

Restauración de señales de audio digitales recortadas con aliasing reducido

*Tesis final presentada para obtener el título de Ingeniero
de Sonido de la Universidad Nacional de Tres de Febrero
(UNTREF)*

TESISTA: Martín Messina (DNI 38532300)

TUTOR: Esp. Ing. Víctor Nasini

AGRADECIMIENTOS

Esta investigación concluye una extensa etapa de aprendizaje, esfuerzo y emociones que fue posible gracias a la colaboración de varias personas.

Principalmente, quiero dar un cálido agradecimiento a la Universidad Nacional de Tres de Febrero: a sus autoridades, personal docente y no docente, personal de la biblioteca y coordinador de la carrera por impulsar el desarrollo personal y profesional de los alumnos que la componen.

Un profundo agradecimiento a mis padres Marcela y Vicente por su enorme cariño, inmensurable sustento a lo largo de toda mi vida, además de constantemente potenciar mi deseo por la educación y el conocimiento. A mis hermanos, Nicolás y Lara por su afecto, apoyo y consideración permanente. A Milena, por ser una compañera excepcional en este viaje, llenando este último trayecto y mi vida de color.

Asimismo, este trabajo fue posible gracias a las correcciones, orientación, mirada crítica y desmesurado tiempo invertido de mi tutor Víctor Nasini. Agradezco también las importantes correcciones del profesor Mariano Girola, y las exhaustivas revisiones de mi padre.

También quisiera hacer una mención especial al profesor Ignacio Mieza, a Andrés Marafioti y Diego Leguizamón y Guillermo Marzik por consejos que me han ahorrado considerables horas de labor.

Agradezco además a mi amigo Federico Bosio por compartir la elegante plantilla de LaTeX en la que se basa el formato de este documento.

A Gastón y Fabricio, por acompañarme y apoyar mis decisiones independientemente de cuales sean. A cada uno de mis amigos, por estar presentes en este trayecto.

Finalmente, y sin restar importancia, quiero dar las gracias a mis compañeros (muchos de los cuales me alegra decir que son ahora amigos) por su ayuda desinteresada y las conversaciones, que han enriquecido mi intelecto y personalidad, a lo largo de este recorrido.

A mi madre, Marcela
Por todo su cariño y amor incondicional

ÍNDICE DE CONTENIDOS

RESUMEN	VI
ABSTRACT	VII
1 INTRODUCCIÓN	1
1.1 FUNDAMENTACIÓN	1
1.2 OBJETIVOS	3
1.2.1 Objetivo general	3
1.2.2 Objetivos específicos	3
1.3 ESTRUCTURA DE LA INVESTIGACIÓN	3
2 ESTADO DEL ARTE	4
2.1 BREVE HISTORIA DE LA RESTAURACIÓN DE AUDIO	4
2.2 AUDIO INPAINTING	5
2.3 RESTAURACIÓN DE SEÑALES RECORTADAS	7
2.4 TÉCNICAS DE REDUCCIÓN DE ALIASING	10
3 MARCO TEÓRICO	17
3.1 MODELO DE CONSISTENCIA	17
3.2 REPRESENTACIONES DISPERSAS	20
3.3 TRANSFORMADAS FRECUENCIALES	21
3.3.1 Reales y complejas	21
3.3.2 Redundancia	23
3.4 NORMAS VECTORIALES	24
3.5 PROBLEMA LINEAL INVERSO	25
3.6 REGULARIZACIÓN	25
3.7 SERIE DE FOURIER	27
3.7.1 Fenómeno de Gibbs	28
3.7.2 Promediado de Fejér	32
3.8 BÚSQUEDA DE EMPAREJAMIENTO ORTOGONAL RESTRINGIDA	35
3.8.1 Reconstrucción y procesamiento basados en segmentos	35
3.8.2 Búsqueda de Emparejamiento Ortogonal (OMP) para «inpainting»	37
3.8.3 Selección de átomo	39
3.8.4 Actualización de la solución	40
3.8.5 Mejoras algorítmicas para restauración de audio recortado	40
3.9 DESCRIPTORES OBJETIVOS Y SUBJETIVOS	43
3.9.1 Relación señal a distorsión	43
3.9.2 Evaluación Perceptiva de Calidad de Audio	44
3.9.3 Percepción auditiva de distorsión no lineal	45
4 DISEÑO DE LA INVESTIGACIÓN	46
4.1 LIMITACIONES DE C-OMP	46
4.2 IMPLEMENTACIÓN DEL ALGORITMO	47
4.3 CONFIGURACIÓN DE PARÁMETROS	50
4.4 SELECCIÓN Y PREPARACIÓN DE LAS MUESTRAS MUSICALES	51
4.4.1 Sobremuestreo	51
4.4.2 Interpolador lineal	52
4.4.3 Recorte	53
4.4.4 Filtro anti-aliasing	53
4.4.5 Diezmado	54
4.5 GENERACIÓN DE LA VERDAD FUNDAMENTAL	59

4.6 VALIDACIÓN	59
5 ANÁLISIS DE LOS RESULTADOS	61
6 CONCLUSIONES	71
7 LÍNEAS FUTURAS DE INVESTIGACIÓN	72
BIBLIOGRAFÍA	73

ÍNDICE DE FIGURAS

1	Forma de onda de tono puro recortado en el dominio digital y de tono puro recortado en el dominio analógico	2
2	Extracto del espectro de un tono puro de 7 kHz recortado 1 dB ($\varphi = 0.891$), con una frecuencia de muestreo de 48 kHz	11
3	Período de tono puro de 7 kHz recortado 1 dB ($\varphi = 0.891$), con una frecuencia de muestreo de 48 kHz	12
4	Espectro de una onda cuadrada y una señal de audio de 10 kHz recortada por tres distintas configuraciones	13
5	Comparación del modelo de recorte fuerte y de recorte suave sobre un tono de 10 kHz	15
6	Distintos modelos de consistencia en un extracto de señal de audio recortada de 5 ms de duración	19
7	La función de onda cuadrada, con período 2π	29
8	Fenómeno de Gibbs en la suma parcial de la serie de Fourier de la onda cuadrada	30
9	Fenómeno de Gibbs en la suma parcial de la serie de Fourier de la onda cuadrada y la aproximación de la misma por medio del promediado de Fejér para atenuarlo	35
10	Formación de un segmento de señal restaurada	36
11	Diagrama en bloque de la cadena de procesamiento de muestras musicales	51
12	Respuesta frecuencial del filtro pasabajos anti-aliasing Chebyshev tipo II de orden 30 utilizado	54
13	Muestra percusiva antes y después de someterse a la cadena de procesamiento	55
14	Aplicación consecutiva de cada una de las etapas de la cadena de procesamiento sobre una muestra de audio percusiva	58
15	Versión amplificada de la señal temporal de la Figura 14g luego de aplicarse el filtro anti-aliasing	59
16	Diagrama en bloque de la cadena de procesamiento de la verdad fundamental	59
17	Métricas obtenidas para distintos niveles de recorte	63

18	ΔSDR_c obtenido por LFC-OMP en cada muestra individual para distintos niveles de recorte.	65
19	Diferencia absoluta entre los descriptores calculados en LFC-OMP y C-OMP para distintos niveles de recorte	67
20	Correlación lineal entre descriptores obtenidos en ambos algoritmos	69

ÍNDICE DE TABLAS

1	Algoritmo OMP para «inpainting»	38
2	Algoritmo C-OMP para restauración de audio recortado	42
3	Posibles valores de Grado de Diferencia Objetiva	44
4	Algoritmo LFC-OMP para restauración de audio recortado con aliasing reducido	49
5	Configuración utilizada de LFC-OMP	50

RESUMEN

En esta investigación, se analiza el problema inverso de la restauración de señales de audio recortadas. Luego, se estudian las señales saturadas libres de aliasing. Ellas comprenden las señales con recorte provenientes del dominio analógico, como pueden ser las captadas en un estudio de grabación, o las saturadas en el dominio digital con posterior tratamiento de reducción de aliasing.

Se propone un algoritmo para la restauración de este tipo de señales. La solución se plantea en términos de un método de aproximación sucesiva basado en el modelo de representaciones dispersas. La restitución de la señal en el algoritmo presentado es posible debido a la estimación del umbral de recorte por medio del promediado de Fejér. El mismo se denomina Búsqueda de Emparejamiento Ortogonal Restringida de Baja Frecuencia o LFC-OMP y es una adaptación del algoritmo tradicional Búsqueda de Emparejamiento Ortogonal Restringida o C-OMP al problema de las señales saturadas con bajo contenido de aliasing.

Los resultados conseguidos indican una correlación lineal positiva entre ambos algoritmos luego de estudiar los descriptores ΔSDR_c , SDR_c , PEAQ ODG y R_{nonlin} obtenidos a partir de reconstruir señales saturadas libres de aliasing con LFC-OMP y señales recortadas en el dominio digital con C-OMP. Además, se concluye que las diferencias absolutas entre cada uno de los descriptores son poco significativas. Estos resultados indican que LFC-OMP puede ser utilizado para la restauración de señales recortadas con aliasing reducido, del mismo modo que C-OMP es usado para restaurar señales que fueron saturadas digitalmente.

Palabras clave: Audio, declipping, inpainting, recorte, aliasing, disperso.

ABSTRACT

In this work, the audio declipping inverse problem is tackled. Then, aliasing-free clipped signals are studied. They comprise both signals clipped in the analog domain, e.g. those coming from a recording studio, and signals clipped in the digital domain subsequently processed with aliasing reduction techniques.

An algorithm to restore this type of signals is proposed. The solution arises in terms of a successive approximation method based on the sparse signals framework. Signal restitution in the presented approach is possible due to the clipping threshold estimation's by means of the Fejér's averaging method. This procedure, coined Low Frequency Constrained Orthogonal Matching Pursuit or LFC-OMP, is a variation of the traditional algorithm Constrained Orthogonal Matching Pursuit or C-OMP to the use case of the aliasing-free clipped signals.

The results state a positive linear correlation between both algorithms after studying the metrics ΔSDR_c , SDR_c , PEAQ ODG y R_{nonlin} obtained from restoring aliasing-free clipped signals with LFC-OMP and digitally clipped signals with C-OMP. Also, the absolute differences between each of these metrics are insignificant. These results imply that LFC-OMP can be used to restore clipped aliasing-free signals, in the same way that C-OMP is used to restore signals which were clipped in the digital domain.

Keywords: Audio, declipping, inpainting, clipping, aliasing, sparse.

1 INTRODUCCIÓN

1.1 FUNDAMENTACIÓN

El recorte o saturación es un problema recurrente en el contexto del procesamiento de señales. El mismo se presenta en audio, imágenes y comunicaciones digitales durante la adquisición, reproducción y conversión analógica-digital.

Este ocurre por la superación del nivel de tensión permitido por un circuito analógico, o por el alcance del máximo o mínimo dígito de representación en sistemas digitales. Los casos más simples comprenden audio con clicks o ligeras explosiones, mientras que los más severos incluyen audio altamente degradado con ruido aditivo causado por la presencia de armónicos en altas frecuencias. Este fenómeno modifica la calidad percibida del audio y perjudica el rendimiento de sistemas de reconocimiento automático de voz [1, 2, 3]. No obstante, también puede ser útil para evitar la sobremodulación en transmisores de audio [4].

Esta investigación trata el problema inverso de la restauración de señales de audio recortadas (en inglés, «declipping»). Es decir, la reconstrucción de una señal de audio a partir de su versión saturada.

Diversos acercamientos fueron utilizados en el pasado para resolver este problema. Entre ellos están la minimización restringida de cuadrados mínimos, los criterios psicoacústicos, el sensado compresivo (o «compressive sensing»), el aprendizaje de diccionarios, la factorización de matrices no negativas, y las representaciones dispersas o codispersas, todos exhaustivamente estudiados por Záviska et al. [2, 5, 6, 7, 8, 9, 10, 11, 12, 13].

Los métodos y algoritmos para reconstrucción de señales anteriormente mencionados reducen también en gran medida la distorsión por aliasing [13]. Sin embargo, todos ellos tratan exclusivamente la restauración de señales recortadas en el dominio digital. Cuando la saturación ocurre en el dominio analógico el proceso es distinto ya que, antes de la conversión analógica-digital, un filtro pasabajos es utilizado para mitigar el aliasing de la señal [13].

En este caso, existe una dificultad en la detección de muestras recortadas a estimar, ya que estas no conforman una secuencia de amplitud constante como cuando el recorte ocurre en el dominio digital (Figura 1). Lo mismo ocurre con señales recortadas en el dominio digital y procesadas con compresores, sobremuestreo o cualquier técnica de reducción de aliasing. Por consiguiente, nuevas metodologías para la reconstrucción de señales recortadas con aliasing reducido deben ser implementadas.

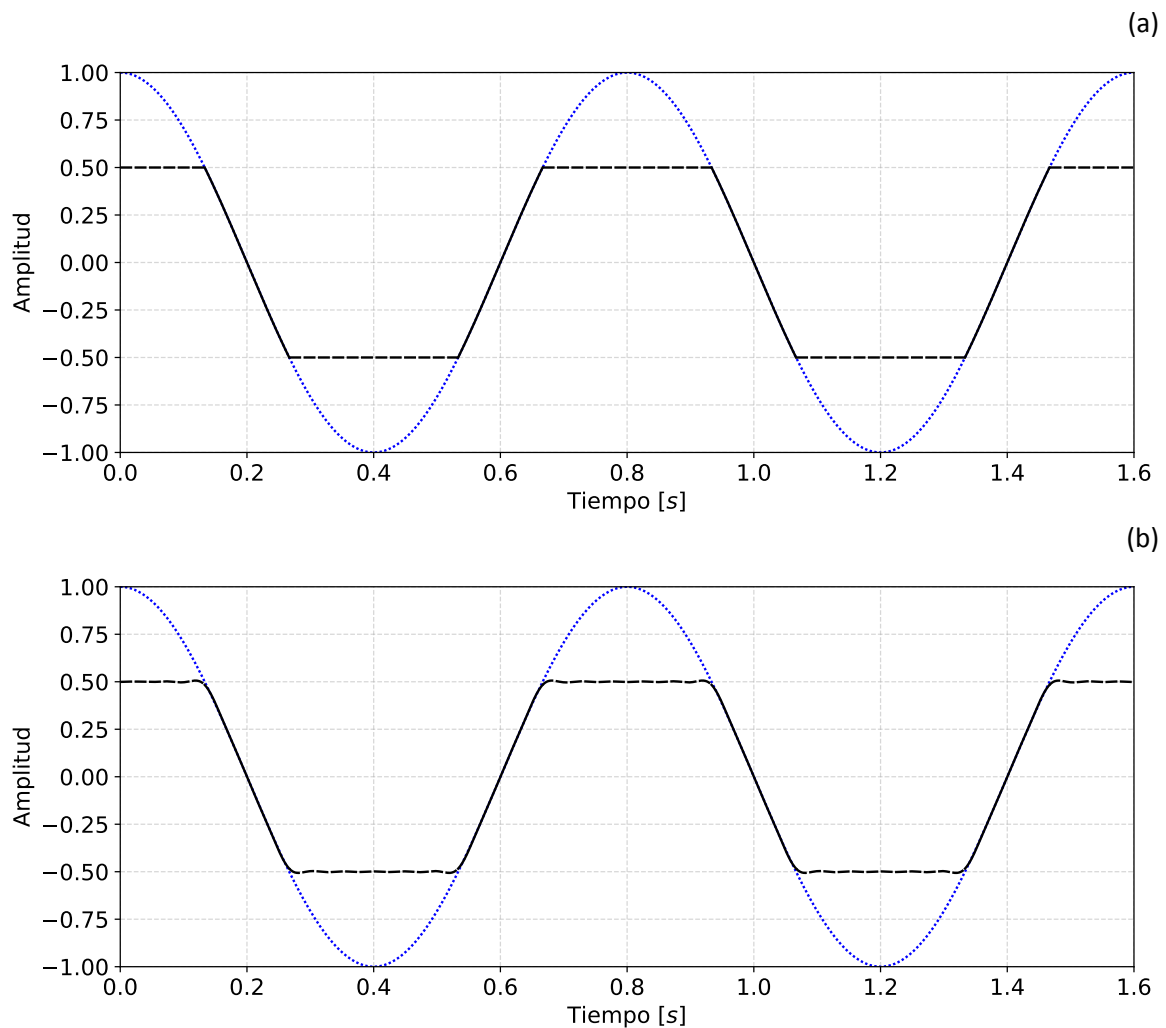


Figura 1: Forma de onda de tono puro recortado en el dominio digital (a) y de tono puro recortado en el dominio analógico (b). En ambos casos se denota el tono original con puntos. Las ondulaciones que se presentan en la zona de recorte del tono saturado analógicamente, delimitada con líneas punteadas, no permiten la detección de las muestras degradadas a restaurar por medio de operaciones triviales.

1.2 OBJETIVOS

1.2.1 Objetivo general

Restaurar señales de audio digitales degradadas por recorte en el dominio analógico, o saturadas en el dominio digital con posterior procesamiento de reducción de aliasing.

1.2.2 Objetivos específicos

1. Estudiar los modelos matemáticos relevantes al problema.
2. Analizar un conjunto de algoritmos utilizados en la actualidad.
3. Estructurar una solución algorítmica acorde al problema.
4. Implementar el algoritmo desarrollado en un lenguaje de programación de alto nivel.
5. Seleccionar y preparar las muestras musicales con las que evaluar el método numérico.
6. Validar el método numérico utilizando métricas acordes y comparando resultados con los de un algoritmo similar.

1.3 ESTRUCTURA DE LA INVESTIGACIÓN

En la sección 2 se expone el estado del arte de la disciplina de reconstrucción de audio. La sección 3 desarrolla el marco teórico sobre el cual se sustenta el presente estudio. El diseño y la metodología de la investigación se detallan en la sección 4, incluyendo el tipo y el alcance de la misma. Finalmente, se analizan los resultados en la sección 5, para derivar de forma posterior las conclusiones del estudio en la sección 6.

2 ESTADO DEL ARTE

2.1 BREVE HISTORIA DE LA RESTAURACIÓN DE AUDIO

Las técnicas de restauración de audio analógicas existen al menos desde la invención de la cinta magnética en 1951. En esta época, era común la utilización de corte y empalme para remover ligeras explosiones y la ecualización frecuencial para reducir el ruido de fondo [14].

Reductores de clicks de la época utilizaban el filtrado de bajas frecuencias para la identificación de este tipo de transitorios, y de altas frecuencias para enmascararlos [15, 16]. Sin embargo, al reducir de forma significativa el ruido de las señales tratadas, estos métodos a la vez perjudican la calidad de la señal a restaurar. En contraste, las técnicas de “pre-énfasis” de Dolby han sido muy exitosas para reducir el ruido de fondo en cintas analógicas [17]. Aún así, su desventaja reside en que el “pre-énfasis” tiene que ser agregado a la señal durante la etapa de grabación.

Los métodos digitales de restauración de audio son más flexibles que los analógicos aunque un exceso en la aplicación de alternativas inapropiadas puede derivar en peores resultados que tratamientos analógicos. Trabajos pioneros en restauración de audio incluyen la deconvolución de señales para mejorar las cualidades de una voz solista de una fuente grabada acústicamente [18, 19].

Todos los fenómenos anteriormente mencionados pertenecen a la clase de degradaciones globales, las cuales afectan a todas las muestras de la señal e incluyen al ruido de fondo y ciertos tipos de distorsión no lineal particulares de reproductores analógicos con componentes rotativos como el «wow» y «flutter», en donde como producto de una falla mecánica se generan cambios de velocidad de larga y corta duración respectivamente. Por otro lado, existe la clase de degradaciones locales que comprenden discontinuidades en la forma de onda y a su vez afectan sólo a algunas muestras. Entre ellas se encuentran los clicks, rayaduras, crujidos, roturas, pérdida de muestras, ruidos impulsivos y recorte [14].

Se ha dedicado un gran esfuerzo a la restauración de señales de audio corrompidas a

partir de vinilos rayados y grabaciones viejas [20]. En este problema, intervalos de muestras degradadas desde 20 μ s a 4 ms ocurren en ubicaciones aleatorias. Otros ejemplos incluyen las rayaduras de CD y la pérdida de paquetes en teléfonos inalámbricos o de voz sobre IP (VoIP). En estos casos se emplea un algoritmo para restaurar las muestras degradadas de una forma similar a la reconstrucción de imágenes [21]. Sin embargo, en el campo del audio cada uno de estos problemas fueron analizados por separado y dependiendo del contexto las técnicas han sido nombradas interpolación, extrapolación, imputación, inducción, extensión de ancho de banda o encubrimiento [20].

2.2 AUDIO INPAINTING

No fue hasta el 2011 que Adler et al. inspirados en la restauración de imágenes unieron las técnicas mencionadas anteriormente para desarrollar la metodología de «Audio Inpainting» [20]. En esta especialidad, las muestras degradadas se consideran perdidas y las mismas son estimadas a partir de una técnica de restauración de audio arbitraria. Cabe remarcar que la estimación en esta disciplina es libre de restricciones, a diferencia de la restauración de audio recortado donde las muestras estimadas están condicionadas por un umbral de recorte de modo que el valor de las mismas sea superior o igual al valor absoluto de este. En otras palabras, cuando se restaura una muestra recortada positiva, el valor de la estimación debe ser superior o igual a la magnitud de la muestra degradada. Contrariamente, cuando se estima una muestra recortada negativa, el valor de la misma debe ser inferior o igual a la magnitud de su contraparte saturada. Este proceso se detalla en la sección 3.1.

Un tratamiento acorde a esta metodología es el de Janssen et al. quienes modelan la señal como un proceso autoregresivo (AR) y estiman las muestras perdidas a partir de una combinación lineal de las muestras anteriores [22]. Los factores de esta operación son denominados coeficientes autoregresivos, y el número total de los mismos define el orden del modelo. El método requiere una gran cantidad de muestras confiables vecinas, en otras palabras, adyacentes no degradadas. Záviska et al. reportan resultados comparables con

métodos de estado del arte cuando esta técnica se utiliza de forma iterativa en hardware moderno [13]. Sin embargo, al obtener las estimaciones a partir de valores de muestras vecinas, el método no es útil para restaurar señales recortadas, cuyos valores de muestras estimadas son distintos a los de las confiables presentes en la señal observada [13].

Por otra parte, Mallat y Zhang introdujeron en 1993 la Búsqueda de Emparejamiento (MP) como un método de aproximación sucesiva para señales dispersas y en 1994 los autores presentaron una versión mejorada, la Búsqueda de Emparejamiento Ortogonal (OMP) [23, 24]. Este último método brinda mejores resultados que el primero a cambio de un mayor coste computacional y fue posteriormente adaptado para la disciplina de «inpainting» por Adler et al. en 2011 [20]. Además, al no depender de valores de muestras anteriores a la muestra a estimar, este método es potencialmente útil para la restauración de señales de audio con recorte.

Cabe remarcar que los métodos de «inpainting» anteriormente presentados son únicamente adecuados para la estimación de segmentos faltantes de corta duración (entre 10 y 100 ms). Adler et al. estudian el rendimiento de ambos métodos a partir de un experimento en el que tratan señales de diferentes duraciones de intervalos de muestras perdidas. La duración de los mismos varía de 1 ms, correspondiente a distorsiones por ruido impulsivo, a 10 ms, correspondiente a una pérdida en transmisión de paquetes. Para esto, utilizan una colección de señales de habla de $f_s = 8$ kHz y $f_s = 16$ kHz y señales musicales de $f_s = 16$ kHz en las cuales generan segmentos de longitud fija de muestras perdidas periódicamente cada 100 ms. En el método de Janssen, el orden del modelo autoregresivo se establece como $3m + 2$, donde m es el número de muestras perdidas de la señal a restaurar, siguiendo la recomendación del autor del método en su artículo original. Los resultados para las señales musicales de $f_s = 16$ kHz, y las señales de habla de $f_s = 8$ kHz son favorables en OMP para duraciones cortas de intervalos perdidos. Mientras que para señales de habla de $f_s = 16$ kHz y para segmentos de 10 ms de duración en todos los tipos de señal, el método de Janssen es preferido. Aún así, ambos métodos tienen dificultades para estimar segmentos que duren más de 10 ms en todos los tipos de señales [20]. Los autores luego proceden a hacer un

estudio, sólo con OMP para las señales de habla de $f_s = 8$ kHz, en el cual los segmentos de muestras perdidas son generados de modo que la duración de cada uno de ellos sea diferente y su ubicación en la señal sea aleatoria, obteniendo resultados similares al experimento anterior [20].

Métodos más modernos utilizan redes neuronales generativas adversarias para la recuperación de señales con segmentos perdidos de larga duración (más de 1000 ms) [25].

2.3 RESTAURACIÓN DE SEÑALES RECORTADAS

Uno de los principales acercamientos en el área de la restauración de audio recortado o «declipping» fue el realizado por Abel y Smith en 1991 [26]. En su artículo, los autores presentan una solución a partir de un algoritmo que resuelve un problema convexo para estimar muestras de una señal saturada. El enfoque usa la técnica de sobremuestreo e interpolación con funciones sinc. Otro de los métodos pioneros en esta disciplina es el presentado por Fong y Gonsill [27]. En su investigación, los autores abordan el problema desde una perspectiva de estadística Bayesiana, considerando además el modelo autoregresivo.

Dentro de los métodos más modernos para esta disciplina se encuentra SParse Audio DEclipper (SPADE), presentado en 2015 por Kitić et al.. La versión A-SPADE del mismo aproxima de forma sucesiva la solución de un problema de optimización inverso NP-complejo. El procesamiento de la señal es secuencial, utilizando segmentos solapados. La solución se aproxima mediante el Método de Multiplicadores de Dirección Alterna (ADMM) [28]. El costo computacional del método se basa principalmente en dos procesos de descomposición llamados síntesis y análisis, explicados en detalle en la sección 3.2. Luego de estos, es aplicado un operador de umbralización fuerte («hard-thresholding»). Este último consiste en una operación trivial cuya finalidad es la de convertir en cero los elementos de un vector que no sean superiores a un umbral previamente definido. Esta operación ocurre en el dominio de los números complejos, por lo que es necesario considerar parejas de números complejos conjugados para cada elemento que compone la señal [29]. Como último paso,

una proyección ortogonal es aplicada para aproximar una solución del bloque o segmento de señal. Estos pasos se reiteran de forma iterativa hasta que una condición de energía residual mínima es alcanzada.

