



Modelado de las etapas de amplificación y distorsión de guitarra eléctrica mediante redes neuronales recurrentes y redes neuronales convolucionales

Calahorrano Franco Steven Esmaylin

Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Trabajo de titulación, previo a la obtención del Título de Ingeniero en Electrónica y Telecomunicaciones

Ing. Enrique Vinicio Carrera, PhD

19 de enero del 2023

Originality report

COURSE NAME

Tesis

STUDENT NAME

STEVEN ESMAYLIN CALAHORRANO FRANCO

FILE NAME

STEVEN ESMAYLIN CALAHORRANO FRANCO - Documento

REPORT CREATED

Jan 18, 2023

Summary

Flagged passages	1	2%
Cited/quoted passages	0	0%

Web matches

docplayer.es	1	2%
--------------	---	----



Firmado electrónicamente por:
**ENRIQUE
VINICIO
CARRERA ERAZO**

Ing. Enrique Vinicio Carrera Erazo, PHD

Director



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Certificación

Certifico que el trabajo de titulación: **"Modelado de las etapas de amplificación y distorsión de guitarra eléctrica mediante redes neuronales recurrentes y redes neuronales convolucionales"** fue realizado por el señor **Calahorrano Franco Steven Esmaylin**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 19 de enero del 2023



Firmado digitalmente por:
**ENRIQUE
VINICIO
CARRERA ERAZO**

Ing. Enrique Vinicio Carrera Erazo, PHD

C. C 1708792104



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Responsabilidad de Autoría

Yo, **Calahorrano Franco Steven Esmaylin**, con cédula de ciudadanía **1750764431**, declaro que el contenido, ideas y criterios del trabajo de titulación: **"Modelado de las etapas de amplificación y distorsión de guitarra eléctrica mediante redes neuronales recurrentes y redes neuronales convolucionales"** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 19 de enero del 2023

Calahorrano Franco Steven Esmaylin

C.C.: 1750764431



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica y Telecomunicaciones

Autorización de Publicación

Yo **Calahorrano Franco Steven Esmaylin**, con cédula de ciudadanía 1750764431, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **"Modelado de las etapas de amplificación y distorsión de guitarra eléctrica mediante redes neuronales recurrentes y redes neuronales convolucionales"** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 19 de enero del 2023

Calahorrano Franco Steven Esmaylin

C.C.: 1750764431

Dedicatoria

Este trabajo de titulación está dedicado a mis padres, ya que su gran esfuerzo e incondicional apoyo han servido como pilares y ejemplo de personas inteligentes, amorosas, trabajadoras y honestas, que han sabido guiarme con sabiduría y gran dedicación.

A la música la cual ha sido fuente de inspiración, creatividad y paz.

Agradecimientos

La vida es un conjunto de momentos que nos definen como personas, lleno de aciertos, errores, fracasos y victorias. Cada paso que damos por pequeño que este sea nos sitúa en un momento diferente en nuestro camino.

Agradezco profundamente a toda mi familia por haberme guiado al camino en el que me encuentro actualmente y poder permitirme ver la persona que soy y en lo que quiero convertirme.

Especialmente a mis padres que me han brindado su incondicional amor, han sido siempre una luz en momentos difíciles, siempre me han apoyado y creído en mí, dándome consejos y ayudándome a encontrar siempre la manera de ser feliz.

Lo único constante en la vida es el cambio y agradezco a todas las personas que de alguna manera han formado parte de ella y me ha permitido llegar a este momento de culminar mi carrera universitaria, que después de tanto esfuerzo y dedicación ha dado sus frutos.

Agradezco a mis compañeros y amigos por haber formado parte de este proceso de formación además de a los profesores que han sabido inspirar y alentar a sus estudiantes.

Y finalmente a mi tutor de tesis quien me ha apoyado pacientemente y me ha guiado de la mejor manera para poder culminar con el presente trabajo de titulación, a todos desde lo más profundo de mi corazón muchas gracias.

Índice De Contenido

Dedicatoria.....	6
Agradecimientos	7
Resumen	16
Abstract.....	17
Capítulo I	18
Planteamiento del problema de investigación	18
Antecedentes	18
Justificación.....	19
Objetivos	21
<i>Objetivo General</i>	21
<i>Objetivos Específicos</i>	21
Capítulo II	22
Fundamento Teórico.....	22
Funcionamiento de la guitarra eléctrica	22
Pastillas.....	22
Amplificadores de guitarra	23
<i>Mesa Boogie Dual Rectifier</i>	24
Pedales de guitarra	25
<i>Ibanez Tube Screamer</i>	25
Modelado de equipo analógico.	26

Redes Neuronales artificiales	27
Deep learning o aprendizaje profundo.....	28
<i>Redes neuronales convolucionales</i>	29
<i>Modelo WaveNet</i>	30
<i>Redes neuronales recurrentes</i>	30
<i>Redes LSTM</i>	31
Frameworks de redes neuronales	32
Optimizadores	33
<i>Optimizador Adam</i>	33
Funciones de pérdida	34
<i>ESR</i>	35
<i>Filtro de pre-énfasis</i>	35
Capítulo III	36
Implementación de los sistemas	36
Introducción.....	36
Entorno de desarrollo	36
Obtención del set de datos	38
Análisis de los datos obtenidos de entrada.....	41
Parámetros de entrenamiento de las redes	45
Preprocesamiento del set de datos	48
Implementación de la red neuronal convolucional	49

	10
<i>Arquitectura de la CNN</i>	49
<i>Parámetros de la red CNN</i>	49
Implementación de la red neuronal recurrente.....	50
<i>Arquitectura de la RNN</i>	50
<i>Parámetros de la red RNN</i>	50
Entrenamiento de las redes.....	51
Capítulo IV	52
Resultados.....	52
Introducción.....	52
Análisis de los modelos convolucionales obtenidos.....	52
<i>Mesa Boogie Dual Rectifier CNN</i>	53
<i>Ibanez Tube Screamer CNN</i>	55
Análisis de los modelos recurrentes obtenidos	57
<i>Mesa Boogie Dual Rectifier RNN</i>	58
<i>Ibanez Tube Screamer RNN</i>	60
Comparación entre los modelos CNN y RNN	62
<i>Amplificador Mesa Boogie Dual Rectifier</i>	62
<i>Pedal de distorsión Ibanez Tube Screamer</i>	66
Resumen objetivo de los modelos	71
Análisis subjetivo.....	72
Conclusiones	77

Recomendaciones	79
Trabajos Futuros	80
Bibliografía.....	82

Índice de Figuras

Figura 1 <i>Campo electromagnético inducido en una pastilla Humbucker con la vibración de las cuerdas.....</i>	23
Figura 2 <i>Amplificador de guitarra Mesa Boogie Dual Rectifier.....</i>	24
Figura 3 <i>Ibanez Tube Screamer TS808.....</i>	26
Figura 4 <i>Representación de sistemas de Caja blanca y Caja negra.....</i>	27
Figura 5 <i>Estructura de una red neuronal.....</i>	28
Figura 6 <i>Estructura tradicional de una CNN.....</i>	29
Figura 7 <i>Capas de convolución dilatadas con diferentes factores de dilatación.....</i>	30
Figura 8 <i>Estructura de una RNN.....</i>	31
Figura 9 <i>Diagrama de una unidad LSTM.....</i>	32
Figura 10 <i>Comparación del costo de entrenamiento de Adam con otros algoritmos para la optimización.....</i>	34
Figura 11 <i>Compatibilidad cuda de la tarjeta gráfica.....</i>	37
Figura 12 <i>Esquema de conexión de guitarra directo a línea.....</i>	38
Figura 13 <i>Esquema de reamping para capturar el sonido del amplificador.....</i>	39
Figura 14 <i>Esquema de reamping para capturar el sonido del pedal de distorsión.....</i>	40
Figura 15 <i>Señales de entrada en el dominio del tiempo.....</i>	41
Figura 16 <i>Señales de entrada en el dominio de la frecuencia.....</i>	43
Figura 17 <i>Espectrograma de la señal limpia.....</i>	44
Figura 18 <i>Espectrograma del amplificador Mesa Boogie Dual Rectifier.....</i>	44
Figura 19 <i>Espectrograma del Ibanez Tube Screamer.....</i>	45
Figura 20 <i>Arquitectura de la red CNN.....</i>	49

Figura 21 <i>Arquitectura de la red RNN.....</i>	50
Figura 22 <i>Gráficos comparativos del modelo de emulación convolucional del amplificador vs el amplificador Mesa Boogie Dual Rectifier real.....</i>	53
Figura 23 <i>Gráficos comparativos a detalle del modelo de emulación convolucional del amplificador vs el amplificador Mesa Boogie Dual Rectifier real.....</i>	54
Figura 24 <i>Gráficos comparativos del modelo de emulación convolucional del pedal de distorsión vs el pedal Ibanez Tube Screamer.....</i>	55
Figura 25 <i>Gráficos comparativos a detalle del modelo de emulación convolucional del pedal de distorsión vs el pedal Ibanez Tube Screamer.....</i>	56
Figura 26 <i>Gráficos comparativos del modelo de emulación recurrente del amplificador vs el amplificador Mesa Boogie Dual Rectifier real.....</i>	58
Figura 27 <i>Gráficos comparativos del modelo de emulación recurrente del amplificador vs el amplificador Mesa Boogie Dual Rectifier real.....</i>	59
Figura 28 <i>Gráficos comparativos del modelo de emulación recurrente del pedal de distorsión vs el pedal Ibanez Tube Screamer.....</i>	60
Figura 29 <i>Gráficos comparativos a detalle del modelo de emulación recurrente del pedal de distorsión vs el pedal Ibanez Tube Screamer.....</i>	61
Figura 30 <i>Comparación de la forma de onda del amplificador real con los modelos emulados.....</i>	63
Figura 31 <i>Comparación en el espectro de frecuencias del amplificador real con los modelos emulados.....</i>	64
Figura 32 <i>Espectrograma del amplificador real.....</i>	65
Figura 33 <i>Espectrograma del modelo convolucional del amplificador.....</i>	65
Figura 34 <i>Espectrograma del modelo recurrente del amplificador.....</i>	66

Figura 35 <i>Comparación de la forma de onda del pedal real con los modelos emulados.....</i>	68
Figura 36 <i>Comparación en el espectro de frecuencias del pedal real con los modelos emulados.....</i>	69
Figura 37 <i>Espectrograma del pedal de distorsión real.....</i>	70
Figura 38 <i>Espectrograma de la emulación convolucional del pedal de distorsión.....</i>	70
Figura 39 <i>Espectrograma de la emulación recurrente del pedal de distorsión.....</i>	71
Figura 40 <i>Pregunta 1 de la encuesta</i>	73
Figura 41 <i>Pregunta 2 de la encuesta.....</i>	74
Figura 42 <i>Pregunta 3 de la encuesta.....</i>	74
Figura 43 <i>Pregunta 4 de la encuesta.....</i>	75
Figura 44 <i>Pregunta 5 de la encuesta</i>	76

Índice de Tablas

Tabla 1 <i>Parámetros de grabación</i>	38
Tabla 2 <i>Parámetros configurados en el amplificador Mesa Boogie Dual Rectifier</i>	39
Tabla 3 <i>Parámetros configurados en el pedal de distorsión</i>	40
Tabla 4 <i>Resultados obtenidos en modelo convolucional</i>	47
Tabla 5 <i>Resultados obtenidos en modelo recurrente</i>	48
Tabla 6 <i>Parámetros de la red CNN</i>	49
Tabla 7 <i>Parámetros de la red RNN</i>	51
Tabla 8 <i>Tabla de resumen de resultados objetivos de análisis</i>	71
Tabla 9 <i>Etiquetas de los audios de la encuesta</i>	72
Tabla 10 <i>Calidad subjetiva de los audios</i>	76

Resumen

Los amplificadores de guitarra y la introducción de efectos analógicos tienen mucha importancia en la historia de la música desde ya hace varias décadas, ya que son dispositivos muy requeridos tanto para grabaciones como sonido en vivo. El presente trabajo de titulación realiza el modelamiento de los amplificadores de guitarra eléctrica y el efecto producido por un pedal, específicamente el efecto de distorsión que es el más comúnmente utilizado. Para esto se emplea aprendizaje automático y de esta manera realizar una comparación del desempeño de esta tecnología en el campo de la producción musical en términos de calidad de sonido, costos de implementación, facilidad de uso, eficiencia en los procesos computacionales, entre otros. Con este estudio se pretende obtener modelos que específicamente emulen un amplificador de guitarra (*Mesa Boogie Dual Rectifier*) y el efecto distorsión producido por un pedal (*Ibanez Tube Screamer*), a partir de redes neuronales recurrentes y redes neuronales convolucionales. Además, realizar una comparación que permita determinar de manera subjetiva la calidad de sonido obtenido con el uso de estos modelos, empleando una encuesta. Y de manera objetiva con herramientas como, cálculos de errores y comparaciones del contenido frecuencial de las señales para poder extraer características que presentan estos dos modelos basados en el aprendizaje automático.

Palabras clave: Redes neuronales recurrentes, Redes neuronales convolucionales, aprendizaje automático, amplificador analógico, efecto de distorsión.

Abstract

Guitar amplifiers and the introduction of analog effects have been very important in the history of music for several decades now, since they are highly required devices for both recording and live sound. The present work performs the modeling of electric guitar amplifiers and the effect produced by a pedal, specifically the distortion effect that is the most commonly used in this field. For this, machine learning is used and, in this way, make a comparison of the performance of this technology in the field of music production in terms of sound quality, the implementation costs, the ease of use, the efficiency in computational processes, among others. This study aims to obtain the models that specifically emulate a guitar amplifier (Mesa Boogie Dual Rectifier) and the distortion effect produced by a pedal (Ibanez Tube Screamer), from recurrent neural networks and convolutional neural networks. In addition, to carry out a comparison that allows to subjectively determine the sound quality obtained with the use of these models, using a survey. And objectively with tools such as the error calculations and comparisons of the frequency content of the signals in order to extract characteristics presented by these two models based on machine learning.

Keywords: Recurrent neural networks, Convolutional neural networks, machine learning, analog amplifier, distortion effect.

Capítulo I

Planteamiento del problema de investigación

Antecedentes

El procesamiento digital de señales en la actualidad está siendo utilizado en múltiples campos del conocimiento, uno de ellos es el campo del procesamiento de las señales de audio. En particular, el procesamiento de audio aplicado en la producción musical se encuentra en constante evolución y desarrollo mediante nuevas técnicas y herramientas que permiten realizar, de manera digital, tareas que antiguamente solo eran posibles con el uso de hardware analógico (Dogorov, 2019). Un ejemplo de ello son los amplificadores de guitarra eléctrica y los pedales de efectos para producir distorsión, que en la actualidad pueden ser emulados digitalmente y para lo que se ocupan diversas técnicas, siendo la más común los modelos matemáticos que aproximan la circuitería del hardware real (Cobas, 2018).

Las diferentes técnicas que pueden ser utilizadas para el modelamiento digital del audio pretenden en la mayoría de los casos, emular con mayor precisión el sonido generado analógicamente. De este modo la señal de la guitarra atraviesa un proceso de digitalización con parámetros comúnmente utilizados de frecuencias de muestreo alrededor de los 44.1 kHz a los 96 kHz, un número de bits de 16, 24 y 32, uso de códec como el LPCM para un formato Wav el cual es el más común utilizado a nivel profesional, técnicas de compresión de audio basadas en codificación sub banda y utilización de filtros QMF, además del uso de filtros activos de segundo orden como el Sallen-Key o incluso filtros de tercer orden como Butterworth, Chebychev, Bessel y Elíptico que son utilizados en el modelado de efectos de guitarra (Ordiales & Barrera, 2019).

Estos modelos digitales se enfocan en mejorar temas como la calidad general del audio, la similitud con el hardware real, reducir el nivel de procesamiento computacional requerido

para su emulación, reducir el costo mediante el uso de software en lugar de hardware, tener mayor versatilidad y facilidad de uso, entre otras (MathWorks, 2021).

Justificación

El funcionamiento de una guitarra eléctrica está basado en una pastilla electromagnética, la cual capta el sonido de las cuerdas y requiere de una etapa de amplificación. Los primeros amplificadores eran de baja potencia y poseían limitados controles de ecualización. Sin embargo, las necesidades de los músicos de la época llevaron a empresas como Fender y Marshall a desarrollar sistemas de amplificación de mayor potencia, así como también experimentaron modificando el sonido natural de la guitarra para generar nuevos tipos de sonidos y efectos (Valenzuela, 2013).

