.5.h corresponden a los estímulos para las pruebas 1 a 8. Se muestran dos gráficos por cada prueba, el superior contiene el espectro y el inferior el histograma de duraciones. Dado que es más habitual el uso del nivel sonoro equivalente L_{eq} que el de nivel de exposición L_{e}, los gráficos de espectro se han convertido, en todos los casos, a nivel sonoro equivalente según

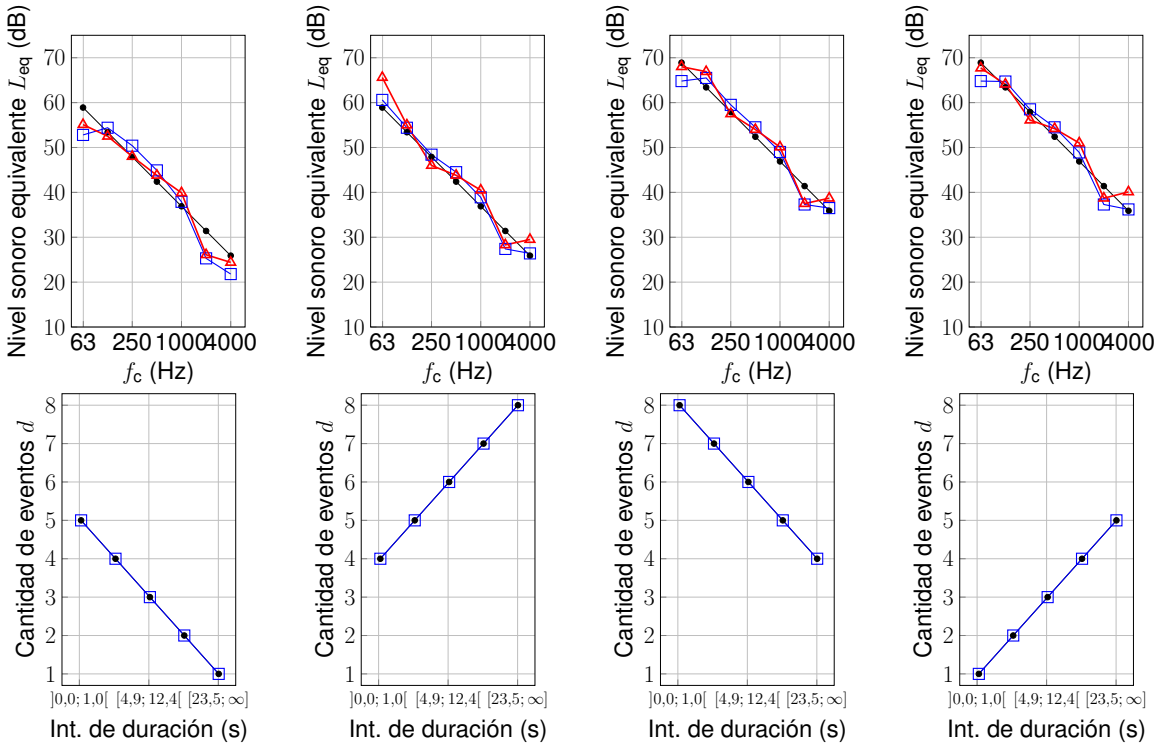
$$L_{\text{eq}} = L_{\text{e}} - 10\log(T_u) \tag{10.3}$$

siendo $T_u = 120$ s la duración total de cada estímulo. Con un círculo relleno se muestran los datos requeridos por el usuario, es decir, aquellos calculados con las ecuaciones 10.2 y 10.1 modificado según (10.3) para mostrar L_{eq}.

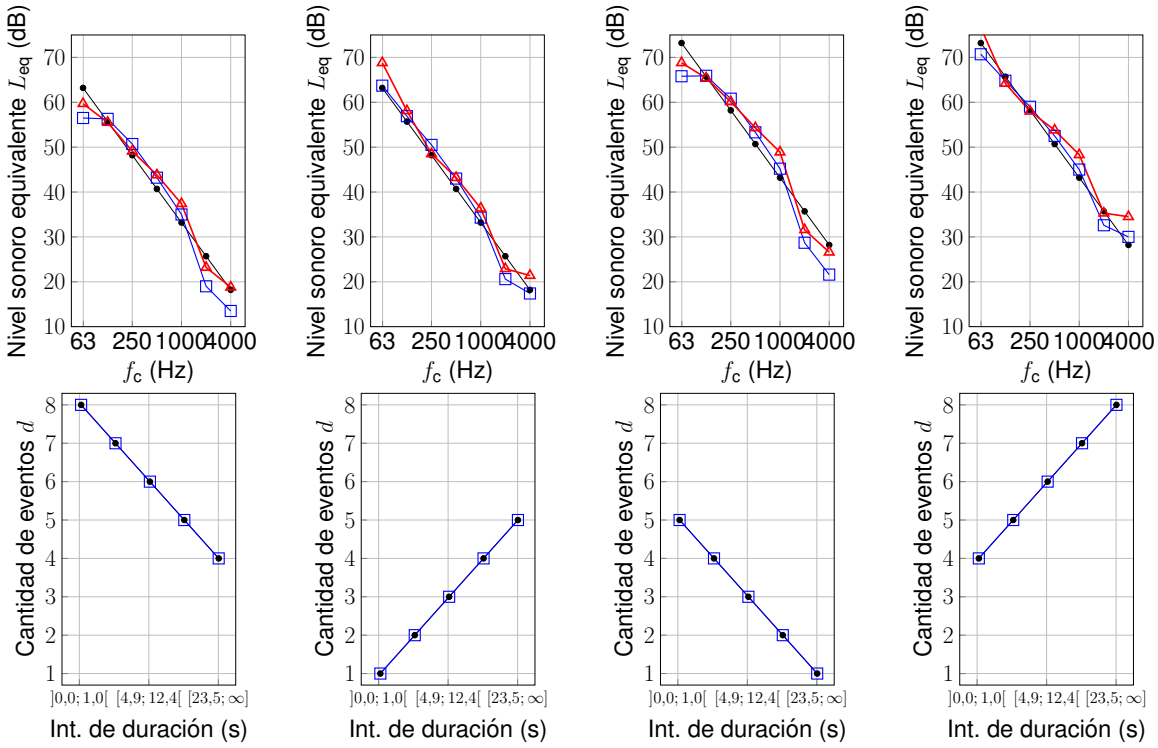
En el apéndice D se detallan los valores reportados en la figura 10.5 y una estima del error para cada prueba para cada banda y para cada intervalo temporal. En el apéndice E se detalla brevemente los sonidos que componen cada uno de los 8 estímulos.

En el caso de los espectros se muestran tres curvas. Con círculos rellenos el espectro requerido por el usuario, con cuadrados el finalmente alcanzado al resolver el problema de combinatoria y con triángulos el medido con un sonómetro clase I en la posición A. La diferencia absoluta entre los niveles equivalentes en bandas de octava especificados por el usuario respecto a los medidos, para las 8 pruebas, tiene un valor promedio de $2{,}6$ dB, una desviación estándar de $2{,}0$ dB y un valor máximo de $7{,}4$ dB. El valor promedio, para todas las pruebas, de la función de costo fue $\bar{v} = 3{,}3521$ o $L_{\bar{v}} = 5{,}3$ dB que es comparable con las diferencias de niveles sonoros equivalentes por banda.

En el caso de los histogramas de duración solo se muestran dos curvas. Con círculos rellenos se muestra el histograma requerido por el usuario y con cuadrados el finalmente alcanzado al resolver el problema de combinatoria (ver ecuación (8.21) y 8.8). Se observa que los estímulos generados alcanzan exactamente el histograma de duración requerido. Este resultado no necesariamente será siempre así, depende



(a) Prueba 1 **(b)** Prueba 2 **(c)** Prueba 3 **(d)** Prueba 4



(e) Prueba 5 **(f)** Prueba 6 **(g)** Prueba 7 **(h)** Prueba 8

Figura 10.5: Espectro e histograma de duración de los estímulos

fuertemente de la base que se utilice y del ancho de los intervalos de duración usados para los histogramas de duración.

10.1.4 Procedimiento

A cada participante del grupo se les solicitó tomar algún material de lectura entre tanto llegaban los demás participantes. Una vez que llegaba el último participante del grupo se lo dejaba escoger algún material de lectura y luego de aproximadamente 2 minutos se les entregaba la encuesta (ver Apéndice C). De forma seguida el encuestador leía la introducción de la encuesta explicitando la organización de las pruebas, el propósito del experimento (proporcionar datos experimentales para esta tesis), las condiciones de seguridad, bioética y protección de datos, su derecho a dar por concluida la prueba en cualquier momento que lo creyeran oportuno y la duración aproximada del ensayo completo incluyendo las pausas, los formularios de molestia y de realismo. Una vez terminado el ensayo los sujetos recibían la remuneración y podían llenar un formulario de sensibilidad al ruido de manera voluntaria vía formularios web. El test de sensibilidad es el test de Weinstein [Weinstein, 1978, 1980] traducido al español [Miyara, 2013b] en el Apéndice B.