También es posible plantear la versión alternativa de A-SPADE, llamada S-SPADE. Aunque esta versión utiliza el algoritmo ADMM, el problema de optimización que se resuelve en este caso es distinto [11, 30]. Sin embargo, la complejidad del algoritmo se basa también en una síntesis, un análisis, una umbralización fuerte y una proyección ortogonal tal como en A-SPADE. Tanto A-SPADE, como S-SPADE demostraron buen desempeño en la mayor cantidad de descriptores en el exhaustivo estudio llevado a cabo por Záviska et al. [13]. Adicionalmente, su baja complejidad y capacidad para brindar buenos resultados sin la necesidad de un ajuste minucioso de parámetros los transforman en dos de los métodos más atractivos en la disciplina de reconstrucción de audio saturado [13].

Otro método actual, presentado por Záviska et al. en 2019, está basado en representaciones dispersas, modelos psicoacústicos y relajación convexa [6]. La idea general de la relajación convexa consiste en la reformulación de un problema no-convexo, como es generalmente la restauración de audio recortado por medio de criterios de dispersión, en uno convexo y, por lo tanto, más fácil de resolver. Para agregar la información de los modelos psicoacústicos, el método penaliza la solución por medio de una multiplicación elemento a elemento con un vector de ponderación. En la formación de este vector, el método utiliza los conceptos psicoacústicos de umbral absoluto de escucha y enmascaramiento frecuencial, a modo de representar el modelo perceptivo humano. Además, en una de sus configuraciones, emplea una parábola como vector de ponderación. La justificación de esta última elección reside en que una señal restaurada tiene un contenido espectral comprendido principalmente por frecuencias bajas, a diferencia de una señal recortada cuyo espectro además tiene un considerable contenido de armónicos de alta frecuencia. Esta última configuración es la que reporta los mejores resultados en el artículo original [6]. El problema es finalmente resuelto por el algoritmo de Douglas-Rachford, especializado en minimizar sumas de funciones convexas.

Cuando la señal es saturada de acuerdo al modelo de recorte fuerte, o en otras palabras, con umbral de recorte de amplitud constante, son utilizados métodos de interpolación para restituir la señal. Por otro lado, cuando en la misma es aplicada un recorte suave, en la forma de un procesamiento dinámico como la compresión, esta es posible de ser recuperada aplicando una función inversa compensadora a la señal saturada. Para lograr esta finalidad, diferentes enfoques a los anteriores deben ser llevados a cabo.

Gorlow y Reiss proponen un esquema de compresión-decompresión para abordar el problema. En su estudio, los autores fundamentan detalladamente un modelo de señal que permite revertir, con alta precisión y baja complejidad, el efecto de compresión en señales procesadas por un compresor de rango dinámico, a partir de conocer los parámetros del mismo [31].

Otro método dentro de esta disciplina es planteado por Ávila et al., quienes desarrollan una técnica de recuperación para señales de audio que hayan sido distorsionadas por funciones suaves y sin memoria, es decir, cuya salida únicamente depende de los valores presentes de la misma [32, 33]. La metodología en cuestión asume la dispersión de la señal y aborda la restauración como un problema de minimización de una suma ponderada de coeficientes de la transformada de coseno discreta sujeto a restricciones, empleando la norma euclídea para encontrar una solución. La magnitud de estos factores de ponderación es extraída de la señal con distorsión. De forma alternativa, Ávila y Biscainho ofrecen otro método similar en el que utilizan la misma suma ponderada pero implementando en cambio la norma $\|\cdot\|_1$ [34].

Cabe recordar que todas estas metodologías son únicamente efectivas cuando el recorte es llevado a cabo de forma digital. Las mismas no contemplan el caso de una señal recortada en el dominio digital que es posteriormente filtrada, indicado en la Figura 1b.

2.4 TÉCNICAS DE REDUCCIÓN DE ALIASING

Como consecuencia del muestreo en la conversión analógica-digital, el espectro de una señal digital es el de la señal analógica trasladado a múltiplos de la frecuencia de muestreo. Cuando la frecuencia de muestreo no es lo suficientemente alta, estos espectros se solapan, generando una distorsión conocida como aliasing [35]. Una forma de visualizar este efecto es considerando que todas las componentes que están por encima de la frecuencia de Nyquist (la mitad de la frecuencia de muestreo) en el espectro analógico, son reflejadas en el espectro digital. Por lo tanto, dado que la distorsión por recorte se extiende por todo el dominio frecuencial de la señal, el aliasing está presente cuando la saturación se realiza de forma digital.

Mapes-Riordan estudia este fenómeno y observa que en un tono de 7 kHz muestreado a 48 kHz sólo es necesario 1 dB de recorte para causar aliasing con niveles de 32 dB por debajo de la frecuencia fundamental, tal como se muestra en la Figura 2 [36]. Además destaca que el noveno armónico tiene menos energía que los subsecuentes decimoprimer y decimotercer armónicos, ambos reflejados tres veces. Esto es debido al solapamiento de armónicos mayores.

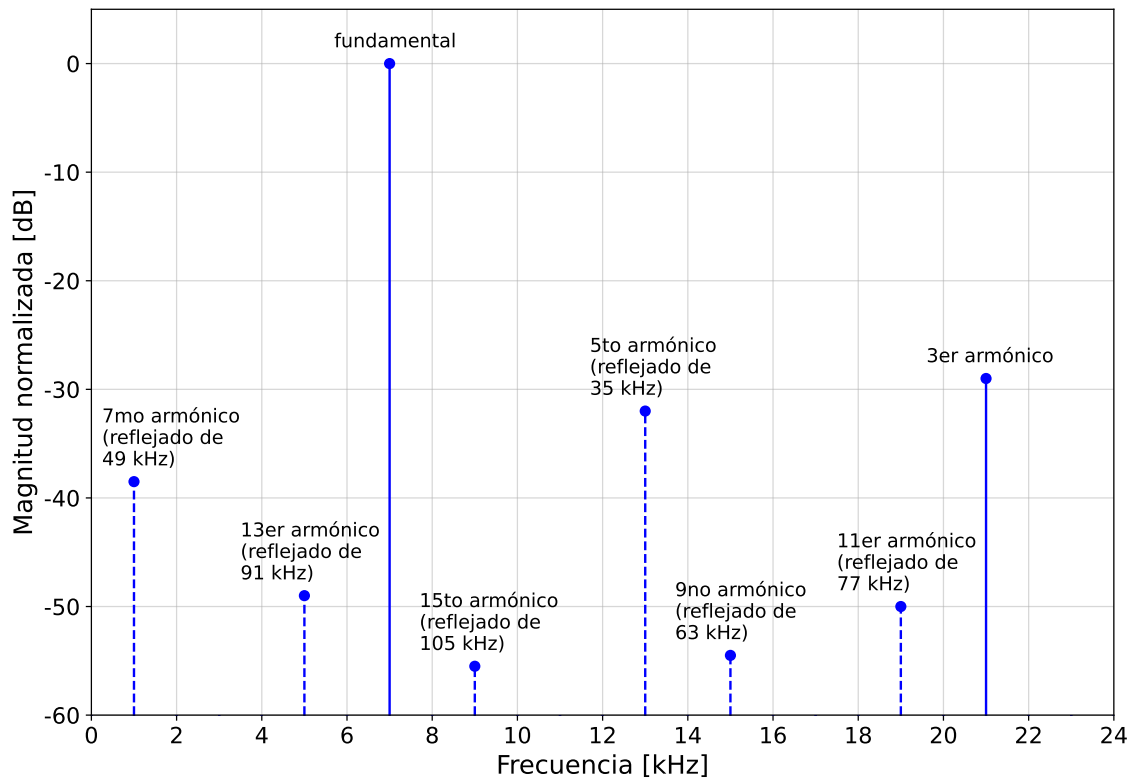


Figura 2: Extracto del espectro de un tono puro de 7 kHz recortado 1 dB ($\varphi = 0.891$), con una frecuencia de muestreo de 48 kHz. El aliasing es indicado con líneas entrecortadas. Extraído de «Mapes-Riordan» [36].

En la Figura 3 se observa la misma señal en el dominio temporal. En ella, puede visualizarse $h(t)$, una función utilizada por el autor para establecer el recorte en la señal original. La intersección sobre muestras positivas de $h(t)$ y el tono puro original denota el valor absoluto del umbral de recorte $|\varphi|$. El período T_0 es el recíproco de la frecuencia fundamental ($f_0 = 7$ kHz).

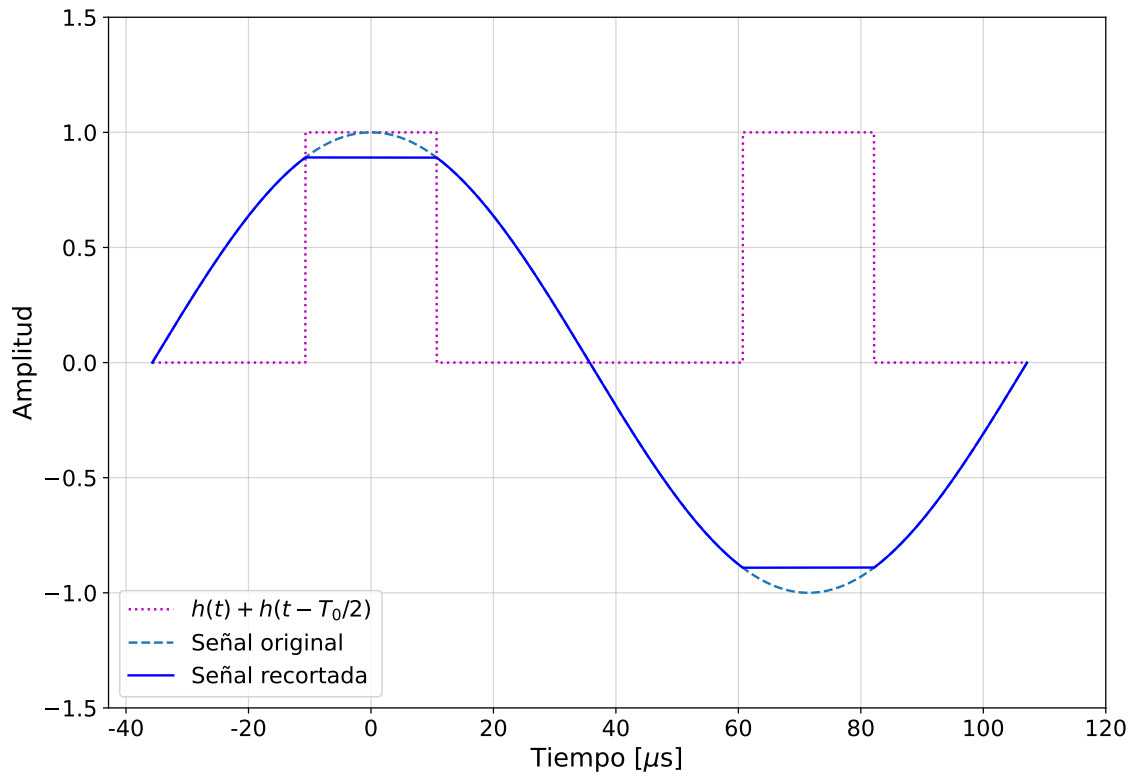


Figura 3: Período de tono puro de 7 kHz recortado 1 dB ($\varphi = 0.891$), con una frecuencia de muestreo de 48 kHz.

Luego extiende su estudio a un tono de 10 kHz recortado por tres distintas configuraciones: 20, 1 y 0.1 dB, comparadas con el espectro de una onda cuadrada para referencia. Estos 4 casos se ilustran en la Figura 4, conformada por los armónicos impares resultantes del cálculo de coeficientes de la serie de Fourier de cada una de las señales. Sus conclusiones principales son que la pendiente de atenuación de los armónicos impares, generados por este tipo de distorsión, se encuentra entre 12 dB para los casos menos afectados por el recorte y 6 dB para los casos más severos, siendo este límite el caso de la onda cuadrada [36].

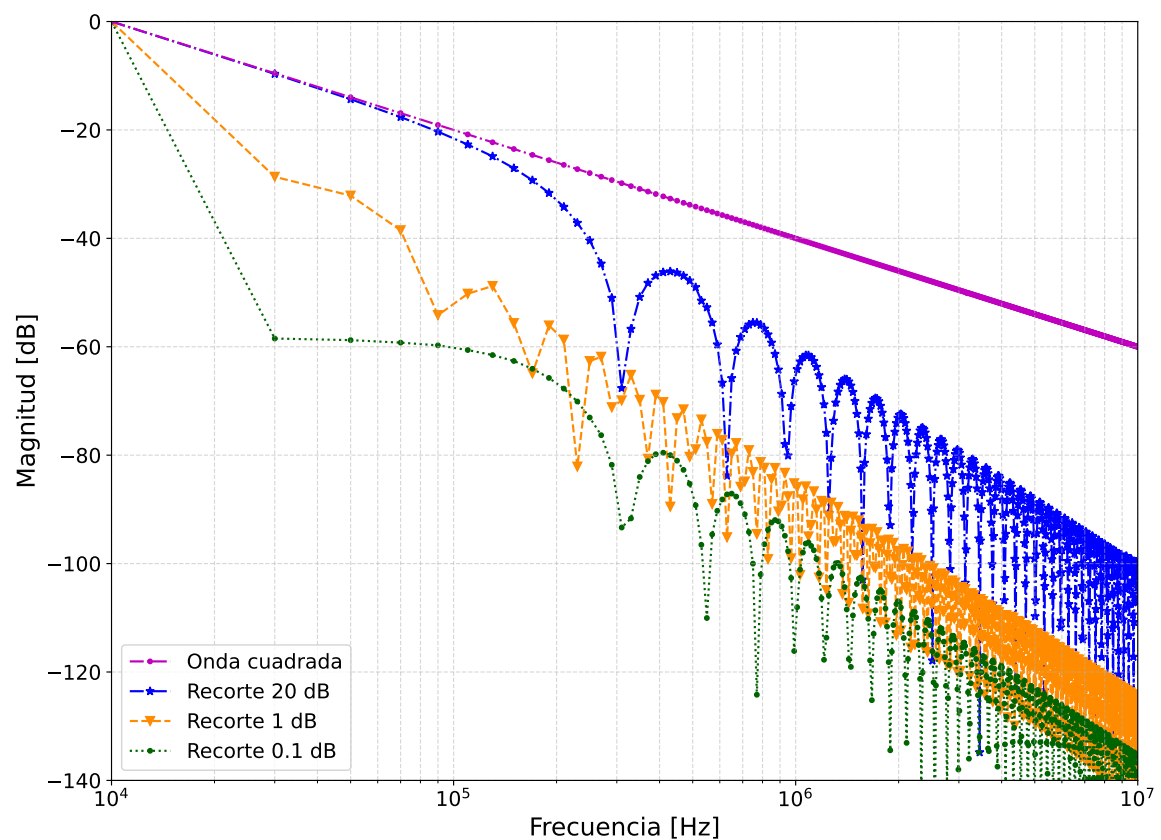


Figura 4: Espectro de una onda cuadrada y una señal de audio de 10 kHz recortada 20 dB ($\varphi = 0.1$), 1 dB ($\varphi = 0.891$) y 0.1 dB ($\varphi = 0.99$). La magnitud es relativa a la frecuencia fundamental (10 kHz) en el eje de ordenadas, mientras que la frecuencia es logarítmica en el eje de abscisas. Extraído de «Mapes-Riordan» [36].

Kraght generaliza esta observación indicando que la atenuación para funciones con la k -ésima derivada discontinua es de $6(k+1)$ dB por octava [4]. Además, afirma que la magnitud de las frecuencias altas depende de la distancia del salto finito de estas discontinuidades y de cuan frecuentemente las mismas ocurren. Es por esto que en la Figura 4 las amplitudes del espectro de la señal recortada 0.1 dB son menores que la procesada con 1 dB, ya que en el primer caso la distancia del salto de las discontinuidades de la primer derivada es menor. En el caso de la señal recortada 20 dB, se observa en la Figura 4 que la atenuación es la misma que la onda cuadrada hasta 100 kHz, a partir de donde comienza a disminuir 12 dB por octava. Este comportamiento es causado por que el tiempo de subida (o en inglés, «rise time») de la señal es de $3.2 \mu\text{s}$ [36]. Al ser este un valor tan bajo, la transición entre valores recortados y no recortados ocurre de forma instantánea y la señal actúa como si fuese discontinua hasta 100 kHz (atenuación de 6 dB por octava), pasando esta frecuencia el tiempo de subida se ajusta y la señal comienza a comportarse como una función continua con primer derivada discontinua.

A partir de estas consideraciones, se deriva la posibilidad de atenuar el aliasing haciendo la primer derivada de la señal continua. Esto puede lograrse con un recortador suave, o en otras palabras, una compresión del rango dinámico. No obstante, esta alternativa está ligada a una relación de compromiso. La aplicación de un modelo de recorte fuerte, esto es de amplitud constante, no genera distorsiones en la señal por debajo del umbral φ , mientras que el suave, o la compresión, comienza a modificar la señal por debajo de este, acortando el rango dinámico del segmento de señal no recortado [4]. Ambos tipos de recorte se observan en la Figura 5.

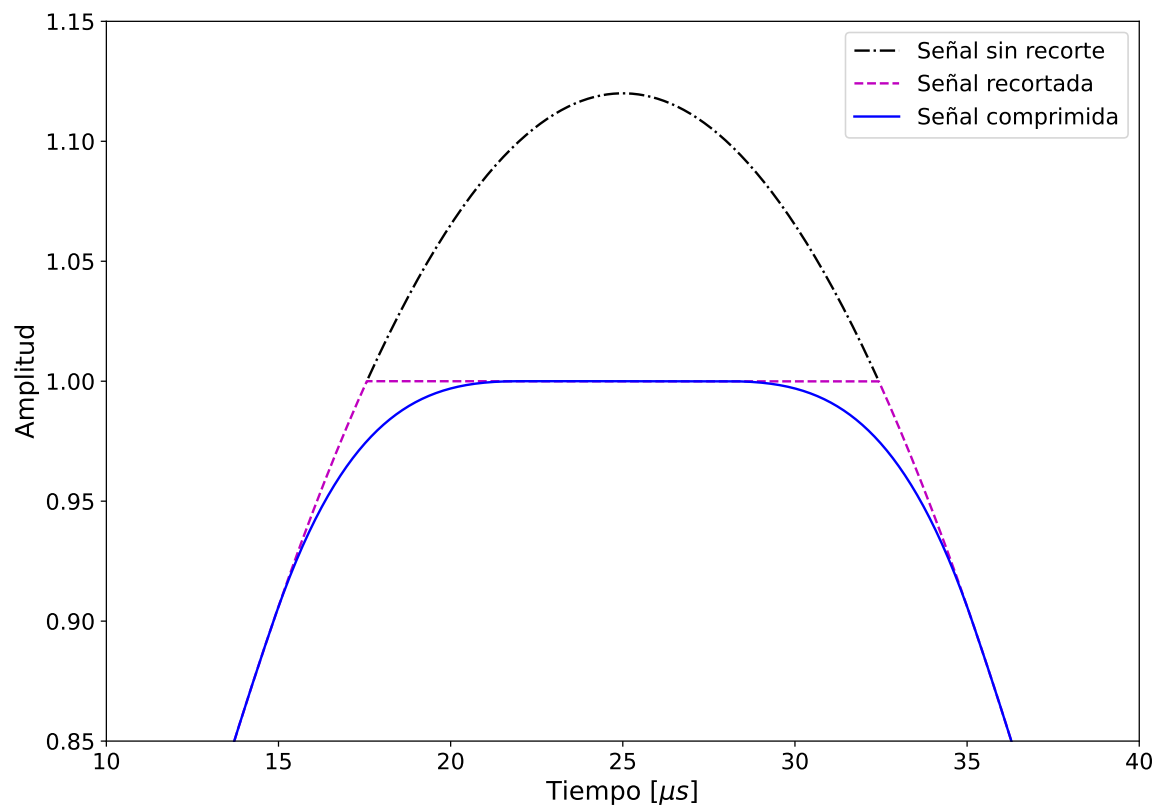


Figura 5: Comparación del modelo de recorte fuerte y de recorte suave sobre un tono de 10 kHz. Extraído de «Mapes-Riordan» [36].

Kraght procede a estudiar un tono de 10 kHz tratado con los dos tipos de recorte, suave y fuerte. Aquí, el autor realiza la observación de que hasta 50 kHz los espectros son iguales debido a que el tiempo de transición que hay entre las muestras no recortadas y recortadas es de 7 μs , el cual es un valor bajo para longitudes de onda menores a 50 kHz, por lo que la transición es prácticamente instantánea para estas frecuencias y la saturación suave no se diferencia de la fuerte. Para valores mayores a 50 kHz, este tiempo se ajusta y el espectro de la señal con saturación fuerte tiene una pendiente de atenuación de 12 dB por octava alcanzando niveles inferiores a los -100 dB, relativos a 10 kHz, para frecuencias mayores a 2330 kHz, mientras que la señal con recorte suave tiene una atenuación de 18 dB por octava (discontinua en la segunda derivada) alcanzando notablemente antes los -100 dB, a los 610 kHz [4]. Si bien esta técnica es útil para reducir el aliasing, requiere la disposición de la señal

original sin distorsión, lo cual en muchas ocasiones no es posible de cumplir.

Otro procedimiento para reducir el aliasing a niveles por debajo de -100 dB con respecto a la frecuencia fundamental, consiste en la aplicación de tasas de muestreo en el orden de los mega-hercios [36]. El mismo se denomina sobremuestreo y conlleva un coste computacional elevado [37].

Asimismo, existen técnicas de atenuación de aliasing que implican no modificar la tasa de muestreo ni el rango dinámico del audio en cuestión, además de no requerir la versión del mismo sin procesar. Esqueda et al. proponen un método numérico de reducción de aliasing para señales recortadas [38]. Esta metodología, denominada polyBLAMP, es beneficiosa para la transmisión de radio y la producción musical ya que en estas disciplinas abunda el procesamiento de audio con limitadores para maximizar el nivel de la señal y su sonoridad, por lo que la restauración no es una prioridad. El método principalmente consiste en la corrección de muestras cercanas a las recortadas a partir de una función rampa de referencia. La primera versión del método corrige 2 muestras y para una sinusoidal con frecuencia fundamental entre 400 y 3100 Hz logra una relación señal a ruido de aproximadamente 12 dB, comparable al obtenido por un sobremuestreo con un factor de 4, es decir, llevada a una tasa de muestreo de 4 veces su frecuencia de muestreo original. La segunda versión del método, por otra parte, alcanza 20 dB de relación señal a ruido corrigiendo 4 muestras. Estas técnicas permiten obtener las ventajas de altas tasas de muestreo, sin su elevado coste computacional. Además, ambas cuentan con la ventaja adicional de que no producen sobrepico (en inglés, «overshoot»), presente en las señales procesadas utilizando sobremuestreo. Aún así, no garantizan resultados para señales con frecuencias fundamentales fuera de la banda de 400 a 3100 Hz [38].

3 MARCO TEÓRICO

En esta sección se expone el cuerpo de conocimiento actual sobre el cual se sustenta la presente investigación. Cabe destacar que el apartado no comprende un desarrollo exhaustivo, sino que únicamente trata de dar al lector una visión general de las herramientas teóricas utilizadas.

3.1 MODELO DE CONSISTENCIA

El modelo de recorte fuerte de la ecuación (1), es una idealización del efecto en cuestión en una señal de audio arbitraria [13]. A partir del mismo se puede deducir la señal observada $\mathbf{y} \in \mathbb{R}^L$ a través de las muestras de la señal original $\mathbf{x} \in \mathbb{R}^L$ y el umbral de recorte φ . Además, mediante el mismo es posible determinar los conjuntos Ω_r , Ω_c^+ y Ω_c^- que nos permiten clasificar los índices de los elementos de la señal saturada. Estos se corresponden respectivamente a los índices de las muestras sin distorsión, y las muestras recortadas positivas y negativas.

$$\mathbf{y}(x_i) = \begin{cases} x_i & |x_i| \leq \varphi \\ \text{sgn}(x_i)\varphi & |x_i| > \varphi \end{cases} \quad (1)$$

El objetivo es estimar una señal \mathbf{x} a partir de otra señal observada \mathbf{y} . La señal estimada $\hat{\mathbf{x}}$ debe cumplir con la condición de las expresiones (2a), (2b) y (2c) para ser consistente con el modelo de la ecuación (1).

$$\mathbf{M}_r \hat{\mathbf{x}} = \mathbf{M}_r \mathbf{y} \quad (2a)$$

$$\mathbf{M}_c^+ \hat{\mathbf{x}} \geq \mathbf{M}_c^+ \mathbf{y} \quad (2b)$$

$$\mathbf{M}_c^- \hat{\mathbf{x}} \leq \mathbf{M}_c^- \mathbf{y} \quad (2c)$$

M_r , M_c^+ y M_c^- llamados operadores de restricción, son matrices reducidas por fila derivadas de una matriz identidad $M \in \mathbb{R}^{L \times L}$ que se utilizan para extraer las muestras indexadas por los conjuntos Ω_r , Ω_c^+ y Ω_c^- respectivamente. Cuando la señal estimada cumple con las condiciones de la ecuación (2) se denota $\hat{\mathbf{x}} \in \Gamma(\mathbf{y})$. Por brevedad, se utiliza Γ para referirse a $\Gamma(\mathbf{y})$, asumiendo de forma implícita la dependencia del conjunto con el vector \mathbf{y} en el resto del documento. Adicionalmente, las restricciones (2a), (2b) y (2c) constituyen los conjuntos Γ_r , Γ_c^+ y Γ_c^- de forma respectiva.

Cuando un modelo cumple con (2) se considera completamente consistente. Aquí, las muestras estimadas coinciden con las muestras sin distorsión de la señal observada y a la vez cumplen con las restricciones impuestas por los umbrales de recorte positivo y negativo en las muestras con distorsión, como se observa en la Figura 6. El incumplimiento de (2) deriva en un modelo completamente inconsistente. Si, en cambio, sólo se cumple con (2b) y (2c), $\hat{\mathbf{x}} \in \Gamma_c^+ \cap \Gamma_c^-$, el mismo es únicamente consistente en el segmento de señal recortada. El cumplimiento de (2a) prescindiendo de (2b) y (2c), $\hat{\mathbf{x}} \in \Gamma_r$, conlleva a un modelo consistente sólo en el segmento confiable de la señal observada, ilustrado en la Figura 6. Este último caso comprende el área de investigación de «inpainting», explicada en la sección 2.2.