Para la construcción de estos amplificadores se emplean diversos mecanismos tales como los amplificadores de válvulas o tubos los cuales funcionan con válvulas de vacío. Estos elementos amplifican la señal eléctrica mediante un mecanismo de control de movimiento de los electrones insertos en gases especialmente seleccionados. Por otro lado, los amplificadores basados en transistores utilizan estos dispositivos basados en silicio para amplificar la señal y son especialmente populares debido a su bajo costo. Finalmente, se tienen a los amplificadores digitales los cuales son de aparición más reciente. Estos amplificadores son básicamente emuladores o recreaciones de modelos de amplificadores reales en la mayoría de los casos con la utilización de filtros digitales (Cayetano Matamoros, 2017).

En la actualidad, la utilización de amplificadores y efectos digitales para guitarra cada vez aumenta debido a que estos presentan varias ventajas frente a los amplificadores de válvulas, transistores o sistemas híbridos, debido a que los costos se ven reducidos, tienen mayor versatilidad en cuanto a la cantidad de efectos que pueden procesar, así como los tonos que pueden reproducir. Además de la capacidad de poder correr como *software* o *plug-ins* en

un computador de propósito general, dejando de lado el hardware específico que requiere de mantenimiento, espacio y una mayor inversión económica (Sanchez, 2005).

El desarrollo de nuevas técnicas y sistemas que permiten la digitalización del audio con menores pérdidas, han permitido que estos sistemas de audio digitales cada vez sean mejores, llegando incluso en ciertos casos a que las diferencias entre un amplificador real y un *software* que emule tal amplificador sean casi imperceptibles. Aun así, la principal desventaja que muchos de estos pedales y amplificadores digitales presentan es la calidad de sonido, ya que muchas veces estos modelos creados no son lo suficientemente buenos o adecuados (Jakub, 2021).

En los últimos años, nuevas técnicas de modelado se utilizan para tratar de replicar con exactitud las características de amplificadores y efectos reales, para que el sonido generado tenga una gran calidad y puedan ejecutarse de manera simple y correcta en un computador. Muchas de estas técnicas se encuentran aún en desarrollo, generando cada vez mejores resultados. Uno de los métodos para modelado de audio que presenta enorme potencial es la utilización de redes neuronales artificiales; es decir, el uso de *Machine Learning* y/o *Deep learning* para extraer las características del hardware y emularlo en un software computacional (Eero-Pekka, Lauri, Etienne, & Vesa, 2019).

El modelamiento digital de las características de amplificadores y efectos de guitarra o efectos en general es el paso primordial previo a la implementación de un *software* que opere en tiempo real, o plug-ins que funcionen en *software* especializado para la producción de música (Fermatta, 2019).

Debido a lo anteriormente expuesto, consideramos que es importante la investigación en las técnicas de modelamiento de efectos digitales, ya que esto proporciona herramientas de apoyo, las cuales serán más versátiles, eficientes, económicas y de gran calidad para la

producción de audio, adicionalmente que sean fáciles de utilizar y están al alcance de músicos, productores, disc-jockeys, ingenieros y personas involucradas en la producción sonora (Wright, Eero-Pekka, Lauri, & Vesa, 2019).

Objetivos

Objetivo General

Modelar un amplificador y un pedal de distorsión de guitarra eléctrica en base al uso de redes neuronales recurrentes y redes neuronales convolucionales.

Objetivos Específicos

- Estudiar el estado del arte y los conceptos básicos correspondientes a las redes neuronales y su uso en el procesamiento de señales de audio.
- Implementar el modelado digital de un amplificador de guitarra eléctrica con redes neuronales recurrentes y redes neuronales convolucionales.
- Implementar el modelado digital de un pedal de distorsión de guitarra eléctrica con redes neuronales recurrentes y redes neuronales convolucionales.
- Comparar el desempeño del audio generado por el amplificador y distorsión real (hardware) con el amplificador y distorsión modelado con redes neuronales recurrentes y con el amplificador y efecto modelado con redes convolucionales.

Capítulo II

Fundamento Teórico

Funcionamiento de la guitarra eléctrica

En las guitarras eléctricas se produce una vibración en las cuerdas la cual produce un efecto sonoro captado por las denominadas pastillas electromagnéticas o micrófonos, estas atraviesan el circuito interno de cada guitarra, el cual generalmente está compuesto de un control de tono; es decir, un filtro de frecuencias y un control de volumen. Luego, esta señal se envía mediante un cable hacia un amplificador el cual se encarga de amplificar la señal que procede de la guitarra hacia una etapa de potencia (Lähdevaara, 2012).

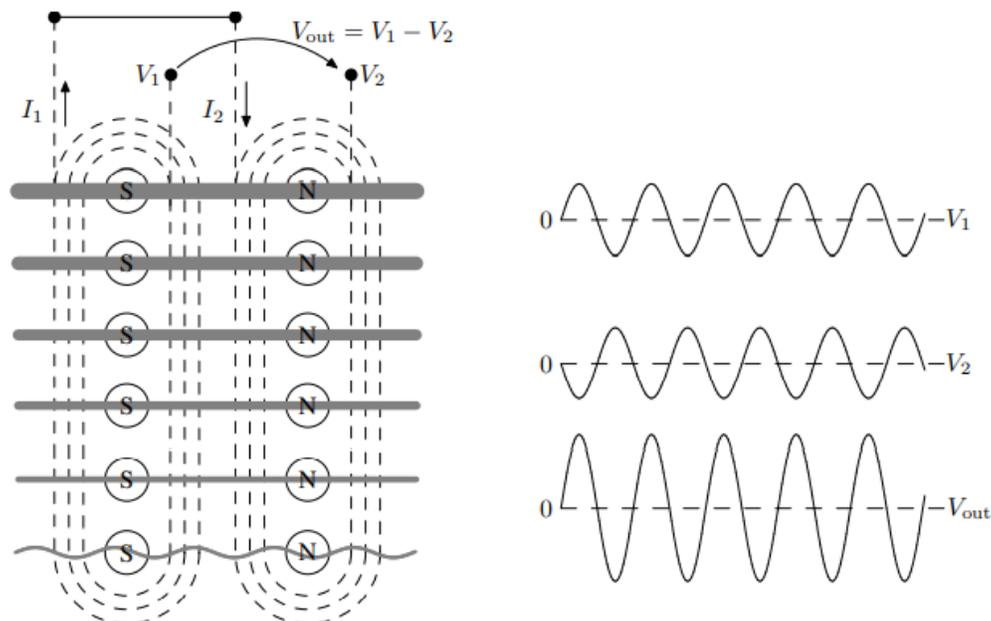
Pastillas

A diferencia de una guitarra acústica o clásica, una guitarra eléctrica requiere de un elemento eléctrico para poder tomar el sonido y transformarlo en un impulso eléctrico que después puede ser amplificado. Las denominadas pastillas, son generadores eléctricos que basan su funcionamiento en la ley de inducción magnética de Faraday como se puede observar en la Figura 1. Están compuestas de imanes que actúan sobre cada cuerda por lo tanto van situados debajo de ellas, de esta manera al producirse un movimiento se altera el campo magnético lo cual produce la corriente eléctrica (Cobas, 2018).

Existen principalmente pastillas simples denominadas *single coil* o bobina simple y las denominadas *Humbucker* las cuales son un sistema compuesto de un par de bobinas las cuales poseen 2 imanes para cada cuerda, ambas poseen cualidades sonoras que las diferencian entre sí y puede destacarse que el tipo de construcción de las pastillas *Humbucker* las hacen menos propensas a ruidos o interferencias (Henshall, 2011).

Figura 1

Campo electromagnético inducido en una pastilla Humbucker con la vibración de las cuerdas



Nota. Tomado de *The science of electric guitars and guitar electronics*, por Lähdevaara, J. 2012.

Amplificadores de guitarra

Los amplificadores esencialmente toman el sonido captado por la pastilla electromagnética o micrófono de guitarra y la amplifican, existen tres tipos de amplificadores los cuales están clasificados por sus mecanismos de construcción: amplificadores de transistores, amplificadores de válvulas y amplificadores digitales (Valenzuela, 2013).

Los amplificadores de transistores utilizan como su nombre lo dice, transistores para amplificar la señal de la guitarra. Estos se caracterizan por tener un costo no tan elevado, no requieren de demasiado mantenimiento y generalmente son livianos. Por otro lado, los amplificadores de válvulas utilizan un mecanismo con tubos de vacío el cual fue el primer sistema de amplificación diseñado para guitarra y considerado por muchos como el sonido más

natural, cálido y preferido para la mayoría de guitarristas, pero de manera general suelen ser los más costosos y requieren de mayor mantenimiento, finalmente se tienen los amplificadores digitales los cuales generalmente simulan otros tipos de amplificadores (Valenzuela, 2013).

Mesa Boogie Dual Rectifier

Es un amplificador diseñado por la empresa *Mesa Engineering* de su línea insignia conocida como *series rectifier*. Este amplificador, el cual se puede observar en la Figura 2, debe su nombre a que tiene dos formas de rectificación eléctrica, por un lado, un sistema basado en diodos de Silicio y otro con tubos de vacío. Es un amplificador de alta ganancia el cual obtiene esta característica de un circuito preamplificador de ganancia en cascada, este posee varios canales los cuales pueden ser seleccionados con los selectores en su parte frontal el cual posee desde sonidos muy limpios hasta distorsiones muy pesadas (Hunter, 2002).

Figura 2

Amplificador de guitarra Mesa Boogie Dual Rectifier



Nota. Tomado del sitio oficial *Mesa Boogie*, por Mesa Boogie, 2022.

Pedales de guitarra

Son dispositivos externos a la guitarra los cuales sirven para modificar el sonido que esta produce. Estos funcionan haciendo un procesamiento de la señal de la misma, la cual puede modificar parámetros como: fase, amplitud y frecuencia, estos pueden interconectarse unos con otros para producir efectos combinados los cuales dependiendo del orden en el cual estén conectados, pueden producir diferentes tipos de sonido (Evangelista, 2010).

Existen varios tipos de efectos, se destacan principalmente los efectos de distorsión entre los que se encuentran los pedales de: *fuzz*, *overdrive*, *drive*, pedales de modulación como *phaser*, *chorus*, *flanger* y pedales con efectos de tiempo como *delay* o reverberación (Rico, 2015).

Estos dispositivos externos o procesadores pueden ser análogos los cuales hacen uso de elementos electrónicos como: transistores y amplificadores operacionales para modificar la señal eléctrica producida por la guitarra. De la misma manera estos dispositivos pueden ser digitales; es decir, el procesamiento de la señal digital por medio de algoritmos y convertidores AD (analógico - digital) y DA (digital - analógico) (Burgos & Vargas, 2016).

Ibanez Tube Screamer

Es un pedal de distorsión producido por la empresa *Ibanez*, como se puede observar en la Figura 3, es un clásico utilizado en música rock desde los 70's hasta la actualidad, existen diferentes versiones del mismo. Este debe su nombre al tipo de *overdrive* que aporta que se asimila a un sonido de distorsión de un amplificador de tubos y esencialmente posee tres controles los cuales se pueden modificar (*Gain*, *Tone*, *Level*) (Roascio, 2007).

Figura 3

Ibanez Tube Screamer TS808



Nota. Tomado del sitio oficial *Ibanez*, por Ibanez, 2022.

Modelado de equipo analógico.

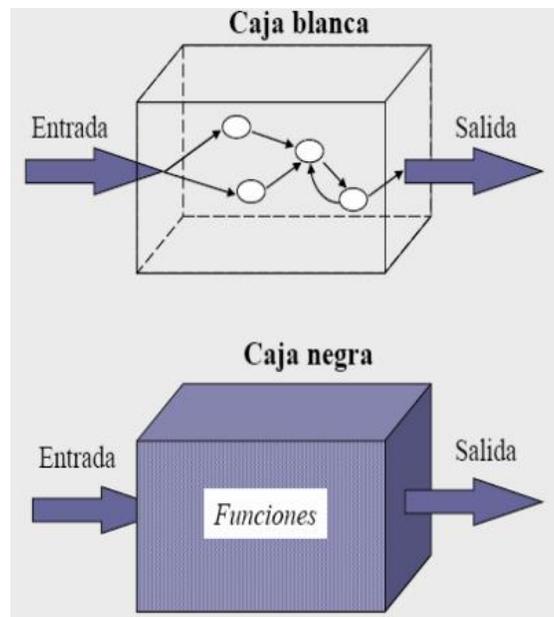
El modelado se basa en las técnicas utilizadas para la obtención de un modelo el cual se define como una representación de manera simple de características y aspectos de un sistema real (Zeigler, Muzy, & Kofman, 2019).

Se tienen modelos matemáticos los cuales se clasifican en modelos de caja blanca y modelos de caja negra como se observa en la Figura 4. Por un lado, los modelos matemáticos de caja blanca pretenden realizar un análisis riguroso del circuito del equipo a emular y se realizan aproximaciones matemáticas de los elementos del circuito, para obtener un modelo el cual se basa principalmente en las relaciones matemáticas entre las variables significativas de un sistema de una manera visible. Por otro lado, el modelo matemático de caja negra tiene un

enfoque inductivo y es utilizado como un sistema predictivo el cual refleja el funcionamiento de un sistema y puede ser implementado con el uso de redes neuronales (Velten, 2009).

Figura 4

Representación de sistemas de Caja blanca y Caja negra



Nota. Tomado de *ISI*, por Universidad De Sevilla, 2022.

Redes Neuronales artificiales

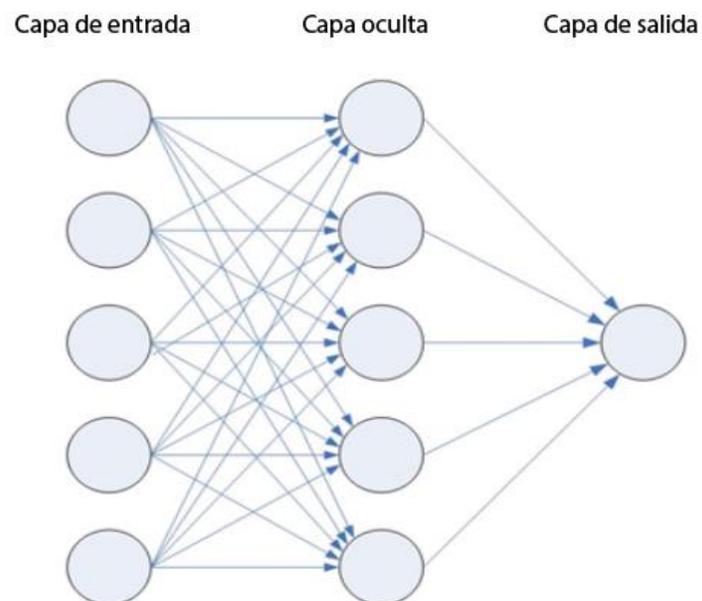
Son una representación de cómo funciona de manera simplificada un cerebro humano, está conformada por unidades denominadas neuronas las cuales son organizadas en capas para formar las redes, estas no pretenden representar el funcionamiento fisiológico real de una neurona sino simplemente el comportamiento global de una red (Izaurieta & Carlos, 2000).

Generalmente las neuronas están organizadas en tres tipos de capas (capa de entrada, capas ocultas y capa de salida) como muestra la Figura 5. Estas unidades se interconectan por medio de ponderaciones que en un principio son aleatorias y van cambiando su peso de acuerdo al entrenamiento de la red, lo cual consiste en dar ejemplos a la red para que trate de

ajustar las predicciones a las respuestas de los ejemplos con resultados conocidos. De esta manera la red tiene un aprendizaje continuo que hace que la red cada vez sea más precisa. Cuando el entrenamiento finaliza la red puede ser capaz de realizar predicciones para entradas desconocidas (IBM, 2021).