Los 8 estímulos se presentaron en orden aleatorio, siguiendo un vector de índices de orden, distinto para cada uno de los 6 grupos de participantes. Los 6 vectores se inicializaron con una semilla distinta del algoritmo de generación aleatoria de índices. El encuestador contaba con un control remoto que le permitió iniciar las pruebas sin tener que salir de la habitación. Durante la presentación de cada estímulo sonoros la mayoría de los participantes continuaba leyendo el mismo material y eventualmente cambiaban a otro material de lectura distinto de manera silenciosa. Después de cada estímulo el encuestador esperaba a que todos los participantes marcaran la respuesta a las dos preguntas (ver C.1) y luego presentaba el siguiente estímulo sonoro. Cada estímulo dura exactamente 2 minutos y el tiempo de pausa para responder no excedió los 30 segundos en ningún caso (esta cota responde a una observación no a una imposición). El tiempo total del ensayo, incluyendo la introducción, las preguntas sobre molestia, y las preguntas sobre realismo y cualquier debate posterior a las pruebas fue de 45 minutos en promedio.

De acuerdo con la norma Nordtest 111 [2002], en condiciones de laboratorio se estudia el potencial de molestia causada por ruido. Es decir, dado que el escenario es generado virtualmente, es de esperar que los resultados de este ensayo representen

algunos aspectos de la realidad pero la molestia causada por ruido es un fenómeno complejo que puede depender de algún aspecto de la realidad que no haya sido reproducido en la condición de laboratorio. Para cada estímulo contestaron dos preguntas sobre molestia.

La primera es *"¿Cuánto le molestó o perturbó el ruido de este ambiente?"* y la respuesta puede tomar una de las 5 opciones siguientes

- NADA
- LIGERAMENTE
- MEDIANAMENTE
- MUCHO
- EXTREMADAMENTE

La segunda es *"¿En qué grado le molestó?"* y la respuesta se marca en una escala numérica de 11 segmentos con referencias verbales en los extremos. Las referencias verbales son NADA para el segmento 0 y EXTREMADAMENTE para el segmento 10.

Las palabras usadas tanto para las preguntas como para las opciones de respuesta responden a las recomendaciones de la norma ISO 15666 [2003] con una pequeña modificación para adaptarlas al contexto regional (por ejemplo la palabra "cuantía" no es de uso habitual en Rosario por lo tanto se reemplazo por "grado").

Las pruebas son de diferencial semántico en tanto las preguntas no buscan una comparación entre estímulos sonoros sino palabras relacionadas con las características del sonido [Guski, 1997]. La primera pregunta además no supone una escala y por lo tanto no son válidas las operaciones para obtener los momentos estadísticos de la variable como si lo es en el caso de la segunda pregunta pues la respuesta se da en una escala numérica.

Una vez finalizada la presentación de los estímulos se le solicitaba a los sujetos llenar el formulario de realismo del Apéndice C.2). Las preguntas son generales sobre las 8 pruebas y cada pregunta admite respuestas en escala de Likert de 5 puntos.

Una vez que todos los participantes del grupo completaban el formulario se les agradecía y se les entregaba la remuneración. Algunos participantes en esta instancia solicitaban más detalles sobre el experimento y solo en este momento (después

de realizadas las pruebas) el encuestador les mostró la configuración del sistema de reproducción sonora. Todos los sujetos que vieron la configuración real manifestaron espontáneamente no haber imaginado tal configuración durante el ensayo. Una posible explicación es que el sonido no solo se propagaba por el camino directo a través del ventanal sino también a través de las vías laterales como el techo, el suelo y las paredes además del efecto envolvente generado por la propia reverberación de la sala.

10.2 Resultados

En las figuras 10.6.a-10.6.h se muestran los resultados obtenidos para la primera pregunta para las pruebas 1 a 8 respectivamente. Los resultados se expresan en términos de porcentaje de participantes que respondieron en cada opción (todos los sujetos respondieron en alguna y solo una de las opciones en cada prueba tal como se les indicó).

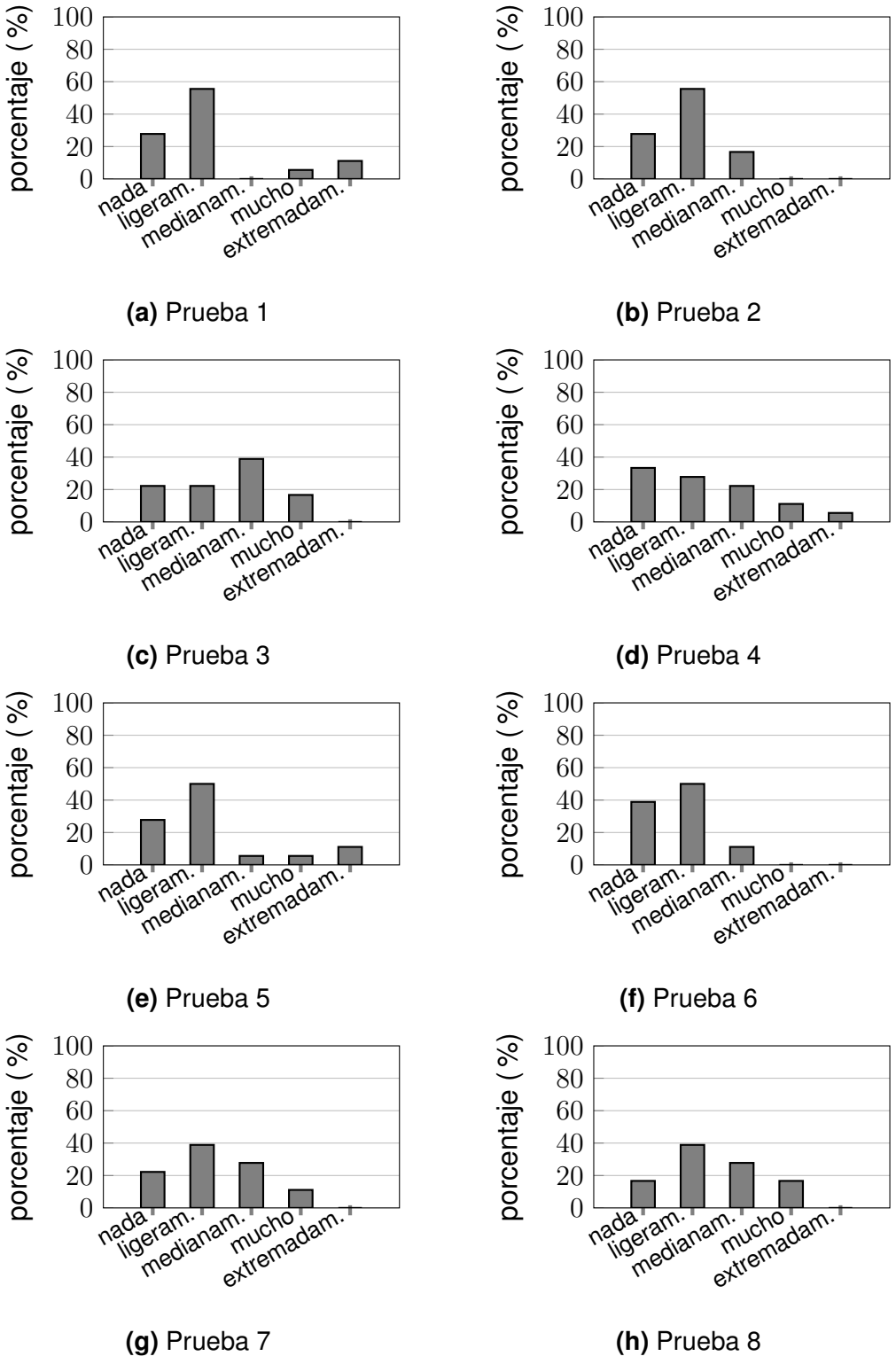
En la tabla 10.6 se muestra el análisis de la varianza de la segunda pregunta del experimento. Como es de esperar el nivel equivalente ponderado A $L_{\text{A,eq}}$ es significativo (para un nivel de significación $\alpha = 0{,}05$). De particular interés resulta que el efecto de la cantidad total de eventos q_e sea también significativo ($\alpha = 0{,}05$).

Tabla 10.6: Análisis de varianza

	Suma cuad.	G. Libertad	Cuad. Medios	F-Fisher	valor p
$L_{\text{A,eq}}$	$148{,}0278$	1	$148{,}0278$	$4{,}4 \times 10^{1}$	$5{,}6 \times 10^{-10}$
$\Delta L_{\text{eq}}/\Delta m$	$0{,}0278$	1	$0{,}0278$	$8{,}3 \times 10^{-3}$	$9{,}9 \times 10^{-1}$
q_{e}	$13{,}4444$	1	$13{,}4444$	$4{,}0 \times 10^{0}$	$4{,}6 \times 10^{-2}$
$\Delta q_{\text{e}}/\Delta l$	$1{,}0000$	1	$1{,}0000$	$3{,}0 \times 10^{-1}$	$5{,}8 \times 10^{-1}$
Error	$462{,}8056$	139	$3{,}3295$	$1{,}0 \times 10^{0}$	$5{,}0 \times 10^{-1}$

En la tabla 10.7 se muestra una estimación del tamaño de los efectos. Se observa que los factores con efectos significativos son positivos, es decir, la molestia crece si crecen $L_{\text{A,eq}}$ o q_{e}.