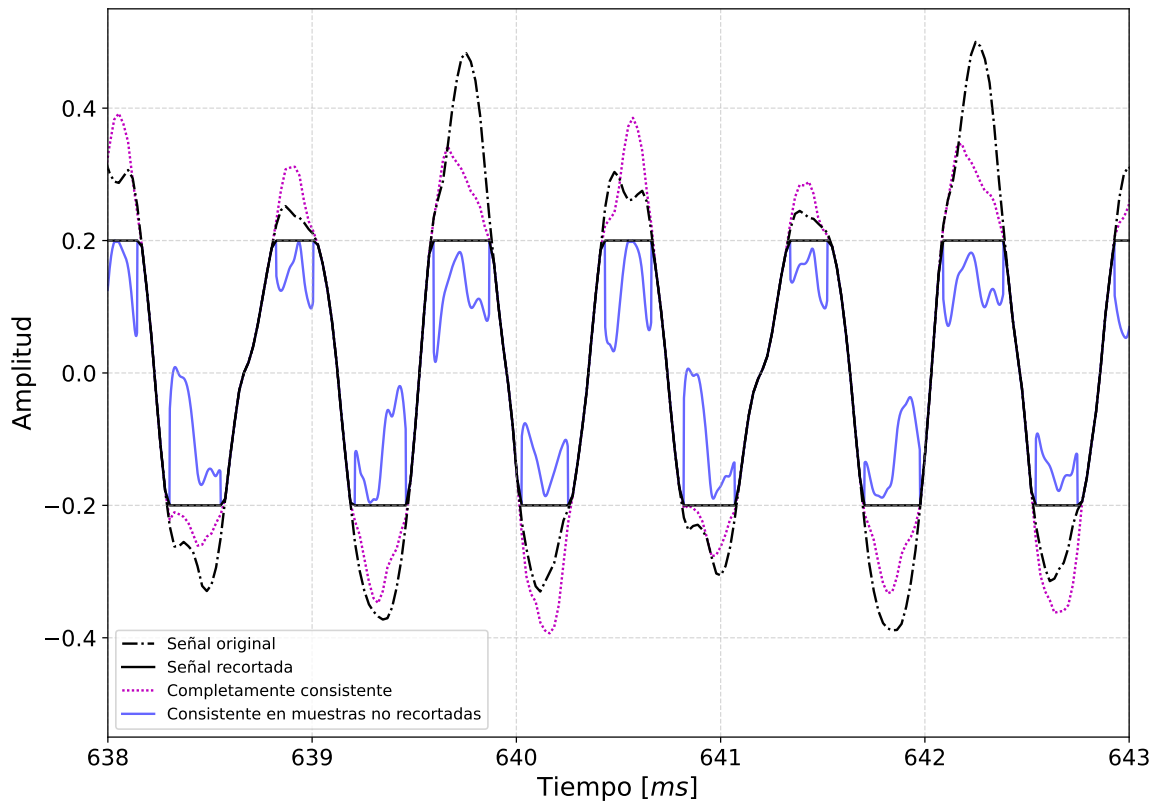


Figura 6: Distintos modelos de consistencia en un extracto de señal de audio recortada de 5 ms de duración ($\varphi = 0.2$).

Hasta ahora, el problema de la ecuación (2) es completamente consistente, pero también está mal planteado en el sentido de Hadamard, por lo que no puede asegurarse la unicidad de $\hat{\mathbf{x}}$ [39]. En otras palabras, el problema está bien planteado si:

- Tiene solución.
- La solución es única.
- La solución depende de las observaciones \mathbf{y} .

En el caso de (2) se tiene una solución que depende de las observaciones, pero la cual no es única. Por lo tanto, es necesario regularizarlo a partir de considerar un modelo de señal.

3.2 REPRESENTACIONES DISPERSAS

Las representaciones dispersas pueden pensarse como una extensión del principio filosófico de la Navaja de Ockham llevada a la disciplina del procesamiento de señales [40]. El mismo plantea que en una situación problemática con igualdad de condiciones, la solución más simple es la preferida. Esta idea no debe confundirse con el razonamiento de que una hipótesis o conjetura debe ser simple, sino más bien, de que el modelo teórico construido en base a ella debe serlo [40].

Del mismo modo, este modelo sugiere que existe un dominio de transformación en el cual la energía de una señal está concentrada en sólo algunos de sus elementos. En otros términos, esta es posible de ser representada a través de un vector compuesto por una mínima cantidad de elementos no nulos. Una señal de audio arbitraria $\mathbf{x} \in \mathbb{R}^L$ puede expresarse a partir de la combinación lineal de las columnas de una matriz de componentes elementales y los coeficientes de un vector disperso. Esta matriz se denomina diccionario u operador de síntesis $\mathbf{D} \in \mathbb{C}^{L \times P}$ con $P \geq L$, mientras que el vector se denota $\mathbf{z} \in \mathbb{C}^P$ tal como se muestra en la ecuación (3).

$$\mathbf{x} \approx \mathbf{D}\mathbf{z} \quad (3)$$

También es posible utilizar el llamado operador de análisis $\mathbf{A} \in \mathbb{C}^{Q \times L}$ con $Q \geq L$ para obtener la representación

$$\mathbf{z} \approx \mathbf{A}\mathbf{x}. \quad (4)$$

Tanto las columnas del diccionario u operador de síntesis \mathbf{D} , como las del operador de análisis \mathbf{A} son denominadas átomos en este contexto. Los problemas (3) y (4) son distintos, a menos que $\mathbf{A} = \mathbf{D}^{-1}$ y $P = Q = L$. La ecuación (3) se denomina modelo disperso de síntesis, mientras que la ecuación (4) es referida como modelo codisperso de análisis. Encontrar la solución más dispersa en términos de síntesis o análisis es por lo general inviable, pero existen algoritmos codiciosos o heurísticas convexas que permiten aproximar una solución [10].

3.3 TRANSFORMADAS FRECUENCIALES

Las transformadas frecuenciales son necesarias ya que, a pesar de que una señal discreta en el dominio temporal es fácil de obtener a partir de la medición con un micrófono, la misma no es dispersa.

3.3.1 Reales y complejas

Entre las de dominio complejo, está la transformada discreta de Fourier (DFT). A partir de considerar una señal en el dominio del tiempo $\mathbf{x} = (x_0, \dots, x_{L-1}) \in \mathbb{C}^L$ y su representación en el dominio frecuencial $\mathbf{z} = (z_0, \dots, z_{L-1}) \in \mathbb{C}^L$, la DFT y su inversa se expresan en las ecuaciones (5) y (6) respectivamente, donde i es el número complejo definido $i^2 = -1$.

$$z_j = \frac{1}{L} \sum_{n=0}^{L-1} x_n \cdot \exp\left(-\frac{i2\pi nj}{L}\right) \quad (5)$$

$$x_n = \sum_{j=0}^{L-1} z_j \cdot \exp\left(-\frac{i2\pi nj}{L}\right) \quad (6)$$

Podemos reescribir (5) y (6) en forma matricial. Por ejemplo, la forma matricial de la DFT se indica en (7) junto con su inversa en la expresión (8).

$$\mathbf{z} = \mathbf{A}\mathbf{x}, \quad \mathbf{A}_{nj} = \frac{1}{L} \exp\left(-\frac{i2\pi nj}{L}\right) \quad (7)$$

$$\mathbf{x} = \mathbf{D}\mathbf{z}, \quad \mathbf{D}_{jn} = \exp\left(-\frac{i2\pi nj}{L}\right) \quad (8)$$

donde $\mathbf{A} \in \mathbb{C}^{L \times L}$ es la transformación frecuencial directa y $\mathbf{D} \in \mathbb{C}^{L \times L}$ es la transformación frecuencial inversa. Se observa que la expresión (7) corresponde al modelo codisperso de análisis previamente definido en (4), mientras que (8) corresponde al modelo disperso de síntesis de (3).

Una herramienta ampliamente utilizada en audio, derivada de la DFT y de dominio real,

es la transformada de coseno discreta (DCT). Dada una señal temporal $\mathbf{x} \in \mathbb{R}^{L_c}$, su DCT se define junto con su inversa en (9) y (10) respectivamente.

$$z_j = \frac{2}{L_c} \sum_{n=0}^{L_c-1} x_n \cdot \cos \left(\frac{\pi}{L_c} \left(j + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) \right) \quad (9)$$

$$x_n = \sum_{j=0}^{L_c-1} z_j \cdot \cos \left(\frac{\pi}{L_c} \left(j + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) \right) \quad (10)$$

Otro ejemplo, es la transformada de seno discreta (DST), siendo esta preferida junto con la DCT y la DFT, debido a que el algoritmo de la transformada rápida de Fourier (FFT) permite procesar las mismas con un costo computacional menor al cuadrático [10].

Si bien utilizando los mecanismos anteriormente mencionados es posible lograr una buena representación de la señal, estos no aportan información sobre los cambios frecuenciales en el tiempo de la misma. Es por este motivo que existen las transformadas tiempo-frecuencia. Estas consideran tanto el contenido frecuencial, como la evolución temporal de la señal. El enfoque de las mismas consiste en usar una ventana deslizante en la señal en el dominio temporal y calcular su espectro para cada porción de la señal ventaneada. Si la longitud de la ventana varía con respecto a la frecuencia, se obtiene la transformación de «wavelet» discreta (DWT), en caso contrario, se consigue la transformación de Fourier de tiempo corto (STFT) de la señal en cuestión. Un ejemplo de esta última clase de representaciones, es la transformada discreta de Gabor (DGT). La misma se define junto a su inversa, para una señal temporal $\mathbf{x} \in \mathbb{R}^{L_g}$ en las ecuaciones (11) y (12) respectivamente, donde las componentes de la representación están indexadas por el conjunto continuo de pares ordenados $\Psi \triangleq \llbracket 0, L_g - 1 \rrbracket \times [0, 2\pi)$.

$$z_j = \frac{2}{L_g} \sum_{(n,\phi) \in \Psi} x_{(n,\phi)} \cdot \cos \left(\frac{\pi}{L_g} \left(j + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) + \phi \right) \quad (11)$$

$$x_n = \sum_{(j,\phi) \in \Psi} z_{(j,\phi)} \cdot \cos \left(\frac{\pi}{L_g} \left(j + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) + \phi \right) \quad (12)$$

Estas son también posibles de expresar en su forma matricial compacta, en la cual (11) y (12) se transforman en (13) y (14),

$$\mathbf{z} = \frac{2}{L_g} \mathbf{A}^g \mathbf{x} \quad (13)$$

$$\mathbf{x} = \mathbf{D}^g \mathbf{z}, \quad (14)$$

donde las columnas de su matriz asociada son $\mathbf{A}^g = \{\mathbf{a}_{(n,\phi)}^g\}_{(n,\phi) \in \Psi}$, y $\mathbf{D}^g = \{\mathbf{d}_{(j,\phi)}^g\}_{(j,\phi) \in \Psi}$ en el caso de su inversa, y $\mathbf{a}_{(n,\phi)}^g$, $\mathbf{d}_{(j,\phi)}^g$ hacen referencia a las componentes de la representación en cada caso.

La incorporación del término de fase en (11) y (12) favorece el modelado de señales de fase inicial arbitraria, aumentando la dispersión de la señal resultante y mejorando significativamente las estimaciones en términos de calidad de reconstrucción física [20]. Los aspectos algorítmicos de esta descomposición se detallan en la sección 3.8.3.

3.3.2 Redundancia

Las transformadas tiempo-frecuencia mencionadas no pueden ser simultáneamente precisas en tiempo y frecuencia. De hecho, cuanto más ancha es la ventana temporal, mayor es la resolución frecuencial, y cuanto menor es el ancho de esta ventana, mayor es la resolución temporal. Esta premisa es a menudo referida como el principio de incertidumbre de Heisenberg en el contexto del procesamiento de señales [41]. Varias soluciones han sido inventadas para reducir los efectos de esta limitación. Entre ellas se encuentra la transformada de Gabor anteriormente presentada. Está representación utiliza una ventana gaussiana para permitir la mayor resolución teóricamente posible en el dominio temporal y en el frecuencial en simultáneo [42]. Otro ejemplo son las descomposiciones de resolución múltiple en tiempo-frecuencia, donde la transformada «wavelet» es de suma importancia [41]. Otra solución a partir de representaciones clásicas es la utilización de transformadas frecuenciales redundantes [40]. La redundancia puede observarse en los modelos disperso de síntesis de (3) y codisperso de análisis de (4) cuando $P > L$ y $Q > L$ respectivamente. En ellas, esta propiedad extiende el número de átomos (columnas del diccionario o matriz de análisis). De

este modo, tanto D como A son matrices rectangulares, contrariamente a la transformación original donde ambas matrices son cuadradas. Tales diccionarios o matrices de análisis redundantes ocasionalmente son llamados “sobrecompletos” (en inglés, «overcomplete»).

Tal como es observado en la expresión (4), los productos con A son realizados usando la transformada frecuencial de longitud Q en una señal temporal \mathbf{x} , con ceros agregados, de longitud inicial L . Del mismo modo, en la ecuación (3) los productos con D son efectuados con su transformación frecuencial inversa \mathbf{z} de longitud P .

3.4 NORMAS VECTORIALES

Para cuantificar la dispersión se emplea la pseudonorma $\|\cdot\|_0$, también llamada norma cero, de un vector $\mathbf{x} \in \mathbb{R}^L$ definida en (15). Esta función otorga la cantidad de elementos vectoriales distintos a cero.

$$\|\mathbf{x}\|_0 = \sum_{n=0}^{L-1} |x_n|^0 \quad (15)$$

Dado a que la expresión (15) no cumple que para cualquier $\lambda \in \mathbb{R}$ con $\lambda \neq 0$, $\|\lambda\mathbf{x}\| \neq |\lambda| \|\mathbf{x}\|$, la misma no constituye una norma vectorial. No obstante, es particularmente útil en el contexto de representaciones dispersas.

Otra norma de interés es la euclídea, utilizada para cuantificar la energía de una señal discreta y obtener el grado de similitud entre vectores. La misma se define en la ecuación (16).

$$\|\mathbf{x}\|_2 = \sum_{n=0}^{L-1} |x_n|^2 \quad (16)$$

De forma más general, la norma $\|\cdot\|_p$ de un vector se define en la ecuación (17). De esta expresión debe considerarse que cuanto más pequeño es el valor de p , mayor es la importancia de los valores pequeños del vector en el cálculo de la norma. Contrariamente, cuando p es más grande, mayor prioridad se les da a los elementos de mayor valor del vector [40]. También debe destacarse que en el límite, cuando $p = \infty$, el valor de la norma es igual al

elemento de mayor magnitud del vector.

$$\|\mathbf{x}\|_p = \begin{cases} \left(\sum_{n=0}^{L-1} |x_n|^p \right)^{\frac{1}{p}} & 0 \leq p \leq \infty \\ \max_n |x_n| & p = \infty \end{cases} \quad (17)$$

3.5 PROBLEMA LINEAL INVERSO

En esta sección se procede a definir formalmente el problema lineal inverso, el cual es de principal importancia para la presente investigación. Dada una señal observada \mathbf{y} y una señal de interés \mathbf{x} , es planteado el problema directo

$$\mathbf{y} = \Phi \mathbf{x} \quad (18)$$

donde Φ es llamada matriz u operador de medición y codifica la relación directa entre \mathbf{x} e \mathbf{y} [40].

Abordar el problema inverso de la ecuación (18) significa estimar la señal \mathbf{x} a partir de las observaciones \mathbf{y} . Tal problema es fácil de resolver si está bien planteado. Desafortunadamente, los problemas en el contexto de la restauración de señales recortadas la mayoría de las veces no cumplen con esta condición. Es por este motivo que para resolver el problema de la ecuación (18) se necesita regularizarlo a partir de un modelo estadístico o de señal.

3.6 REGULARIZACIÓN

La regularización es utilizada para agregar información adicional a la señal que se quiere estimar. Por ejemplo, es posible estimar la señal \mathbf{x} a partir de las observaciones \mathbf{y} , con el modelo

$$\min_{\mathbf{x}} f_r(\mathbf{x}) + f_f(\mathbf{y}, \mathbf{x}) \quad (19)$$

en donde la función $f_f(\cdot)$ es el término de fidelidad y cuantifica una medida de ajuste entre las observaciones \mathbf{y} y la solución \mathbf{x} . Por otro lado, $f_r(\cdot)$ es el regularizador. Este término agrega información adicional sobre \mathbf{x} al problema.

Otro beneficio de la regularización es la prevención del sobreajuste. Este fenómeno consiste en la imposibilidad de generalización de un modelo estadístico debido a una correspondencia demasiado precisa con un conjunto de datos particular [43].

Una medida de fidelidad ampliamente usada se muestra en la ecuación (20). Esta diferencia cuadrática asegura que la estimación $\hat{\mathbf{x}}$ sea compatible con las observaciones \mathbf{y} . En otras palabras, se busca la solución con la menor distancia euclídea con respecto a la señal observada, por medio de la minimización del error cuadrático medio entre dichas señales. Si este término cuadrático es utilizado, el regularizador debe ser seleccionado de acuerdo a la aplicación y el modelo de la señal que se quiere estimar.

$$f_d(\mathbf{x}, \mathbf{y}) = \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \quad (20)$$

Dado a que, como se mencionó anteriormente, el problema de restauración de audio recortado admite infinitas soluciones y considerando el modelo de dispersión (o codispersión) en cualquier señal \mathbf{z} a recuperar, el regularizador elegido para abordar el problema en cuestión es la pseudonorma. De este modo, la tarea de optimización resultante es la estimación de una señal $\hat{\mathbf{x}}$ que tenga a la vez el menor error cuadrático medio con respecto a las observaciones \mathbf{y} y la menor norma cero, tal como se muestra en la ecuación (21).

$$\min_{\mathbf{z}} \|\mathbf{y} - \Phi\mathbf{z}\|_2^2 + \|\mathbf{z}\|_0 \quad (21)$$

Sin embargo, la minimización de la pseudonorma es un problema NP-complejo, que no es convexo ni suave, por lo que su solución directa no es viable. Es por esto que existen algoritmos codiciosos y métodos de relajación convexa que permiten obtener una solución aproximada para \mathbf{z} . Sin embargo, existe evidencia empírica que demuestra que los métodos

convexos no tienen buen rendimiento cuando son aplicados a la disciplina de restauración de audio saturado [5, 44].

3.7 SERIE DE FOURIER

Una función periódica arbitraria $f(x)$ es capaz de descomponerse en una suma convergente de funciones trigonométricas, llamada serie de Fourier, dadas las condiciones que se presentan a continuación [45]. La expansión en serie de Fourier exponencial compleja de $f(x)$ en el intervalo simétrico $(-l, l)$, periódica con período $2l$, es

$$f(x) = \sum_{k=-\infty}^{\infty} c_k \cdot \exp\left(\frac{ik\pi x}{l}\right), \quad -l < x < l \quad (22)$$

donde c_k es el coeficiente de Fourier,

$$c_k = \frac{1}{2l} \int_{-l}^l f(x) \cdot \exp\left(-\frac{ik\pi x}{l}\right) dx \quad (23)$$

Es posible usar las identidades de Euler para demostrar que las ecuaciones (22) y (23) son equivalentes a la siguiente forma trigonométrica de la serie,

$$f(x) = a_0 + \sum_{k=1}^{\infty} a_k \cos \frac{k\pi x}{l} + b_k \sin \frac{k\pi x}{l} \quad -l < x < l \quad (24)$$

donde,

$$a_0 = \frac{1}{2l} \int_{-l}^l f(x) dx \quad (25)$$

$$a_k = \frac{1}{l} \int_{-l}^l f(x) \cos \frac{k\pi x}{l} dx \quad (26)$$

$$b_k = \frac{1}{l} \int_{-l}^l f(x) \sin \frac{k\pi x}{l} dx \quad (27)$$

En el siguiente teorema, se ilustra como la convergencia de esta serie depende, en forma general, de la suavidad de la función.

Teorema 1 (Convergencia puntual de la serie de Fourier) *Si $f(x)$ es una función a trozos suave en el intervalo $(-l, l)$, su serie de Fourier converge en cada punto x del mismo y,*

$$a_0 + \sum_{k=1}^{\infty} a_k \cos \frac{k\pi x}{l} + b_k \sin \frac{k\pi x}{l} = \frac{f(+x) - f(-x)}{2} \quad (28)$$

para toda x dentro del intervalo, es decir $x \in (-l, l)$. Para los extremos $x \pm l$, la serie converge a $(f(-l+) + f(l-))/2$. Cuando $f(x)$ es continua para toda $x \in (-l, l)$ la serie de la expresión (28) converge a $(f(x) + f(x))/2 = f(x)$.

La convergencia de la serie está asegurada mientras la suma de la ecuación (28) sea infinita. En la práctica, sólo es posible calcular un número finito de coeficientes. Esto es denominado una suma parcial de serie de Fourier, o una suma de Fourier truncada, indicada en la expresión (29).

$$s_n(x) = a_0 + \sum_{k=1}^n a_k \cos \frac{k\pi x}{l} + b_k \sin \frac{k\pi x}{l} \quad (29)$$

3.7.1 Fenómeno de Gibbs

El fenómeno de Gibbs representa la dificultad de una suma parcial de la serie de Fourier en aproximar funciones en sus discontinuidades. Se manifiesta como sobrepicos e infrapicos cerca a los saltos de discontinuidad [45].

El efecto fue explicado por primera vez por Wilbraham en 1848, utilizando la onda cuadrada $f(x)_{oc}$ (donde el subíndice OC hace referencia a “onda cuadrada”) de la ecuación (30) por simplicidad, ya que las publicaciones de la época no contenían ilustraciones inmediatamente después de las ecuaciones y el autor debía describirlas cuidadosamente con palabras [46]. En la Figura 7 es posible visualizar la onda cuadrada estudiada por Wilbraham. Medio

siglo después, la función periódica diente de sierra fue analizada por Michelson, y Michelson y Stratton en 1898 por su utilidad para generar la visualización de la coordenada x en osciloscopios [47, 48]. Fue Gibbs en el mismo año, quien explicó los sobrepicos e infrapicos que la aproximación por sumas parciales de la serie de Fourier de esta función presentaba, redescubriendo el acontecimiento que llevaría su nombre [49, 50].

$$f(x)_{oc} = \begin{cases} -1 & -\pi < x < 0 \\ 1 & 0 < x < \pi \end{cases} \quad (\text{período } 2\pi) \quad (30)$$

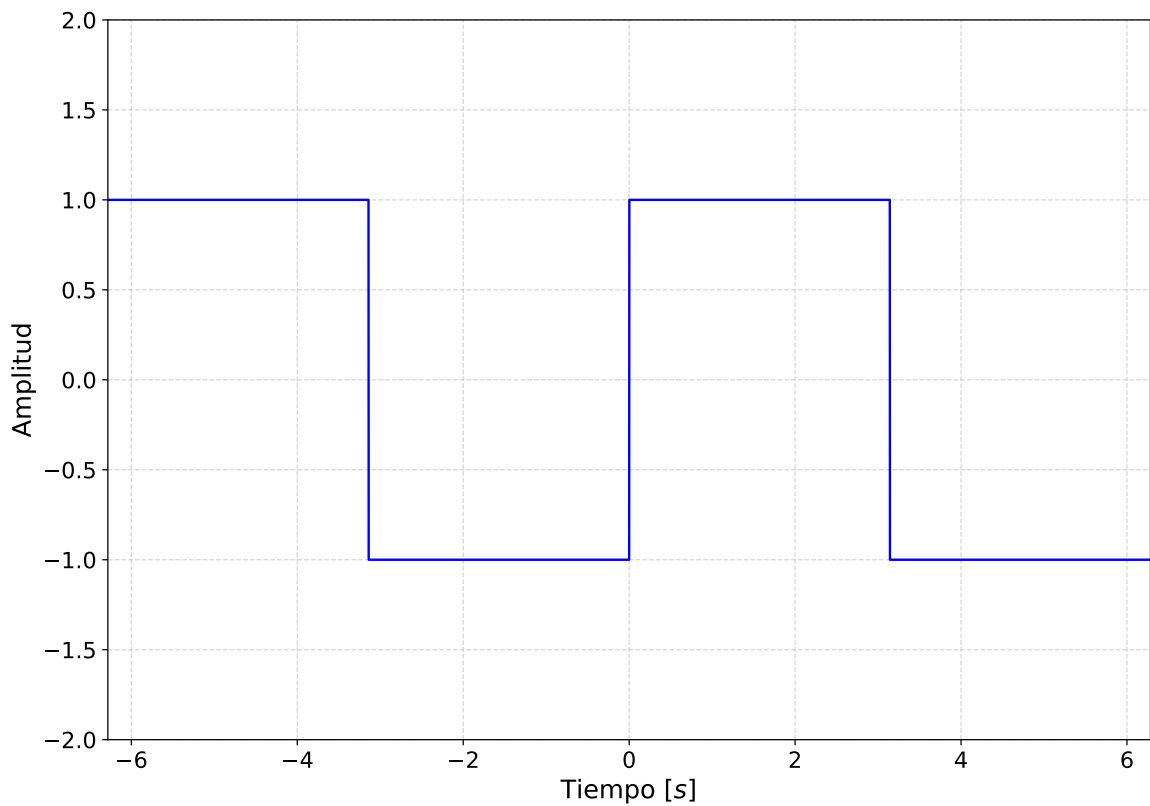


Figura 7: La función de onda cuadrada, con período 2π . Extraído de Jerri [45].