Figura 5

Estructura de una red neuronal



Deep learning o aprendizaje profundo

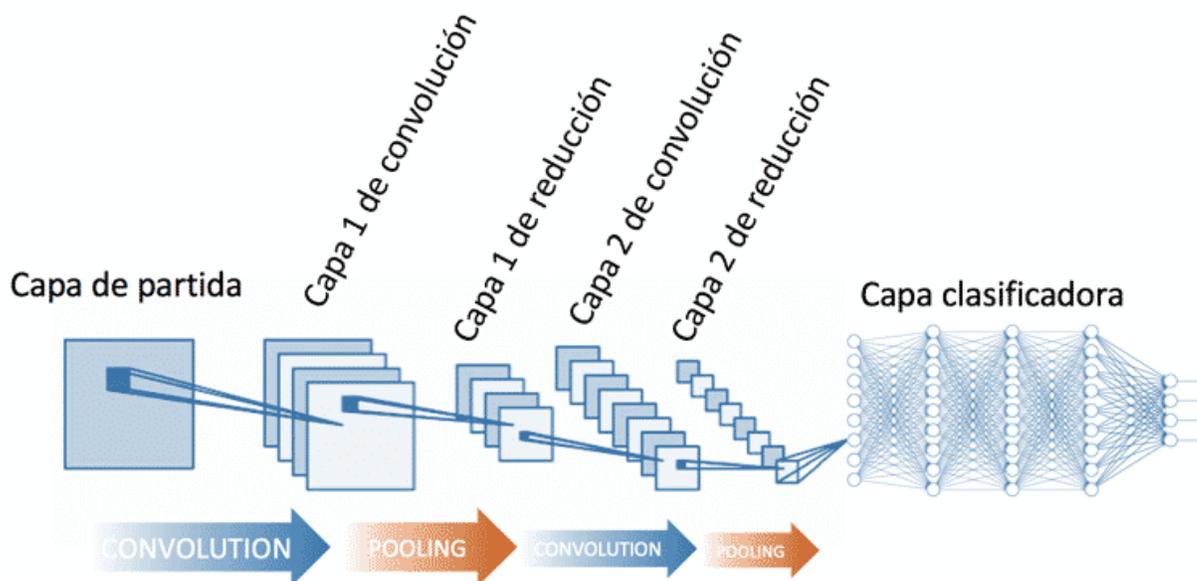
Son estructuras de redes neuronales que presentan múltiples capas ocultas y presentan diversos niveles de abstracción de los datos de entrada que se les otorga, esto permite que automáticamente este tipo de redes pueda extraer características relevantes. Estas redes también denominadas DNN (*Deep Neural Networks*) son capaces de representar con gran exactitud modelos de funciones altamente variables y no lineales (Larochelle, Bengio, Louradour, & Lamblin, 2009).

Redes neuronales convolucionales

También denominadas CNN (*Convolutional Neural Networks*) son un tipo de redes neuronales de aprendizaje profundo las cuales están especializadas en el procesamiento de datos en forma matricial, las cuales pueden ser aplicadas en la representación del muestreo de una señal de audio o también en los píxeles que componen imágenes o una combinación de ambas para componer videos. Las CNN están compuestas de múltiples capas convolucionales que pueden ser unidimensionales o multidimensionales, como se observa en la Figura 6, las cuales tienen funciones de activación en principio lineales y luego no lineales. Las capas convolucionales como su nombre indica, realizan la operación matemática convolución (Goodfellow, Bengio, & Courville, 2016).

Figura 6

Estructura tradicional de una CNN



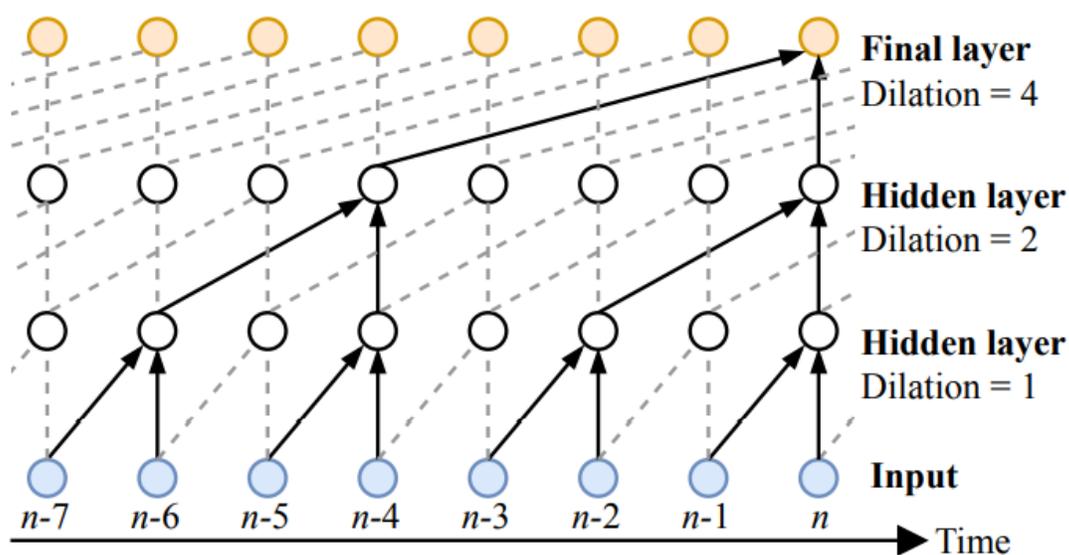
Nota. Tomado de *Diego Calvo*, por Calvo, D, 2017.

Modelo WaveNet

El modelo *WaveNet* es un tipo de CNN en el que básicamente se utilizan una pila de capas convolucionales causales 1D con un factor de dilatación como se observa en la Figura 7. Esto se hace con el fin de incrementar el campo receptivo sin incrementar demasiado el costo computacional, para ello se omiten valores de entrada con un cierto paso lo que provoca que se necesiten menos capas aumentando la eficiencia computacional (Oord & Sander, 2016).

Figura 7

Capas de convolución dilatadas con diferentes factores de dilatación.



Nota. Tomado de *WaveNet: A Generative Model for Raw Audio*, por Oord, A. v., & Sander, D, 2016, Deepmind.

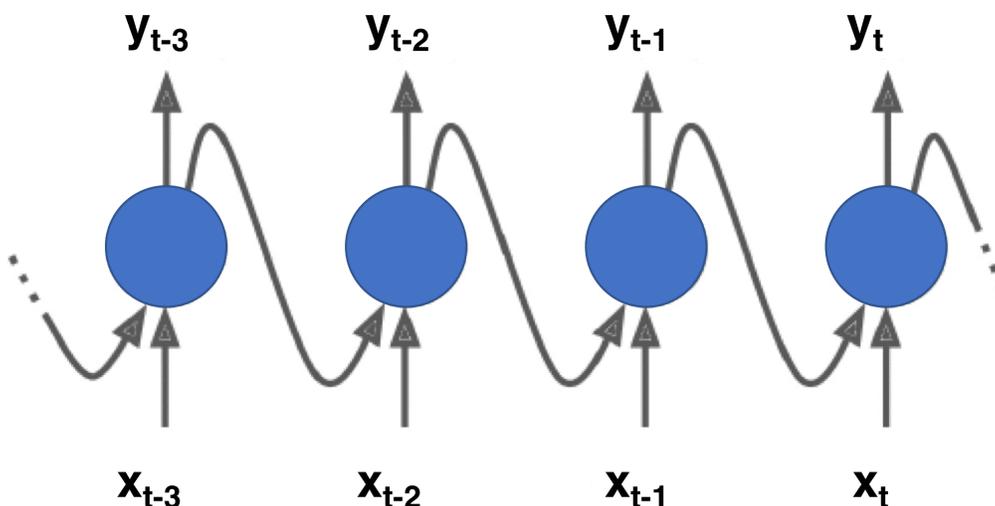
Redes neuronales recurrentes

También denominadas RNN (*Recurrent Neural Networks*) son un tipo de redes neuronales de aprendizaje profundo las cuales operan con datos secuenciales. Es decir, estas tienen información sobre valores temporales previos lo que les permite tratar con la dimensión del tiempo, por lo tanto, trabaja con neuronas recurrentes las cuales en cada instante de tiempo

reciben información de la entrada actual así mismo reciben información sobre su misma salida en un estado anterior como se ve en la Figura 8 (López, 2019).

Figura 8

Estructura de una RNN



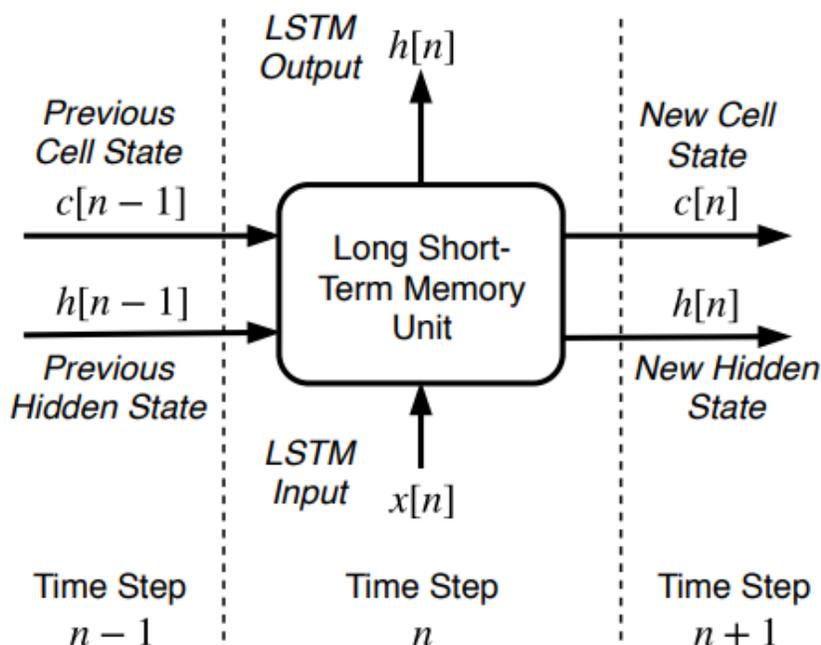
Nota. Tomado de *Jordi Torres AI*, por Torres, J, 2019.

Redes LSTM

Las RNNs de este tipo utilizan como unidad recurrente una unidad LSTM como en la Figura 9, la cual tiene un estado el cual consiste en dos vectores, el *cell state*, c , y el *hidden state*, h . Por lo tanto, en cada paso de tiempo la unidad LSTM tiene como entradas la entrada actual de las muestras en el tiempo presente, el estado inicial de la *cell state* y el estado inicial de la *hidden state* y además esta unidad produce como salidas la *hidden state* actualizada y la *cell state* actualizada en las cuales se utiliza la función *sigmoide* para la *cell state* y la función *tanh* para la *hidden state* (Wright, Damskögg, & Välimäki, 2019).

Figura 9

Diagrama de una unidad LSTM



Nota. Tomado de *Real-time black-box modelling with recurrent neural networks*, por Wright, Damaskagg, & Välimäki, 2019, International Conference on Digital Audio Effects.

Frameworks de redes neuronales

Existen muchos entornos para desarrollar, entrenar y poner a prueba redes neuronales de diferentes tipos, estos poseen herramientas las cuales permiten implementar con código reducido todo lo necesario para el desarrollo de una red neuronal. Entre los más utilizados se encuentran *Tehano* la cual es una biblioteca de Python de código abierto, *Torch* un *framework* de *Machine Learning* y *Deep Learning* basado en Lua, *Pytorch* una biblioteca de código abierto para Python, *Keras*, *MxNet*, *TensorFlow*, entre otras (Pontificia Universidad Católica del Ecuador, 2022).

Optimizadores

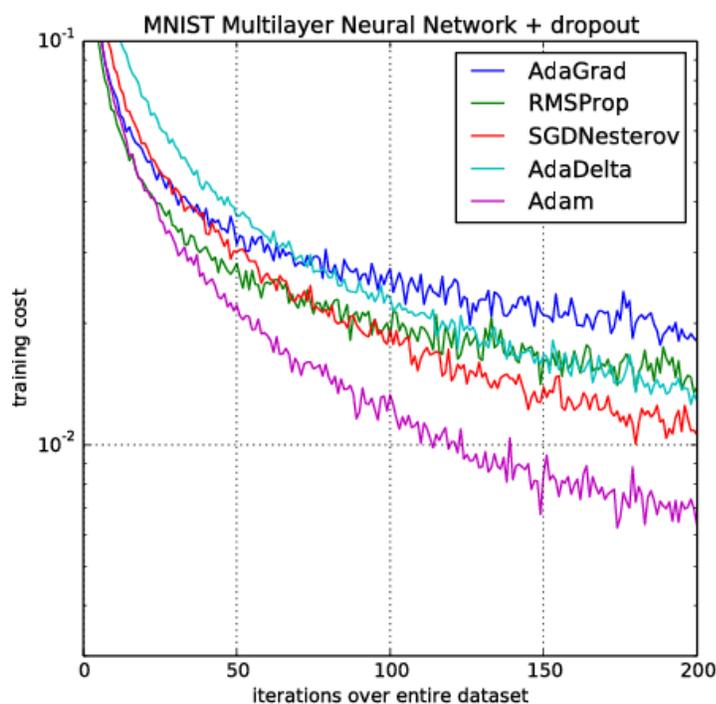
Las redes neuronales pretenden obtener las relaciones de pesos o ponderaciones adecuadas para que se ajuste mejor a su propósito, estas relaciones se establecen en el entrenamiento de la red mediante el algoritmo de *backpropagation*. Para ello, un elemento denominado optimizador se encarga de generar los pesos que se vayan ajustando de mejor manera. Esto lo hace calculando el gradiente de la función de coste que determina el error entre el valor estimado y el valor real por cada peso y se multiplica al vector del gradiente por un parámetro denominado *learning rate* con lo que se pretende tener una convergencia más rápida (Velasco, 2020).

Optimizador Adam

Es uno de los algoritmos de optimización más utilizados en las redes neuronales, toma su nombre debido a la estimación del momento adaptativo. Este presenta varios beneficios de otros algoritmos como el algoritmo de gradiente adaptativo denominado *AdaGrad* y el algoritmo de propagación cuadrática media *RMSProp*. Además, es muy eficiente en términos de facilidad de implementación, tiene un costo computacional bajo y además permite manejar problemas con una gran cantidad de datos de manera rápida, adicionalmente se ha demostrado que este algoritmo se compara favorablemente con otros métodos de optimización como muestra la Figura 10 (Velasco, 2020).

Figura 10

Comparación del costo de entrenamiento de Adam con otros algoritmos para la optimización.



Nota. Imagen tomada de *Adam: A Method For Stochastic Optimization*, por Kingma, D., & Lei Ba, J, 2015, ICLR 2015.

Funciones de pérdida

Las funciones de pérdida son utilizadas en el campo del aprendizaje automático para medir que tan cerca se encuentra la red de alcanzar el resultado que se considera deseado. Existen una variedad entre las más utilizadas se encuentra el Error cuadrático medio utilizado en problemas de regresión, la entropía cruzada categórica que suele ser utilizada en problemas de clasificación, la entropía cruzada binaria utilizada para clasificaciones binarias o la entropía cruzada categórica la cual trabaja con números enteros (Gavilán, 2020).

ESR

Denominado *Error-to-Signal Ratio* es una función de pérdida la cual es utilizada para medir señales de audio, es básicamente una medida similar a la del error cuadrático dividido para la energía de la señal objetivo, es una función de pérdida utilizada para la regresión de audio (Jakub, 2021).

Filtro de pre-énfasis

Por lo general la mayor cantidad de energía en el audio se concentra en la parte baja del espectro de frecuencias. Los efectos producidos por la aplicación de equipos para guitarra eléctrica como en este caso los pedales de distorsión o amplificadores tienen efectos importantes en frecuencias altas también las cuales deben ser tomadas en cuenta en el entrenamiento de las redes, por lo tanto, se puede utilizar un filtro de paso alto para amplificar estas frecuencias y no se vean enmascaradas por la cantidad de energía en la parte baja del espectro. Este filtro se aplica únicamente para el cálculo de la ESR en la función de pérdida más la señal no es procesada a través de este filtro (Jakub, 2021).

Capítulo III

Implementación de los sistemas

Introducción

En este capítulo se detallan los pasos realizados para obtener los modelos tanto convolucional y recurrente para la emulación del amplificador *Mesa Boogie Dual Rectifier* y el pedal de distorsión *Ibanez Tube Screamer*. Desde la correcta instalación del entorno de desarrollo, pasando por las características, la topología y arquitectura de cada red, la generación del set de datos, la etapa de entramiento y los modelos obtenidos.