Los efectos de las pendientes del contenido espectral y del histograma de duración no son significativos ($\alpha = 0{,}05$). Sin embargo no sería correcto descartar estos parámetros en tanto los valores-p tampoco son muy elevados y podría deberse al error



(a) Prueba 1 **(b)** Prueba 2

(c) Prueba 3 **(d)** Prueba 4

(e) Prueba 5 **(f)** Prueba 6

(g) Prueba 7 **(h)** Prueba 8

Figura 10.6: Respuestas sobre molestia en 5 categorías verbales

Tabla 10.7: Tamaño del efecto

Efecto	Tamaño
Intersección	$2{,}93$
$L_{\text{A,eq}}$	$1{,}01$
$\Delta L_{\text{eq}}/\Delta m$	$0{,}01$
q_{e}	$0{,}31$
$\Delta q_{\text{e}}/\Delta l$	$0{,}08$

o bien a la elección de niveles no muy distantes para cada uno de estos factores.

En las figuras 10.7.a-10.7.h se reportan los resultados de las preguntas de realismo (ver C.2). Los resultados se expresan en barras que representan el porcentaje de participantes que respondieron en cada opción. Las líneas de segmentos representan el valor promedio, respecto a la respuesta de cada participante, para cada una de las preguntas.

No se pudo rechazar la hipótesis nula mediante la prueba t de Student ($\alpha = 0{,}05$), para ninguno de los valores medios reportados con lineas de segmentos en la figura 10.7 correspondientes a las 8 preguntas. Sin embargo los valores medios de las respuestas son en general favorables al realismo de los escenarios propuestos.

Idealmente, para una validación subjetiva de la herramienta, se debería separar los efectos sobre el realismo causados por la herramienta respecto a los demás sistemas involucrados. En la práctica no es posible separar de la herramienta en sí el efecto del sistema de auralización, incluyendo la parte electroacústica y la acústica en si. Tampoco es posible hacer una evaluación que asegure independencia de los valores que se exijan a cada parámetro, por ejemplo, para cualquier espectro sonoro de salida o cualquier pendiente del histograma de duración. Solo se puede validar el conjunto de sistemas que incluyen a la herramienta y a las pruebas en si.

En la figura 10.7.a se muestran las respuestas a la pregunta *¿Fueron realistas los 8 ambientes propuestos?*. La respuesta promedio (linea de segmentos) es $65\,\%$ del extremo derecho, es decir, “sí, fueron realistas”. Por supuesto que al superar el $50\,\%$ ya se trata de un valor favorable al realismo.

En la figura 10.7.b se muestran las respuestas a la pregunta *¿Desde donde percibía que provenían los sonidos?*. La respuesta promedio (linea de segmentos) es

$35\,\%$ del extremo derecho, es decir, “dentro”. La intención de la prueba era simular que las fuentes sonras estaban fuera de la habitación. Por lo tanto, que la respuesta se encuentre debajo del $50\,\%$ de estar “dentro”, es favorable al realismo.

En la figura 10.7.c se muestran las respuestas a la pregunta *¿Le costó mucho esfuerzo imaginar la situación propuesta?*. La respuesta promedio es $23\,\%$ del extremo derecho, es decir, “sí, totalmente”. La respuesta se encuentra por debajo del $50\,\%$ de “sí, totalmente” por lo tanto es favorable al realismo o bien más cercana a no haber costado mucho esfuerzo imaginar la situación propuesta.

En la figura 10.7.d se muestran las respuestas a la pregunta *¿Qué tan rápido se adaptó al ambiente propuesto?*. La respuesta promedio (linea de segmentos) es $35\,\%$ del extremo derecho, es decir, “al finalizar”. La respuesta se encuentra por debajo del $50\,\%$ de adaptarse “al finalizar” por lo tanto es favorable al realismo o bien más cercano a haberse adaptado al iniciar la prueba.

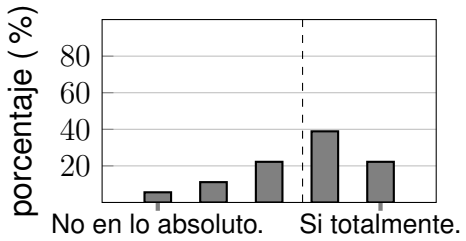
En la figura 10.7.e se muestran las respuestas a la pregunta *¿Qué tan lejos estaban las fuentes sonoras?*. La respuesta promedio (linea de segmentos) es $50\,\%$ del extremo derecho, es decir, “lejos/no visibles”. La respuesta se encuentra exactamente a $50\,\%$ lo cuál puede deberse a varias razones pero en principio es contrario a la pregunta de la figura 10.7.b que es similar. La pregunta 10.7.b no hace referencia directa a la distancia pero al especificar los extremos dentro y fuera es de esperar que si la fuente estaba dentro de la sala fuese más cercana que si estuviese fuera de la sala (en cuyo caso estaría lejos). Sin embargo puede haber ocurrido otra interpretación por parte de los participantes.

En la figura 10.7.f se muestran las respuestas a la pregunta *¿Qué tan desorientado se sintió en las pausas al inicio y al final de cada prueba?*. La respuesta promedio (linea de segmentos) es $23\,\%$ del extremo derecho, es decir, “las que más influyeron”. La respuesta se encuentra por debajo del $50\,\%$ de haber sido “las que más influyron” por lo tanto es favorable al realismo en tanto una posible desorientación al inicio y al fin de cada prueba no influyó en gran medida sobre las respuestas.

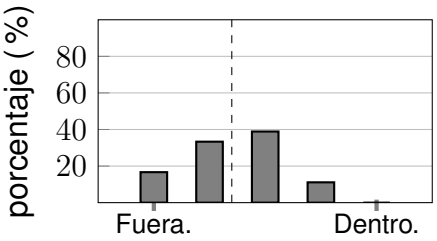
En la figura 10.7.g se muestran las respuestas a la pregunta *¿Qué tanto influyeron sus respuestas los eventos de la realidad?*. La respuesta promedio (linea de segmentos) es $18\,\%$ del extremo derecho, es decir, “las que más influyeron”. La respuesta se encuentra por debajo del $50\,\%$ de haber sido *las que más influyron* por lo tanto es favo-

rable al realismo en tanto una posible desorientación debida a eventos de la realidad no influyó en gran medida sobre las respuestas.

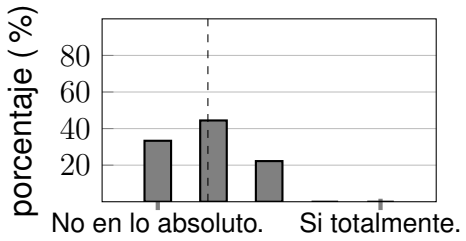
En la figura 10.7.h se muestran las respuestas a la pregunta *¿Qué tan similar fue la prueba comparada con sus experiencias de la vida real?*. La respuesta promedio (linea de segmentos) es $40\,\%$ del extremo derecho, es decir, “totalmente diferente”. La respuesta se encuentra por debajo del $50\,\%$ de haber sido pruebas *totalmente diferentes* a las de la vida real. Por lo tanto es favorable al realismo.



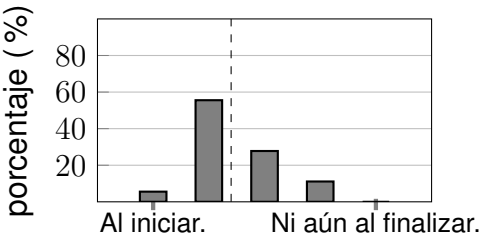
(a) ¿Fueron realistas los 8 ambientes propuestos?



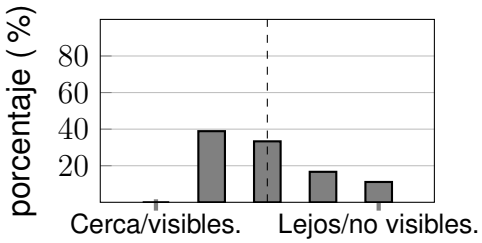
(b) ¿Desde donde percibía que provenían los sonidos?



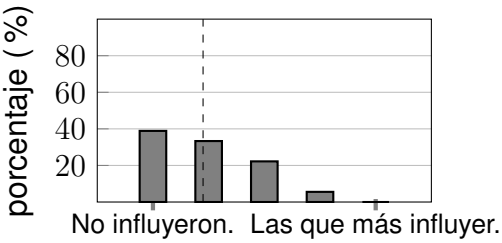
(c) ¿Le costó mucho esfuerzo imaginar la situación propuesta?



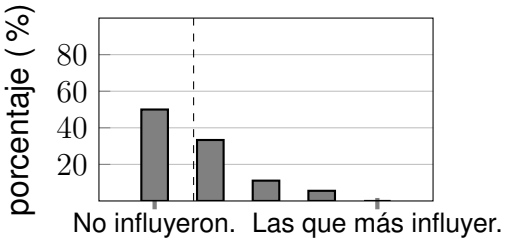
(d) ¿Qué tan rápido se adaptó al ambiente propuesto?



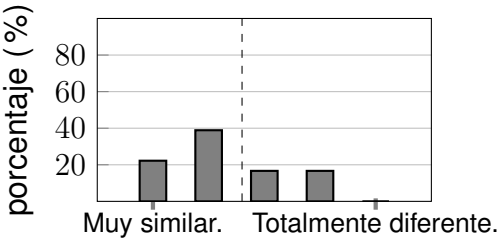
(e) ¿Qué tan lejos estaban las fuentes sonoras?



(f) ¿Qué tan desorientado se sintió en las pausas al inicio y al final de cada prueba?



(g) ¿Qué tanto influyeron sus respuestas los eventos de la realidad?