La suma de Fourier truncada de la onda cuadrada en cuestión se describe en (31). En la

Figura 8 se visualiza el fenómeno de Gibbs de la discontinuidad en $x = 0$, para $n = 10$.

$$s_n(x)_{oc} = \frac{4}{\pi} \sum_{k=1}^n \frac{1}{2k-1} \sin(2k-1)x \quad -\pi < x < \pi \quad (31)$$

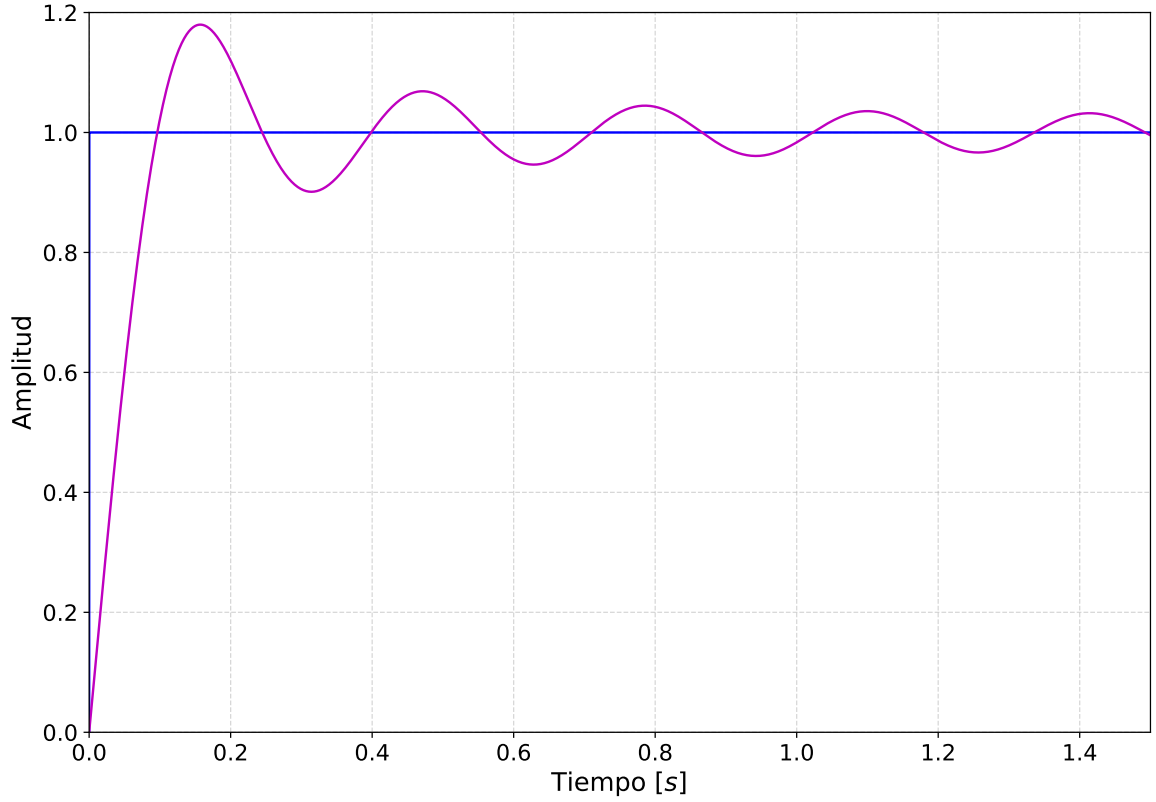


Figura 8: Fenómeno de Gibbs en la suma parcial de la serie de Fourier de la onda cuadrada, $n = 10$. Extraído de Jerri [45].

Se destaca también que para la curva de la ecuación (30) la convergencia puntual (tal como se determina en el teorema 1) de sus sumas parciales de serie de Fourier en el límite $\lim_{x \rightarrow \infty} s_n(x)_{oc}$ converge para todo x , incluyendo los saltos en las discontinuidades, $x = 0$ y $x = \pm\pi$. En este caso, el fenómeno de Gibbs no se presenta. Es únicamente cuando se lidia con las sumas parciales $s_n(x)$ de una función discontinua que el mismo se manifiesta, tal como se muestra en la Figura 8 alrededor de $x = 0$. Esto ocurre sin importar cuan grande, pero finito, sea n .

En la Figura 8 se observan los sobrepicos e infrapicos de la aproximación de una onda cuadrada en sumas parciales de una serie de Fourier para la discontinuidad en $x = 0$. Es posible demostrar que las ubicaciones de estos máximos y mínimos locales se encuentran en $x_k = k\pi/2n$ [51]. Además, la magnitud del primer sobrepico alcanza 1.17898... cuando n es lo suficientemente grande [51]. Por lo tanto, el primer sobrepico alcanza $(1.17898... - 1)/2 \approx 0.08949...$ (o aproximadamente 8.95%) de la longitud del salto en la discontinuidad. En este caso particular, la longitud del salto es de $J = 2$, como se observa en la Figura 7, por lo que la amplitud del sobrepico en $x = 0$ es de 1.17898... para valores grandes de n , ilustrado en la Figura 8 [52, 53, 54, 55]. El próximo extremo relativo es el primer infrapico. Este, en cambio, logra aproximadamente el 4.86% de la distancia del salto [53, 54, 55].

Desde la perspectiva del procesamiento de señales, una serie de Fourier truncada $s_n(x)$ es esencialmente la transformada de Fourier inversa de un filtro pasabajos $g(x)$ aplicado sobre la transformada de Fourier de la señal a expandir $f(x)$ [56]. Esto puede ser representado a partir de una convolución entre la respuesta al impulso de un filtro pasabajos ideal y la señal original, tal como se expresa en la ecuación (32)¹. En consecuencia, las oscilaciones causantes del fenómeno de Gibbs son producto de la función $\sin(nx)/x$.

$$\mathcal{F}^{-1}[f(\omega)g(\omega)] = \frac{1}{\pi} \int_{-\infty}^{\infty} f(x - \tau) \frac{\sin(n\tau)}{\tau} d\tau \quad (32)$$

Un resultado importante que se deriva de la ecuación (32) parte de definir x_n , la coordenada x del primer máximo local, el primer sobrepico, a la derecha del eje y en $s_n(x)_{oc}$. A medida que el número de términos aumenta hacia infinito, la descomposición en sumas parciales de la serie de Fourier de $f(x_n)_{oc}$ tiende a ser constante, como se demuestra en la ecuación (33).

$$\lim_{n \rightarrow \infty} s_n(x_n) = \frac{J}{\pi} \int_0^{\pi} \frac{\sin(\mu)}{\mu} d\mu = \frac{J}{\pi} G' \quad (33)$$

Este resultado, anteriormente mencionado, demuestra que el sobrepico más cercano a la discontinuidad de una serie de Fourier truncada, sólo depende del salto de la discontinuidad J y no de la función que está siendo evaluada. De este modo, el fenómeno de Gibbs no

¹Para lograr la equivalencia con $s_n(x)$ se debe adicionar $\frac{b_{1n}}{2} \sin nx$ en (32), donde para n par $b_{1n} = 0$, y para n impar $b_{1n} = 1/n$ [56].

puede remediarse a partir del mero incremento de términos en las sumas parciales de la serie de Fourier. El mismo análisis puede realizarse para el mínimo local que se encuentra del otro lado de la discontinuidad. El valor $G' = 1.8519\dots$ de la expresión (33) es denominado constante de Gibbs [56]. Otro factor relacionado, y anteriormente mencionado, que lleva el nombre de constante de Gibbs es $G = (2/\pi)G' = 1.17898\dots$ [57].

3.7.2 Promediado de Fejér

Fejér, en 1904, descubrió que sumando la serie de Fourier de una forma novedosa es posible obtener la convergencia en ocasiones donde la serie original de la ecuación (24) no es convergente. Asimismo, esta técnica es de principal utilidad para atenuar los sobrepicos e infrapicos que se manifiestan en el fenómeno de Gibbs. El promediado de Fejér de la suma parcial de una serie de Fourier, también denominado criterio de sumabilidad de Césaró $(C, 1)$, en lugar de considerar la convergencia de la suma parcial directa $s_n(x)$ para $k = 1, 2, \dots, n$ utiliza el promedio de las mismas $S_n(x)$ tal como se describe en la ecuación (34).

$$S_n(x) = \frac{s_1 + s_2 + \dots + s_n}{n + 1} = \frac{1}{n + 1} \sum_{k=0}^n s_n(x) \quad (34)$$

Posteriormente, se procede a considerar el límite del promedio,

$$\lim_{n \rightarrow \infty} S_n(x) = \lim_{n \rightarrow \infty} \frac{1}{n + 1} \sum_{k=0}^n s_n(x) \quad (35)$$

El límite de la ecuación (35) es convergente aún en la descomposición de funciones que no cumplan con el teorema 1, tal como se describe en el teorema 2. Este límite se denota $S^{(1)}$.

Teorema 2 (Convergencia de la media aritmética de Fejér) *Si una serie converge en el sentido usual, en otras palabras si $\lim_{n \rightarrow \infty} s_n = a$, entonces también es convergente mediante el promediado de Fejér, es decir $S^{(1)} = \lim_{n \rightarrow \infty} S_n = a$. Sin embargo, en ocasiones $\lim_{n \rightarrow \infty} s_n$ puede no existir mientras que $S^{(1)}$ si lo hace.*

La secuencia del promedio de Fejér de (35) es equivalente a aplicar un factor de multiplicación a los coeficientes originales de Fourier a_k y b_k de la expresión (29),

$$S_n(x) = a_0 + \sum_{k=1}^n \left(1 - \frac{k}{n+1}\right) \left[a_k \cos \frac{k\pi x}{l} + b_k \sin \frac{k\pi x}{l} \right]. \quad (36)$$

Además, $1 - \frac{k}{n+1} \leq 1$, lo cual se destaca por ser una mejora que contribuye con la convergencia de $S^{(1)}$.

La expresión (36) constituye una versión de $s_n(x)$ con un menor grado de oscilaciones. Para explicar esta afirmación, es necesario introducir el concepto de núcleo en el contexto de la serie de Fourier. A modo de simplificar la notación, se considera $l = \pi$ en las siguientes deducciones.

Es posible utilizar las expresiones de los coeficientes de Fourier (26) y (27) para reducir la suma parcial de la serie de Fourier $s_n(x)$ a la integral

$$s_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\xi) D_n(\xi - x) d\xi \quad (37)$$

donde el término $D_n(x)$ es el núcleo de Dirichlet,

$$D(n, x) \equiv D_n(x) = \frac{\sin \left(n + \frac{1}{2} \right) x}{2 \sin \frac{1}{2} x} \quad (38)$$

y como puede visualizarse en la ecuación (38), el mismo oscila entre valores negativos y positivos. Este comportamiento de $D_n(x)$ tiene gran influencia en la aparición del fenómeno de Gibbs en la descomposición por sumas parciales $s_n(x)$ de funciones con saltos de discontinuidad. Contrariamente, a partir del promediado de Fejér puede derivarse un núcleo “alisado” que reduzca la magnitud de tal efecto. Tomando el promediado de Fejér de las sumas parciales de Fourier $s_n(x)$ como se indica en la ecuación (34), y combinando las ecuaciones (35) y (37) se deduce que dicho promediado es equivalente a promediar núcleos

de Dirichlet. Por tanto, $S_n(x)$ está asociado con el núcleo de Fejér dictado por (39), a partir de considerar promedios de la expresión (38), referente al núcleo de Dirichlet.

$$M_n(x) = \frac{1}{n+1} \sum_{k=0}^n D_k(x) = \frac{1}{n+1} \sum_{k=0}^n \frac{\sin(n+\frac{1}{2})x}{2 \sin \frac{1}{2}x} \quad (39)$$

A partir de racionalizar la expresión (39), y hacer converger la serie resultante se obtiene el nuevo núcleo de Fejér expresado en la ecuación (40) [45].

$$M_n(x) = \frac{2}{n+1} \cdot \left[\frac{\sin \frac{1}{2} (n+1) x}{2 \sin \frac{1}{2} x} \right]^2 \quad (40)$$

Luego, el promediado de Fejér puede reescribirse de la forma de la ecuación (37) como

$$S_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\xi) M_n(\xi - x) d\xi \quad (41)$$

donde $M_n(x)$ está definido por la expresión (40). En contraposición con el núcleo de Dirichlet de (38), este nuevo núcleo de Fejér es no-negativo, y debido al factor $1/(n+1)$ en (40) decae hacia cero a medida que $n \rightarrow \infty$.

La positividad del núcleo de Fejér $M_n(x)$ se traduce en la eliminación de los sobrepicos e infrapicos del fenómeno de Gibbs, asociados al núcleo de Dirichlet. De hecho, es posible demostrar que $S_n(x)$ es monótono, por lo que no tiene extremos [52]. Sin embargo, como se ilustra en la Figura 9, la descomposición mediante $S_n(x)$ de la onda cuadrada yace debajo del valor de 1, que debe aproximar, en todo momento. Esta característica del promediado de Fejér está presente siempre que n es finito, y se ajusta a medida que $n \rightarrow \infty$.

En este punto, debe destacarse que el promediado de Fejér tiene, obviamente, una complejidad adicional debida al cálculo de las sumas parciales de la serie de Fourier de la señal y su posterior promedio, como se observa en la expresión (34). Aún así, es posible emplear la recursión en el cómputo de tal promedio, ya que, para obtener la suma parcial s_{n+1} es necesario primero calcular s_n .

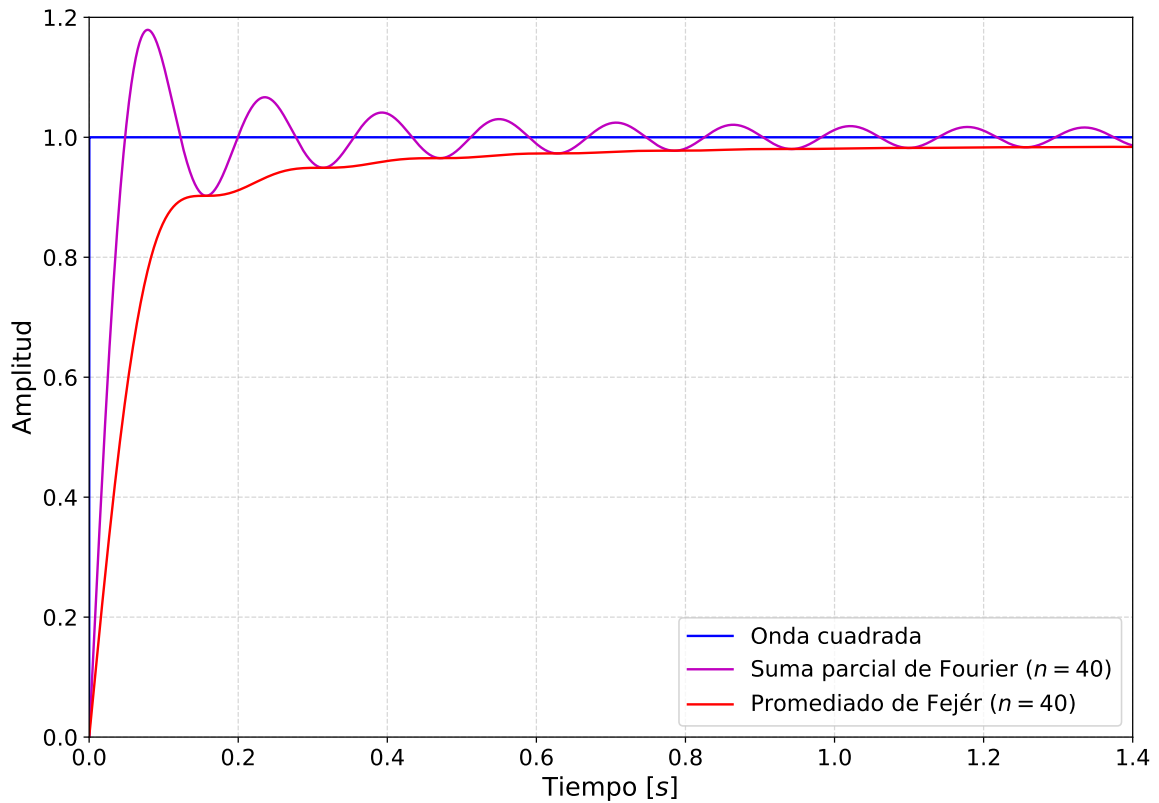


Figura 9: Fenómeno de Gibbs en la suma parcial de la serie de Fourier de la onda cuadrada y la aproximación de la misma por medio del promediado de Fejér para atenuarlo, $n = 40$.
Extraído de Jerri [45].

3.8 BÚSQUEDA DE EMPAREJAMIENTO ORTOGONAL RESTRINGIDA

La Búsqueda de Emparejamiento Ortogonal Restringida (C-OMP) es una metodología de aproximación sucesiva a la solución del problema (21) en el contexto de la reconstrucción de señales por medio del modelo de representaciones dispersas. En esta sección se deducen, paso a paso, las bases sobre las que se sustenta este procedimiento.

3.8.1 Reconstrucción y procesamiento basados en segmentos

De forma previa a la restauración, la señal se descompone en segmentos de igual longitud que son independientemente procesados por un algoritmo. Posteriormente, la señal

resultante completa es sintetizada a partir de cada uno de los segmentos restaurados utilizando el popular método de solapamiento y adición [58]. La señal observada se descompone en segmentos solapados $\mathbf{y}_i \in \mathbb{R}^N$ tal que $N \ll L$, donde i refiere al segmento en cuestión. Cada segmento i comienza en un tiempo t_i y es multiplicado por una ventana rectangular de longitud N de modo tal que $\mathbf{y}_i(t) \triangleq \mathbf{y}(t + t_i)$ para $0 \leq t \leq N - 1$. Adaptando el problema a la formulación de segmentos locales, se definen las muestras confiables del segmento i como

$$\mathbf{y}_i^r = \mathbf{M}_i^r \mathbf{z}_i \quad (42)$$

donde \mathbf{M}_i^r es la matriz de medición confiable del segmento i obtenida de \mathbf{M}^r y luego \mathbf{z}_i es el segmento restaurado i tal que $\mathbf{z}_i(t) \triangleq \mathbf{z}(t + t_i)$, definido también para $0 \leq t \leq N - 1$. Ahora, se proceden a definir los conjuntos de índices $\Omega_i^r, \Omega_i^{c+}$ y Ω_i^{c-} , que contienen los índices de las muestras confiables r , degradadas positivas $c+$ y degradadas negativas $c-$ en el segmento i respectivamente. Estos se obtienen a partir de los conjuntos de índices generales Ω_r, Ω_c^+ y Ω_c^- . Esta restauración local a partir de segmentos se ilustra para un algoritmo de restauración de señales recortadas arbitrario B en la Figura 10.

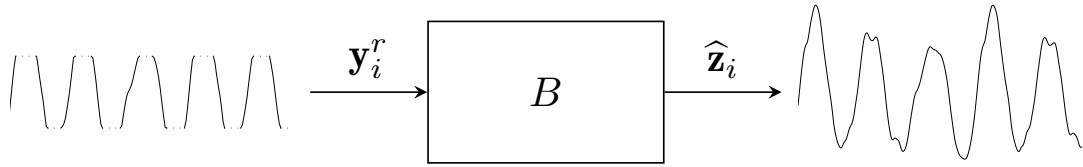


Figura 10: Formación de un segmento de señal restaurada. Se ilustran las muestras recortadas de la señal de entrada con líneas punteadas a modo de referencia.

Una vez que la estimación $\hat{\mathbf{z}}_i$ es realizada por medio de un algoritmo de restitución de

señales saturadas, la reconstrucción de la señal completa se obtiene como

$$\hat{\mathbf{z}}(t) \triangleq \frac{\sum_{i=0}^{N-1} \mathbf{w}_s(t - t_i) \cdot \hat{\mathbf{z}}_i(t - t_i)}{\sum_{i=0}^{N-1} \mathbf{w}_s(t - t_i)} \quad (43)$$

donde \mathbf{w}_s corresponde a la ventana de síntesis, para la cual se elige una ventana sinusoidal. Cuando el segmento seleccionado no tiene muestras degradadas, se utiliza el segmento original \mathbf{y}_i^r para la sintetización final.

3.8.2 Búsqueda de Emparejamiento Ortogonal (OMP) para «inpainting»

El algoritmo C-OMP, surge del más básico problema de «inpainting»:

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{sujeto a} \quad \|\mathbf{y}_i^r - \mathbf{M}_i^T \mathbf{D} \mathbf{z}\|_2 \leq \epsilon_i \quad (44)$$

para un umbral de error aproximado dado ϵ_i . La ecuación (44) es un problema NP-complejo, por lo que no puede resolverse de forma directa [59, 60]. Una solución aproximada se obtiene aplicando el algoritmo OMP con la utilización de un diccionario \mathbf{D} . El mismo aproxima de forma sucesiva la solución más dispersa de (44), y se detalla en la Tabla 1. Este es una modificación del algoritmo clásico de «inpainting» dado a que todas las columnas del diccionario $\tilde{\mathbf{d}}_j$ son multiplicadas, elemento a elemento, por una ventana de ponderación \mathbf{w}_d y normalizadas, a modo de que todas tengan la misma energía, usando la matriz diagonal \mathbf{W} . El algoritmo deja de iterar cuando su energía residual cae por debajo del umbral ϵ_i^{OMP} o alcanza el máximo nivel de dispersión K^{OMP} .

Tabla 1: Algoritmo OMP para «inpainting».

<p>Entrada:</p> $\mathbf{y}_i^r, \mathbf{M}_i^r, \mathbf{D} = \{\mathbf{w}_d \odot \mathbf{d}_j\}_{j \in \Psi}, K^{\text{OMP}}, \epsilon_i^{\text{OMP}}$
<p>Inicialización:</p> <ul style="list-style-type: none"> • Diccionario $\tilde{\mathbf{D}} = \{\tilde{\mathbf{d}}_j\}_{j \in \Psi} = \mathbf{M}_i^r \times \mathbf{D} \times \mathbf{W}$, donde $\mathbf{W}_{jj'} = 0$ para $j \neq j'$ y $\mathbf{W}_{jj} = \ \mathbf{M}_i^r \mathbf{d}_j\ _2^{-1}$ • Contador de iteraciones $k = 0$ • Conjunto soporte $\Omega = \emptyset$ • Residuo $\mathbf{r}_0 = \mathbf{y}_i^r$
<p>Selección de soporte y estimación de coeficientes:</p> <p>Repetir hasta que $k = K^{\text{OMP}}$ ó $\ \mathbf{r}_k\ _2^2 < \epsilon_i^{\text{OMP}}$</p> <ul style="list-style-type: none"> • Incrementar contador de iteración $k = k + 1$ • Seleccionar átomo: encontrar $j = \max_{j \in \Psi} \left \left\langle \mathbf{r}_{k-1}, \tilde{\mathbf{d}}_j \right\rangle \right \quad (45)$ <ul style="list-style-type: none"> • Actualizar conjunto soporte $\Omega_k = \Omega_{k-1} \cup j$ • Actualizar solución actual $\mathbf{z}_k = \min_{\mathbf{u}} \left\ \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{u} \right\ _2 \quad (46)$ <ul style="list-style-type: none"> • Actualizar residuo $\mathbf{r}_k = \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{z}_k \quad (47)$
<p>Salida:</p> $\hat{\mathbf{z}}_i = \mathbf{W} \mathbf{z}_k$

3.8.3 Selección de átomo

La etapa para la selección de átomo consiste simplemente en calcular la correlación expresada en la ecuación (45).

En el caso del diccionario de Gabor, esta etapa tiene complejidad adicional. La descomposición con átomos continuamente indexados \mathbf{d}^g (en referencia al intervalo de valores tomados por la variable ϕ) puede ser expresado usando pares de átomos en un diccionario discreto con L_g columnas. Los pares de átomos pueden ser exponenciales complejas conjugadas, o pares de cosinusoides y sinusoides en la misma frecuencia y con fase nula [23, 61, 62]. A modo de emplear esta última opción, se introducen los átomos sinusoides \mathbf{d}_j^s ,

$$\mathbf{d}_j^s \triangleq \sin \left(\frac{\pi}{L_g} \left(j + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) \right), \quad (48)$$

además de usar los previamente mencionados átomos cosinusoides \mathbf{d}_j^c ,

$$\mathbf{d}_j^c \triangleq \cos \left(\frac{\pi}{L_g} \left(j + \frac{1}{2} \right) \left(n + \frac{1}{2} \right) \right), \quad (49)$$

y se define la versión de norma $\|\cdot\|$ $\tilde{\mathbf{d}}_j^c$ y $\tilde{\mathbf{d}}_j^s$ de los átomos \mathbf{d}_j^c y \mathbf{d}_j^s respectivamente, tal como se define en el algoritmo de la Tabla 1. En cada iteración k , la selección del átomo mejor correlacionado \mathbf{d}^g es equivalente a elegir el par $(\tilde{\mathbf{d}}_j^c, \tilde{\mathbf{d}}_j^s)$ de modo que se cumpla con

$$j = \min_{j \in \llbracket 1, L_g/2 \rrbracket} \left\| \mathbf{r}_{k-1} - \tilde{\mathbf{d}}_j^c \hat{x}_j^c - \tilde{\mathbf{d}}_j^s \hat{x}_j^s \right\|_2^2 \quad (50)$$

donde

$$\begin{cases} \hat{x}_j^c = \frac{\langle \tilde{\mathbf{d}}_j^c, \mathbf{r}_{k-1} \rangle - \langle \tilde{\mathbf{d}}_j^c, \tilde{\mathbf{d}}_j^s \rangle \langle \tilde{\mathbf{d}}_j^s, \mathbf{r}_{k-1} \rangle}{1 - \langle \tilde{\mathbf{d}}_j^c, \tilde{\mathbf{d}}_j^s \rangle^2} \\ \hat{x}_j^s = \frac{\langle \tilde{\mathbf{d}}_j^s, \mathbf{r}_{k-1} \rangle - \langle \tilde{\mathbf{d}}_j^c, \tilde{\mathbf{d}}_j^s \rangle \langle \tilde{\mathbf{d}}_j^c, \mathbf{r}_{k-1} \rangle}{1 - \langle \tilde{\mathbf{d}}_j^c, \tilde{\mathbf{d}}_j^s \rangle^2} \end{cases} \quad (51)$$

En el caso en donde los átomos de un par determinado no están correlacionados (matemá-

ticamente, $\langle \tilde{\mathbf{d}}_j^c, \tilde{\mathbf{d}}_j^s \rangle = 0$), la ecuación (50) puede ser simplificada a

$$j = \max_{j \in \llbracket 1, L_g/2 \rrbracket} \left\langle \tilde{\mathbf{d}}_j^c, \mathbf{r}_{k-1} \right\rangle^2 + \left\langle \tilde{\mathbf{d}}_j^s, \mathbf{r}_{k-1} \right\rangle^2 \quad (52)$$

3.8.4 Actualización de la solución

La etapa de actualización de la solución, correspondiente a (46), es realizada en cada iteración por medio de una proyección ortogonal.