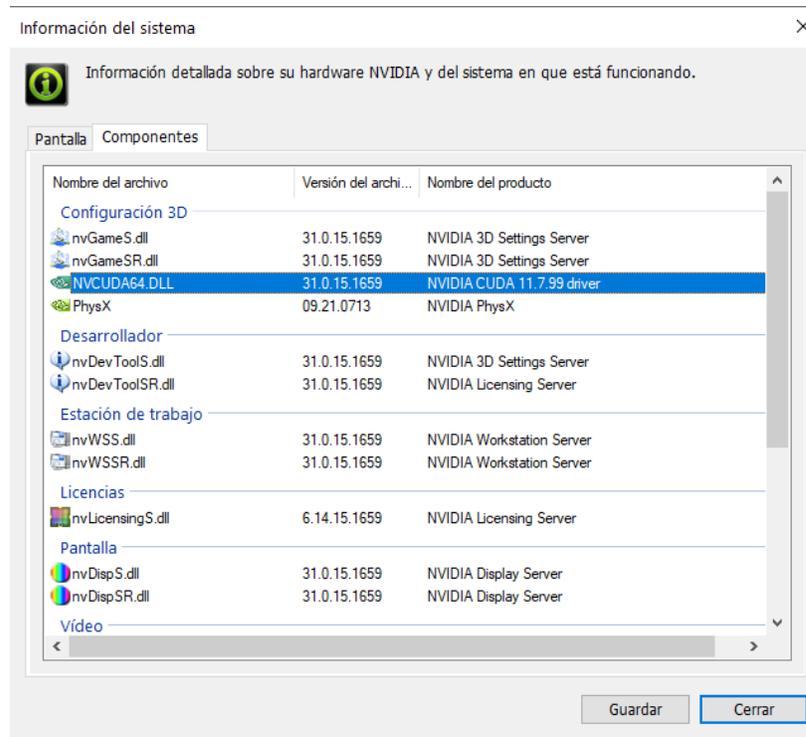
Entorno de desarrollo

El desarrollo, implementación, entrenamiento y pruebas de las redes neuronales se realizó con herramientas y librerías de *Python*, específicamente el *framework* para desarrollo de redes neuronales denominado *Pytorch* el cual permite el uso de los *cuda cores* de la tarjeta gráfica para el proceso de entrenamiento de las redes. Esto permite que el entrenamiento se realice de una manera más rápida que realizarlo utilizando la capacidad de procesamiento de la CPU con la ventaja de que es un *framework* de uso libre y código abierto y permite la implementación y uso de una herramienta conocida como *Pytorch Lightning* la cual permite simplificar el código para la implementación y el entrenamiento de las redes. Adicionalmente se hizo uso de otras herramientas para el procesamiento de señales para el análisis como la librería *Numpy*, y librerías para graficar y representar datos como es el caso de *PyPlot*. Además del software Matlab, para lo cual previo a la implementación, se realizó la instalación de todas las herramientas necesarias para lo cual se siguieron los siguientes pasos:

1. Se verificó la disponibilidad y compatibilidad cuda de la GPU desde el panel de control de la tarjeta gráfica como se observa en la Figura 11.

Figura 11

Compatibilidad cuda de la tarjeta gráfica.



2. Instalar Anaconda: Se accedió a la página web <https://www.anaconda.com/products/distribution>, se descargó e instaló la última versión disponible del instalador gráfico para Windows x64 la cual incluye el entorno de *Python*, además de varias herramientas para el desarrollo de redes neuronales y el análisis de datos y señales.
3. Instalar *Pytorch*: se accedió al sitio <https://pytorch.org> y se seleccionó la última versión disponible compatible para la versión del driver cuda que en este caso como se pudo observar en la Figura 11 es la versión 11.7.99 además se accedió también al sitio <https://www.pytorchlightning.ai> y se instaló la última versión del mismo.
4. Se verificó la correcta instalación de los entornos de *pytorch*, *pytorch lightning*, *numpy*, *scipy* y *matplotlib*.

Obtención del set de datos

El primer paso para entrenar una red es tener información que servirá a modo de ejemplo para que esta pueda aprender. En este caso en particular se requiere de una señal de guitarra limpia sin ningún tipo de procesamiento, esto se puede lograr grabando la señal de la guitarra por medio de un cable conectada directamente a una entrada de línea como muestra la Figura 12, para lo cual se utilizó una guitarra eléctrica (*Ibanez RG premium*) con la pastilla *Humbucker* (*DiMarzio Ibz*) en la posición del puente, conectada a una interfaz de audio (*Focusrite Scarlett 2i2*) y se grabó 2 minutos y 39 segundos de audio con pequeños fragmentos de varias canciones, los parámetros de grabación esta descritos en la Tabla 1.

Figura 12

Esquema de conexión de guitarra directo a línea



Tabla 1

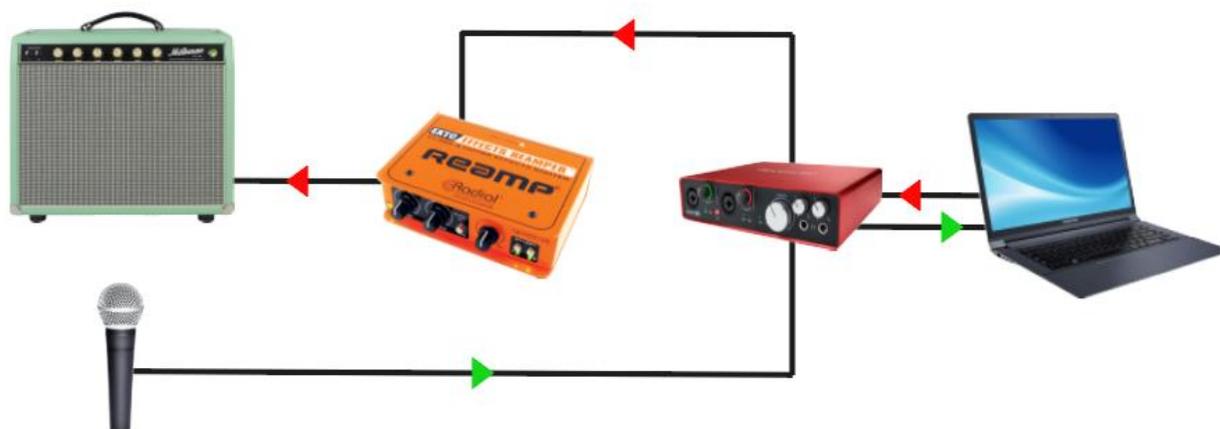
Parámetros de grabación

Frecuencia de muestreo	44100 Hz
Tasa de Bits	32 (float)
Formato de audio	WAV
Duración	2:39 min

Adicionalmente, se requiere de los datos que van a servir de ejemplo del sonido del amplificador y del pedal, para lo cual se utilizó una técnica denominada *reamping* descrita en la Figura 13, la cual nos permite extraer el audio limpio que ya fue grabado, acondicionar la señal y pasar esa señal por el amplificador (*Mesa Boogie Dual Rectifier*) con un gabinete 4 x 12 *Recto Standard* a través de un micrófono *shure sm57* con los parámetros descritos en la Tabla 2, de esa manera se obtiene el set de datos de entrenamiento para la emulación del amplificador.

Figura 13

Esquema de reamping para capturar el sonido del amplificador



Nota. El audio se graba a través de un micrófono para poder captar el sonido producido por los altavoces mejor conocidos como gabinete del amplificador.

Tabla 2

Parámetros configurados en el amplificador Mesa Boogie Dual Rectifier

PARÁMETROS	NIVEL
CHANNEL	CH2
PRESENCE	50%
MASTER	50%

PARÁMETROS	NIVEL
GAIN	50%
BASS	50%
MID	50%
TREBLE	50%
MODE	Vintage

De la misma manera con la técnica de *reamping* descrita anteriormente se procesó la señal a través del pedal de distorsión (*Ibanez Tube Screamer*) como muestra la Figura 14, con los parámetros descritos en la Tabla 3, por lo tanto, se obtienen los datos de entrenamiento para la emulación del pedal de distorsión.

Figura 14

Esquema de reamping para capturar el sonido del pedal de distorsión



Tabla 3

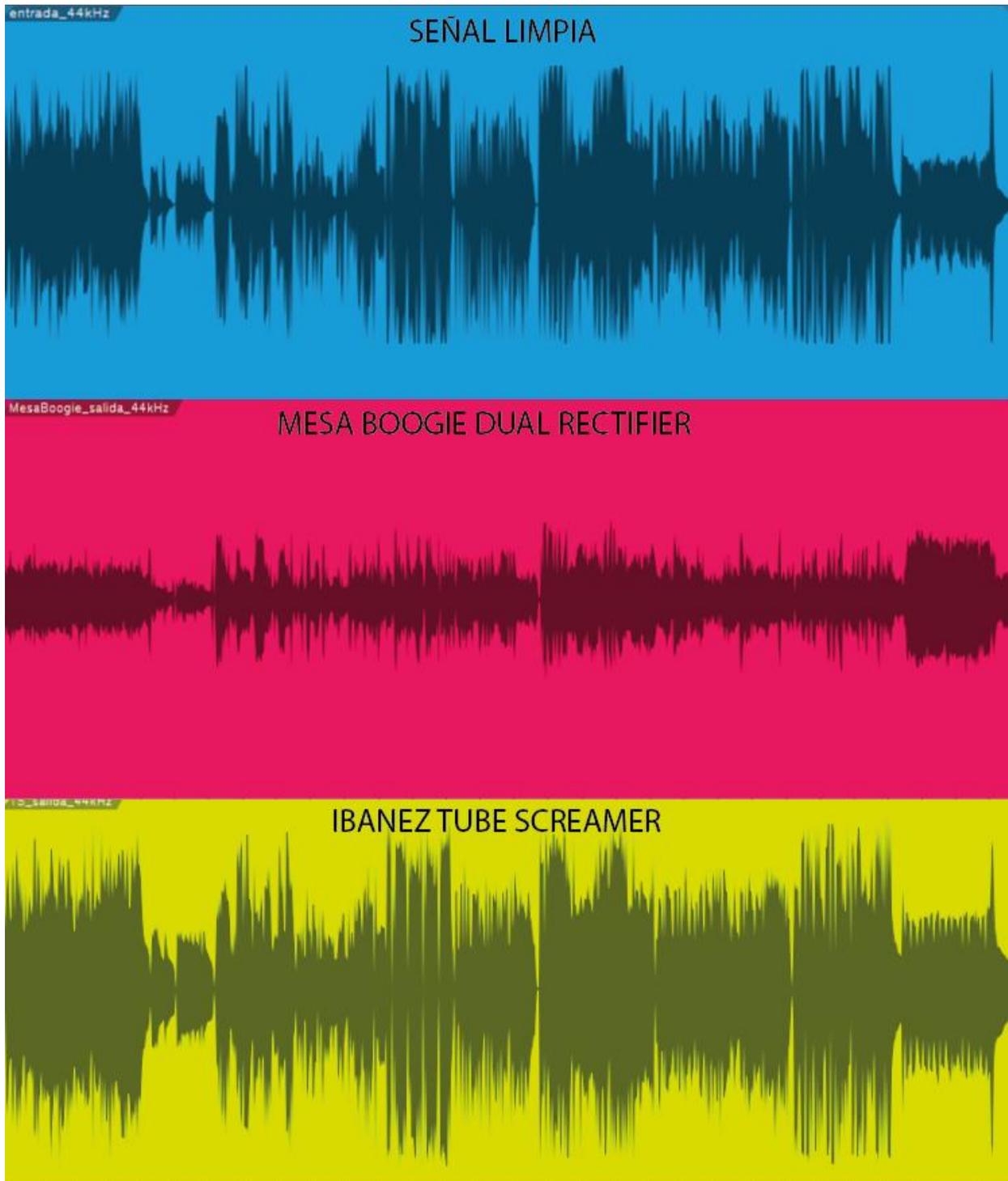
Parámetros configurados en el pedal de distorsión

PARÁMETROS	NIVEL
OVERDRIVE	0%
TONE	50%
LEVEL	100%

Análisis de los datos obtenidos de entrada

Figura 15

Señales de entrada en el dominio del tiempo



En la Figura 15 se puede apreciar la señal limpia la cual fue grabada directamente conectada a línea con los pasos descritos previamente. Es una señal de audio tomada directamente de la guitarra eléctrica y por la amplitud se puede tener una idea de que fue una ejecución dinámica ya que fue tocada con diferentes intensidades, lo cual es coherente debido a que se tratan de fragmentos de varias canciones, por lo tanto, se tienen varios tramos en los que se cambia de intensidad en la ejecución. Por otro lado, se pueden observar en las señales de salida del amplificador y del pedal de distorsión algunas diferencias con la señal original.

La primera es que se nota un gran cambio en la amplitud de la señal correspondiente a la del amplificador *Mesa Boogie Dual Rectifier*, en la cual se puede observar una señal bastante atenuada respecto a la original. Se aprecia como la diferencia entre los picos máximos y mínimos de la señal se ve disminuida, ya que el canal 2 en el que fue grabado el amplificador es un canal con una distorsión bastante marcada lo cual produce un efecto de compresión.

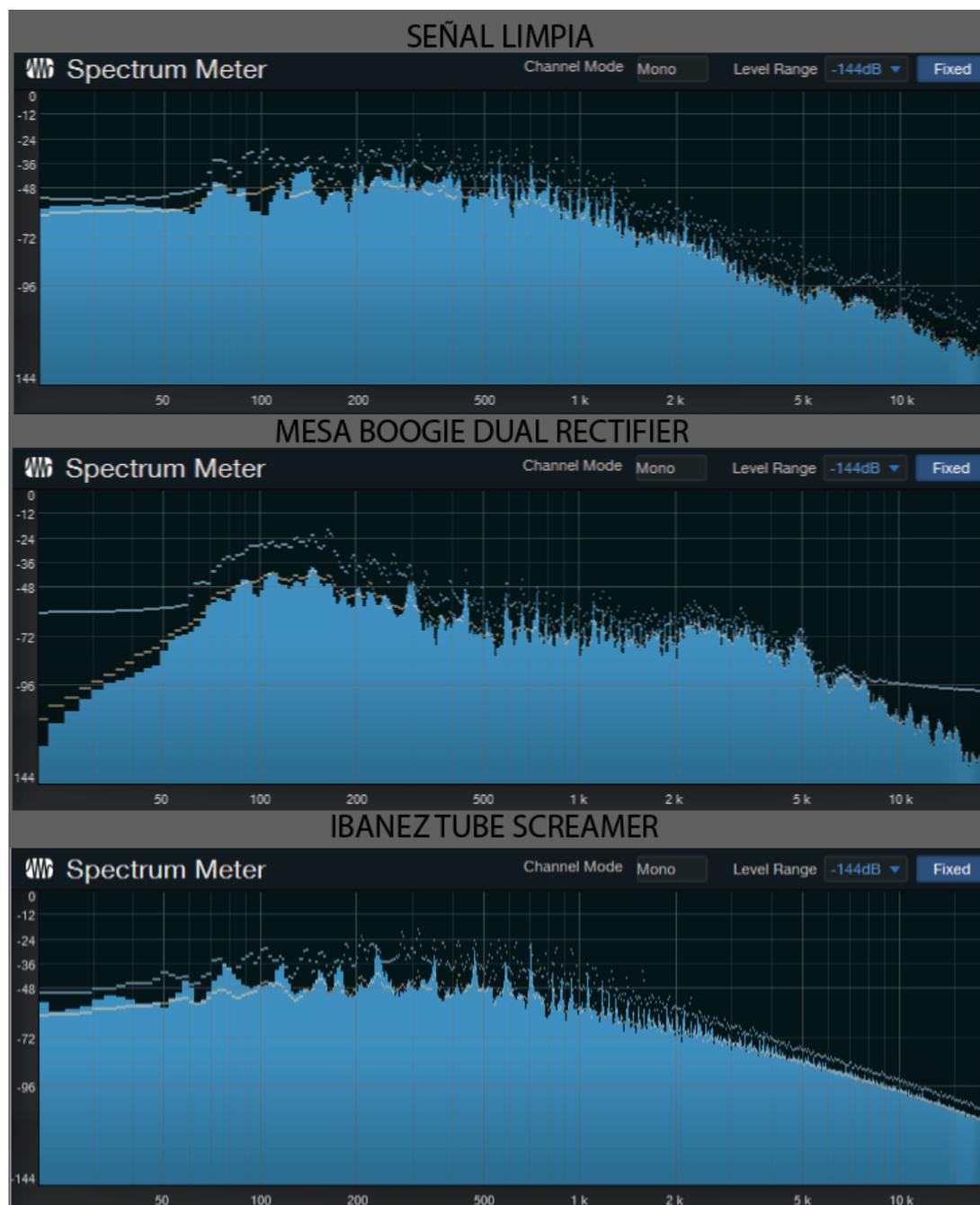
Respecto a la última señal presentada correspondiente a la señal del *Ibanez Tube Screamer* se puede ver una diferencia muy sutil respecto a la señal limpia. Esto se debe a que la distorsión producida por este pedal no es demasiado alta, sino que más bien actúa como un *booster* el cual amplifica ligeramente la señal, esto se puede apreciar al ver que en general la señal tiene una amplitud ligeramente mayor respecto de la original a excepción de los picos donde la señal se comprime ligeramente.

Además, se realizó un gráfico en el dominio de la frecuencia es decir un gráfico de la transformada rápida de Fourier FFT de las señales de entrada como se puede observar en la Figura 16, en el cual se puede apreciar claramente las diferencias entre la respuesta frecuencial que presentan estas señales la cual nos da una idea de que en las 3 señales la mayor cantidad de información se encuentra contenida en frecuencias menores a 5 kHz.

Se puede decir también que tanto el pedal de distorsión como el amplificador actúan como una especie de combinación de filtros ya que modifican de cierto modo las frecuencias originales de la señal de audio.