(h) ¿Qué tan similar fue la prueba comparada con sus experiencias de la vida real?

Figura 10.7: Respuestas sobre realismo y otros aspectos generales del experimento. Barras: Porcentaje de participantes que respondieron en ese segmento. Línea de segmentos: Promedio.

Capítulo 11

Discusión final y conclusiones

11.1 Trabajos futuros

En 8.2 se muestran varias soluciones al problema de la combinación controlada. Aunque esas soluciones no son cubiertas en detalle en esta tesis, pueden ser implementadas si previamente se desarrollan algunos métodos para poder representar con realismo ciertos sucesos de la naturaleza. En efecto la representación de estos sucesos constituye una línea de investigación interesante que se encuentra actualmente en desarrollo.

La distribución temporal de eventos sonoros es un interesante tópico de investigación. (Ver 3.2.1) Eventos sonoros de clases específicas, por ejemplo el ruido de tránsito rodado, pueden admitir modelos de distribución temporal actualmente implementados mediante algoritmos de simulación microscópica (Ver Cameron y Duncan [1996] para un algoritmo comercial fuertemente desarrollado o Behrisch et al. [2011] para código libre y abierto). Este campo de investigación es prometedor para las aplicaciones que comprenden un sistema de realidad virtual para investigar los efectos del ruido [Ruotolo et al., 2012]. Una de las dificultades en la aplicación de estos algoritmos es desarrollar una modificación que permita usar aspectos temporales y espectrales como datos de entrada pues actualmente los algoritmos se basan en datos de la infraestructura de ciudades como datos de entrada.

Otro interesante tema de investigación que actualmente está en desarrollo es la mejora de los algoritmos de Programación Lineal con Solución en los Enteros (IP) con un gran número de variables y ecuaciones de tal forma que las soluciones se alcancen

en un tiempo, y con un costo computacional, razonable. El problema de la combinación controlada completamente definido como un problema IP (Ver ecuación 8.15 en 8.2) puede ser solucionado pero sería necesario mejorar los algoritmos existentes. Además, un gran número de problemas de ingeniería se pueden resolver como problemas IP con un gran número de variables y ecuaciones (incluyendo restricciones). Para una clase específica de complejidad llamada problemas solucionables en tiempo polinomial ($\mathcal{P}$), en problemas de decisión , la tecnología actual está bien desarrollada. Desafortunadamente, el problema de la ecuación 8.15 pertenece a una clase llamada NP-difícil incluido en la clase NP-completo ($\mathcal{NPC}$), para los cuales no es sencillo predecir el tiempo de cómputo y los métodos de solución se basan en técnicas de reducción (por ejemplo ver algoritmos de ramificación y corte en Wolsey [1998]). El estado actual se basa en suponer la existencia de la clase $\mathcal{NPC}$ y polinómicamente reducir los problemas nuevos hasta problemas con soluciones conocidas en $\mathcal{NPC}$. Volviendo al problema de la combinación controlada, este enfoque promete ser muy realista porque permite introducir atenuación por frecuencias al repetir las columnas de $\mathbf{A}$ en la ecuación (8.16).

La encuesta de realismo usada en la prueba piloto (ver Capítulo 10) arroja unos resultados preliminares. No es el propósito de esta tesis desarrollar una encuesta normalizada pero en si puede ser útil para esta tarea. Es habitual en este tipo de tareas el trabajo con paneles de expertos y grupos de estudio. En Argentina existen una gran cantidad de expertos en acústica en general y varias de sus subramas pero, esta tesis es pionera en su temática: la experimentación en laboratorio con escenarios mixtos donde la parte virtual corresponde solo a un sonido ambiente realista. Se utilizaron elementos de otras encuestas similares, de otros países y desarrolladas para otras temáticas. La encuesta de realismo utilizada en esta tesis es en si un pequeño primer paso hacia el desarrollo de un instrumento normalizado. No obstante los resultados, sobre todo de la sección de pregunta abierta, sugieren que la herramienta tiene un alto grado de realismo.

11.2 Conclusiones

Se implementó con éxito una herramienta que permite componer estímulos sonoros realistas para la experimentación en laboratorio sobre los efectos del ruido en el ser

humano en función del patrón temporal, el tipo de fuentes y eventos sonoros y el contenido espectral del ruido. La herramienta fue desarrollada mediante códigos propios y la utilización del algoritmo para resolver problemas de optimización con soluciones en enteros y reales, de uso académico, CPLEX.

Se validó la herramienta desde el punto de vista objetivo y subjetivo mediante la prueba piloto. La parte objetiva se relaciona con la generación de los estímulos sonoros y la parte subjetiva con la evaluación de esos estímulos por parte de personas.

La validación objetiva no asegura que la herramienta sea capaz de generar absolutamente cualquier estímulo sino simplemente los del experimento propuesto. La posibilidad de generar los estímulos sonoros está ligada a un criterio de aceptabilidad respecto a las diferencias entre los valores requeridos por el usuario y los finalmente alcanzados por el estímulo. Un criterio posible puede ser el de las diferencias apenas perceptibles (JNDs) y queda como un trabajo futuro el estudio de esas diferencias para el nivel de exposición espectral a mediano plazo y el histograma de duraciones. En los estímulos generados se logró ajustar de manera exacta el histograma de duraciones pero los niveles de exposición sonora por bandas de frecuencia de 1/1 octava se ajustaron de manera aproximada. Esta aproximación alcanzó una diferencia máxima de $7{,}4$ dB en una banda que si bien puede ser perceptible en términos de nivel sonoro de sonidos estacionarios es posible que quede debajo de las JNDs para valores promedio de largo plazo de sonidos ambientales. Las posibilidades de reducir esa diferencia y también de validar otros estímulos está ligada a la variedad y cantidad de eventos sonoros que compongan la base de datos. Con una base apropiada es posible lograr concordancias mucho mayores, por ejemplo con una base que contuviera un conjunto de sonidos cuasi tonales con energía concentrada en cada una de las bandas espectrales, ya que en la figura 8.1 sería equivalente a tomar vectores cercanos a los ejes, por lo tanto existirán buenas aproximaciones a cualquier vector, habiendo soluciones exactas a muchos más casos pues la “hipercuña” cubriría casi la totalidad del “hiperoctante”.

La validación subjetiva es favorable. Si bien en ninguna de las preguntas se alcanzó una conclusión con un nivel de significación estadística aceptable de todos modos las respuestas promedio fueron favorables en cuanto al realismo de la experiencia. Se destaca los comentarios posteriores a las pruebas, por parte de los participantes,

respecto al realismo y a no haber imaginado la configuración real del sistema de auralización.

Apéndice A

Descriptores básicos de ruido

En este apéndice se muestran los descriptores básicos del ruido, según IRAM 4113-1 [2009] equivalente a ISO 1996-1 [2003], que se utilizan en esta tesis.

La **presión sonora** p se define (A.1) como la presión relativa a la presión atmosférica (P_0), causada por una perturbación (u onda) sonora. Es decir, para un punto dado del espacio, es variable en el tiempo.

$$p(t) = P(t) - P_0 \tag{A.1}$$

donde $P(t)$ es la presión total en función del tiempo.

A su vez el valor cuadrático de la **presión sonora eficaz** o rms (A.2) es de especial interés por guardar una estrecha relación con la energía sonora y la exposición sonora, ambos descriptores relacionados en cierta medida con los efectos de la exposición a ruidos del ser humano.

$$p_{\text{rms}}^2 = \frac{1}{T}\int_0^T p(t)^2 \mathrm{d}t \tag{A.2}$$

Los valores típicos de p_{rms} van desde los 20 µPa a los 20 Pa. Para un tono, o señal acústica sinusoidal, de frecuencia 1 kHz, una presión rms de 20 µPa es conocida como el umbral de audición, es decir, aquella presión sonora que es apenas perceptible por el ser humano en un sentido estadístico. También para un tono de 1 kHz, una presión sonora eficaz de 20 Pa es conocida como el umbral del dolor, es decir, que un tono con esa frecuencia y presión sonora comienza a producir dolor en el oído. Habitualmente el ser humano se expone a presiones sonoras entre estos dos límites sin llegar a dichos

extremos. Notar que la presión atmosférica es habitualmente del orden de los 100 kPa, por lo cual la presión sonora ofrece un ajuste de rango respecto a la presión total. Además el oído humano detecta las variaciones respecto a la presión atmosférica de forma independiente al valor que la presión atmosférica pueda tener en un sitio particular. Esto ocurre debido a un mecanismo de regulación de presión en el oído medio.

La ecuación (A.2), define la presión eficaz para un periodo T determinado. En casos de ruidos continuos se puede estimar rápidamente a partir de T cortos, pero en casos de ruidos con mayor variabilidad se debe utilizar un T largo para obtener una buena estimación. Sin embargo, para obtener mayores detalles sobre la variabilidad de un ruido, es de interés obtener la presión sonora móvil en lugar de la presión sonora total en el intervalo T. Dicha presión sonora móvil se obtiene reemplazando la integral y la división, ambas en T, por un bloque de promedio móvil. Dicho bloque corresponde a un filtro pasabajos de primer orden con constante de tiempo τ. Finalmente la señal contendrá un ancho de banda menor por lo cual habitualmente es submuestreada. Esta presión sonora móvil se denomina **presión sonora con constante temporal** τ. En la figura A.1 se muestra un diagrama en bloques de cómo obtener la p_τ.