En el caso del diccionario de Gabor, este paso se efectúa utilizando el par equivalente de sinusoides y cosinusoides $\{\mathbf{d}_j^c, \mathbf{d}_j^s\}_{j \in \Omega_k}$ en la proyección. De este modo, no sólo las amplitudes sino también las fases son actualizadas en cada iteración [20].

3.8.5 Mejoras algorítmicas para restauración de audio recortado

Información adicional debe ser agregada a la ecuación (44) para adaptar el algoritmo de la Tabla 1 a la disciplina de restauración de señales recortadas. A partir de considerar las matrices \mathbf{M}_i^{c+} y \mathbf{M}_i^{c-} , derivadas respectivamente de \mathbf{M}_c^+ y \mathbf{M}_c^- , se define la formulación basada en segmentos de las expresiones (2a) y (2b),

$$\mathbf{M}_i^{c+} \hat{\mathbf{x}} \geq \mathbf{M}_i^{c+} \mathbf{y} \quad (53a)$$

$$\mathbf{M}_i^{c-} \hat{\mathbf{x}} \leq \mathbf{M}_i^{c-} \mathbf{y} \quad (53b)$$

En la expresión (53) es posible reemplazar $\mathbf{M}_i^{c+} \mathbf{y}$ y $\mathbf{M}_i^{c-} \mathbf{y}$ por las estimaciones del umbral de recorte $\hat{\varphi}$ y $-\hat{\varphi}$ respectivamente,

$$\mathbf{M}_i^{c+} \hat{\mathbf{x}} \geq \hat{\varphi} \quad (54a)$$

$$\mathbf{M}_i^{c-} \hat{\mathbf{x}} \leq -\hat{\varphi} \quad (54b)$$

en donde $\hat{\varphi}$ es obtenido fácilmente por medio de la norma $p = \infty$.

El conjunto de restricciones de la expresión (53) puede ser extendido introduciendo un límite superior,

$$\mathbf{M}_i^{c+} \hat{\mathbf{x}} \leq \hat{\varphi}_{\max} \quad (55a)$$

$$\mathbf{M}_i^{c-} \hat{\mathbf{x}} \geq -\hat{\varphi}_{\max} \quad (55b)$$

El límite superior $\hat{\varphi}_{\max}$ es opcional, y puede ser aproximadamente estimado como $\hat{\varphi}_{\max} \triangleq V \times \hat{\varphi}$, donde V es un escalar mayor a 1. Según Adler et al., este límite debe ser ajustado por el usuario del algoritmo, hasta encontrar la mejor calidad de audio en la reconstrucción [20].

De esta forma, se deriva el problema de restauración de audio recortado restringido por límite inferior y superior, dado por

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{sujeto a} \quad \|\mathbf{M}_i^r \mathbf{y} - \mathbf{M}_i^r \mathbf{D} \mathbf{z}\|_2 \leq \epsilon, \quad \mathbf{D} \mathbf{z} \in \Gamma_{\min} \cap \Gamma_{\max} \quad (56)$$

donde Γ_{\min} y Γ_{\max} son los conjuntos de vectores que cumplen con las ecuaciones (54) y (55) respectivamente. La solución aproximada de (56), incorporando al algoritmo OMP para «inpainting» las restricciones dadas por (54) y (55) en una etapa que actualiza por última vez la solución, es presentada en la Tabla 2.

En otras palabras, el algoritmo OMP de la Tabla 1 es aplicado para obtener el conjunto soporte disperso Ω_k . Una vez que el mismo es extraído, los coeficientes de la descomposición dispersa son estimados nuevamente a partir de resolver el problema de optimización convexa de la ecuación (60). Las restricciones de (60) son lineales, por lo que es posible utilizar un solucionador de optimización convexa para esta etapa. Teóricamente, (60) puede no tener solución. Cuando esto ocurre, el segmento es restaurado utilizando la minimización sin restricciones $\mathbf{z}_k = \min_{\mathbf{u}} \left\| \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{u} \right\|_2$.

Tabla 2: Algoritmo C-OMP para restauración de audio recortado.

<p>Entrada:</p> <p>$\mathbf{y}_i^r, \mathbf{M}_i^r, \mathbf{D} = \{\mathbf{w}_d \odot \mathbf{d}_j\}_{j \in \Psi}, K^{\text{OMP}}, \epsilon_i^{\text{OMP}}, \hat{\varphi}, \hat{\varphi}_{\max}$</p>
<p>Inicialización:</p> <ul style="list-style-type: none"> • Diccionario $\tilde{\mathbf{D}} = \{\tilde{\mathbf{d}}_j\}_{j \in \Psi} = \mathbf{M}_i^r \times \mathbf{D} \times \mathbf{W}$, donde $\mathbf{W}_{jj'} = 0$ para $j \neq j'$ y $\mathbf{W}_{jj} = \ \mathbf{M}_i^r \mathbf{d}_j\ _2^{-1}$ • Contador de iteraciones $k = 0$ • Conjunto soporte $\Omega = \emptyset$ • Residuo $\mathbf{r}_0 = \mathbf{y}_i^r$
<p>Selección de soporte y estimación de coeficientes:</p> <p>Repetir hasta que $k = K^{\text{OMP}}$ ó $\ \mathbf{r}_k\ _2^2 < \epsilon_i^{\text{OMP}}$</p> <ul style="list-style-type: none"> • Incrementar contador de iteración $k = k + 1$ • Seleccionar átomo: encontrar $j = \max_{j \in \Psi} \left \left\langle \mathbf{r}_{k-1}, \tilde{\mathbf{d}}_j \right\rangle \right \quad (57)$ <ul style="list-style-type: none"> • Actualizar conjunto soporte $\Omega_k = \Omega_{k-1} \cup j$ • Actualizar solución actual $\mathbf{z}_k = \min_{\mathbf{u}} \left\ \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{u} \right\ _2 \quad (58)$ <ul style="list-style-type: none"> • Actualizar residuo $\mathbf{r}_k = \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{z}_k \quad (59)$
<p>Actualización final de la solución:</p> $\mathbf{z}_k = \min_{\mathbf{u}} \left\ \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{u} \right\ _2 \text{ sujeto a } \begin{cases} \hat{\varphi}_{\max} & \geq \mathbf{M}_i^{c+} \mathbf{D} \mathbf{W} \mathbf{u} \geq \hat{\varphi} \\ -\hat{\varphi}_{\max} & \leq \mathbf{M}_i^{c-} \mathbf{D} \mathbf{W} \mathbf{u} \leq -\hat{\varphi} \end{cases} \quad (60)$
<p>Salida: $\hat{\mathbf{z}}_i = \mathbf{W} \mathbf{z}_k$</p>

3.9 DESCRIPTORES OBJETIVOS Y SUBJETIVOS

En esta sección, se consideran los descriptores objetivos y subjetivos relevantes a la tarea de restauración de señales de audio recortadas.

3.9.1 Relación señal a distorsión

La primer y más usada medida es la relación señal a distorsión o «signal-to-distortion ratio» (SDR), definida en la ecuación (61).

$$\text{SDR}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \quad (61)$$

Este descriptor mide la calidad de reconstrucción física de una señal estimada $\hat{\mathbf{x}}$ con respecto a su versión original \mathbf{x} , también denominada verdad fundamental. No obstante, en métodos como OMP, que son inconsistentes en la estimación de las muestras no degradadas, el SDR podría arrojar resultados poco confiables [13]. Por este motivo, es necesario un descriptor que evalúe únicamente la estimación sobre las muestras recortadas. Para esta empresa, se utiliza el SDR_c definido en la ecuación (62).

$$\text{SDR}_c(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \frac{\|\mathbf{x}_c\|_2^2}{\|\mathbf{x}_c - \hat{\mathbf{x}}_c\|_2^2} \quad (62)$$

Donde \mathbf{x}_c y $\hat{\mathbf{x}}_c$ son las muestras degradadas de la señal original y la estimada respectivamente. Estas cumplen con las restricciones de la ecuación (2b) y de la ecuación (2c) ($\mathbf{x}_c, \hat{\mathbf{x}}_c \in \Gamma_c^+ \cap \Gamma_c^-$).

Dado que el grado de mejora de una señal restaurada es de interés, un descriptor relevante es el ΔSDR_c expresado en la ecuación (63). El mismo consiste en la diferencia entre el SDR_c de la señal estimada y la señal recortada. Asimismo, la métrica ΔSDR_c funciona como una referencia para SDR_c .

$$\Delta\text{SDR}_c(\mathbf{x}, \hat{\mathbf{x}}) = \text{SDR}_c(\mathbf{x}, \hat{\mathbf{x}}) - \text{SDR}_c(\mathbf{x}, \mathbf{y}) \quad (63)$$

3.9.2 Evaluación Perceptiva de Calidad de Audio

La Evaluación Perceptiva de Calidad de Audio (PEAQ) es una recomendación estandarizada que permite medir la calidad de una señal de audio desde la perspectiva del sistema auditivo humano [63]. El algoritmo forma parte de la recomendación ITU-R BS.1387 y utiliza modelos para simular el oído. La salida del mismo es el descriptor Grado de Diferencia Objetiva (ODG) que califica la diferencia percibida entre la señal original y la estimación. Los valores de ODG, mostrados en la Tabla 3, forman parte de una escala de 5 valores y van desde -4 (muy molesto) hasta 0 (imperceptible).

Tabla 3: Posibles valores de Grado de Diferencia Objetiva.

ODG	Descripción del deterioro
0	Imperceptible
-1	Perceptible, pero no molesto
-2	Levemente molesto
-3	Molesto
-4	Muy molesto

Los autores no presentan estudios de respuestas subjetivas relacionadas a ODG en el artículo original. Este análisis es llevado a cabo por Dafraene et al. quienes exponen a sujetos con entrenamiento musical a 48 pares de estímulos sonoros, con y sin procesamiento, consistentes en 4 muestras musicales sometidas a 12 diferentes escenarios de distorsión artificial [64]. Además, en el artículo se investigan algoritmos de precompensación para reducir los efectos de la distorsión no lineal y lineal. El coeficiente de correlación lineal obtenido entre las calificaciones subjetivas y ODG es de 0.67, además de que sea rechazada la hipótesis de que el coeficiente de correlación lineal de la población sea 0 con un nivel de significación $\alpha=0.05$. La investigación concluye afirmando la validez de utilizar ODG para cuantificar la percepción de la distorsión lineal y no lineal, aunque cabe remarcar que este descriptor se inventa para poder medir el rendimiento de codecs de audio, cuyas características de

distorsión pueden ser distintas a las producidas por el recorte [63].

3.9.3 Percepción auditiva de distorsión no lineal

Aunque PEAQ provee una predicción razonable de una evaluación perceptiva, y es a menudo usado en estudios relacionados a la restauración de audio, fue diseñado para la evaluación de sistemas de reducción de tasas de bits en codificación de audio. Por este motivo se incluye el descriptor R_{nonlin} , inventado específicamente para medir la distorsión de sistemas no lineales en audio [65].

El descriptor es utilizado para predecir la calidad perceptiva de señales distorsionadas de forma no lineal. La predicción es un valor en el rango de 0 a 1, y está basada en la correlación cruzada ponderada de la salida de un arreglo de filtros gammatone.

Los autores además realizaron mediciones subjetivas para corroborar la validez del modelo propuesto. Para esto, expusieron sujetos a estímulos sonoros en 20 configuraciones de altoparlantes y recintos acústicos. Adicionalmente, se agregaron al experimento 12 condiciones de saturación artificial junto con una sin distorsión. Los resultados indicaron un coeficiente de correlación lineal de 0.92 para estímulos de habla y de 0.93 para programas musicales. A partir de un polinomio ajustado a los datos obtenidos para formar un predictor de las respuestas subjetivas, logran extender estos coeficientes de correlación lineal a 0.93 para muestras de habla y 0.98 para muestras musicales. Luego, concluyen en que el ajuste de un polinomio es necesario debido a que no existe una relación única entre R_{nonlin} y las calificaciones subjetivas, ya que estas se encuentran influenciadas por las características de la distorsión y de los sujetos participantes del estudio. Sin embargo, aseguran que la métrica relaciona estas cantidades de forma monótona y que la misma puede utilizarse para predecir con exactitud la percepción de distorsión no lineal en sistemas no lineales reales y artificiales [65].

4 DISEÑO DE LA INVESTIGACIÓN

En la sección anterior se desarrolló la base del conocimiento sobre la cual se apoya la presente investigación, además del análisis y la comprensión de los modelos matemáticos y algoritmos relacionados a la temática bajo estudio, correspondientes a los objetivos específicos 1 y 2. En esta sección, se detallan el tipo y alcance de la investigación, como también se discuten los límites de C-OMP en la restitución de señales recortadas con alto y bajo contenido de aliasing, a modo de establecer el motivo de la presente investigación, y finalmente se abordan los objetivos específicos no cubiertos en la sección anterior.

La presente investigación es cuantitativa por su naturaleza probatoria, deductiva, secuencial y acotada a la realidad objetiva. Es de alcance exploratorio, ya que se estudia un tema reciente y poco investigado, siendo este la restauración de señales recortadas con bajos niveles de aliasing. Se trata también de un estudio descriptivo ya que pretende caracterizar este tipo de señales. Además, es correlacional dado a que se procura estudiar el grado de relación entre cada uno de los descriptores evaluados en distintos algoritmos. Finalmente también es explicativo debido a que se exponen las causas justificativas de cada algoritmo o modelo en cuestión.

4.1 LIMITACIONES DE C-OMP

Záviska et al. evalúan a C-OMP para 7 distintos niveles de recorte en un conjunto de extractos musicales con un amplio rango de características de señal de audio [13]. En su investigación, encuentran que C-OMP obtiene niveles negativos de ΔSDR_c en la restauración de señales recortadas con niveles de SDR inferiores a 10 dB. Considerando la perspectiva del descriptor PEAQ ODG, su estudio obtiene resultados menores a -3 (molesto) para señales recortadas con niveles de SDR inferiores a 20 dB. Esto se traduce en que C-OMP no es efectivo para umbrales de recorte que generen un SDR menor a 10 dB en señales de audio musicales.

Además, como fue anteriormente mencionado, C-OMP no es útil para la restauración de

señales recortadas con niveles de aliasing reducidos. Esto es debido a que la estimación del umbral de recorte $\hat{\varphi}$ es generalmente efectuada por medio de la norma infinito sobre una señal saturada y , tal como se muestra en la ecuación (64).

$$\hat{\varphi} = \|y\|_{\infty} \quad (64)$$

La aplicación de la ecuación (64) para la estimación del umbral de recorte, no es efectiva para señales como las de la Figura 1b, haciendo que consecuentemente el algoritmo falle al intentar la reconstrucción. Por tanto, los métodos de restauración de señales recortadas de estado del arte ignoran un vasto grupo de señales digitales. Esto da la invitación al desarrollo de nuevas metodologías de restitución de señales digitales que, en cambio, consideren esta situación particular.

4.2 IMPLEMENTACIÓN DEL ALGORITMO

El algoritmo propuesto, llamado Búsqueda de Emparejamiento Ortogonal Restringida en Baja Frecuencia (LFC-OMP), se origina a partir de la versión condicionada por límite inferior del método de Búsqueda de Emparejamiento Ortogonal Restringida (C-OMP) presentado por Adler et al., y mostrado en la Tabla 2 [9]. Específicamente, se opta por utilizar como base la versión restringida por límite inferior de C-OMP con un diccionario de Gabor dado a presentar la mejor relación de compromiso entre mejor rendimiento y baja complejidad [13]. Asimismo, la elección del diccionario de Gabor está principalmente motivada en la localización tiempo-frecuencial. Este permite además la representación con la mínima incertidumbre de Heisenberg posible, favoreciendo así la dispersión de la señal estimada [9, 20]. Además, esta transformada presenta los mejores resultados en la mayor cantidad de métricas en estudios de restauración de audio recortado anteriores [13]. La metodología es implementada en el lenguaje de programación Python, y está públicamente disponible en un repositorio de código remoto.² Para solucionar el problema de programación convexa de la última etapa, se utiliza el módulo de código abierto CVXPY [66]. En caso de lidiar con

²<https://github.com/martinmessina13/lfc-omp>

recorte plano, el algoritmo brinda como alternativa la posibilidad de utilizar C-OMP.

La solución propuesta es una adaptación de C-OMP al problema de las señales saturadas con aliasing reducido. En otras palabras, señales recortadas con armónicos de alta frecuencia atenuados en las cuales el recorte no es de amplitud constante [38]. El enfoque sugerido, adaptado a la formulación por segmentos y explicado en la Tabla 4, aproxima iterativamente una solución al problema de restauración de audio recortado dado por

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{sujeto a} \quad \|\mathbf{M}_i^r \mathbf{y}_i^{\text{LF}} - \mathbf{M}_i^r \mathbf{D} \mathbf{z}\|_2 \leq \epsilon, \quad \mathbf{D} \mathbf{z} \in \Gamma_{\min} \quad (65)$$

donde \mathbf{y}_i^{LF} es la versión con aliasing reducido de \mathbf{y}_i . En otras palabras, \mathbf{y}_i^{LF} es el segmento de señal observada \mathbf{y}_i luego de ser procesado por el filtro anti-aliasing. El umbral de recorte $\hat{\phi}_F$ es estimado por medio del promediado de Fejér sobre la señal completa \mathbf{y}^{LF} como

$$\hat{\phi}_F = \|\mathcal{S}_n(\mathbf{y}^{\text{LF}})\|_{\infty} \quad (66)$$

cuyo razonamiento se deriva a partir de lo detallado en la sección 3.7.2: el máximo valor absoluto de la descomposición por medio del promediado de Fejér de la señal recortada con aliasing reducido, es el más cercano a la amplitud del umbral de recorte de la señal saturada antes de ser procesada por el filtro anti-aliasing. La diferencia entre ambos valores, a saber, el máximo valor absoluto de la descomposición de Fejér y el umbral de recorte de amplitud constante, se reduce a medida que $n \rightarrow \infty$.

Asimismo, por lo establecido en el Teorema 2, la descomposición de la señal a través del promediado de Fejér también es posible de ser utilizada en las situaciones de recorte plano. Sin embargo, esta decisión conlleva a una metodología de mayor complejidad que la simple estimación del umbral de recorte por medio de la aplicación de la norma $p = \infty$, debido a la cantidad de sumas implicadas (ecuación (34)).

Tabla 4: Algoritmo LFC-OMP para restauración de audio recortado con aliasing reducido.

<p>Entrada:</p> <p>$\mathbf{y}_i^r, \mathbf{M}_i^r, \mathbf{D} = \{\mathbf{w}_d \odot \mathbf{d}_j\}_{j \in \Psi}, K^{\text{OMP}}, \epsilon_i^{\text{OMP}}, \hat{\varphi}_F$</p>
<p>Inicialización:</p> <ul style="list-style-type: none"> • Diccionario $\tilde{\mathbf{D}} = \{\tilde{\mathbf{d}}_j\}_{j \in \Psi} = \mathbf{M}_i^r \times \mathbf{D} \times \mathbf{W}$, donde $\mathbf{W}_{jj'} = 0$ para $j \neq j'$ y $\mathbf{W}_{jj} = \ \mathbf{M}_i^r \mathbf{d}_j\ _2^{-1}$ • Contador de iteraciones $k = 0$ • Conjunto soporte $\Omega = \emptyset$ • Residuo $\mathbf{r}_0 = \mathbf{y}_i^r$
<p>Selección de soporte y estimación de coeficientes:</p> <p>Repetir hasta que $k = K^{\text{OMP}}$ ó $\ \mathbf{r}_k\ _2^2 < \epsilon_i^{\text{OMP}}$</p> <ul style="list-style-type: none"> • Incrementar contador de iteración $k = k + 1$ • Seleccionar átomo: encontrar $j = \max_{j \in \Psi} \left \left\langle \mathbf{r}_{k-1}, \tilde{\mathbf{d}}_j \right\rangle \right \quad (67)$ <ul style="list-style-type: none"> • Actualizar conjunto soporte $\Omega_k = \Omega_{k-1} \cup j$ • Actualizar solución actual $\mathbf{z}_k = \min_{\mathbf{u}} \left\ \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{u} \right\ _2 \quad (68)$ <ul style="list-style-type: none"> • Actualizar residuo $\mathbf{r}_k = \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{z}_k \quad (69)$
<p>Actualización final de la solución:</p> $\mathbf{z}_k = \min_{\mathbf{u}} \left\ \mathbf{y}_i^r - \tilde{\mathbf{D}}_{\Omega_k} \mathbf{u} \right\ _2 \text{ sujeto a } \begin{cases} \mathbf{M}_i^{c+} \mathbf{D} \mathbf{W} \mathbf{u} \geq \hat{\varphi}_F \\ \mathbf{M}_i^{c-} \mathbf{D} \mathbf{W} \mathbf{u} \leq -\hat{\varphi}_F \end{cases} \quad (70)$
<p>Salida: $\hat{\mathbf{z}}_i = \mathbf{W} \mathbf{z}_k$</p>

4.3 CONFIGURACIÓN DE PARÁMETROS

Los resultados obtenidos dependen significativamente de la configuración de los parámetros del algoritmo. Estos son establecidos con la finalidad de obtener la restauración con el valor más alto de SDR posible. Se emplea este descriptor como métrica de evaluación debido a ser recurrente su uso en estudios de estado del arte para calificar métodos de restauración de señales recortadas [2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13]. Los parámetros utilizados para LFC-OMP fueron obtenidos manualmente, sin ajuste fino, usando un conjunto de muestras musicales de prueba específico y considerando de base los empleados por C-OMP en el estudio presentado por Záviska et al. referente a la evaluación de distintos métodos de restauración de audio recortado [13]. Estos se indican en la Tabla 5.

Tabla 5: Configuración utilizada de LFC-OMP.

Parámetro	Valor
Longitud del segmento	$N = 1024$ muestras (aprox. 23.22 ms para $f_s = 44.1$ kHz)
Solapamiento	75%
Ventana de síntesis \mathbf{w}_s	Sinusoidal
Ventana de ponderación de átomo \mathbf{w}_d	Rectangular
Diccionario	DGT
Redundancia	2 (doble para DGT)
K^{OMP}	$\frac{N}{4}$
ϵ_i^{OMP}	$\epsilon \times \#I_i^r$ donde $\epsilon \triangleq 10^{-6}$ es un valor fijo y $\#I_i^r$ es el número de muestras confiables en el segmento i
$\hat{\varphi}$	$\ S_{500}(\mathbf{y}^{\text{LF}})\ _{\infty}$

4.4 SELECCIÓN Y PREPARACIÓN DE LAS MUESTRAS MUSICALES

Los segmentos de audio utilizados en la investigación se extrajeron de la base de datos EBU SQAM [67]. Los mismos son mono y cuentan con una duración aproximada de 7 segundos, una frecuencia de muestreo de 44.1 kHz y cubren un amplio espectro musical, incluyendo sonidos de fagot, arpa, clarinete, violín, acordeón, guitarra, piano y celesta. Dada la alta complejidad de los algoritmos empleados en el presente estudio, sólo se consideran 10 muestras para la evaluación de los mismos.

Cabe destacar que en esta investigación sólo se trata la restauración de señales mono. Sin embargo, actualmente existen métodos de restauración de audio multicanal que aprovechan la correlación entre distintas fuentes de distintos canales para mejorar la estimación [68, 69].

Se recuerda al lector que el propósito del actual estudio es la restauración de señales recortadas en el dominio analógico o recortadas en el dominio digital con posterior procesamiento de reducción de aliasing. En este estudio, se opta por la primer alternativa para evaluar el algoritmo desarrollado. De este modo, las muestras deben ser procesadas en pos de simular un recorte en el dominio analógico a partir de técnicas digitales, de forma previa a su restauración. A continuación, se explican cada una de las etapas de la cadena de procesamiento, observada en forma de diagrama en la Figura 11.

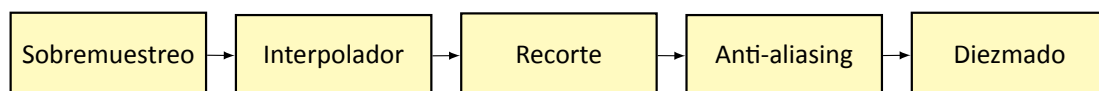


Figura 11: Diagrama en bloque de la cadena de procesamiento de muestras musicales.

4.4.1 Sobremuestreo

Como ya se ha mencionado anteriormente, el recorte produce una gran cantidad de armónicos de alta frecuencia y consecuentemente, aliasing. A modo de reducir este último efecto, la técnica elegida es el sobremuestreo. Como se mencionó anteriormente, la misma

consiste en llevar la señal a tasas de muestreo elevadas, alejando los espectros generados en los múltiplos de la frecuencia de muestreo original de forma que no interfieran en la banda base [37]. De este modo, si se desea aumentar la frecuencia de muestreo por un factor de L , son agregados $L - 1$ ceros entre cada muestra de la señal. Cabe recordar en este punto el estudio realizado por Mapes-Riordan, en el que se explica que para reducir los componentes de aliasing de un tono de 10 kHz a niveles cercanos a los -100 dB, con respecto a la frecuencia fundamental, son necesarios 5 MHz de frecuencia de muestreo [36]. Considerando que la tasa de muestreo original de las señales analizadas es de 44.1 kHz, esto corresponde a un sobremuestreo por factor de 114, resultando en una nueva frecuencia de muestreo de 5.0274 MHz para cada una de las muestras musicales.

4.4.2 Interpolador lineal

Cuando la frecuencia de muestreo es incrementada a partir del agregado de ceros, los espectros trasladados a múltiplos de la frecuencia de muestreo de la señal original se hacen visibles en el espectro de la señal con frecuencia de muestreo incrementada. A fin de reducir la influencia de los mismos en la banda base, se aplica un interpolador. Al mismo tiempo, este cumple la función de designar valores a las muestras nulas resultantes del sobremuestreo. En este caso particular, se usa la interpolación lineal para aproximar funciones continuas dada por

$$L(x) = \frac{f(x_1) - f(x_2)}{x_1 - x_2}(x - x_1) + f(x_1) \quad (71)$$

donde $f(x)$ es la función que se desea aproximar, en este caso comprendida por los valores de la señal de audio discretizada, y x_1, x_2 son dos puntos dados tales que $L(x_1) = f(x_1)$ y $L(x_2) = f(x_2)$ [70]. Asimismo, considerando que la segunda derivada de la función a aproximar es continua es posible aplicar el teorema de Rolle para encontrar el error de esta estimación, expresado por

$$f(x) - L(x) = \frac{\partial^2 f(\varepsilon)}{\partial x^2} \cdot \frac{(x - x_1)(x - x_2)}{2} \quad (72)$$

donde $a < \varepsilon < b$, siendo a y b los extremos del intervalo cerrado en el cual se aproxima la función [71]. La ecuación (72) indica que la calidad de aproximación empeora a medida que aumenta la segunda derivada de la función. Esto es de esperar, ya que mientras más curvada es una función, más difícil es estimarla a partir de una interpolación lineal.