Figura 16

Señales de entrada en el dominio de la frecuencia



La señal del pedal de distorsión es muy similar a la original, pero con mayor amplitud mientras que el amplificador tiende a atenuar algunas frecuencias y amplificar otras lo cual puede dar un indicio de tener una respuesta no lineal.

Figura 17

Espectrograma de la señal limpia

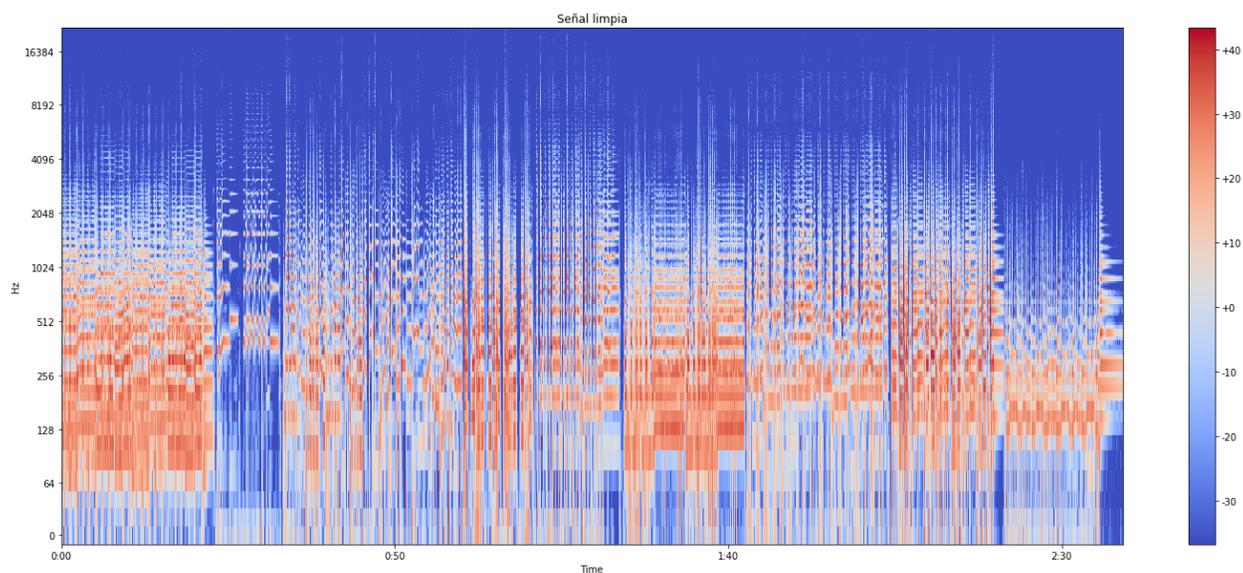


Figura 18

Espectrograma del amplificador Mesa Boogie Dual Rectifier

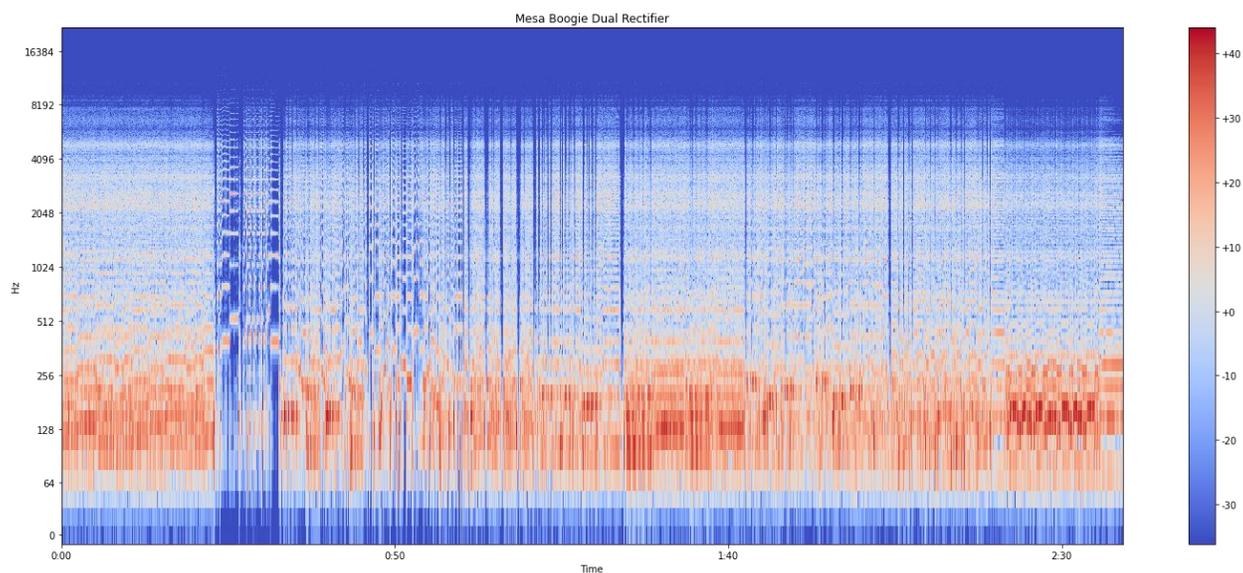
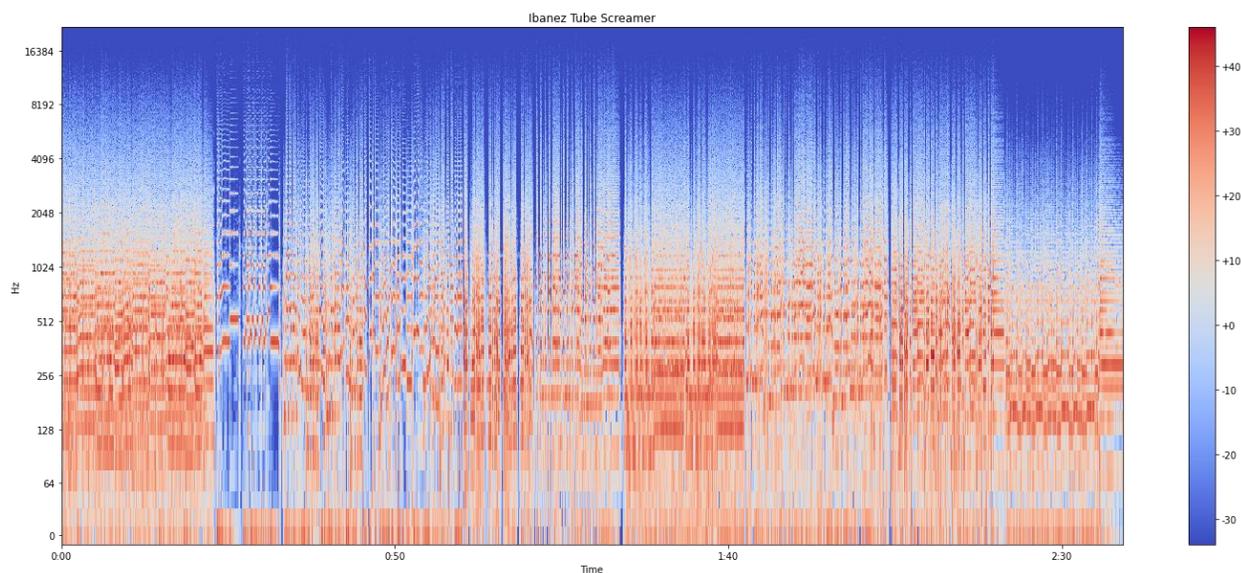


Figura 19

Espectrograma del Ibanez Tube Screamer



Como se puede observar en las Figuras 17, 18 y 19 tenemos los espectrogramas de las señales de entrada. Estos nos muestran información en el dominio tiempo-frecuencia de estas señales. Estos gráficos son muy representativos del comportamiento y del tipo de sonido que estos dispositivos presentan y se puede apreciar claramente las diferencias entre estas señales.

Parámetros de entrenamiento de las redes

Medir el impacto que tiene un equipo analógico o dispositivo sobre una señal puede ser clasificado como un problema de regresión en el cual se debe estimar la relación entre la señal limpia de la guitarra y las señales con la aplicación de los efectos en este caso específicamente el amplificador y el efecto distorsión. Para esto se debe trabajar con las señales de audio las cuales generalmente son muestreadas a 44.1 kHz a nivel de audio profesional orientado a música lo cual significa que por cada minuto de audio se tienen más de 2.5 millones de datos por lo tanto se tiene la ventaja de poder tener gran cantidad de información necesaria para el

entrenamiento con unos pocos minutos de audio, para este ejemplo se utiliza un audio con una duración de 2 minutos y 39 segundos.

Además, para el modelado de equipos analógicos con respuesta no lineal con redes neuronales es importante determinar su tiempo de respuesta a diferentes estímulos. Este se encuentra alrededor de los 80 ms para equipos de válvulas lo que significa que se van a necesitar al menos esa cantidad de tiempo para tener una estimación correcta por lo tanto se ha establecido la duración de las muestras en 100 ms (Eero-Pekka, Lauri, Etienne, & Vesa, 2019).

Debido a la gran cantidad de datos que se maneja en este problema en específico se ha seleccionado el optimizador de Adam ya que permite manejar estos datos a un costo computacional más bajo.

La función de pérdida a utilizar es el ESR que como se vio previamente es muy útil en señales de audio y sumado a esto la utilización de un filtro de pre-énfasis para las frecuencias agudas lo cual evita que las frecuencias graves enmascaren las frecuencias agudas en el cálculo del error.

Para escoger los parámetros de la red CNN se ha tomado en cuenta la velocidad de procesamiento la cual es un factor de cuantas veces puede ejecutarse el modelo en tiempo real, es decir un valor igual o mayor a 1 quiere decir que se ejecuta en tiempo real, menor a 1 no se puede ejecutar en tiempo real. Además, se toma en cuenta también el ESR el cual es una medida del error. Los resultados se presentan en la Tabla 4.

Tabla 4*Resultados obtenidos en modelo convolucional*

10 capas		
Canales	2	16
Velocidad RT	6.1	3
ESR 1000 epochs	1.97	0.058
ESR 1700 epochs	1.12	0.0381
24 capas		
Canales	2	16
Velocidad RT	1.87	0.92
ESR 1000 epochs	1.95	0.057
ESR 1700 epochs	1.07	0.0372

Como se puede observar en la Tabla 4, la mayor velocidad de procesamiento se obtiene al utilizar un menor número de capas, así como de canales, es decir 10 Capas y 2 Canales, sin embargo, el menor ESR se presenta en el extremo contrario ya que requiere de 24 capas y 16 canales, además se puede ver que el ESR disminuye ligeramente al aumentar el número de epochs. Por lo tanto, es importante obtener un modelo que tenga una buena velocidad de procesamiento, así como un ESR lo más bajo posible, por ello, se ha escogido 10 capas para garantizar la mayor velocidad de procesamiento además de que el ESR no se reduce significativamente al aumentar el número de capas, 12 canales convolucionales y 1500 epochs, lo cual permite un equilibrio entre velocidad de procesamiento y ESR.

De manera similar, para la red RNN se seleccionan los valores basados en los resultados de la Tabla 5.

Tabla 5*Resultados obtenidos en modelo recurrente*

LSTM Hidden size	16	80
Velocidad RT	15	2.5
ESR 1000 epochs	2.3200	0.1341
ESR 1700 epochs	2.2300	0.1032

Como se puede observar en la Tabla 5, la mayor velocidad de procesamiento se obtiene con el LSTM Hidden size más bajo es decir 16. Sin embargo, el menor ESR se obtiene al aumentar el valor de LSTM Hidden size y con mayor número de epochs, por lo tanto, al igual que en el caso anterior se buscan parámetros que den un buen equilibrio entre velocidad de procesamiento y ESR, en este caso se va a utilizar un LSTM hidden size de 64 y al igual que en el caso anterior 1500 epochs.

Preprocesamiento del set de datos

Los datos deben ser ajustados para el ingreso a las redes neuronales. Como todos los datos fueron obtenidos en base a la misma señal de entrada grabada con los parámetros descritos previamente, cada audio de ejemplo contiene un total de 7017159 datos. Al ser una señal de audio la cual va a ser procesada en el dominio del tiempo se requiere trabajar con muestras de datos de duraciones más cortas con el fin de reducir la aleatoriedad de estas muestras, por lo tanto, se ha utilizado una duración de 100 ms para cada muestra.

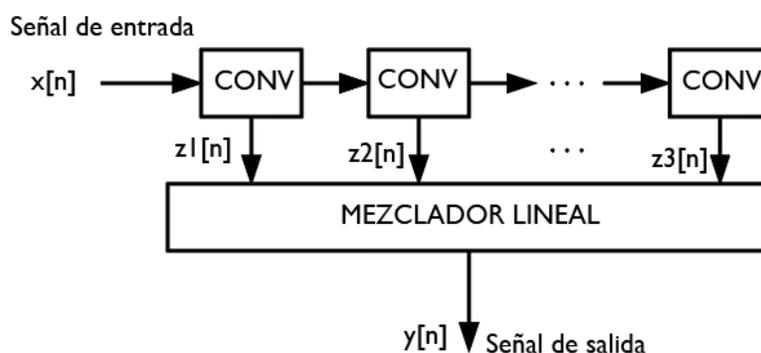
Adicionalmente se requieren datos de validación y datos de prueba los cuales se extraen del mismo set de datos, para lo cual de las 1591 muestras resultantes se utilizan, 954 muestras para entrenamiento, 318 para validación y 319 para prueba, es decir 60% - 20% - 20% respectivamente. además, para reducir la carga computacional durante el entrenamiento se han normalizado los datos de las muestras.

Implementación de la red neuronal convolucional

Arquitectura de la CNN

Figura 20

Arquitectura de la red CNN



La red neuronal CNN de la Figura 20 es un modelo WaveNet adaptado el cual está conformada por una serie de capas convolucionales. La señal de audio sin procesar es la entrada de la primera capa convolucional, estas capas se combinan por medio de un mezclador lineal para generar la salida. Cada capa convolucional procesa la señal de entrada con un filtro FIR causal dilatado al cual se le aplica una función de activación *gated activation function*.

Parámetros de la red CNN

El resumen de los parámetros de la red y de entrenamiento de la misma están descritos en la Tabla 6.

Tabla 6

Parámetros de la red CNN

Tamaño de la muestra	100 ms
Número de capas	10
Factores de dilatación	$dk=\{1,2,4,16,32,64,128,256,512\}$
Convolutional channels	12
Función de activación	Gated Activation (Tanh,

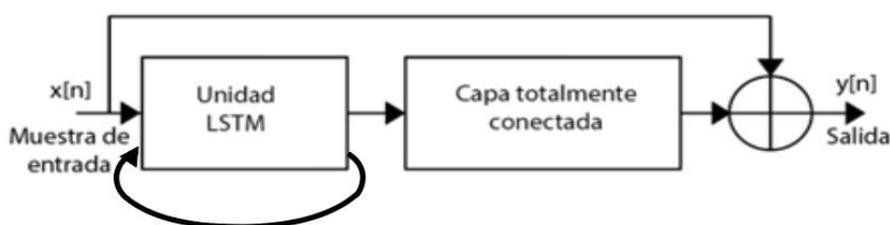
	Sigmoide)
Optimizador	Adam
Tasa de aprendizaje	3e-3
Max epochs	1500
Batch size	64
Función de pérdida	ESR con filtro de pre-énfasis

Implementación de la red neuronal recurrente

Arquitectura de la RNN

Figura 21

Arquitectura de la red RNN



La red neuronal de la Figura 21 consta de una sola unidad LSTM seguida de una capa *Fully Connected*. En cada paso de tiempo ingresa una muestra de la señal de audio, procesada por la unidad LSTM y luego esta se conecta a una capa totalmente conectada en la cual el número de entradas de la capa totalmente conectada es igual al número de estados ocultos en la unidad LSTM, lo cual representa una sola muestra de salida para cada paso de tiempo. además, se tiene una conexión residual lo cual ayuda a disminuir el problema de desvanecimiento de gradiente de unidades LSTM por tanto la salida es la suma de la muestra de entrada con la salida de la neurona de la capa totalmente conectada.

Parámetros de la red RNN

El resumen de parámetros de entrenamiento y de la red están descritos en la Tabla 7.

Tabla 7*Parámetros de la red RNN*

Tamaño de la muestra	100 ms
LSTM hidden size	64
Fully connected inputs	64
Función de activación	Tanh y sigmoide utilizadas en la unidad LSTM
Optimizador	Adam
Tasa de aprendizaje	3e-3
Max epochs	1500
Batch size	64
Función de pérdida	ESR con filtro de pre-énfasis

Entrenamiento de las redes

Las redes se entrenaron con los parámetros descritos previamente en los apartados anteriores.

La red CNN fue entrenada individualmente para obtener los modelos convolucionales tanto del amplificador como del pedal de distorsión, de la misma manera con la red RNN para obtener los modelos recurrentes del amplificador y pedal de distorsión por lo tanto se obtienen cuatro modelos.