Figura A.1: Presión sonora con constante temporal τ

Según el valor de la constante temporal, la presión sonora así estimada se denomina presión sonora con constante lenta $\tau = 1$ s, rápida $\tau = 125$ ms, impulsiva $\tau = 35$ ms. En psicoacústica es habitual el uso de una constante diferente, $\tau = 2$ ms, que permite modelar de manera no lineal la integral temporal que simula la respuesta temporal del sistema auditivo. Así, por ejemplo, la constante de $\tau = 2$ ms se suele utilizar en modelos de enmascaramiento temporal (ver Chalupper y Fastl [2002]).

Se define el **nivel de presión sonora** L_p mediante la ecuación

$$L_\mathsf{p} = 10 \times \log\left(\frac{p_\mathsf{rms}^2}{p_0^2}\right) \tag{A.3}$$

donde $p_0 = 20$ µPa se denomina presión sonora de referencia y corresponde al umbral de audición para un tono puro de 1 kHz. Es decir, el nivel de presión sonora es una

medida relativa en torno a la presión de referencia. Presiones sonoras eficaces por debajo de p_0 corresponden a niveles negativos y valores por sobre los 20 Pa exceden los 120 dB. Es decir, la escala de esta variable está más comprimida que en el caso de la presión sonora. Por esta razón es habitual, y de uso en este texto, mostrar los valores en términos de nivel aunque las ecuaciones y la formulación de problemas se haga en términos de presión u otros descriptores que tengan un descriptor equivalente en términos de nivel.

Si en la ecuación (A.3) se reemplaza p_{rms} por p_τ, se obtiene un patrón temporal del nivel de presión sonora. Utilizando ese patrón temporal se definen los niveles estadísticos en Argentina [IRAM 4113-1, 2009], también denominados internacionalmente niveles percentiles como en la versión oficial en español de la norma ISO 1996-1 [2003]. El **nivel estadístico** se define como aquel nivel que es superado un N % del intervalo de tiempo considerado[1].

Se define el **nivel estadístico** N **%** $L_{\tau,N}$ como el nivel de presión sonora, calculado con un algoritmo que simula un circuito de promedio móvil con constante temporal τ, como el nivel de presión sonora L_{p_τ} que es superado el N % del periodo de medición T. En otros países se llama nivel percentil, pero en Argentina se utiliza la locución *nivel estadístico* para evitar confusión con la definición de percentiles en estadística pues no son equivalentes. En particular en esta tesis se utiliza el nivel estadísitco 5 % con constante temporal $\tau = 2$ ms, denotado L_5.

Se define la **exposición sonora** según la ecuación (A.4).

$$E = \int_0^T p(t)^2 \mathrm{d}t \tag{A.4}$$

donde T es la duración de la integración.

Se define el **nivel de exposición sonora** L_{E} mediante la ecuación

$$L_{\text{E}} = 10 \times \log\left(\frac{E}{E_0}\right) \tag{A.5}$$

donde $E_0 = p_0^2$ es la exposición de referencia. Se debe informar la duración T.

[1] Por esa razón en Argentina se define como nivel estadístico, para evitar confusión con los percentiles cuya definición es el valor que es *alcanzado* el N % del total de muestras

Apéndice B

Test de Weinstein sobre sensibilidad al ruido

La mayoría de los ítems se presentan en una escala de Likert de 6 puntos que van desde “estoy totalmente de acuerdo” (1) hasta “estoy totalmente en desacuerdo” (6). Los ítems con * se gradúan en dirección opuesta antes de sumar las respuestas

1. No me importaría vivir en una calle ruidosa si tuviera un departamento acogedor.
2. Soy mas consciente del ruido que antes *
3. A nadie debería importarle mucho si alguien sube el volumen de su estéreo una vez cada tanto.
4. En el cine los cuchicheos y el ruido de los envoltorios de caramelos me perturban.*
5. El ruido me despierta fácilmente. *
6. Si hay ruido donde estoy estudiando trato de cerrar la puerta o la ventana o ir a otra parte. *
7. Me molesta cuando mis vecinos son ruidosos. *
8. Me acostumbro a los ruidos sin mucha dificultad.
9. ¿Cuánto le preocuparía si un departamento que está interesado en alquilar estuviera ubicado frente a una estación de bomberos? *

10. A veces los ruidos me ponen nervioso y me irritan. *
11. Aun la música que normalmente me gusta me molesta si estoy tratando de concentrarme. *
12. No me molestaría escuchar los ruidos de la vida diaria de los vecinos (pasos, agua corriendo, etc.).
13. Cuando quiero estar solo me perturba escuchar los ruidos de afuera. *
14. Soy bueno para concentrarme sin importar lo que esté sucediendo alrededor mío.
15. En una biblioteca no me molesta si la gente conversa si lo hace suavemente.
16. Frecuentemente hay momentos en que deseo completo silencio. *
17. Se debería exigir que las motos tuvieran silenciadores más grandes. *
18. Me cuesta relajarme en un lugar que es ruidoso. *
19. Me vuelve loco la gente que hace ruido impidiéndome dormirme o realizar mi trabajo. *
20. No me importaría vivir en un departamento con paredes delgadas.
21. Soy sensible al ruido. *

El test originalmente se diseñó en inglés y ha sido traducido a diversos idiomas. A continuación se muestran las preguntas en el idioma original.

1. I wouldn't mind living on a noisy street if the apartment I had was nice.
2. I am more aware of noise than I used to be.
3. No one should mind much if someone turns up his stereo fiull blast ones in a while.
4. At movies, whispering and crinkling candy wrappers disturb me.
5. I am easily awakened by noise.

6. If it's noisy where I'm studying, I try to close the door or window or move somewhere else.
7. I get annoyed when my neighbors are noisy.
8. I get used to most noises without much difficulty.
9. How much would it matter to you if an apartment you were interested in renting was located across from a fire station?
10. Sometimes noises get on my nerves and get me irritated.
11. Even music I normally like will bother me if I'm trying to concentrate.
12. It wouldn't bother me to hear the sounds of everyday living from neighbors (footsteps, running water, etc.).
13. When I want to be alone, it disturbs me to hear outside noises.
14. I'm good at concentrating no matter what is going on around me.
15. In a library, I don't mind if people carry on a conversation if they do it quietly.
16. There are often times when I want complete silence.
17. Motorcycles ought to be required to have bigger mufflers.
18. I find it hard to relax in a place that's noisy.
19. I get mad at people who make noise that keeps me from falling asleep or getting my work done.
20. I wouldn't mind living in an apartment with thin walls.
21. I am sensitive to noise.

Apéndice C

Encuesta

ENCUESTA PARA ESTUDIO EXPERIMENTAL EN EL MARCO DE TESIS DE DOCTORADO EN INGENIERÍA.

Introducción: Esta encuesta tiene como propósito proporcionar datos experimentales para la tesis de Doctorado en Ingeniería (de la Facultad de Ingeniería de la UNR) del doctorando Ernesto Accolti dirigido por Federico Miyara y Ernesto Kofman. Este trabajo respeta estrictamente los estándares bioéticos internacionales y legislación sobre protección de datos. Ningún resultado intermedio o final será asociado a sus datos personales en ninguna publicación. Los datos serán tratados estadísticamente y la información de contacto será confidencial y utilizada únicamente en caso de requerirse alguna prueba futura suplementaria, en un lapso no mayor a un año. La participación en cualquier etapa ulterior dentro de ese lapso será estrictamente voluntaria. Los resultados globales serán publicados oportunamente e informados a los participantes.

Encuesta: Imagine que se encuentra en un lugar donde pasa el tiempo libre. En cada prueba el lugar estará en un ambiente distinto.

En cada prueba deberá contestar dos preguntas. La pregunta a es ¿Cuánto le molestó o perturbó el ruido de este ambiente? Y su respuesta es verbal entre las categorías: nada, ligeramente, medianamente, mucho y extremadamente. A continuación, en el punto b se da una escala de opinión de cero a diez para que Vd. pueda expresar en qué grado le molesta o perturba el ruido producido por este ambiente. Por ejemplo, si Vd. no está NADA molesto por el ruido debería escoger el cero, y si Vd. está EXTREMADAMENTE molesto debería escoger el diez.

C.1 Para cada prueba

A continuación se muestran las preguntas formuladas solo para la prueba 1. La misma se repite para las 7 pruebas restantes.

Prueba 1

1.a) ¿Cuánto le molestó o perturbó el ruido de este ambiente?

NADA

LIGERAMENTE

MEDIANAMENTE

MUCHO

EXTREMADAMENTE

1.b) ¿En qué grado le molestó?

NADA EXTREMADAENTE

0	1	2	3	4	5	6	7	8	9	10

C.2 Para las 8 pruebas

Pensando en las 8 pruebas anteriores por favor responder:

1. ¿Fueron realistas los 8 ambientes propuestos?

1	2	3	4	5

No, en lo absoluto Si, totalmente

2. ¿Desde donde percibía que provenían los sonidos?

1	2	3	4	5

Fuera de la sala Dentro de la sala

3. ¿Le costó mucho esfuerzo imaginar la situación propuesta?

1	2	3	4	5

No, en lo absoluto Si, totalmente

4. ¿Qué tan rápido se adaptó al ambiente propuesto?