4.4.3 Recorte

Para representar el recorte digital, el modelo de recorte fuerte de la ecuación (1) es aplicado sobre cada una de las muestras musicales luego de haber aumentado su tasa de muestreo por medio de la interpolación lineal. La elección del umbral de recorte se hace en función del SDR calculado en la señal en este punto de la cadena de procesamiento, tal como se realiza en el estudio llevado a cabo por Záviska et al. sobre evaluación de métodos de restauración de audio recortado [13].

Estudios previos normalizan la señal de audio bajo estudio y establecen un umbral de recorte en función de un valor absoluto de amplitud [20]. Sin embargo, Gaultier detalla que la amplitud del umbral de recorte o el porcentaje de muestras recortadas no es un buen indicador de la severidad del recorte para el sistema perceptivo humano [40]. Por otra parte, dado a que el recorte se realiza en base al SDR deseado, no es necesario normalizar la señal [13].

4.4.4 Filtro anti-aliasing

Posteriormente, se recrea el filtro anti-aliasing encontrado en la entrada de los convertidores analógico-digital. Para este propósito, se utiliza un filtro pasabajos Chebyshev tipo II debido a su rápida transición entre la banda de paso y la banda de atenuación, necesaria para el caso de uso en cuestión [72, 73]. El mismo es de trigésimo orden y su frecuencia de corte es de 20 kHz. Cabe destacar que el mismo alcanza los 100 dB de atenuación a los 22.048 kHz, aproximadamente la mitad de la frecuencia de muestreo de las señales analizadas. La respuesta frecuencial del filtro se observa en la Figura 12.

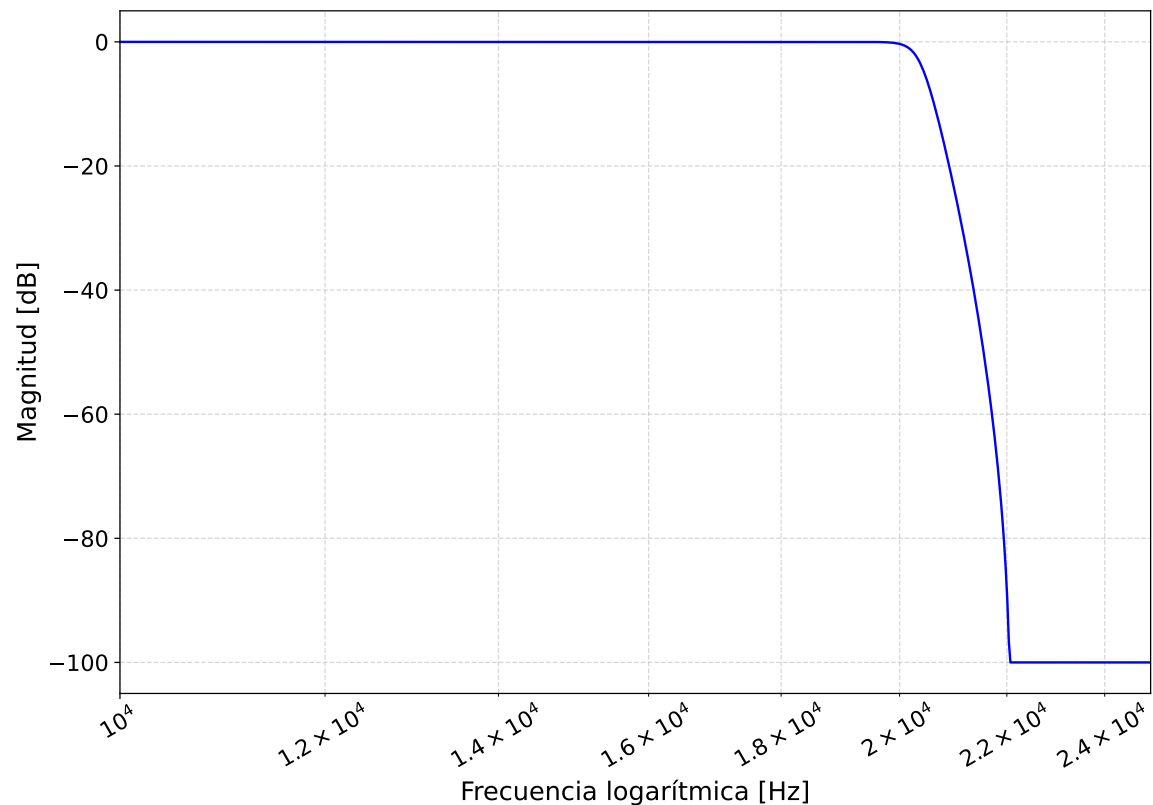


Figura 12: Respuesta frecuencial del filtro pasabajos anti-aliasing Chebyshev tipo II de orden 30 utilizado.

4.4.5 Diezmado

Finalmente, se vuelve a la tasa de muestreo original a partir de diezmar muestras en la señal filtrada. Esto es equivalente a reducir la frecuencia de muestreo por un factor de L . En este caso particular, $L = 114$. Con este último paso, se logra obtener una señal con características similares a una señal recortada por un circuito electrónico analógico.

Para mostrar los efectos de la cadena de procesamiento, en la Figura 13 se expone un audio de ejemplo antes y después de ser modificado por la misma. El segmento de audio en cuestión es una muestra de batería con 1 segundo de duración, y una frecuencia de muestreo de 44.1 kHz.

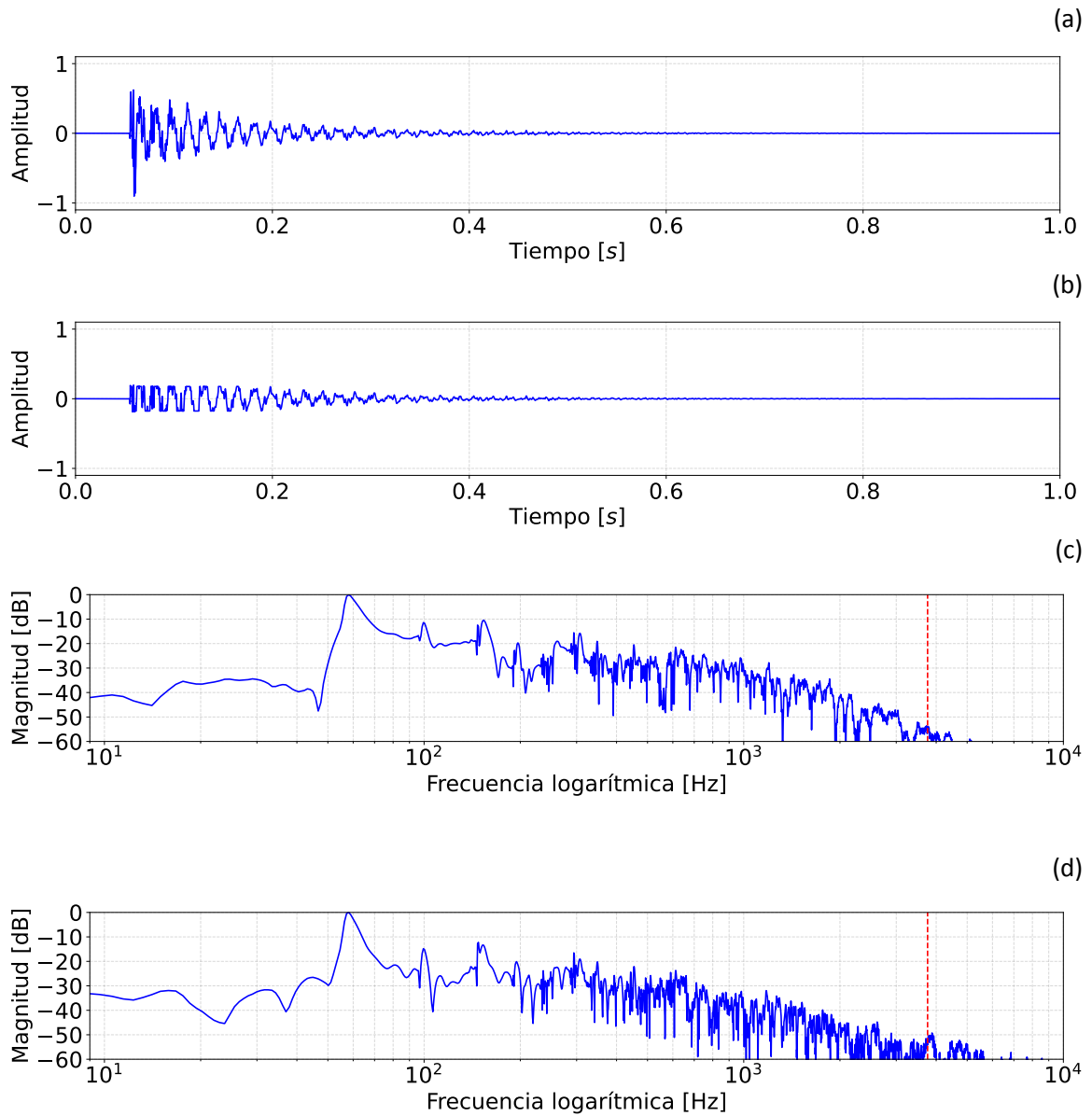


Figura 13: Muestra percusiva antes y después de someterse a la cadena de procesamiento. En (a) y (b) se observa en el dominio temporal antes y después de ser procesada, de forma respectiva. Luego en (c) y (d) se ilustra respectivamente el contenido frecuencial del segmento de audio antes y después de la cadena de procesamiento. Es posible visualizar el cambio en magnitud de los armónicos en frecuencias anteriores y posteriores a la fundamental de 58 Hz, como consecuencia de la saturación ($\varphi = 0.176$).

A primera vista, comparando las Figuras 13c y 13d pueden observarse varias diferencias en el espectro de las señales, como resultado del recorte. Entre las más significativas se encuentran las descritas a continuación. Dentro del rango comprendido por la fundamental (58 Hz) y el tercer armónico (153 Hz), hay una reducción significativa de la magnitud del espectro. Esta es de aproximadamente 10 dB, alcanzando los -26 dB de magnitud, alrededor de los 90 Hz. Luego, alrededor de los 105 Hz es de 20 dB, resultando en una magnitud de -40 dB, previamente en -20 dB antes de la saturación. Entre 105 y 150 Hz el decremento llega a 10 dB, dando lugar a una magnitud aproximada de -30 dB en los 118 Hz. Adicionalmente, también puede verse un decremento de 10 dB en 270 Hz, resultando en una magnitud de aproximadamente -31 dB en dicha frecuencia. Esta situación se repite alrededor de 550 Hz, en donde el incremento consecuente del recorte supera los 10 dB, dando lugar a una magnitud de más de -30 dB en la frecuencia mencionada. Finalmente, a partir de 3760 Hz en adelante, indicado con líneas punteadas, es posible ver otro aumento, el cual resulta en un valor de -50 dB de magnitud en frecuencias cercanas a 4000 Hz.

Por otra parte, la Figura 14 contiene una visualización gráfica de las primeras cuatro etapas de procesamiento sobre la muestra de audio ejemplar anteriormente mencionada, que se ilustra en las Figuras 13a y 13c. En las Figuras 13b y 13d se observa el efecto de la quinta y última etapa, a saber, el diezmado, sobre la muestra en cuestión.

Varios sucesos son observados en la Figura 14: principalmente, como consecuencia de la etapa de sobremuestreo, en la Figura 14b se pueden visualizar los espectros copia del analógico, trasladados a múltiplos de la frecuencia de muestreo. Como se dijo anteriormente, este suceso ocurre debido al aumento de la frecuencia de muestreo. Para remover los mismos, se utiliza el interpolador lineal, resultando en una señal cuyo espectro, observado en la Figura 14d, contiene sólo las componentes frecuenciales de la señal original, pero en una frecuencia de muestreo superior. Del mismo modo, es posible observar en la Figura 14d las sutiles imperfecciones del interpolador lineal en 44.1 kHz, en forma de un leve incremento de magnitud en la frecuencia en cuestión. La señal resultante del recorte visualizada en la Figura 14e, muestra los cambios de magnitud en su respuesta frecuencial ya comentados

(Figura 14f). Luego, en la anteúltima etapa de la cadena, el filtro anti-aliasing atenúa agresivamente los armónicos mayores a 22.05 kHz, como se demuestra en la Figura 14h. Como consecuencia de esta acción, el umbral de recorte de la señal en el dominio temporal ya no es de amplitud constante, tal como se da en señales cuyo recorte es llevado a cabo en el dominio analógico. Este es el fenómeno de Gibbs detallado en la sección 3.7.1, posible de apreciar en la Figura 15, la cual constituye una versión amplificada de la Figura 14g. Asimismo, este efecto constituye la causa que impide la aplicación de algoritmos tradicionales de restauración de señales recortadas, en este tipo de señales [13].

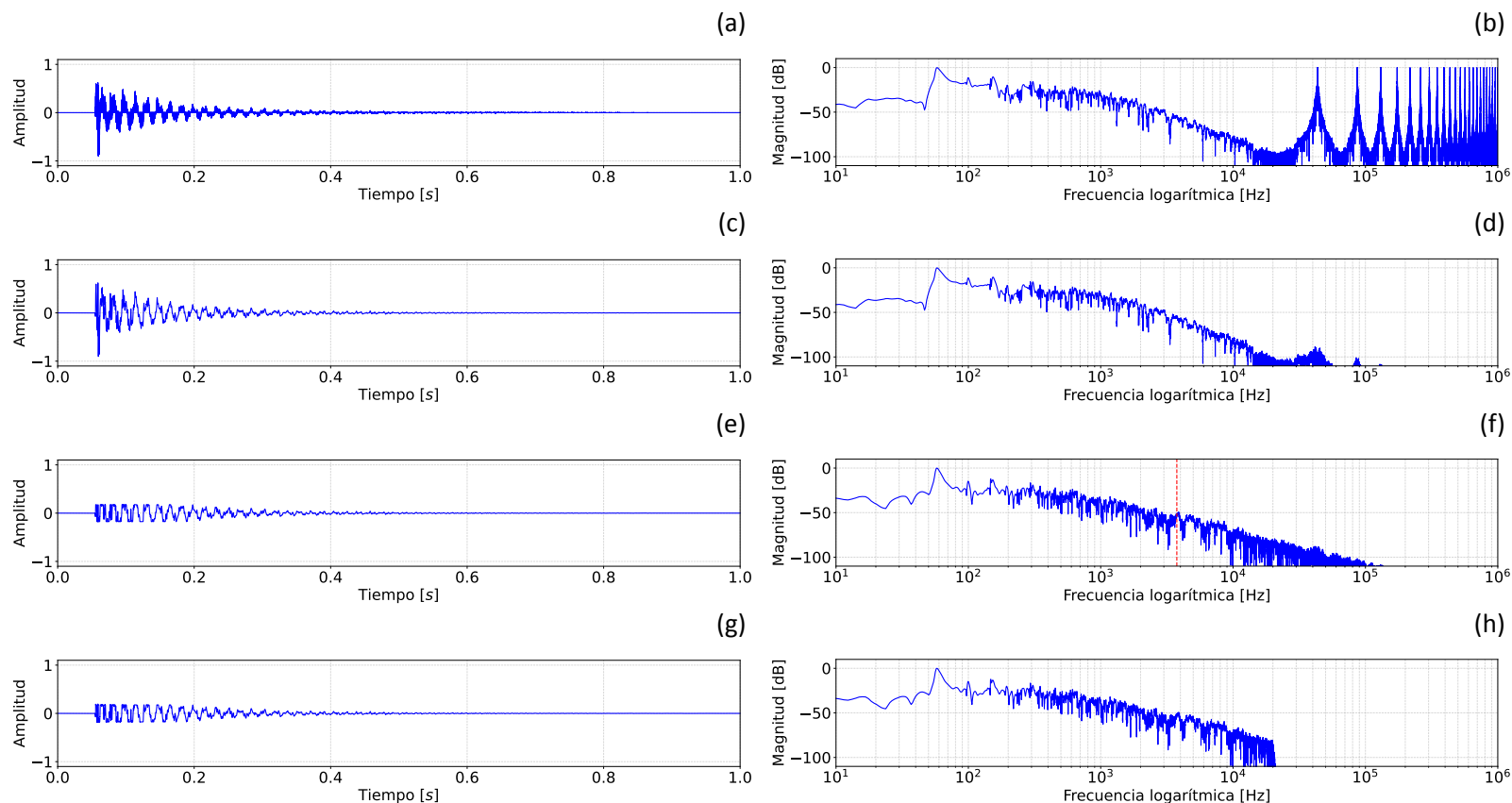


Figura 14: Aplicación consecutiva de cada una de las etapas de la cadena de procesamiento sobre una muestra de audio percusiva. En (a) y (b) se exponen respectivamente el dominio temporal y frecuencial de la señal resultante del sobremuestreo. Es posible observar en (b) los espectros trasladados a los múltiplos de la frecuencia de muestreo de la señal original de la Figura 13c. (c) y (d) demuestran, de forma respectiva, el contenido temporal y frecuencial consecuente de la interpolación lineal. (e) y (f) se corresponden respectivamente con la señal temporal y frecuencial resultado de la etapa de recorte. Se visualiza en (f) una reducción en la magnitud de parte del espectro de baja frecuencia y un aumento en el rango de los 40 a 50 Hz, de frecuencias aledañas a 550 Hz y de los armónicos de alta frecuencia posteriores a la línea punteada, como resultado de la saturación. La señal temporal y frecuencial obtenidas luego de la aplicación del filtro anti-aliasing se visualizan en las Figuras (g) y (h), respectivamente. En (h) se aprecia la eliminación de armónicos superiores a 22.05 kHz, como consecuencia del filtrado.

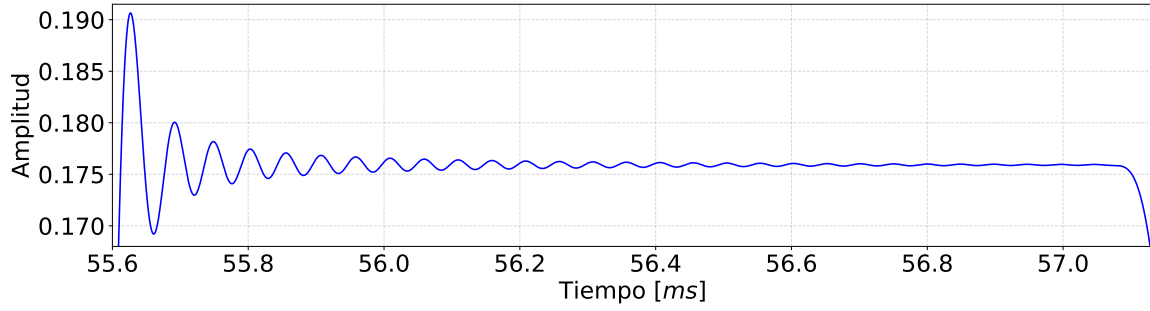


Figura 15: Versión ampliada de la señal temporal de la Figura 14g luego de aplicarse el filtro anti-aliasing. El truncado de armónicos de alta frecuencia se manifiesta como el fenómeno de Gibbs.

4.5 GENERACIÓN DE LA VERDAD FUNDAMENTAL

Para poder utilizar los descriptores expuestos en la sección 3.9 es necesaria la creación de una señal que represente la verdad fundamental x . Con este fin, las muestras musicales sin procesar son sometidas a la cadena de procesamiento indicada en la Figura 16. La misma es similar a la cadena de la Figura 11 con la omisión de la etapa de recorte. De este modo, se logra una señal de referencia con aliasing reducido.

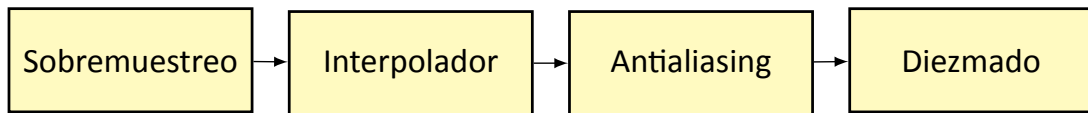


Figura 16: Diagrama en bloque de la cadena de procesamiento de la verdad fundamental.

4.6 VALIDACIÓN

La señal resultante del algoritmo LFC-OMP es evaluada a partir de los descriptores ΔSDR_c , SDR_c , PEAQ y R_{nonlin} , definidos en la sección 3.9. Son utilizados 7 diferentes niveles de recorte en el presente estudio. Se consideran los promedios de los 10 programas musicales para cada descriptor y nivel de recorte dado.

La versión de PEAQ utilizada es la implementada por P. Kabel en 2004 de acuerdo a la

recomendación ITU-R BS.1387 y a la fecha de escritura del documento se encuentra disponible públicamente en línea junto con una versión detallada del método.³ La frecuencia de muestreo de los programas musicales fue aumentada a 48 kHz para la obtención de este descriptor, dado a que el modelo aplicado en PEAQ fue diseñado para este tipo de señales [74]. La implementación de R_{nonlin} empleada está disponible en la página principal de intercambio de archivos de MATLAB.⁴

El algoritmo C-OMP brindado por Adler et al. en el paquete de herramientas de MATLAB SMALLbox es utilizado como referencia para comparar resultados con los obtenidos por LFC-OMP.⁵ Este es aplicado con la misma configuración que para LFC-OMP, con la excepción de aplicar la norma $p = \infty$ para la estimación del umbral de recorte.

³<https://www-mmsep.ece.mcgill.ca/Documents/Software/>

⁴https://www.mathworks.com/matlabcentral/fileexchange/50230-rnonlin_calc

⁵<http://small.inria.fr/software-data/smallbox/>

5 ANÁLISIS DE LOS RESULTADOS

Los resultados obtenidos en las 10 muestras musicales para las métricas SDR_c , ΔSDR_c , PEAQ ODG y R_{nonlin} se ilustran para 7 diferentes niveles de recorte en los gráficos de barra de las Figuras 17a, 17b, 17c y 17d respectivamente.

Principalmente, la Figura 17 exhibe para LFC-OMP resultados similares a los del algoritmo de referencia, C-OMP, en los 4 descriptores estudiados. Además, la reducción del valor de ΔSDR_c , SDR_c , PEAQ ODG y R_{nonlin} al mismo tiempo que decrece el nivel de recorte indica un incremento de la dificultad de ambos algoritmos para restaurar señales a medida que el umbral de recorte es más bajo en las mismas. Las Figuras 17a, 17b y 17c exponen altos valores de desvío estándar para las respectivas medidas ΔSDR_c , SDR_c y PEAQ ODG en 15 y 20 dB. Además, un decrecimiento de los desvíos al mismo tiempo que se reduce el nivel de recorte es observado para ambos algoritmos en estas cantidades. Del mismo modo que con la magnitud de los descriptores, esto manifiesta la incapacidad de los algoritmos para restaurar señales con bajos umbrales de recorte, donde existen altos niveles de distorsión.

La Figura 17a muestra que para los niveles de recorte de 1, 3, 5, 7 y 10 dB se obtienen ΔSDR_c inferiores a 2 dB para el C-OMP y LFC-OMP, lo cual sugiere que tales metodologías son más adecuadas para situaciones de recorte leve tales como 15 y 20 dB. En estos niveles, se obtienen altos valores de desvío estándar tanto para C-OMP como para LFC-OMP. Esto es debido a la utilización de programas musicales de género variado con diversas características en la investigación. A modo de estudiar dichas variaciones con más detalle, se presentan los resultados de ΔSDR_c para cada muestra individual obtenidos por LFC-OMP en la Figura 18. Aquí es posible visualizar, en los niveles de 15 y 20 dB de recorte, el buen rendimiento del algoritmo propuesto para instrumentos como la celesta, el arpa y el fagot y el deplorable rendimiento del mismo para los casos del acordeón y la guitarra. Esta última, demostrando niveles de ΔSDR_c inferiores a 5 dB para todos los grados de recorte. Asimismo, se remarca el caso particular de la muestra de clarinete, la cual registra valores altos de ΔSDR_c para los dos niveles de recorte de 15 y 20 dB para luego obtener un resultado inferior que la mayor

parte de las muestras en el recorte de 10 dB.

En el caso del SDR_c , mostrado en la Figura 17b, obtenido por LFC-OMP los valores para todos los niveles de recorte, con excepción de 20 dB, son levemente mayores a los de C-OMP. Las magnitudes de esta métrica para LFC-OMP, varían de forma similar a los de ΔSDR_c , con valores de 22.7 dB compuestos por 8.8 dB de mejora en 20 dB de recorte, 16.8 dB con 5 dB de mejora en 15 dB de recorte, y resultados inferiores a 10 dB en los niveles de recorte restantes. Del mismo modo, para C-OMP, en la Figura 17b se exhiben magnitudes aproximadas de 23 dB con 9.8 dB de mejora en 20 dB de recorte, 15.5 dB con 4.6 dB de mejora en 15 dB de recorte, y resultados inferiores que LFC-OMP en los demás niveles de recorte.

Entrando al campo de los medidores de calidad perceptiva PEAQ ODG y R_{nonlin} , se observa en la Figura 17c una preferencia hacia C-OMP en cuanto al PEAQ ODG, con niveles cercanos a -1 (perceptible, pero no molesto) para 20 dB de recorte y -2 (levemente molesto) para 15 dB de recorte. Esto es frente a un valor aproximado de -2 para un recorte de 20 dB y -3 (molesto) para 15 dB de recorte en LFC-OMP. Sin embargo, si el mismo análisis se hace considerando al descriptor R_{nonlin} , los resultados de LFC-OMP son sutilmente superiores a C-OMP en los niveles de recorte severos, 1, 3, 5 y 7 dB, y similares para 10, 15 y 20 dB de saturación. Esto sugiere que LFC-OMP pueda ser útil para las restauraciones desde la perspectiva de las distorsiones no lineales, como las estudiadas en la presente investigación.