Cada carpeta contiene el archivo *ckpt* con el modelo entrenado, una transformación a *json* para el uso del modelo en tiempo real y un archivo con los datos de entrenamiento que se utilizaron en la red, los cuales son producto del pre-procesamiento de las señales de entrada para ser admitidos dentro de las redes neuronales.

En el siguiente capítulo se tiene un análisis de los modelos obtenidos para su comparación y verificar su desempeño en la emulación del equipamiento analógico propuesto.

Capítulo IV

Resultados

Introducción

En este capítulo se presentan los resultados obtenidos con el entrenamiento de las redes detalladas en el capítulo anterior. Se hace la utilización de librerías de análisis de señales en Python como *Numpy* y *Matplotlib* para generar gráficos utilizados para comparar el desempeño de los modelos obtenidos, así como también el cálculo del ESR de cada modelo para poder valorar objetivamente el funcionamiento de los modelos para encontrar similitudes en las señales que sirven de referencia para medir el grado en que las emulaciones se asemejan al equipo real. Por último, también se hace un análisis subjetivo mediante la utilización de una encuesta realizada en *Google forms*.

Análisis de los modelos convolucionales obtenidos

Como se puede observar en la Figura 22 se ve como la señal de la emulación a comparación con la del amplificador real tienen diferencias sutiles en el gráfico de la forma de onda, pero se puede apreciar claramente como la señal se comprime lo cual es un efecto propio de la distorsión ocasionada por el amplificador además presenta un ESR de 0.0451 y un error absoluto con picos alrededor de 0.175 pero que por lo general se mantiene alrededor de 0.075, estos datos se corroboran con los gráficos presentados en la Figura 23 donde se realiza un acercamiento a las señales para verlas a mayor detalle.

En la Figura 24 se ve como la señal de emulación del pedal de distorsión presenta diferencias sutiles en los picos más altos de la forma de onda del pedal real pero cuando la señal tiene menor intensidad la diferencia es prácticamente imperceptible, presenta un ESR de

0.0057 y un error absoluto pico alrededor de 0.3 pero que por lo general se mantiene alrededor de 0.05.

Mesa Boogie Dual Rectifier CNN

Figura 22

Gráficos comparativos del modelo de emulación convolucional del amplificador vs el amplificador Mesa Boogie Dual Rectifier real

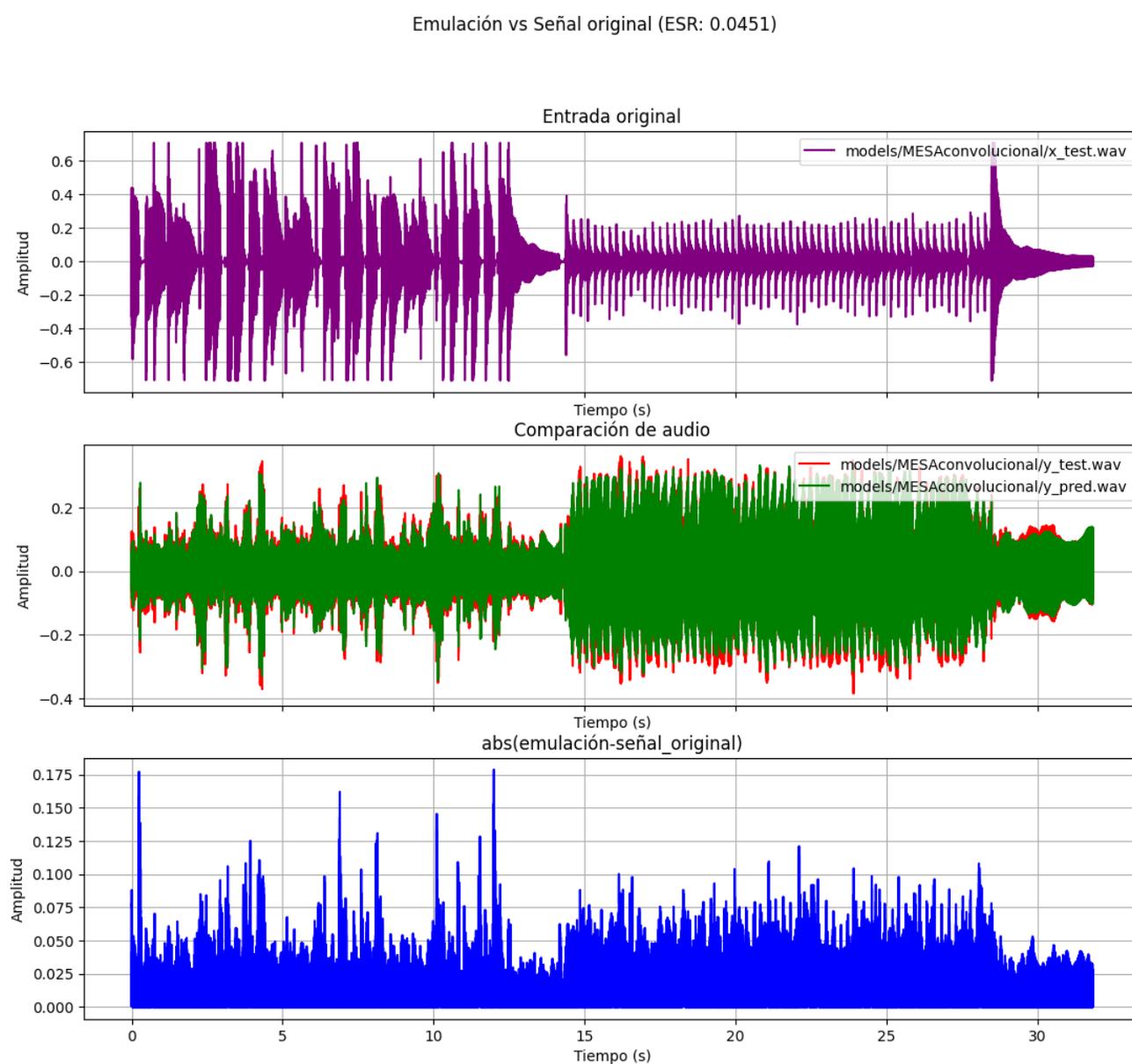
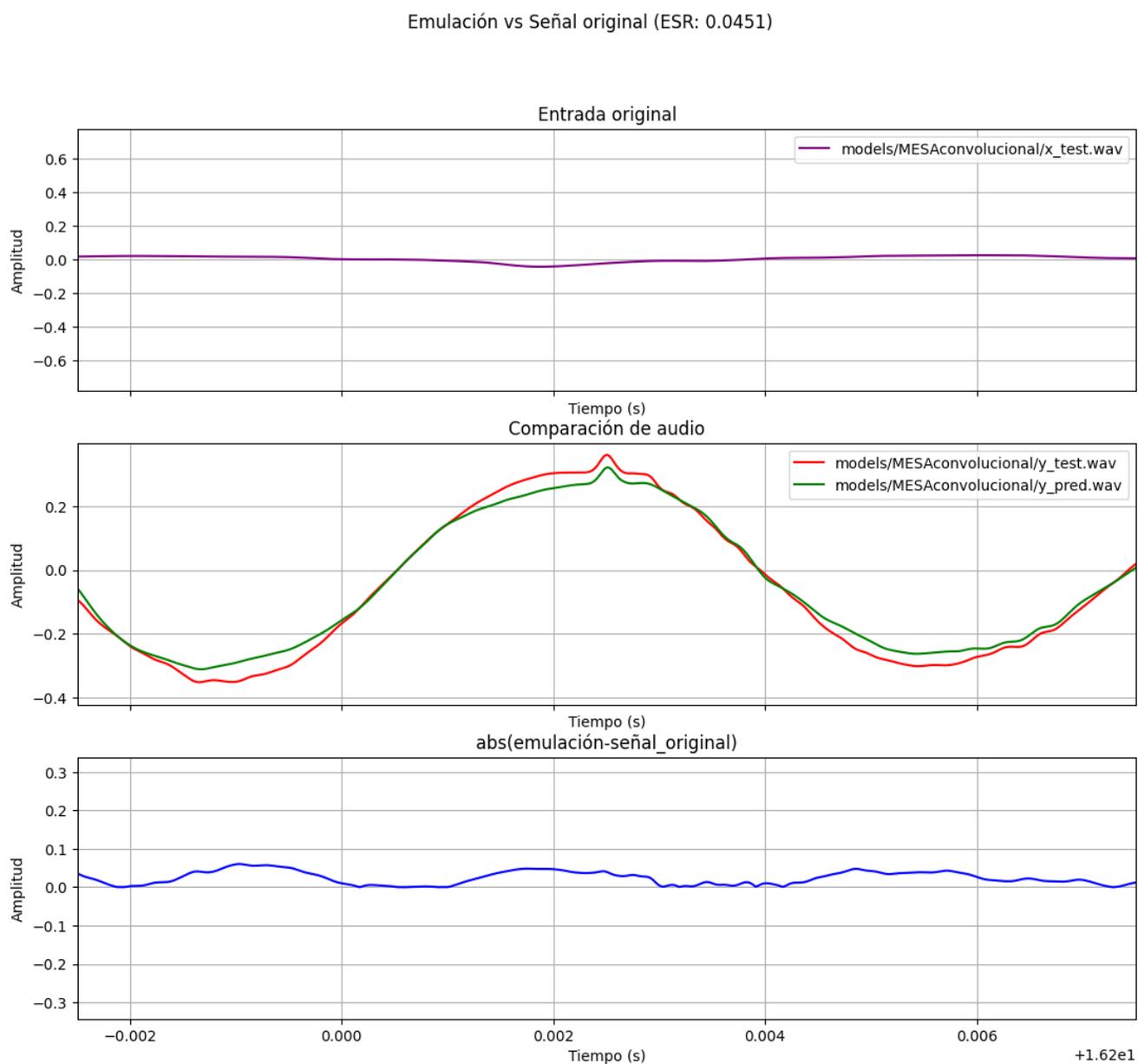


Figura 23

Gráficos comparativos a detalle del modelo de emulación convolucional del amplificador vs el amplificador Mesa Boogie Dual Rectifier real



Ibanez Tube Screamer CNN

Figura 24

Gráficos comparativos del modelo de emulación convolucional del pedal de distorsión vs el pedal Ibanez Tube Screamer

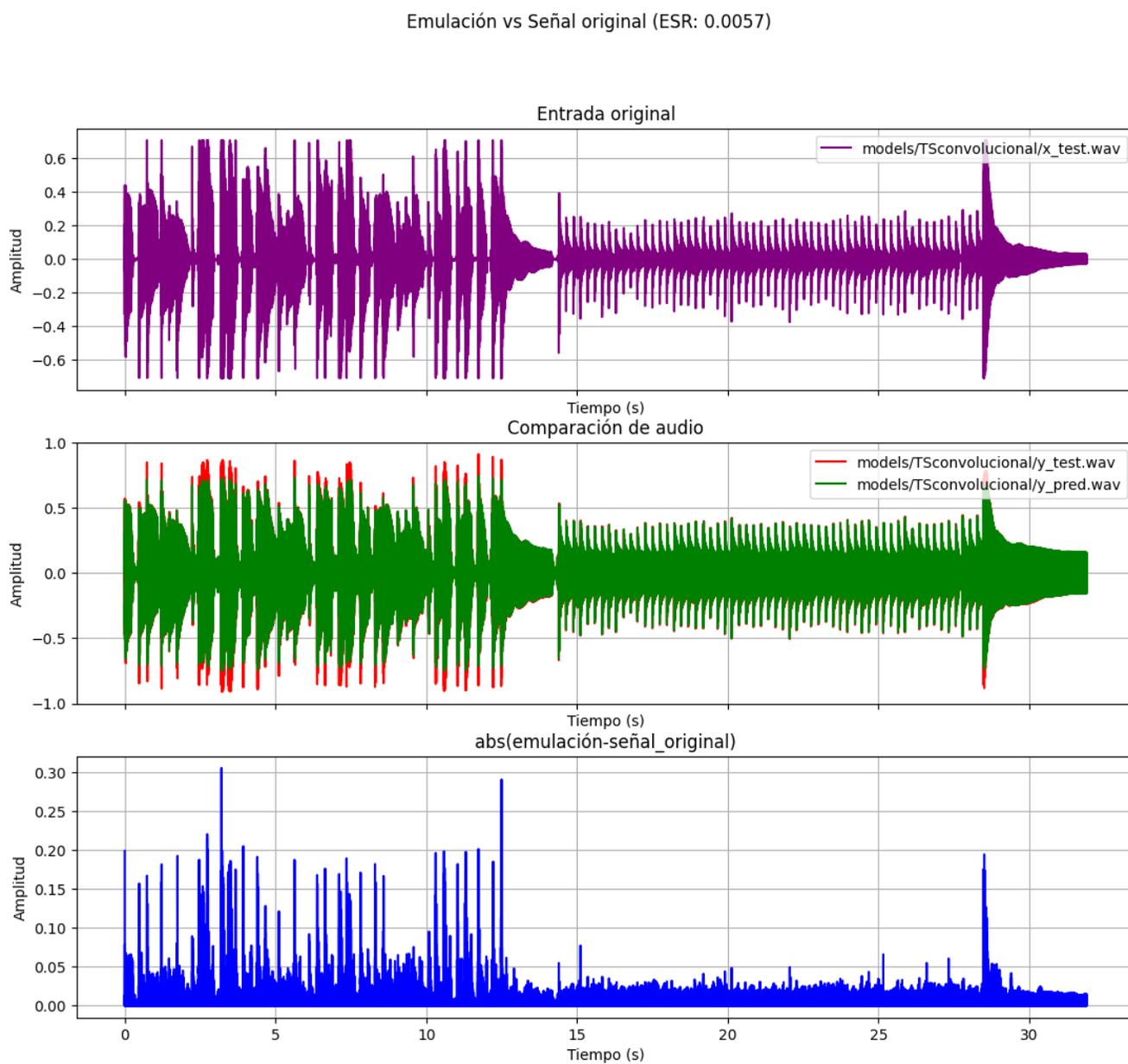
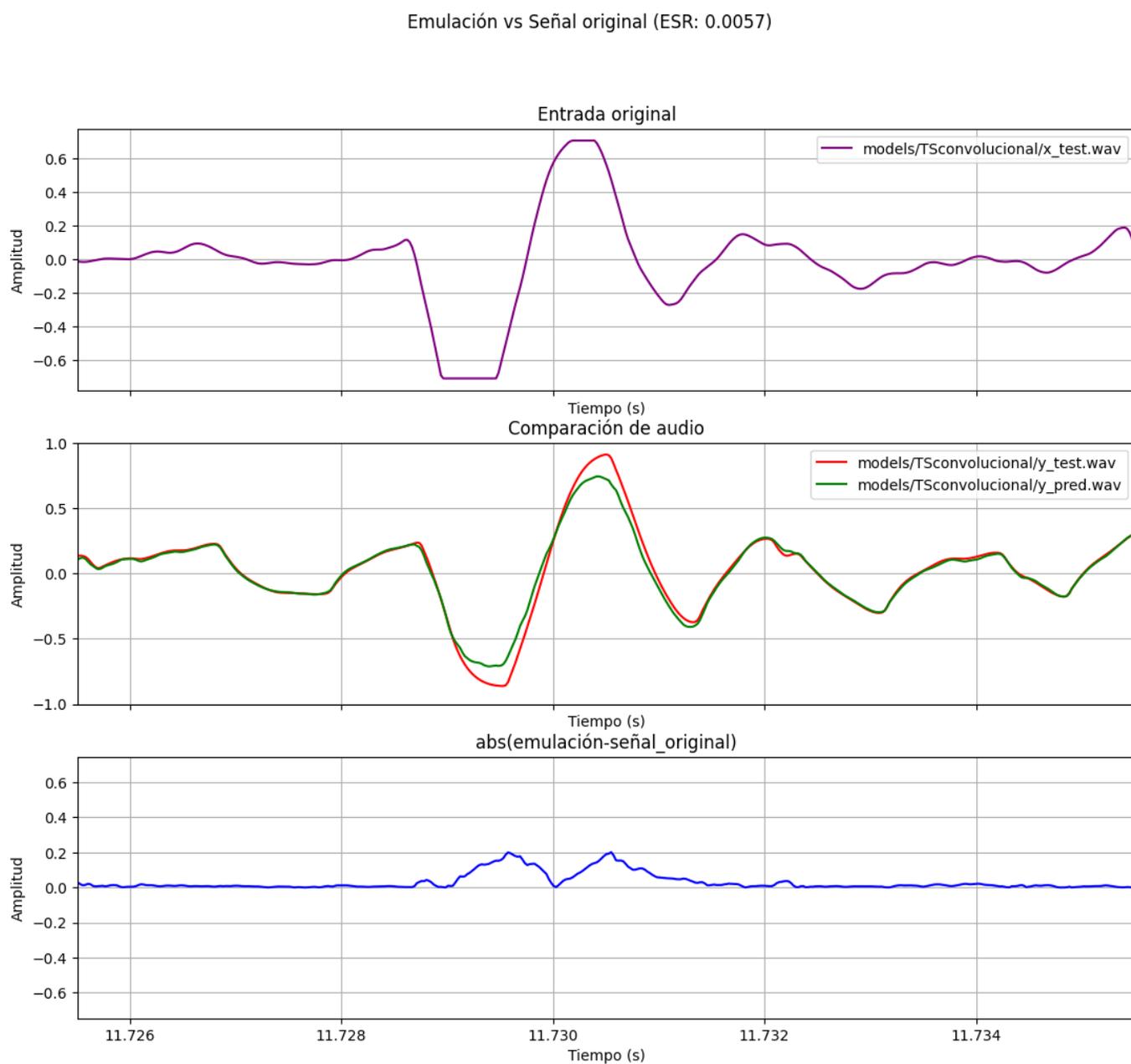


Figura 25

Gráficos comparativos a detalle del modelo de emulación convolucional del pedal de distorsión vs el pedal Ibanez Tube Screamer



Análisis de los modelos recurrentes obtenidos

Como se puede observar en la Figura 26 se ve como la señal de la emulación a comparación con la del amplificador real tienen ligeras diferencias en la amplitud de la forma de onda que parecen reducir al ingresar señales de intensidad más baja, de igual manera que en el amplificador emulado por CNN se puede apreciar la compresión ocasionada por la distorsión del amplificador lo cual muestra un comportamiento similar a la del amplificador real, además presenta un ESR de 0.1097 y un error absoluto con picos alrededor de 0.20 pero que por lo general se mantiene alrededor de 0.1, estos datos se corroboran con los gráficos presentados en la Figura 27 donde se hace un acercamiento a las señales para mirarlas a mayor detalle.