1	2	3	4	5

Al iniciar la prueba Ni aún finalizada la prueba

5. ¿Qué tan lejos estaban las fuentes sonoras? ¿Deberían haber sido visibles dentro de la sala?

1	2	3	4	5

Cerca/visibles Lejos/ no visibles

6. ¿Qué tan desorientado se sintió en las pausas al inicio y al final de la prueba?

1	2	3	4	5

No influyeron Las que más influyeron

7. ¿Qué tanto influyeron sus respuestas los eventos de la realidad que no pertenecían a la simulación? (Considere que el interior de la sala si pertenece a la simulación)

1	2	3	4	5

No influyeron Las que más influyeron

8. ¿Qué tan similar fue la prueba comparada con sus experiencias de la vida real?

1	2	3	4	5

Muy similar Totalmente diferente

9. Si lo desea puede escribir más información sobre sus sensaciones respecto a las pruebas.

Apéndice D

Valores requeridos por usuario y finalmente alcanzados en prueba piloto

En la tabla D.1 se muestra el historgrma de duraciones y en la tabla D.2 se muestra el espectro para cada una de las 8 pruebas enumeradas en la tabla 10.4 del capítulo 10.

En la tabla D.1 se muestran dos versiones del histograma de duraciones, es decir, la versión d que corresponde al histograma requerido por el usuario y la versión $\widehat{d}$ que representa el valor alcanzado por la solución del problema de optimización. En la tercera fila de cada prueba se muestra el error ($\widehat{d} - d$). El error es nulo para todos los casos.

En la tabla D.1 se muestran tres versiones del histograma de duraciones, es decir, la versión L_{eq} que corresponde al espectro requerido por el usuario, la versión $\widehat{L_{eq}}$ que representa el valor alcanzado por la solución del problema de optimización y la versión $\widetilde{L}_{eq}$ que corresponde a la medición en la sala con un sonómetro clase 1. En la cuarta fila de cada prueba se muestra el error ($\widetilde{L}_{eq} - L_{eq}$).

Tabla D.1: Histograma de duración requerido y alcanzado

Prueba	parámetro	$]0{,}0;1{,}0[$	$[1{,}0;4{,}9[$	$[4{,}9;12{,}4[$	$[12{,}4;23{,}5[$	$[23{,}5;\infty]$
4	d	1,0	2,0	3,0	4,0	5,0
	$\widehat{d}$	1,0	2,0	3,0	4,0	5,0
	$\widehat{d}-d$	0,0	0,0	0,0	0,0	0,0
3	d	8,0	7,0	6,0	5,0	4,0
	$\widehat{d}$	8,0	7,0	6,0	5,0	4,0
	$\widehat{d}-d$	0,0	0,0	0,0	0,0	0,0
2	d	4,0	5,0	6,0	7,0	8,0
	$\widehat{d}$	4,0	5,0	6,0	7,0	8,0
	$\widehat{d}-d$	0,0	0,0	0,0	0,0	0,0
1	d	5,0	4,0	3,0	2,0	1,0
	$\widehat{d}$	5,0	4,0	3,0	2,0	1,0
	$\widehat{d}-d$	0,0	0,0	0,0	0,0	0,0
2	d	4,0	5,0	6,0	7,0	8,0
	$\widehat{d}$	4,0	5,0	6,0	7,0	8,0
	$\widehat{d}-d$	0,0	0,0	0,0	0,0	0,0
1	d	5,0	4,0	3,0	2,0	1,0
	$\widehat{d}$	5,0	4,0	3,0	2,0	1,0
	$\widehat{d}-d$	0,0	0,0	0,0	0,0	0,0
4	d	1,0	2,0	3,0	4,0	5,0
	$\widehat{d}$	1,0	2,0	3,0	4,0	5,0
	$\widehat{d}-d$	0,0	0,0	0,0	0,0	0,0
3	d	8,0	7,0	6,0	5,0	4,0
	$\widehat{d}$	8,0	7,0	6,0	5,0	4,0
	$\widehat{d}-d$	0,0	0,0	0,0	0,0	0,0

Tabla D.2: Composición espectral requerida, alcanzada y medida

Prueba	parámetro	63	125	250	500	1 000	2 000	4 000
1	L_{eq}	58,9	53,4	47,9	42,4	36,9	31,4	25,9
	$\widehat{L}_{eq}$	52,8	54,4	50,4	44,9	37,9	25,3	21,8
	$\widetilde{L}_{eq}$	55,1	52,5	48,0	43,8	39,9	26,1	24,4
	$\widetilde{L}_{eq} - L_{eq}$	−3,9	−1,0	0,1	1,4	3,0	−5,4	−1,6
2	L_{eq}	58,9	53,4	47,9	42,4	36,9	31,4	25,9
	$\widehat{L}_{eq}$	60,6	54,4	48,4	44,5	38,9	27,4	26,4
	$\widetilde{L}_{eq}$	65,6	55,0	46,0	43,8	40,5	28,3	29,5
	$\widetilde{L}_{eq} - L_{eq}$	6,7	1,5	−1,9	1,4	3,5	−3,1	3,6
3	L_{eq}	68,9	63,4	57,9	52,4	46,9	41,4	35,9
	$\widehat{L}_{eq}$	64,8	65,5	59,5	54,5	49,0	37,3	36,5
	$\widetilde{L}_{eq}$	68,0	66,9	57,5	54,0	50,1	37,5	38,7
	$\widetilde{L}_{eq} - L_{eq}$	−0,9	3,5	−0,5	1,6	3,2	−4,0	2,7
4	L_{eq}	68,9	63,4	57,9	52,4	46,9	41,4	35,9
	$\widehat{L}_{eq}$	64,8	64,7	58,5	54,5	48,9	37,3	36,2
	$\widetilde{L}_{eq}$	67,7	64,1	56,1	54,2	51,0	38,7	40,1
	$\widetilde{L}_{eq} - L_{eq}$	−1,3	0,6	−1,8	1,8	4,0	−2,8	4,2
5	L_{eq}	63,2	55,7	48,2	40,7	33,2	25,7	18,2
	$\widehat{L}_{eq}$	56,5	56,3	50,7	43,2	35,0	19,0	13,5
	$\widetilde{L}_{eq}$	59,7	55,5	49,1	43,8	37,4	23,2	18,8
	$\widetilde{L}_{eq} - L_{eq}$	−3,5	−0,2	0,9	3,1	4,2	−2,5	0,6
6	L_{eq}	63,2	55,7	48,2	40,7	33,2	25,7	18,2
	$\widehat{L}_{eq}$	63,7	56,9	50,5	43,0	34,3	20,6	17,4
	$\widetilde{L}_{eq}$	68,8	58,1	48,5	43,2	36,4	22,9	21,4
	$\widetilde{L}_{eq} - L_{eq}$	5,6	2,4	0,3	2,4	3,2	−2,8	3,2
7	L_{eq}	73,2	65,7	58,2	50,7	43,2	35,7	28,2
	$\widehat{L}_{eq}$	65,8	65,9	60,8	53,3	45,2	28,7	21,6
	$\widetilde{L}_{eq}$	68,8	65,4	60,1	54,3	48,9	31,6	26,6
	$\widetilde{L}_{eq} - L_{eq}$	−4,5	−0,3	1,9	3,6	5,7	−4,1	−1,6
8	L_{eq}	73,2	65,7	58,2	50,7	43,2	35,7	28,2
	$\widehat{L}_{eq}$	70,7	64,8	59,0	52,5	45,0	32,6	30,0
	$\widetilde{L}_{eq}$	76,7	64,2	58,0	53,8	48,3	35,3	34,5
	$\widetilde{L}_{eq} - L_{eq}$	3,5	−1,5	−0,2	3,1	5,1	−0,4	6,3

Apéndice E

Archivos usados para generar estímulos de la prueba piloto

En las tablas E.1 a E.8 se muestra la cantidad de archivos y repeticiones usadas en función del cuarto nivel de la estructura de clasificación jerárquica de los eventos sonoros que intervinieron en cada estímulo.

Tabla E.1: Prueba 1. Cantidad de repeticiones y archivos usados

Fuente	# repeticiones	# archivos
Herramientas grandes	1	1
Automóviles	3	2
Taladrar	1	1
Motocicletas	1	1
Máquinas de coser	1	1
Camiones	1	1
Lluvia	2	1

Tabla E.2: Prueba 2. Cantidad de repeticiones y archivos usados

Fuente	# repeticiones	# archivos
Aviones	5	2
Herramientas grandes	2	2
Motocicletas	4	2
Taladrar	2	1
Camiones	3	1
Lluvia	6	1
Máquina de fábrica	2	1

Tabla E.3: Prueba 3. Cantidad de repeticiones y archivos usados

Fuente	# repeticiones	# archivos
Aviones	1	1
Herramientas grandes	3	1
Taladrar	2	2
Monos	1	1
Motocicletas	4	2
Taladrar	2	1
Lluvia	2	1
Viento	2	1

Tabla E.4: Prueba 4. Cantidad de repeticiones y archivos usados

Fuente	# repeticiones	# archivos
Aviones	1	1
Herramientas grandes	3	1
Motocicletas	2	2
Taladrar	3	1
Lluvia	2	1
Viento	2	1

Tabla E.5: Prueba 5. Cantidad de repeticiones y archivos usados

Fuente	# repeticiones	# archivos
Herramientas grandes	7	3
Motosierra	1	1
Taladrar	1	1
Monos	1	1
Camiones	4	2
Equipo de aire acondicionado	2	1
Viento	3	1

Tabla E.6: Prueba 6. Cantidad de repeticiones y archivos usados

Fuente	# repeticiones	# archivos
Herramientas grandes	3	3
Taladrar	1	1
Autos deportivos	1	1
Motocicletas	2	2
Camiones	2	1
Máquina de fábrica	6	1

Tabla E.7: Prueba 7. Cantidad de repeticiones y archivos usados

Fuente	# repeticiones	# archivos
Aviones	1	1
Herramientas grandes	2	2
Automóviles	2	2
Taladrar	1	1
Autos deportivos	3	1
Camiones	1	1
Viento	2	1

Tabla E.8: Prueba 8. Cantidad de repeticiones y archivos usados

Fuente	# repeticiones	# archivos
Aviones	1	1
Herramientas grandes	3	1
Taladrar	3	2
Motocicletas	3	2
Máquina de coser	1	1
Taladrar	1	1
Camiones	3	1
Lluvia	6	1
Lluvia	2	1
Máquina de fábrica	2	1

Bibliografía

Accolti, E., Maffei, L., y Miyara, F. (2010a). Controlling temporal factors of aural stimulus for assessment of environmental noise effects on human being. *Acta Acustica united with Acustica*, 96(Sup.1):S17.4.