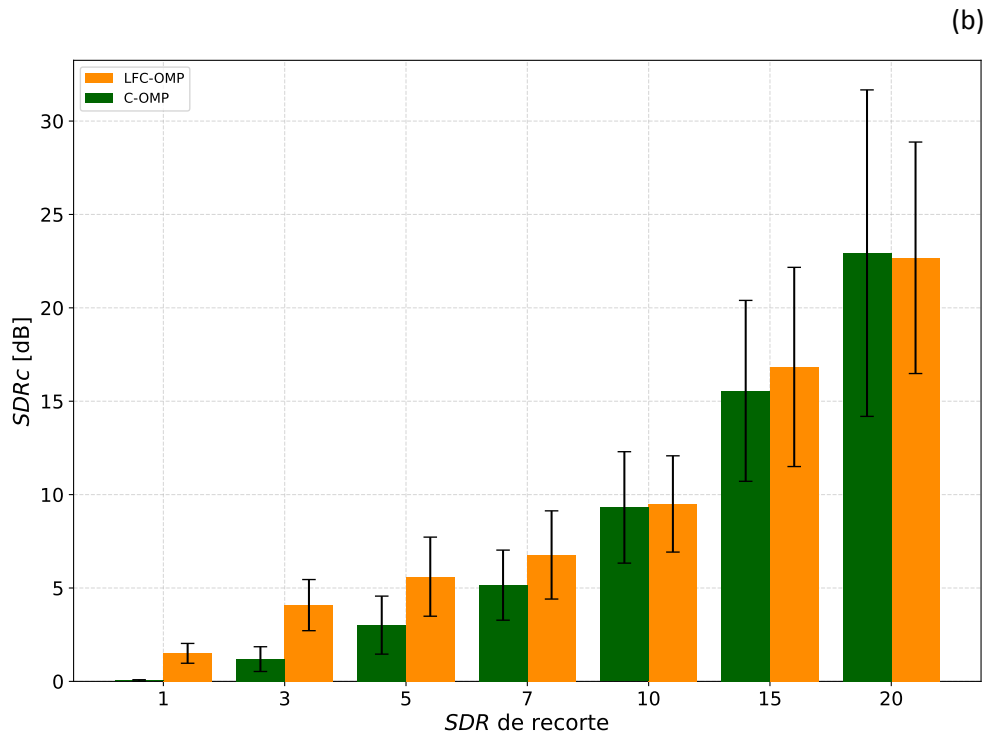
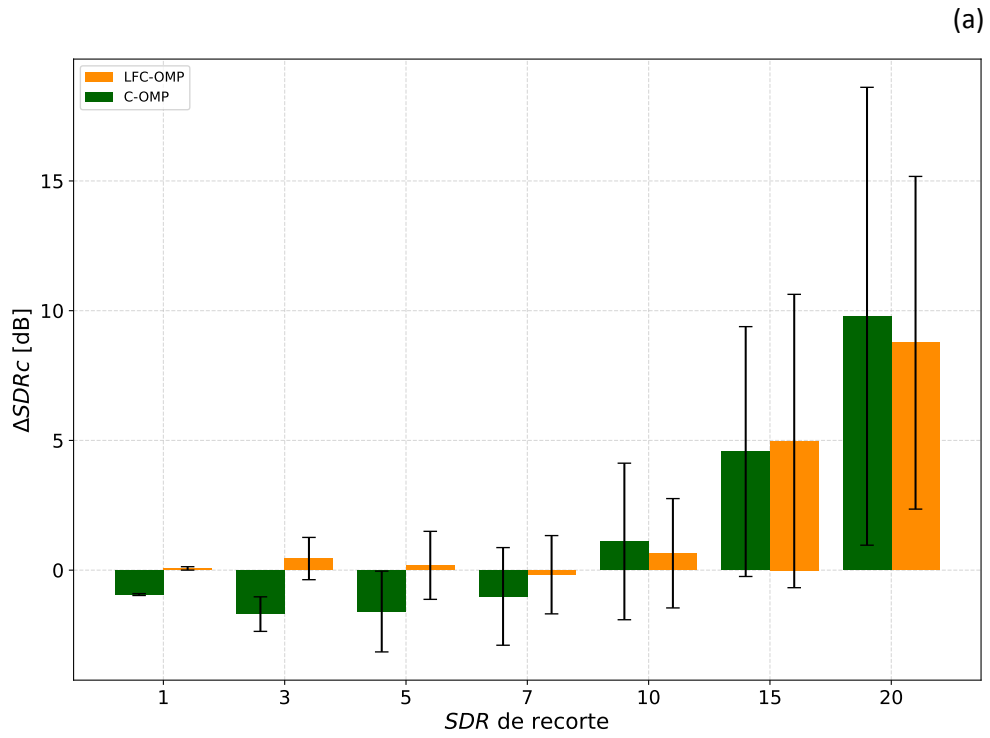


Figura 17: Métricas obtenidas para distintos niveles de recorte: (a) ΔSDR_c , (b) SDR_c , (c) PEAQ ODG, (d) R_{nonlin} .

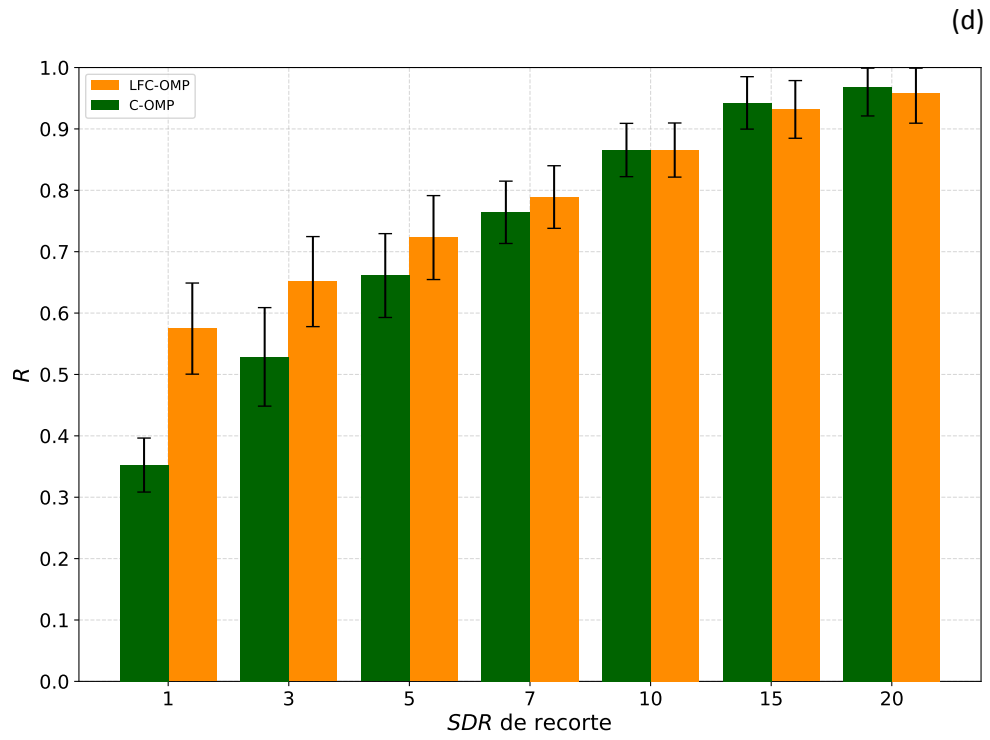
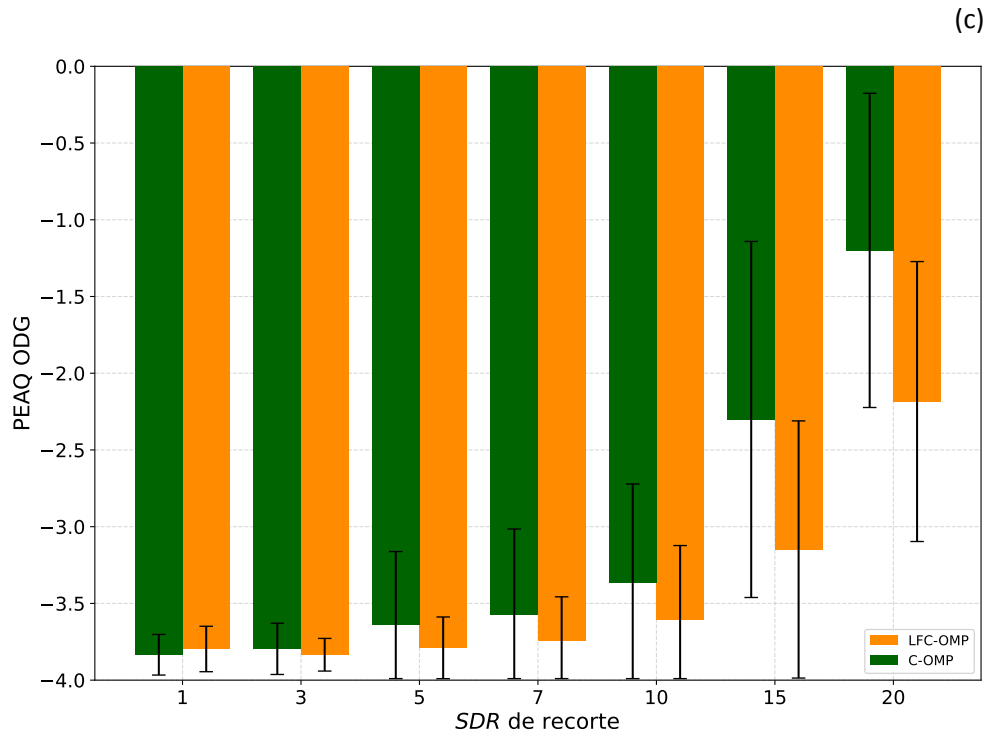


Figura 17: Métricas obtenidas para distintos niveles de recorte: (a) ΔSDR_c , (b) SDR_c , (c) PEAQ ODG, (d) R_{nonlin} .

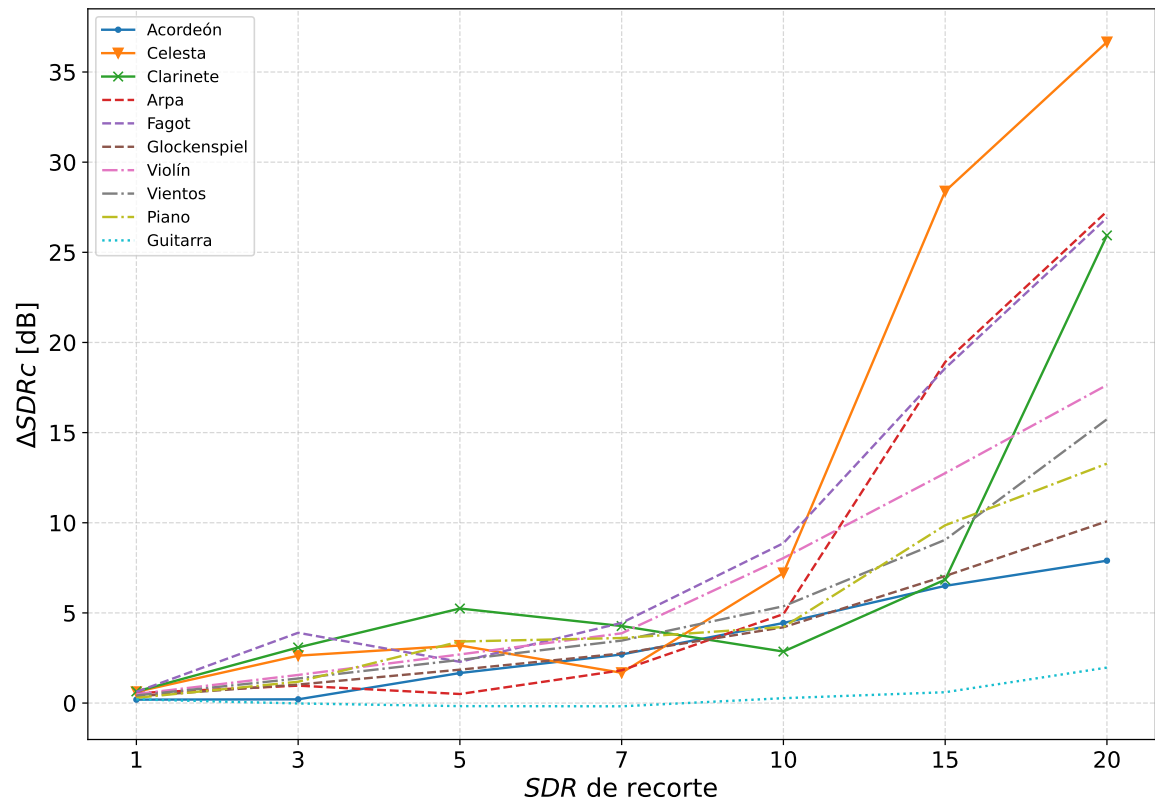


Figura 18: ΔSDR_c obtenido por LFC-OMP en cada muestra individual para distintos niveles de recorte.

A modo de estudiar la similitud entre LFC-OMP y C-OMP se consideran los valores absolutos de las diferencias de ΔSDR_c , SDR_c , PEAQ ODG y R_{nonlin} entre algoritmos. En el análisis, también se tienen en cuenta los valores absolutos de las diferencias del desvío estándar. Principalmente, la diferencia de ΔSDR_c entre algoritmos, mostrada en la Figura 19a es de un máximo de 2.1 dB dado en 3 dB de recorte. Considerando el desvío estándar dentro del análisis, la máxima diferencia de ΔSDR_c es alcanzada en el recorte de 20 dB, con un valor total de 3.4 dB. De igual modo, la máxima diferencia de SDR_c es de 2.9 dB con 3 dB de recorte, y este nivel sigue siendo el mayor si además se adiciona el valor del desvío estándar en el cálculo del máximo SDR_c . El máximo de la diferencia entre el PEAQ ODG obtenido en LFC-OMP y C-OMP, indicado en la Figura 19c se logra en 20 dB de recorte, con un valor de 0.98, o en 15 dB de recorte si se considera además el desvío estándar, con un valor de 1.17. Para

la métrica R_{nonlin} , esto ocurre en 1 dB de recorte con un valor absoluto de 0.22, sin tomar en cuenta el desvío estándar. Estos resultados, especialmente los obtenidos para ΔSDR_c , denotan un grado de semejanza entre LFC-OMP y C-OMP.

Luego, se realiza un análisis de correlación lineal comparando entre ambos algoritmos cada uno de los 4 descriptores analizados. A modo de establecer una forma de comparación equitativa entre descriptores, cada muestra del conjunto de datos se traslada al valor mínimo y se escala por la diferencia entre el máximo y el mínimo valor del mismo. Así, se logra que todos los puntos evaluados se encuentren dentro del intervalo $x \in [0, 1]$ y $y \in [0, 1]$.

Principalmente, el grado de mejora ΔSDR_c obtiene un coeficiente de correlación lineal de $\hat{r}_{\Delta\text{SDR}_c} = 0.98$, Figura 20a, mientras que el de SDR_c es de $\hat{r}_{\text{SDR}_c} = 0.99$, Figura 20b. En las Figuras 20c y 20d, ambas métricas de percepción de calidad, PEAQ ODG y R_{nonlin} , obtienen un coeficiente de correlación lineal de 0.99. Promediando estos últimos resultados, se obtiene un coeficiente de correlación global de $\hat{r} = 0.99$ advirtiendo de una alta similitud entre LFC-OMP y C-OMP, y remarcando la utilidad del promediado de Fejér para la estimación de umbrales de recorte en el contexto de las señales recortadas libres de aliasing.

A partir del coeficiente de correlación global muestral obtenido de $\hat{r} = 0.99$, se prueba la hipótesis H_0 contra su alternativa H_a

$$H_0 : r = 0 \quad (73a)$$

$$H_a : r > 0 \quad (73b)$$

donde r es el coeficiente de correlación de Pearson de la población. Esta hipótesis estadística fue probada con un nivel de significación de $\alpha = 0.01$ por medio de una prueba t de una cola con $N_r - 2$ grados de libertad, donde $N_r = 7$ es la cantidad de niveles de recorte, para el estadístico $T = |\hat{r}| \sqrt{N_r - 2} / \sqrt{1 - \hat{r}^2}$ [75]. El p-valor resultante es de $9.55 \cdot 10^{-6} < \alpha$, por ende, la hipótesis nula H_0 puede ser confiadamente rechazada a favor de la hipótesis alternativa H_a .

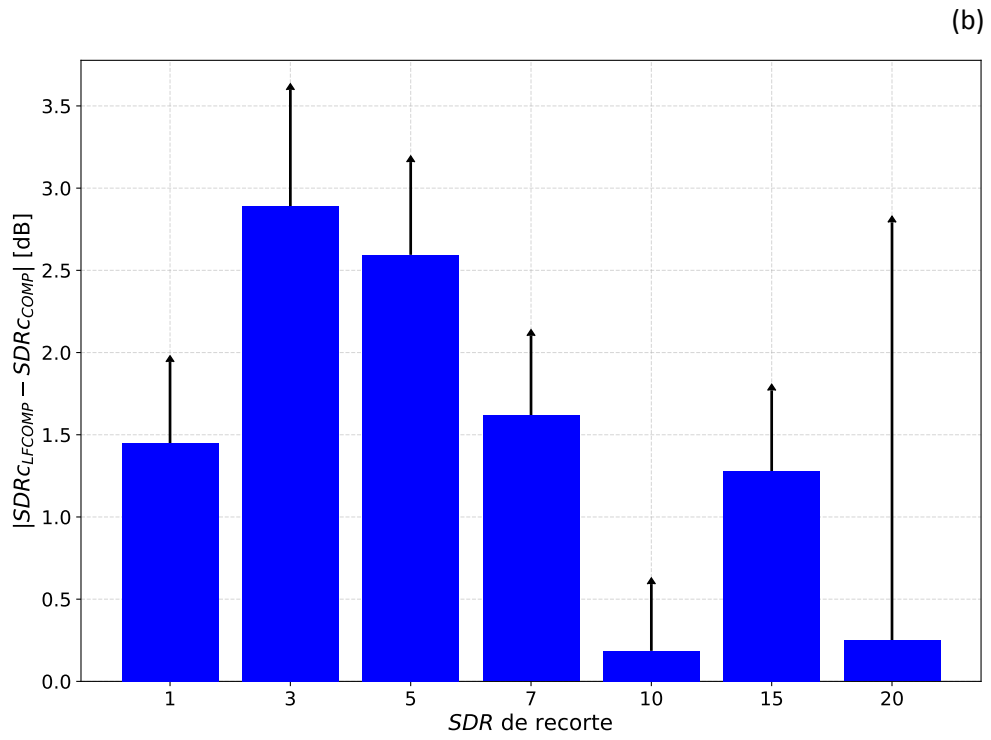
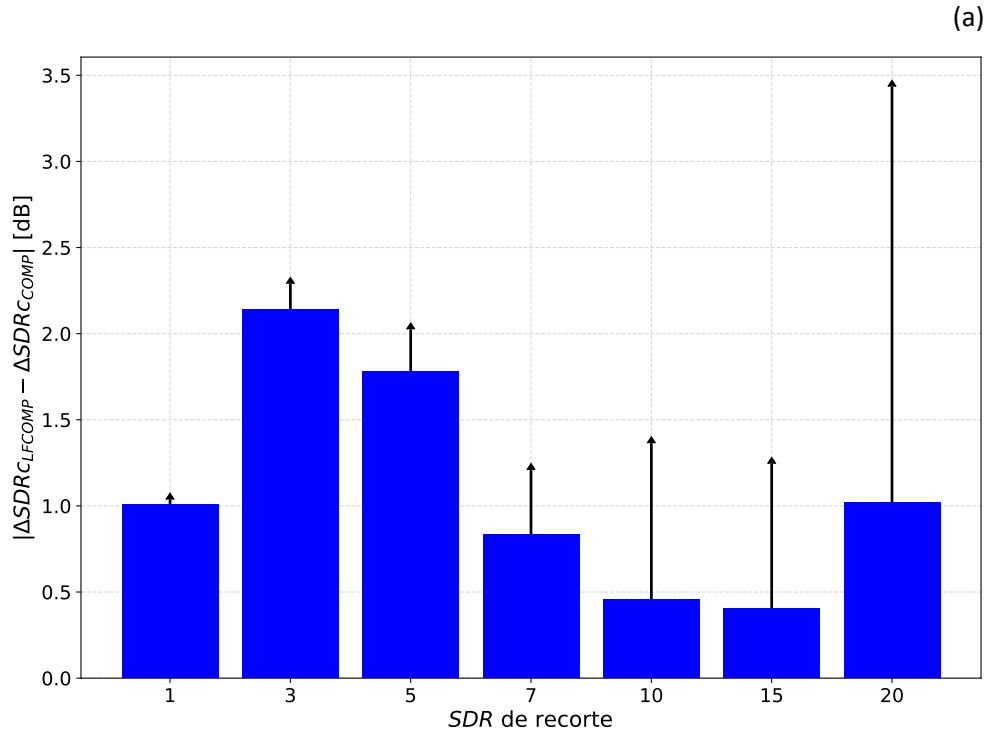


Figura 19: Diferencia absoluta entre los descriptores calculados en LFC-OMP y C-OMP para distintos niveles de recorte: (a) ΔSDR_c , (b) SDR_c , (c) PEAQ ODG, (d) R_{nonlin} .

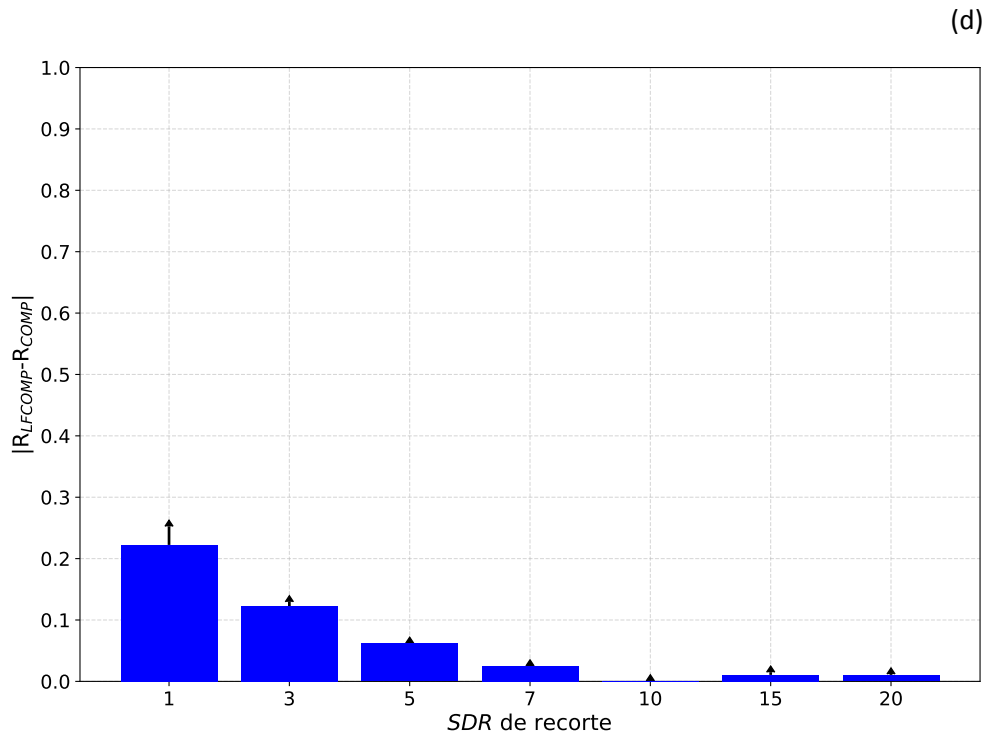
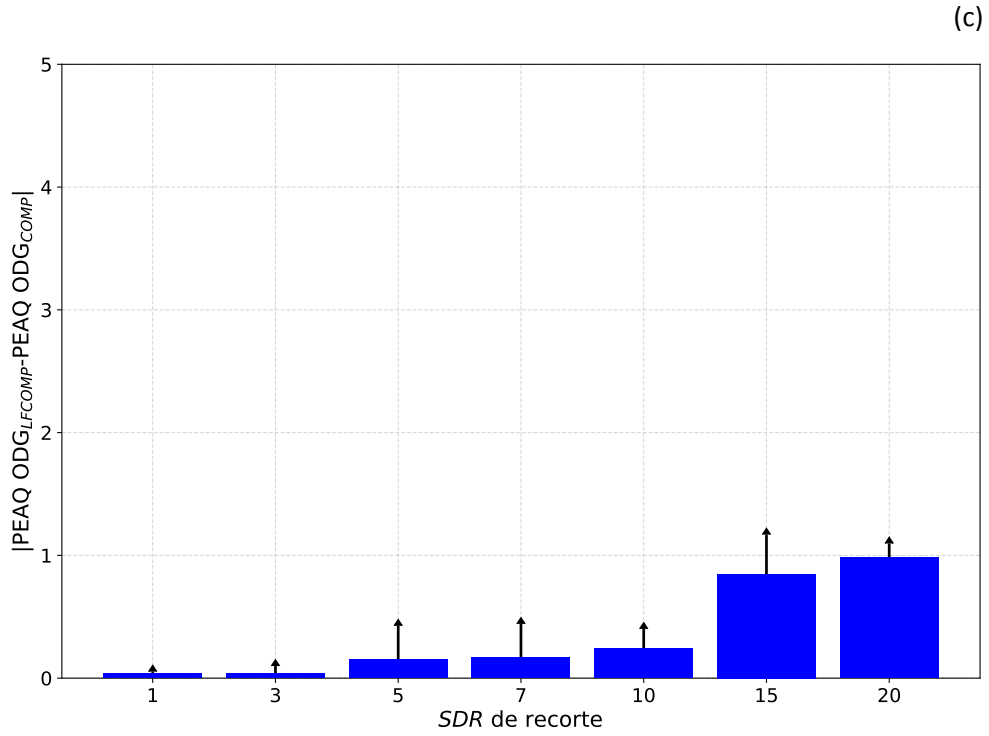


Figura 19: Diferencia absoluta entre los descriptores calculados en LFC-OMP y C-OMP para distintos niveles de recorte: (a) ΔSDR_c , (b) SDR_c , (c) PEAQ ODG, (d) R_{nonlin} .