En la Figura 28 se ve como la señal de emulación del pedal de distorsión se ve muy similar a la de la emulación CNN del pedal de distorsión vista en la Figura 24 en las cuales presentan diferencias casi imperceptibles con los niveles de señal más bajos y con diferencias sutiles en las partes con mayor intensidad de señal, presenta un ESR de 0.022 y un error absoluto pico alrededor de 0.32 pero que por lo general se mantiene alrededor de 0.1.

De manera general con los resultados presentados en la diferencia de forma de onda de las señales de audio se aprecia que el error es bastante pequeño en el caso de los modelos convolucionales, así como en los recurrentes, sin embargo, el ESR y el error absoluto tanto para la emulación del amplificador como para la emulación del pedal es más pequeño en el caso de la red convolucional.

Mesa Boogie Dual Rectifier RNN

Figura 26

Gráficos comparativos del modelo de emulación recurrente del amplificador vs el amplificador

Mesa Boogie Dual Rectifier real

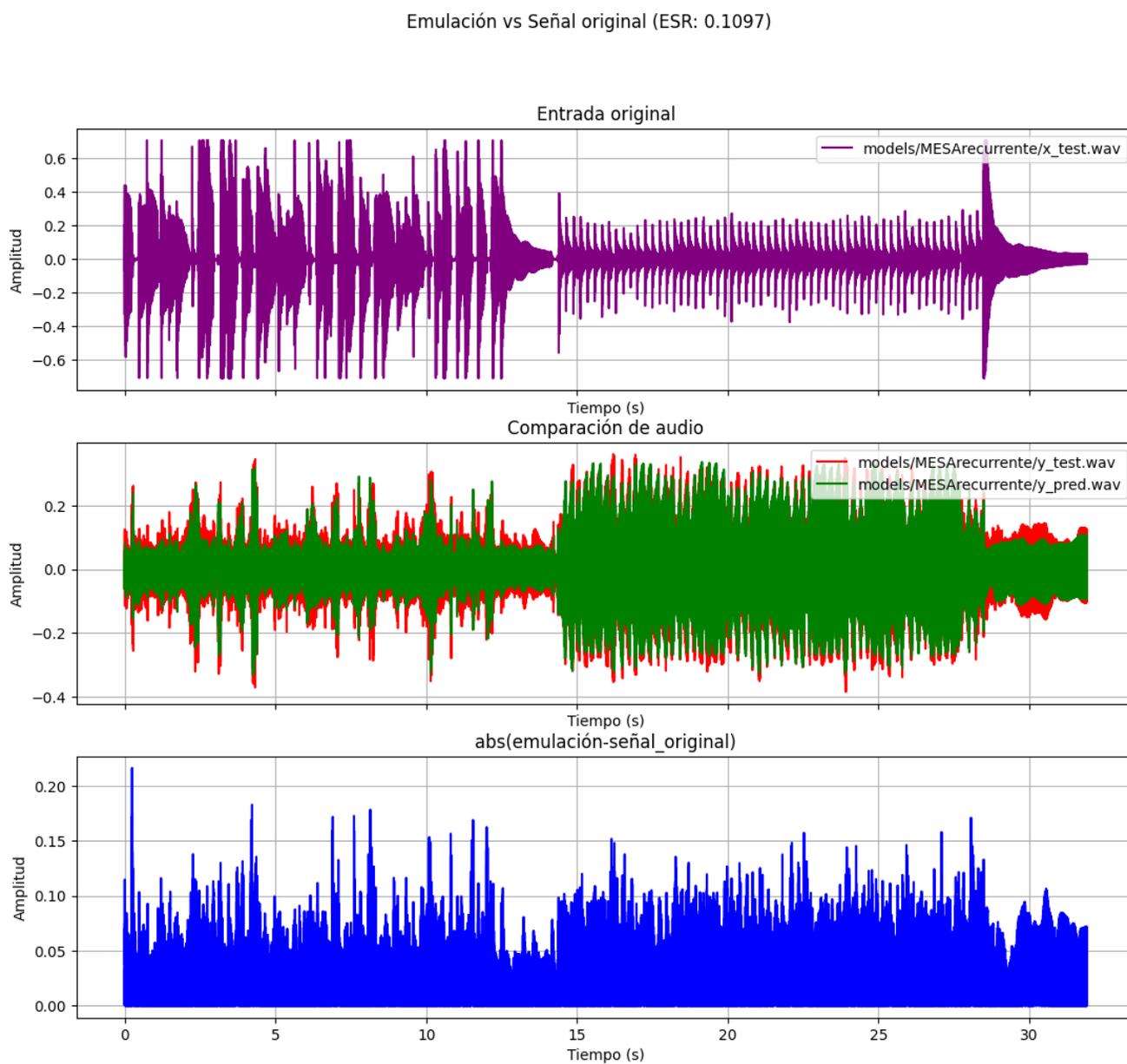
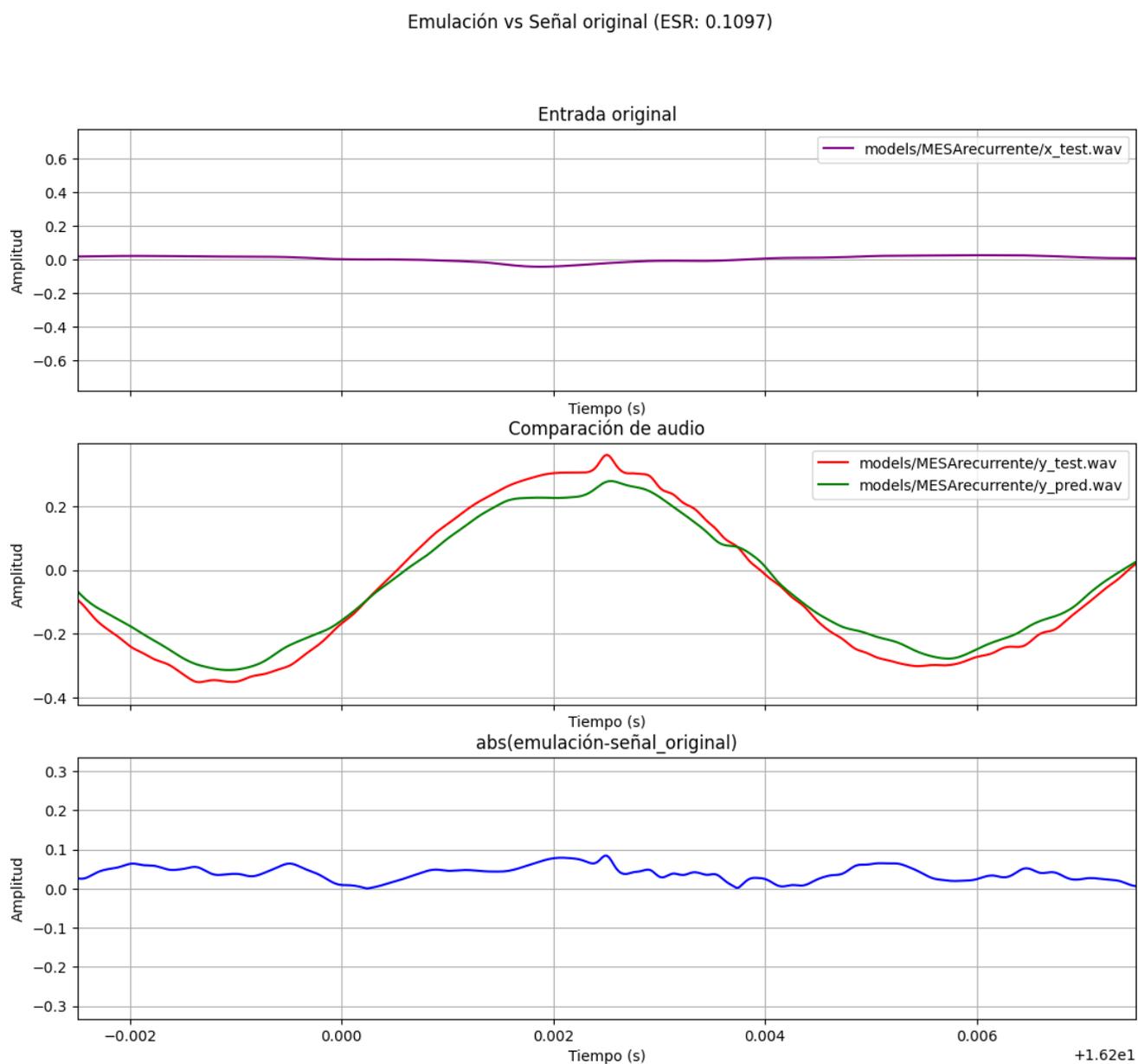


Figura 27

Gráficos comparativos a detalle del modelo de emulación recurrente del amplificador vs el amplificador Mesa Boogie Dual Rectifier real



Ibanez Tube Screamer RNN

Figura 28

Gráficos comparativos del modelo de emulación recurrente del pedal de distorsión vs el pedal Ibanez Tube Screamer

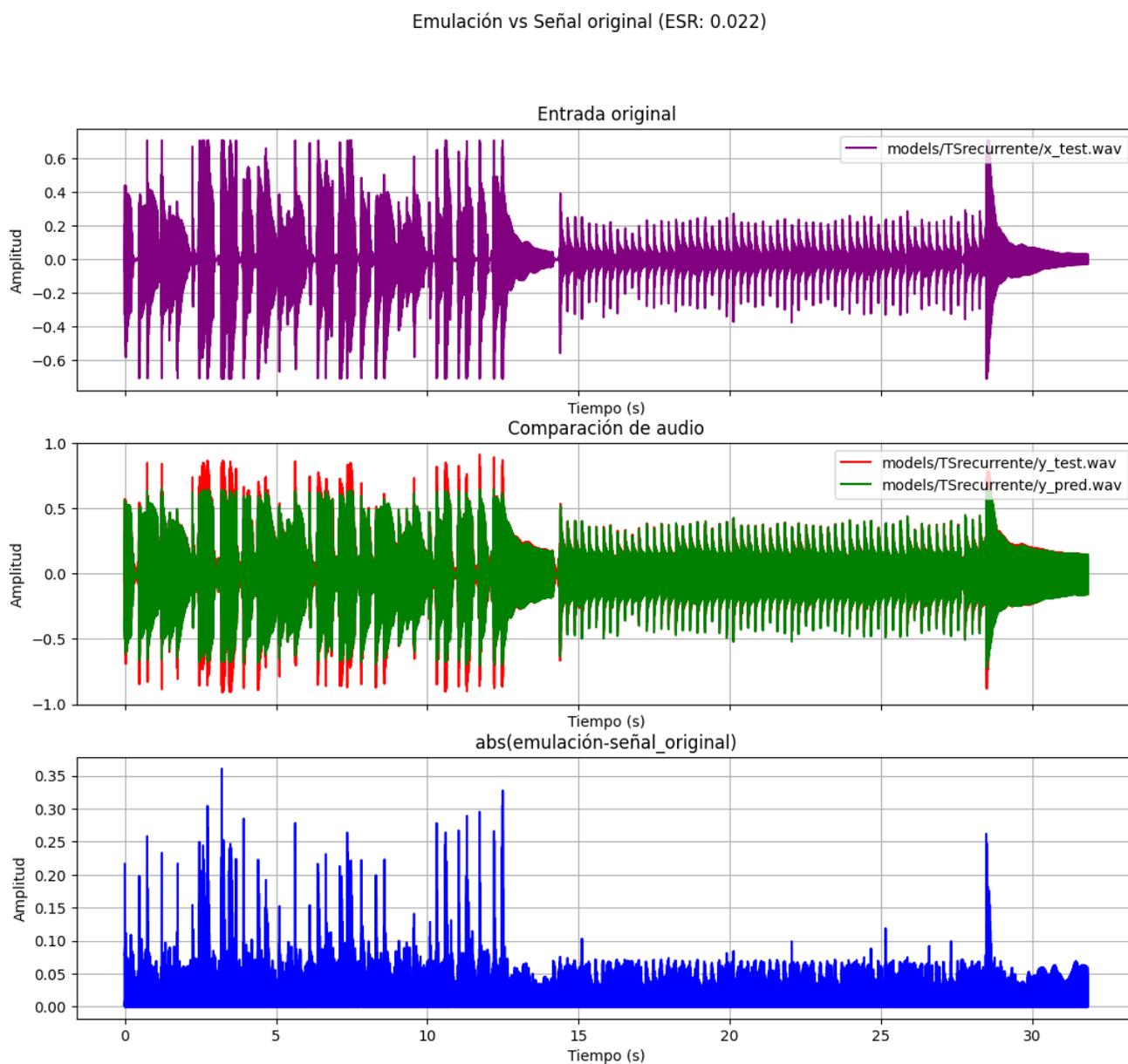
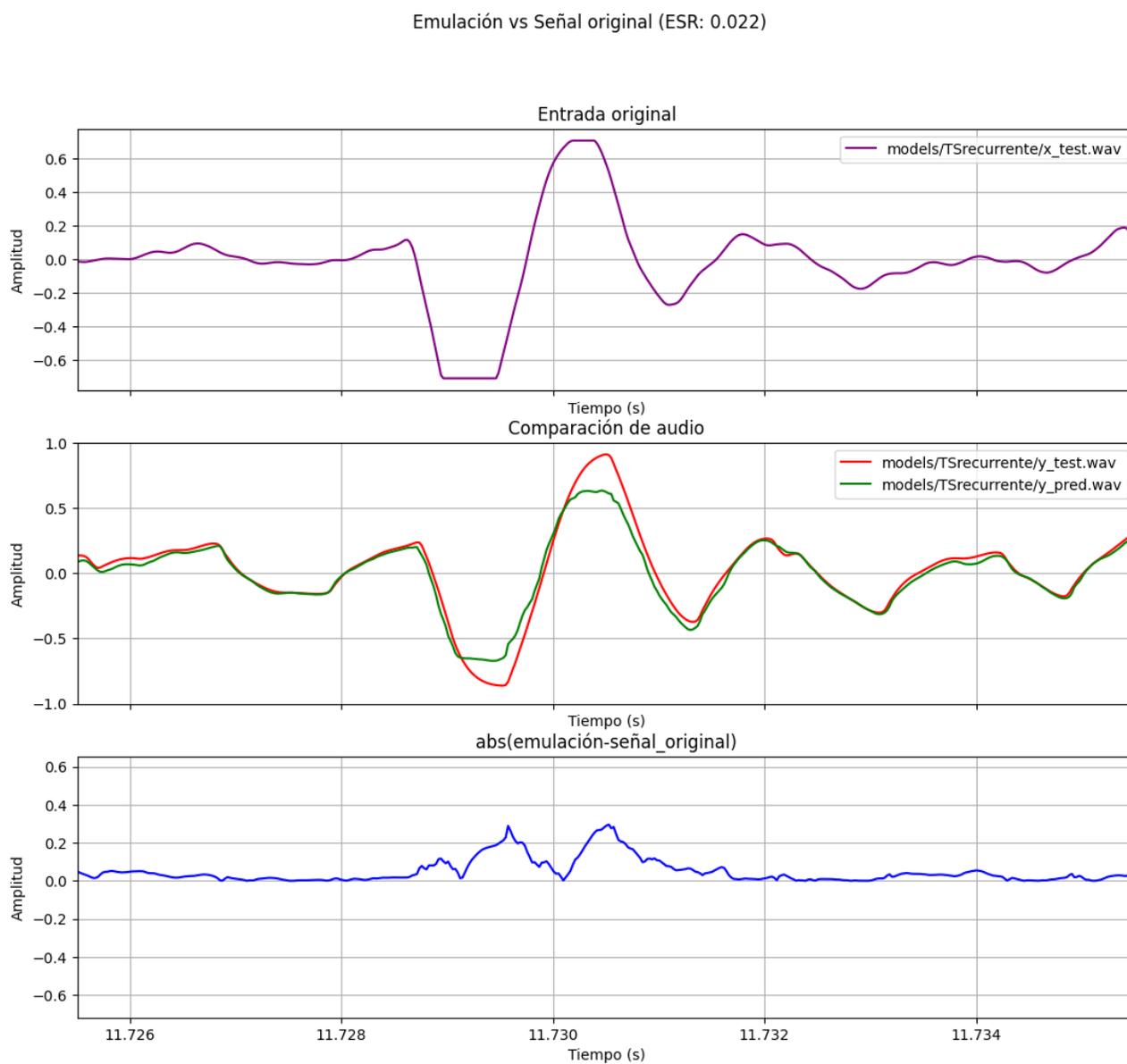