Accolti, E., Maffei, L., y Miyara, F. (2010b). Influence of traffic factors in the slope descriptor. In *II Congreso sobre Acústica, UNTREF*, Caceros, Buenos Aires.

Accolti, E. y Miyara, F. (2008). Combinación digital controlada de ruidos diversos. In *Proceedings of the VI Congreso Iberoamericano de Acústica (FIA2008)*, Buenos Aires, Argentina. Asociación de Acústicos Argentinos y Federación Iberoamericana de Acústica. Cód. A090.

Accolti, E. y Miyara, F. (2009). Fluctuation strength of mixed fluctuating sound sources. *Mecanica Computacional*, XXVIII (2):9–22.

Accolti, E. y Miyara, F. (2010). Tools for studying noise effects based on spectral and temporal content. In *Proceedings of the 39th International Congress and Exposition on Noise Control Engineering, INTERNOISE 2010*, page 51, Lisbon, Portugal. International Institute of Noise Control Engineering (I-INCE), Portuguese Acoustical Society (SPA) and the Spanish Acoustical Society (SEA). Invited Article.

Accolti, E. y Miyara, F. (2015). Method for generating realistic sound stimuli with given characteristics by controlled combination of audio recordings. *The Journal of the Acoustical Society of America. Express Letter.*

Ahrens, J., Rabenstein, R., y Spors, S. (2008). The theory of wave field synthesis revisited. In *Audio Engineering Society Convention 124*.

Alayrac, M., Marquis-Favre, C., y Viollon, S. (2011). Total annoyance from an industrial noise source with a main spectral component combined with a background noise. *The Journal of the Acoustical Society of America*, 130(1):189–199.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., y Tang, Z. (2002). Approximating the head-related transfer function using simple geometric models of the head and torso. *The Journal of the Acoustical Society of America*, 112(5):2053–2064.

Algazi, V. R., Duda, R. O., Morrison, R. P., y Thompson, D. M. (2001a). Structural composition and decomposition of hrtfs. In *Proc. of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 103–106, New Paltz, NY.

Algazi, V. R., Duda, R. O., Thompson, D. M., y Avendano (2001b). The cipic hrtf database. In *Proc. of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 99–102, New Paltz, NY.

Alvarsson, J. J., Nordström, H., Lundén, P., y Nilsson, M. E. (2014). Aircraft noise and speech intelligibility in an outdoor living space. *The Journal of the Acoustical Society of America*, 135(6):3455–3462.

Ando, Y. (1998). *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners*. Springer Verlag.

Behrisch, M., Bieker, L., Erdmann, J., y Krajzewicz, D. (2011). Sumo - simulation of urban mobility: An overview. In *SIMUL 2011, The Third International Conference on Advances in System Simulation*, Barcelona, Spain.

Berglund, B. y Lindvall, T. (1995). Community noise. *Archives of the Center for Sensory Research*, 2(1):1–195.

Berkhout, A. J., de Vries, D., y Vogel, P. (1993). Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, 93(5):2764–2778.

Box, G. E., Hunter, J. S., y Hunter, W. G. (1988). *Estadística para investigadores: introducción al diseño de experimentos, análisis de datos y construcción de modelos*. Reverté.

Boyd, S. y Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Bunting, O., Stammers, J., Chesmore, D., Bouzid, O., Tian, G., Karatsovis, C., y Dyne, S. (2009). Instrument for soundscape recognition, identification and evaluation (isrie): technology and practical uses. In *EURONOISE, Edinburgh, Scotland*.

Cameron, G. y Duncan, G. (1996). Paramics. parallel microscopic simulation of road traffic. *The Journal of Supercomputing*, 10:25–53.

Chalupper, J. y Fastl, H. (2002). Dynamic loudness model (dlm) for normal and hearing-impaired listeners. *Acta Acustica united with Acustica*, 88(3):378–386.

De Coensel, B. y Botteldooren, D. (2010). A model of saliency-based auditory attention to environmental sound. In *Proc. of ICA 2010 congress*, Sydney, Asutralia.

Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *108th AES Convention*, pages 18–22.

Fastl, H. (1977). Subjective duration and temporal masking patterns of broadband noise impulses. *The Journal of the Acoustical Society of America*, 61(1):162–168.

Fastl, H., Büeler, R., y Fruhmann, M. (2002). Different implementations of a model for subjective duration. In *Tagungsband Fortschritte der Akustik - DAGA - 2002, Bochum*, pages 470–471. 4.-8.3.2002.

Fastl, H. y Zwicker, E. (2005). *Psychoacoustics : Facts and Models*. Springer.

Fritschi, L., Brown, A. L., Kim, R., Schwela, D., y Kephalopoulos, S., editores (2011). *Burden of Disease from Environmental Noise : Quantification of healthy life years lost in Europe.* WHO Europe.

Guski, R. (1997). Psychological methods for evaluating sound quality and assessing acoustic information. *Acta Acustica united with Acustica*, 83(5):765–774.

Hong, J. y Jeon, J. (2013). Designing sound and visual components for enhancement of urban soundscapes. *The Journal of the Acoustical Society of America*, 134(3):2026–2036.

IEC 61260 (1995). *Octave-Band and Fractional-Octave-Band Filters*.

IEC 61672-1 (2002). *Electroacoustics - Sound level meters - Part 1: Specifications*.

Iida, K., Ishii, Y., y Nishioka, S. (2014). Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae. *The Journal of the Acoustical Society of America*, 136(1):317–333.

ILOG (2011). IBM ILOG CPLEX Optimizer. `www-01.ibm.com/software/integration/optimization/cplex-optimizer/`.

IRAM 4062 (2001). *Ruidos molestos al vecindario. Método de medición y clasificación*. En revisión.

IRAM 4081 (1977). *Filtros de banda de octava, de media octava y de tercio de octava destinados al análisis de sonidos y vibraciones*.

IRAM 4109-2 (2011). *Acústica - Medición de parámetros acústicos en recintos. Parte 2 - Tiempo de reverberación de recintos comunes.*

IRAM 4113-1 (2009). *Acústica - Descripción, medición y evaluación del ruido ambiental - Parte 1: Magnitudes básicas y métodos de evaluación*.

IRCAM (2003). Listen hrtf database. `http://recherche.ircam.fr/equipes/salles/listen/`. última actualización 25/05/2003, consultado 03/07/2014.

ISO 11904-1 (2002). *Acoustics - Determination of sound immission from sound sources placed close to the ear - Part 1: Technique using a microphone in a real ear (MIRE technique)*.

ISO 11904-2 (2004). *Acoustics - Determination of sound immission from sound sources placed close to the ear - Part 2: Technique using a manikin*.

ISO 15666 (2003). *Acoustics - Assessment of noise annoyance by means of social and socio-acoustic surveys*.

ISO 18233 (2006). *Acoustics - Application of new measurement methods in building and room acoustics*.

ISO 1996-1 (2003). *Acoustics - Description, measurement and assessment of environmental noise - Part 1: Basic quantities and assessment procedures*.

ISO 3382-2 (2008). *Acoustics - Measurement of room acoustic parameters - Part 2: Reverberation time in ordinary rooms*.

ISO 9613-2 (1996). *Acoustics - Attenuation of sound during propagation outdoors - Part 2: General method of calculation*.

Itti, L. y Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203.

Itti, L., Koch, C., y Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.

Katz, B. F. G. y Parseihian, G. (2012). Perceptually based head-related transfer function database optimization. *The Journal of the Acoustical Society of America*, 131(2):EL99–EL105.

Kayser, C., Petkov, C. I., Lippert, M., y Logothetis, N. K. (2005). Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology*, 15(21):1943–1947.

Kim, J., C. Lim, J. H., y Lee, S. (2010). Noise-induced annoyance from transportation noise: Short-term responses to a single noise source in a laboratory. *The Journal of the Acoustical Society of America*, 127(2):804–814.

Kinsler, L. E., Frey, A. R., Coppens, A. B., y Sander, J. V. (2000). *Foundamentals of Acoustics*. Wiley, 4 edition.