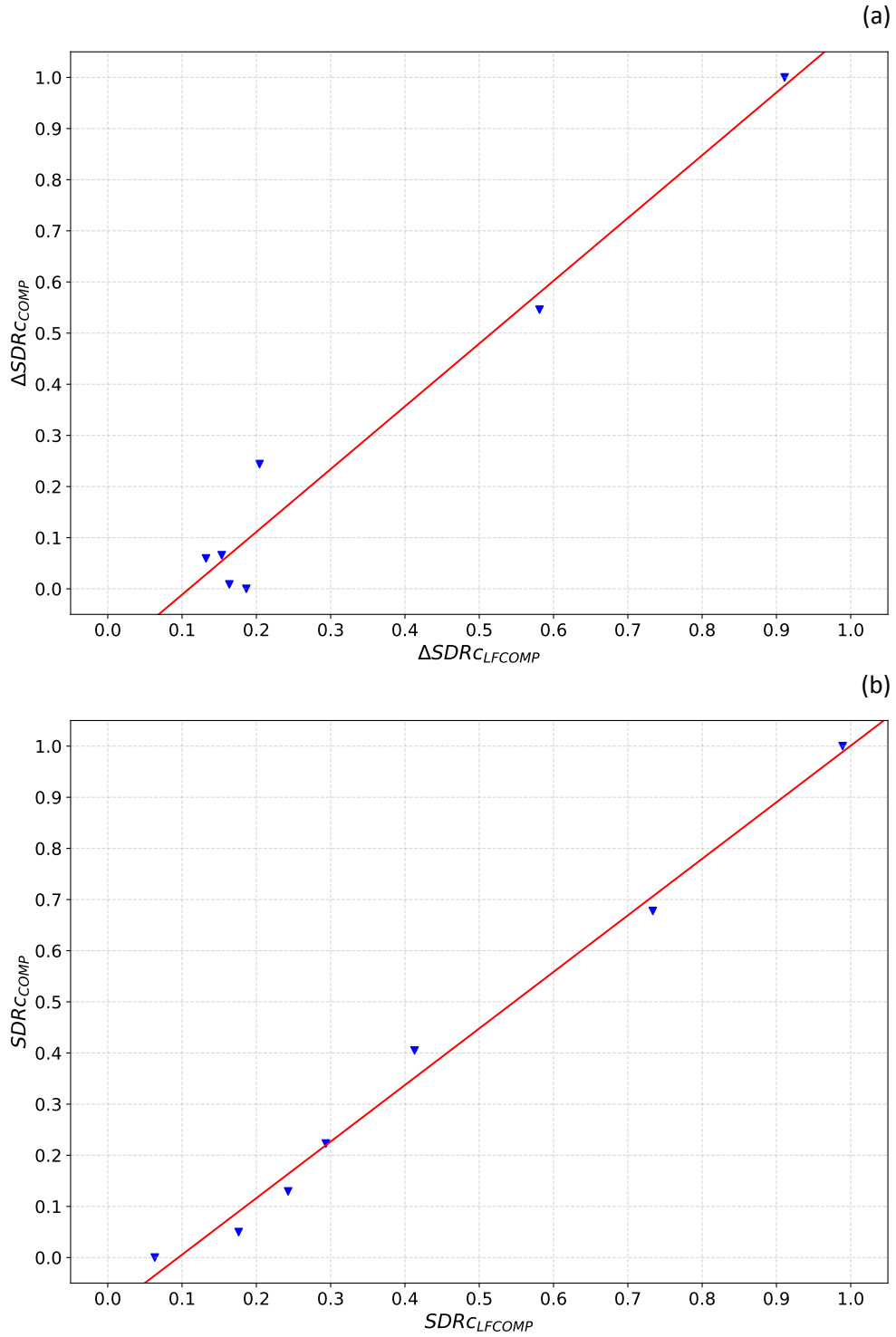


Figura 20: Correlación lineal entre descriptores obtenidos en ambos algoritmos: (a) ΔSDR_c ($\hat{r}_{\Delta\text{SDR}_c} = 0.98$), (b) SDR_c ($\hat{r}_{\text{SDR}_c} = 0.99$), (c) PEAQ ODG ($\hat{r}_{R_{\text{PEAQ ODG}}} = 0.99$), (d) R_{nonlin} ($\hat{r}_{R_{\text{nonlin}}} = 0.99$).

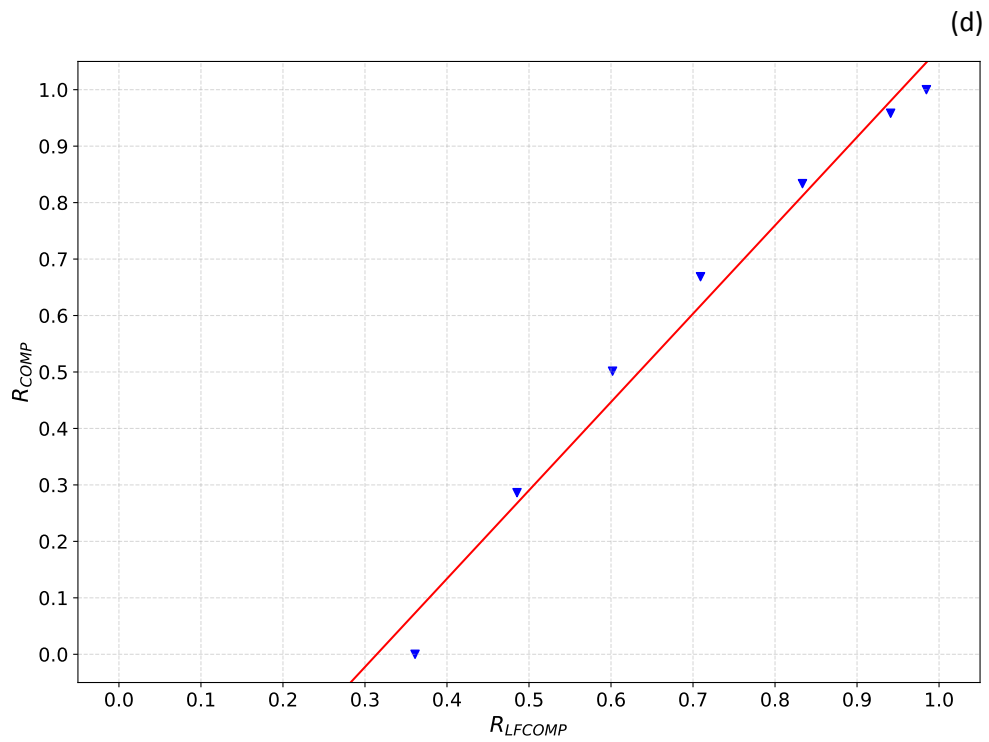
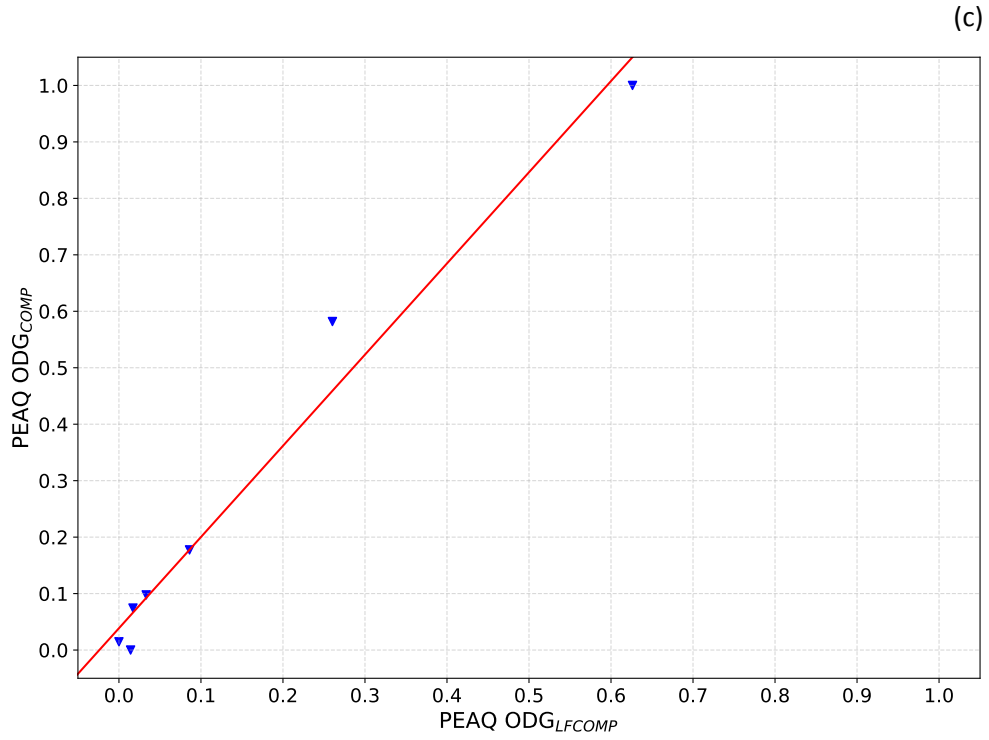


Figura 20: Correlación lineal entre descriptores obtenidos en ambos algoritmos: (a) ΔSDR_c ($\hat{r}_{\Delta SDR_c} = 0.98$), (b) SDR_c ($\hat{r}_{SDR_c} = 0.99$), (c) PEAQ ODG ($\hat{r}_{R_{PEAQ\ ODG}} = 0.99$), (d) R_{nonlin} ($\hat{r}_{R_{nonlin}} = 0.99$).

6 CONCLUSIONES

En la presente investigación, se ha estudiado la disciplina de restauración de señales recortadas y el algoritmo C-OMP como metodología de aproximación sucesiva para la reconstrucción de las mismas. Dentro de esta modalidad, se ha indagado particularmente en la recomposición de señales saturadas libres de aliasing.

El algoritmo propuesto, LFC-OMP, introduce la posibilidad de restaurar señales con recorte leve provenientes de un contexto analógico o de un tratamiento de reducción de aliasing. Las diferencias absolutas entre cada uno de los descriptores ΔSDR_c , SDR_c , PEAQ ODG y R_{nonlin} evaluados en LFC-OMP y su contraparte especializada en umbral de recorte plano, C-OMP, son poco significativas. Del mismo modo, existe una alta correlación lineal entre las mismas métricas evaluadas en ambos algoritmos. Estas últimas características, la mínima diferencia absoluta y la alta correlación lineal entre descriptores, son las que hacen a LFC-OMP una alternativa a considerar para la restauración de señales ligeramente recortadas con aliasing reducido.

En términos más generales, el alto grado de similitud alcanzado entre LFC-OMP y C-OMP es debido a la utilización del promediado de Fejér en la estimación del umbral de recorte en señales saturadas con aliasing reducido. La exploración de dichas técnicas y la proposición de nuevas metodologías similares deben tener el objetivo de lograr valores más altos de ΔSDR_c , SDR_c , PEAQ ODG y R_{nonlin} además de reducir la complejidad temporal que el algoritmo sugerido presenta.

7 LÍNEAS FUTURAS DE INVESTIGACIÓN

Esta investigación abre las puertas a varios potenciales estudios. Como se menciona en la conclusión de la misma, es necesario encontrar técnicas para reducir la complejidad temporal del método de estimación de umbrales de recorte propuesto. Esto podría derivarse además en un análisis exhaustivo de las limitaciones del método sugerido. Aquí, se podría contemplar la utilización muestras distorsionadas de forma analógica para la restauración, ya que en el presente estudio sólo se utilizan muestras distorsionadas artificialmente.

En la investigación presentada únicamente se trata el promediado de Fejér para la estimación del umbral de recorte en señales saturadas con bajo contenido de aliasing. Otro acercamiento, para este propósito, podría sugerir la utilización de métodos de aprendizaje de máquina, como puede ser el aprendizaje profundo a partir de redes neuronales. Asimismo, podrían utilizarse otros criterios de sumabilidad de Césaró con el mismo objetivo.

Adicionalmente, la exploración de nuevas técnicas para el tratamiento de señales basadas en los mencionados criterios de sumabilidad de Césaró es inminente. Por ejemplo, se podrían elaborar metodologías que usen los mismos para tratar la señal recortada libre de aliasing antes de ser procesada por un algoritmo como C-OMP. De este modo, se atenúa el fenómeno de Gibbs sobre la misma previo a la aplicación de un método tradicional para la restauración de recortes.

Por otra parte, en la investigación se usa el algoritmo C-OMP como base para el alcance del objetivo general. No obstante, la técnica basada en el promediado de Fejér, tratada en esta investigación, puede ser aplicada en otros algoritmos similares a C-OMP como, por ejemplo, SPADE. Sistemas de menor complejidad temporal aplicados a esta cuestión permitirían el procesamiento de una mayor cantidad de muestras en menor tiempo, lo cual resultaría en estudios estadísticos de mayor rigor.

Otros métodos para la reconstrucción de señales recortadas libres de aliasing pueden utilizar la ponderación de las columnas por medio de un vector de penalización, propuesta por Záviska et al. y mencionada en la presente investigación, que otorgue un peso signifi-

vo a frecuencias más bajas que las que reportan en su artículo. Esto es debido a que, en este caso, las altas frecuencias son mitigadas por el filtro anti-aliasing. De este modo, dado a que se brinda mayor importancia a los átomos del diccionario correspondientes a frecuencias graves, podrían lograrse estimaciones más similares a la verdad fundamental [6].

Aún considerando que en la presente investigación se evalúa el algoritmo propuesto con métricas de calidad perceptiva, sería prudente realizar pruebas subjetivas con muestras de sujetos con y sin entrenamiento musical, para contrastar los resultados obtenidos por estos descriptores en el contexto de los tipos de distorsión presentados en este estudio. Asimismo, otro posible tema de investigación relacionado es la creación de métricas de percepción de calidad de audio, similares a R_{nonlin} , pero dedicadas exclusivamente a la medición subjetiva de distorsión por aliasing.

Este estudio se ha dedicado exclusivamente a la reconstrucción de señales musicales con recorte. Sin embargo, las señales de habla son igualmente sensibles a los problemas mencionados en esta investigación. Otro tipo de señales en las que también frecuenta el problema estudiado, son las multicanal. Por lo tanto, estas deben ser sometidas a un análisis similar al llevado a cabo en este ensayo para consecuentemente elaborar nuevas técnicas que traten las mismas, en caso de ser necesario.

BIBLIOGRAFÍA

- [1] Tan, C.-T., Moore, B. C. y Zacharov, N., The effect of nonlinear distortion on the perceived quality of music and speech signals, *Journal of the Audio Engineering Society*, 51(11), 1012-1031, (2003).
- [2] Harvilla, M. y Stern, R., Least squares signal declipping for robust speech recognition, *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2073-2077, (2014).

- [3] Tachioka, Y., Narita, T. e Ishii, J., Speech recognition performance estimation for clipped speech based on objective measures, *Acoustical Science and Technology*, 35(6), 324-326, (2014).
- [4] Kraght, P. H., Aliasing in digital clippers and compressors, *Journal of the Audio Engineering Society*, 48(11), 1060-1065, (2000).
- [5] Defraene, B., Mansour, N., De Hertogh, S., Van Waterschoot, T., Diehl, M. y Moonen, M., Declipping of audio signals using perceptual compressed sensing, *IEEE Transactions on Audio, Speech, and Language Processing*, 21(12), 2627-2637, (2013).
- [6] Závíška, P., Rajmic, P. y Schimmel, J., Psychoacoustically motivated audio declipping based on weighted l_1 minimization, *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, 338-342, (2019).
- [7] Rencker, L., Bach, F., Wang, W. y Plumbley, M. D., Consistent dictionary learning for signal declipping, *International Conference on Latent Variable Analysis and Signal Separation*, 446-455, (2018).
- [8] Bilen, Ç., Ozerov, A. y Pérez, P., Audio declipping via nonnegative matrix factorization, *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 1-5, (2015).
- [9] Adler, A., Emiya, V., Jafari, M. G., Elad, M., Gribonval, R. y Plumbley, M. D., A constrained matching pursuit approach to audio declipping, *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 329-332, (2011).
- [10] Kitić, S., Bertin, N. y Gribonval, R., Sparsity and cosparsity for audio declipping: a flexible non-convex approach, *International Conference on Latent Variable Analysis and Signal Separation*, 243-250, (2015).
- [11] Závíška, P., Rajmic, P., Mokřý, O. y Prša, Z., A proper version of synthesis-based sparse audio declipper, *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 591-595, (2019).

- [12] Siedenburg, K., Kowalski, M. y Dörfler, M., Audio declipping with social sparsity, 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1577-1581, (2014).
- [13] Závíška, P., Rajmic, P., Ozerov, A. y Rencker, L., A survey and an extensive evaluation of popular audio declipping methods, IEEE Journal of Selected Topics in Signal Processing, 15(1), 5-24, (2020).
- [14] Godsill, S. y Rayner, P., Digital audio restoration, Department of Engineering, University of Cambridge, 5-7, (1997).
- [15] Kinzie Jr, G. R. y Gravereaux, D. W., Automatic detection of impulse noise, Journal of the Audio Engineering Society, 21(3), 181-184, (1973).
- [16] Carey, M. y Buckner, I., A system for reducing impulsive noise on gramophone reproduction equipment, Radio and Electronic Engineer, 50(7), 331-336, (1980).
- [17] Dolby, R., An audio noise reduction system, Journal of the audio engineering society, 15(4), 383-388, (1967).
- [18] Miller, N. y Stockham, T., Recovery of singing voice from noise by synthesis (development of speech analysis-synthesis system for extracting singing voice signal from background noise on recording device), 678-692, (1972).
- [19] Stockham, T. G., Cannon, T. M. e Ingebretsen, R. B., Blind deconvolution through digital signal processing, Proceedings of the IEEE, 63(4), 678-692, (1975).
- [20] Adler, A., Emiya, V., Jafari, M. G., Elad, M., Gribonval, R. y Plumbley, M. D., Audio inpainting, IEEE Transactions on Audio, Speech, and Language Processing, 20(3), 922-932, (2011).
- [21] Bertalmio, M., Sapiro, G., Caselles, V. y Ballester, C., Image inpainting, Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 417-424, (2000).

- [22] Janssen, A., Veldhuis, R. y Vries, L., Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(2), 317-330, (1986).
- [23] Mallat, S. G. y Zhang, Z., Matching pursuits with time-frequency dictionaries, *IEEE Transactions on signal processing*, 41(12), 3397-3415, (1993).
- [24] Davis, G. M., Mallat, S. G. y Zhang, Z., Adaptive time-frequency decompositions with matching pursuit, *Wavelet Applications*, volumen 2242, 402-413, (1994).
- [25] Marafioti, A., Majdak, P., Holighaus, N. y Perraudin, N., Gacela: a generative adversarial context encoder for long audio inpainting of music, *IEEE Journal of Selected Topics in Signal Processing*, 15(1), 120-131, (2020).
- [26] Abel, J. S., Restoring a clipped signal, *Acoustics, Speech, and Signal Processing, IEEE International Conference On*, 1745-1748, (1991).
- [27] Fong, W. y Godsill, S., Monte carlo smoothing for non-linearly distorted signals, 2001 *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, volumen 6, 3997-4000, (2001).
- [28] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J. y col., Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends® in Machine learning*, 3(1), 11-21, (2011).
- [29] Mokrý, O., Závíška, P., Rajmic, P. y Veselý, V., Introducing spain (sparse audio inpainter), 2019 27th European Signal Processing Conference (EUSIPCO), 1-5, (2019).
- [30] Rajmic, P., Závíška, P., Veselý, V. y Mokrý, O., A new generalized projection and its application to acceleration of audio declipping, *Axioms*, 8(3), 105, (2019).
- [31] Gorlow, S. y Reiss, J. D., Model-based inversion of dynamic range compression, *IEEE Transactions on Audio, Speech, and Language Processing*, 21(7), 1434-1444, (2013).
- [32] Manolakis, D. G. y Proakis, J. G., *Digital signal processing*, Pearson Education, 62, (2014).

- [33] Ávila, F. R., Tcheou, M. P. y Biscainho, L. W., Audio soft declipping based on constrained weighted least squares, *IEEE Signal Processing Letters*, 24(9), 1348-1352, (2017).
- [34] Avila, F. R. y Biscainho, L. W., Audio soft declipping based on weighted l_1 -norm, 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 299-303, (2017).
- [35] Eargle, J. M., *Handbook of recording engineering*, Springer Science & Business Media, 189, (2012).
- [36] Mapes-Riordan, D., A worst-case analysis for analog-quality (alias-free) digital dynamics processing, *Journal of the Audio Engineering Society*, 47(11), 948-952, (1999).
- [37] Hauser, M. W., Principles of oversampling a/d conversion, *Journal of the Audio Engineering Society*, 39(1/2), 3-26, (1991).
- [38] Esqueda, F., Bilbao, S. y Välimäki, V., Aliasing reduction in clipped signals, *IEEE Transactions on Signal Processing*, 64(20), 5255-5267, (2016).
- [39] Hadamard, J., Sur les problèmes aux dérivées partielles et leur signification physique, *Princeton university bulletin*, 49-52, (1902).
- [40] Gaultier, C., Design and evaluation of sparse models and algorithms for audio inverse problems, 128(9), Tesis doctoral, 9-39, (2019).
- [41] Mallat, S., *A wavelet tour of signal processing*, Elsevier, 89-432, (1999).
- [42] Gabor, D., Theory of communication. part 1: the analysis of information, *Journal of the Institution of Electrical Engineers-part III: radio and communication engineering*, 93(26), 429-441, (1946).
- [43] Bishop, C. M., *Pattern recognition, Machine learning*, 128(9), 4-25, (2006).
- [44] Weinstein, A. J. y Wakin, M. B., Recovering a clipped signal in sparseland, *Sampling Theory in Signal and Image Processing*, 12, 55-69, (2013).

- [45] Jerri, A. J., The gibbs phenomenon in fourier analysis, splines and wavelet approximations, Springer Science & Business Media, 8-31, (1998).
- [46] Wilbraham, H., On a certain periodic function, The Cambridge and Dublin Mathematical Journal, 3, 198-201, (1848).
- [47] Michelson, A. A., Fourier's series, Nature, 59(1522), 200-200, (1898).
- [48] Michelson, A. y Stratton, S., Vi. a new harmonic analyser, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 45(272), 85-91, (1898).
- [49] Gibbs, J. W., Fourier's series, Nature, 59(1522), 200, (1898).
- [50] Gibbs, J. W., Fourier's series, Nature, 59(1522), 606, (1898).
- [51] Jerri, A., Integral and discrete transforms with applications and error analysis, CRC Press, 626, (2021).
- [52] Lanczos, C. y Boyd, J., Discourse on fourier series, SIAM, 51-61, 67, (2016).
- [53] Lanczos, C., Applied analysis, Courier Corporation, 225, (1988).
- [54] Hamming, R., Numerical methods for scientists and engineers, Courier Corporation, 534, (2012).
- [55] Hamming, R. W., Digital filters, Courier Corporation, 75, (1998).
- [56] Fay, T. H. y Schulz, K. G., The gibbs' phenomenon from a signal processing point of view, International Journal of Mathematical Education in Science and Technology, 32(6), 863-872, (2001).
- [57] Weisstein, E. W., *Wilbraham-Gibbs Constant*, Extraído el 22 de Octubre de 2022, <https://mathworld.wolfram.com/Wilbraham-GibbsConstant.html>.
- [58] Schafer, R. W. y Rabiner, L. R., Digital representations of speech signals, Proceedings of the IEEE, 63(4), 274-277, (1975).
- [59] Natarajan, B. K., Sparse approximate solutions to linear systems, SIAM journal on computing, 24(2), 227-234, (1995).

- [60] Davis, G., Mallat, S. y Avellaneda, M., Adaptive greedy approximations, Constructive approximation, 13(1), 57-98, (1997).
- [61] Davies, M. E. y Daudet, L., Sparse audio representations using the mclt, Signal processing, 86(3), 457-470, (2006).
- [62] Gribonval, R., Fast matching pursuit with a multiscale dictionary of gaussian chirps, IEEE Transactions on signal Processing, 49(5), 994-1001, (2001).
- [63] Thiede, T., Treurniet, W. C., Bitto, R., Schmidmer, C., Sporer, T., Beerends, J. G. y Colomes, C., Peaq-the itu standard for objective measurement of perceived audio quality, Journal of the Audio Engineering Society, 48(1/2), 3-29, (2000).
- [64] Defraene, B., Waterschoot, T. van, Diehl, M. y Moonen, M., Subjective audio quality evaluation of embedded-optimization-based distortion precompensation algorithms, The Journal of the Acoustical Society of America, 140(1), EL101-EL106, (2016).
- [65] Tan, C.-T., Moore, B. C., Zacharov, N. y Mattila, V.-V., Predicting the perceived quality of nonlinearly distorted music and speech signals, Journal of the Audio Engineering Society, 52(7/8), 699-711, (2004).
- [66] Diamond, S. y Boyd, S., CVXPY: a Python-embedded modeling language for convex optimization, Journal of Machine Learning Research, (2016).
- [67] EBU, EBU SQAM, Extraído el 28 de Marzo de 2022, <https://tech.ebu.ch/publications/sqamcd>.
- [68] Ozerov, A., Bilen, Ç. y Pérez, P., Multichannel audio declipping, 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 659-663, (2016).
- [69] Gaultier, C., Bertin, N. y Gribonval, R., Cascade: channel-aware structured cosparse audio declipper, 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 571-575, (2018).
- [70] Encyclopedia of Mathematics, *Linear interpolation*, Extraído el 13 de Julio de 2022, https://encyclopediaofmath.org/index.php?title=Linear_interpolation.

- [71] Davis, P. J., Interpolation and approximation, Courier Corporation, 56-57, (1975).
- [72] Instruments, N., *Anti-Aliasing Filters and Their Usage Explained*, Extraído el 27 de Julio de 2022, <https://www.ni.com/en-us/innovations/white-papers/18/anti-aliasing-filters-and-their-usage-explained.html>.
- [73] Smith, S. W. y col., The scientist and engineer's guide to digital signal processing, California Technical Pub. San Diego, 52-59, (1997).
- [74] Kabal, P. y col., An examination and interpretation of itu-r bs. 1387: perceptual evaluation of audio quality, TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University, 1-89, (2002).
- [75] Cohen, J., Cohen, P., West, S. G. y Aiken, L. S., Applied multiple regression/correlation analysis for the behavioral sciences, Routledge, 310, (2013).