Larsson, P., Västfjäll, D., y Kleiner, M. (2003). On the quality of experience: A multimodal approach to perceptual ego-motion and sensed presence in virtual environments. In *First ITRW on Auditory Quality of Systems, Akademie Mont-Cenis*, pages 23–25, Alemania.

Maffei, L. (2008). Urban and quiet areas soundscape preservation. conferencia plenaria. In *VI Congreso Iberoamericano de Acústica, FIA, Bs As*, page CP1.

Maffei, L., Masullo, M., Sorrentino, F., y Di Gabriele, M. (2014). Preliminary studies on the relation between the audio-visual cues' perception and the approaching speed of electric vehicles. *Proceedings of Meetings on Acoustics*, 20(1):–.

Marengo-Rodriguez, F., Accolti, E., y Miyara, F. (2011). Blind doppler shift compensation of vehicle noise and its characterization for traffic noise simulation. *Mecanica Computacional*, XXX (41):3187–3199.

Matsinos, Y., Mazaris, A., Papadimitriou, K., Mniestris, A., Hatzigiannidis, G., Maioglou, D., y Pantis, J. (2008). Spatio-temporal variability in human and natural sounds in a rural landscape. *Landscape Ecology*, 23(8):945–959.

Miedema, H. y H., V. (1998). Exposure-response relationships for transportation noise. *The Journal of the Acoustical Society of America*, 104(6):3432–3445.

Miyara, F. (2013a). *Mediciones acústicas basadas en software*. Asociación de Acústicos Argentinos.

Miyara, F. (2013b). *Ruido, arte y sociedad*. UNR Editora.

Miyara, F., Accolti, E., y Marengo-Rodriguez, F. (2012). Estudio de impacto acústico del aeropuerto sobre funes city. *Internal Technical Report. Laboratorio de Acústica y Electroacústica.*

Miyara, F., Accolti, E., Pasch, V., Cabanellas, S., Yanitelli, M., Miechi, P., Marengo-Rodriguez, F.-A., y Mignini, E. (2010). Suitability of a consumer digital recorder for use in acoustical measurements. In *Proceedings of the 39th International Congress and Exposition on Noise Control Engineering, INTERNOISE 2010*, page 993, Lisboa, Portugal. International Institute of Noise Control Engineering (I-INCE), Portuguese Acoustical Society (SPA) and the Spanish Acoustical Society (SEA).

Miyara, F., Cabanellas, S., Pasch, V., Yanitelli, M., Accolti, E., y Miechi, P. (2009a). Contrastación de algoritmos de análisis de espectro sonoro con un instrumento normalizado. *Mecanica Computacional*, XXVIII (2):199.

Miyara, F., Cabanellas, S., Pasch, V., Yanitelli, M., Accolti, E., y Miechi, P. (2009b). Contrastación de algoritmos de análisis de espectro sonoro con un instrumento nor-

malizado. In *Proceedings of the 1as Jornadas Regionales de Acústica AdAA 2009*, Rosario, Argentina. Asociación de Acústicos Argentinos. Cód. A032.

Moorhouse, A., Waddington, D., y Adams, M. (2005). Proposed criteria for the assessment of low frequency noise disturbance.

Nordtest 111 (2002). *Acoustics - Human sound perception – Guidelines for listening tests*.

Okamoto, T., Enomoto, S., y Nishimura, R. (2014). Least squares approach in wavenumber domain for sound field recording and reproduction using multiple parallel linear arrays. *Applied Acoustics*, 86(0):95 – 103.

Oldfield, R. (2013). *The analysis and improvement of focused source reproduction with wave field synthesis*. Tesis Doctoral, University of Salford.

Oppenheim, A. V. y Schafer, R. W. (1975). *Digital Signal Processing*. Prentice Hall.

Raimbault, M. y Dubois, D. (2005). Urban soundscapes: Experiences and knowledge. *Cities, Elsevier*, 22(5):339–350.

Res 201, SEMADS (2004). Resolución sobre preservación, protección y recuperación de la calidad del aire en el ámbito de la provincia de santa fe. *Secretaría de Estado de Medio Ambiente y Desarrollo Sustentable de la Provincia de Santa Fe*.

Riva, M. (2008). *Síntesis en Tiempo Real de Sonido Tridimensional sobre Auriculares*. Proyecto Final Ing. Electrónica, Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario.

Ruotolo, F., Senese, V., Ruggiero, G., Maffei, L., Masullo, M., y Iachini, T. (2012). Individual reactions to a multisensory immersive virtual environment: the impact of a wind farm on individuals. *Cogn Process*.

Schafer, R. M. (1977). *The tuning of the world*. Knopf.

Small, R. (1971). Direct-radiator loudspeaker system analysis. *J. Audio Eng. Soc.*, 20:383–395.

Stammers, J. y Chesmore, D. (2008). Instrument for soundscape recognition, identification and evaluation (isrie): Signal classification. *The Journal of the Acoustical Society of America*, 123(5):3081–3081.

Thiele, A. (1971a). Loudspeakers in vented boxes, part i. *J. Audio Eng. Soc.*, 19:382–392.

Thiele, A. (1971b). Loudspeakers in vented boxes, part ii. *J. Audio Eng. Soc.*, 19:471–483.

Tommasini, F. C. (2012). *Sistema de simulación acústica virtual en tiempo real*. Tesis Doctoral, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba.

Vos, J. y Houben, M. (2013). Enhanced awakening probability of repetitive impulse sounds. *The Journal of the Acoustical Society of America*, 134(3):2011–2025.

Weinstein, N. (1978). Individual differences in reactions tonoise: A longitudinal study in a college dormitory. *Journal of Applied Psychology*, 63:458–466.

Weinstein, N. (1980). Individual differences in critical tenden-ciesand noise annoyance. *Journal of Sound and Vibration*, 68:241–248.

Wierstorf, H., Raake, A., Geier, M., y Spors, S. (2013). Perception of focused sources in wave field synthesis. *J. Audio Eng. Soc*, 61(1/2):5–16.

Witmer, B. G. y Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240.

Wolsey, L. A. (1998). *Integer Programming*. Wiley.

Notación

ℓ_1	Norma uno o norma Cityblock
ℓ_2	Norma dos o norma Euclidea
ℓ_∞	Norma infinito
$\mathbb{N}$	Naturales
$\mathbb{N}_0$	Enteros no negativos
$\mathbb{R}$	Reales
$\mathbb{R}^+$	Reales positivos
$\mathbb{R}_{\geq 0}$	Reales positivos y valor nulo
$\mathbb{Z}$	Enteros
$\mathcal{NPC}$	Clase de problemas solucionables mediante la iteración de soluciones en $\mathcal{P}$
$\mathcal{P}$	Clase de problemas solucionables en tiempo polinómico
$\mathbf{d}_\mathsf{u}$	histograma de duración subjetiva definido por el usuario
$\mathbf{d}$	histograma de duración subjetiva
$\mathbf{e_i}$	vector de exposición sonora por bandas de frecuencia para el archivo w_i de la base
$\mathbf{e}_\mathsf{u}$	vector de exposición sonora por bandas de frecuencia, definido por el usuario
$\mathbf{e}$	vector de exposición sonora por bandas de frecuencia

$\mathbf{h}_\mathsf{o}$ respuesta total, por bandas de frecuencia, del sistema de auralización completo que tiene en cuenta respuestas simuladas y reales

$\mathbf{q}$ vector objetivo del problema MILP

$\mathbf{u}$ vector de datos especificados por usuario

$\mathbf{x}$ vector solución exacta o vector de variables a optimizar

$\mathbf{x}^*$ vector solución óptima

b denominador de fracción de banda de octava

d_a duración de un archivo de audio (s)

d_s duración subjetiva de un evento sonoro (s)

E exposición sonora (Pa^2 s)

f frecuencia (Hz)

f_0 función objetivo en problema de optimización

f_c frecuencia central de una banda (Hz)

f_e frecuencia de aparición de eventos (Hz)

f_inf frecuencia de corte inferior de una banda (Hz)

f_sup frecuencia de corte superior de una banda (Hz)

F_S Tasa de muestreo (Hz)

g ganancia (unidad relativa de amplitud)

L_p nivel de presión sonora (dB)

L_E nivel de exposición sonora (dB)

$L_\mathsf{p,cal}$ nivel de presión sonora del calibrador (dB)

$L_\mathsf{p,FS}$ nivel de presión sonora correspondiente a una señal digital de fondo de escala para un determinado sistema de transducción y digitalización (dB)

$L_{\text{wav,FS}}$ nivel de señal digital relativo a fondo de escala (dBFS)

N_{FFT} Cantidad de muestras usadas para el cómputo de la FFT

n_{FFT} índice de frecuencia en un espectro de líneas FFT

P presión total (Pa)

p presión sonora (Pa)

P_0 presión atmosférica (Pa)

p_0 presión sonora de referencia (20 µPa)

p_{rms} presión sonora eficaz o rms (Pa)

r distancia micrófono a fuente sonora (m)

S señal en función de la frecuencia

s señal en función del tiempo

T Periodo de integración (s)

t_x instante en el cual el nivel sonoro es máximo (s)

v variable objetivo agregada en el problema MILP

x_i coeficiente de combinación de exposición

y_i coeficiente de combinación de repeticiones

z razón de banda crítica (bark)

z_{c} razón de banda crítica central (bark)

z_{inf} corte inferior de una banda crítica (bark)

z_{sup} corte superior de una banda crítica (bark)