Figura 29

Gráficos comparativos a detalle del modelo de emulación recurrente del pedal de distorsión vs el pedal Ibanez Tube Screamer



Comparación entre los modelos CNN y RNN

Amplificador Mesa Boogie Dual Rectifier

Como se puede observar en la Figura 30 la forma de onda en azul representa la señal del amplificador real la cual no presenta una gran diferencia con los modelos emulados tanto con el modelo convolucional mostrado en rojo. Así como con el modelo recurrente mostrado en verde, en la última parte del grafico se puede ver la comparación entre ambos modelos en lo cual se puede destacar que la amplitud del modelo convolucional es ligeramente mayor a la del modelo recurrente lo cual hace que el modelo convolucional se asemeje en forma de onda ligeramente más al amplificador real que el modelo convolucional.

En la Figura 31 se puede ver la comparación de las mismas tres señales, pero en el dominio de la frecuencia en la que igual que en el caso de la forma de onda, no presentan diferencias realmente apreciables. El modelo convolucional introduce algunas frecuencias en la parte baja del espectro que no están muy presentes en la señal del amplificador original, sin embargo, en la parte de las frecuencias superiores a aproximadamente 1200 Hz hacia arriba se ajusta mejor que el modelo recurrente.

En las Figuras 32, 33 y 34 se presentan los espectrogramas del amplificador real, de la emulación convolucional y de la emulación recurrente respectivamente las cuales se puede ver que tienen un comportamiento muy cercano entre sí. Lo cual es muestra de que las emulaciones se acercan bastante a la versión real, las diferencias más apreciables se ven en las frecuencias bajas donde en las emulaciones en general se agrega algo de energía.

Figura 30

Comparación de la forma de onda del amplificador real con los modelos emulados

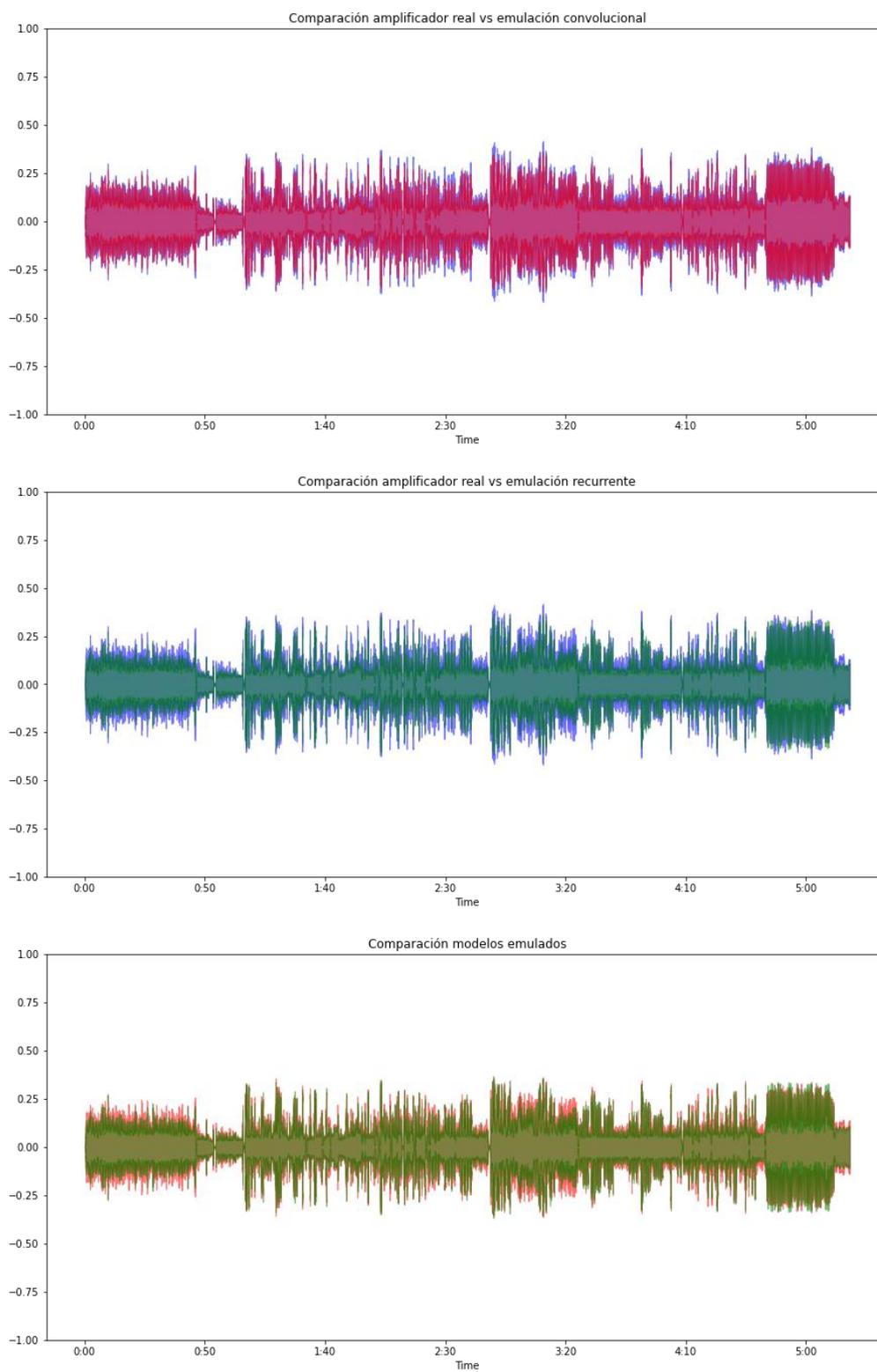


Figura 31

Comparación en el espectro de frecuencias del amplificador real con los modelos emulados

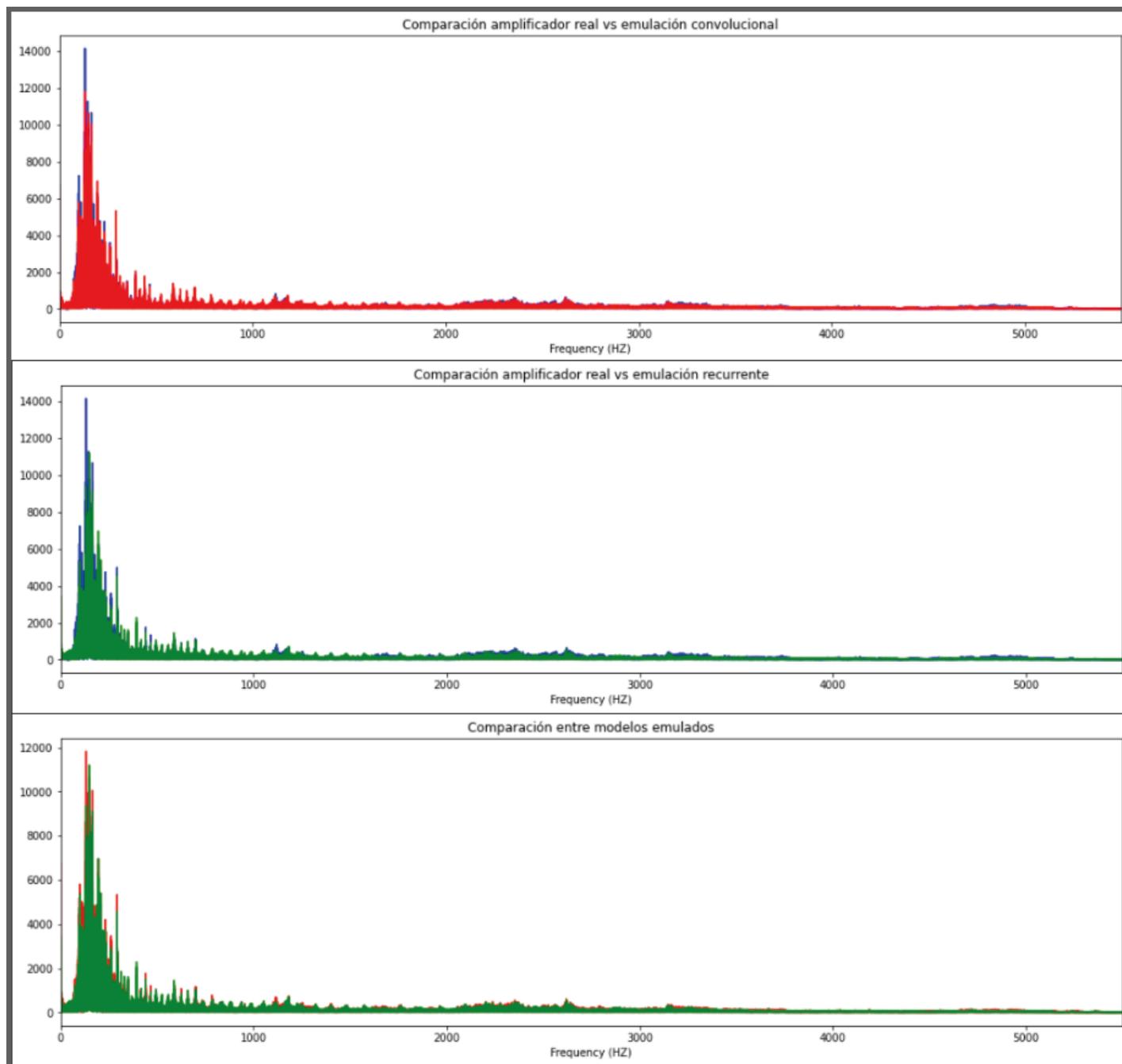
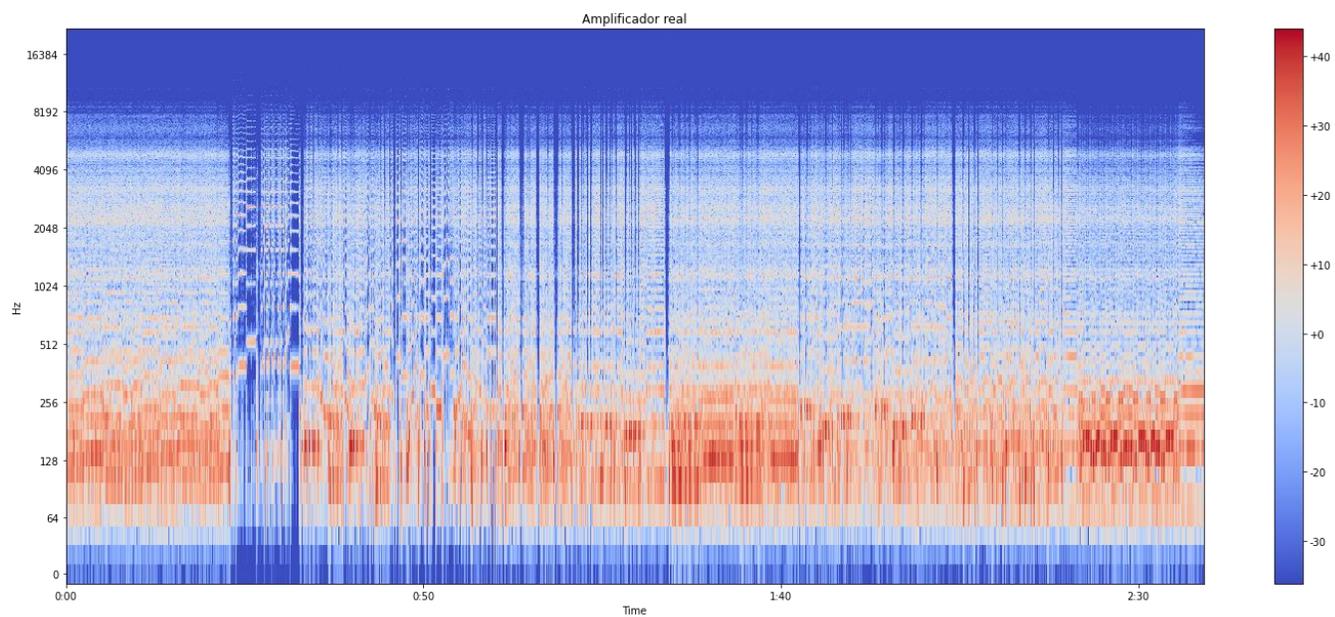


Figura 32

Espectrograma del amplificador real

**Figura 33**

Espectrograma del modelo convolucional del amplificador

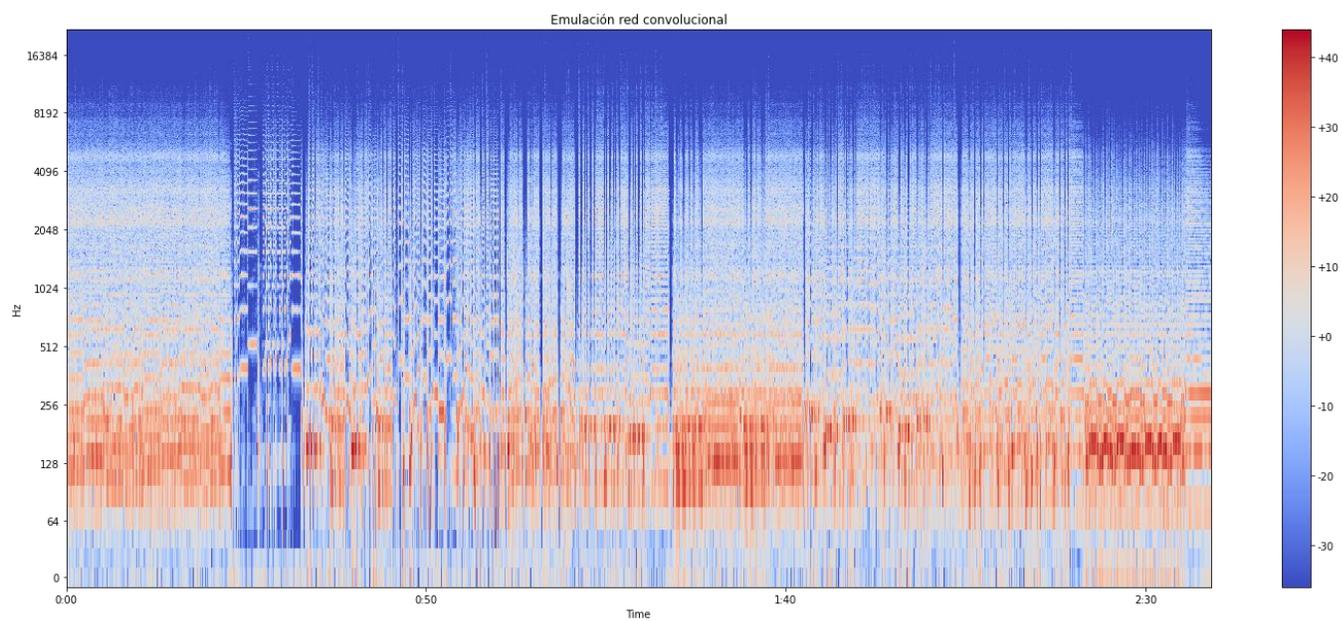
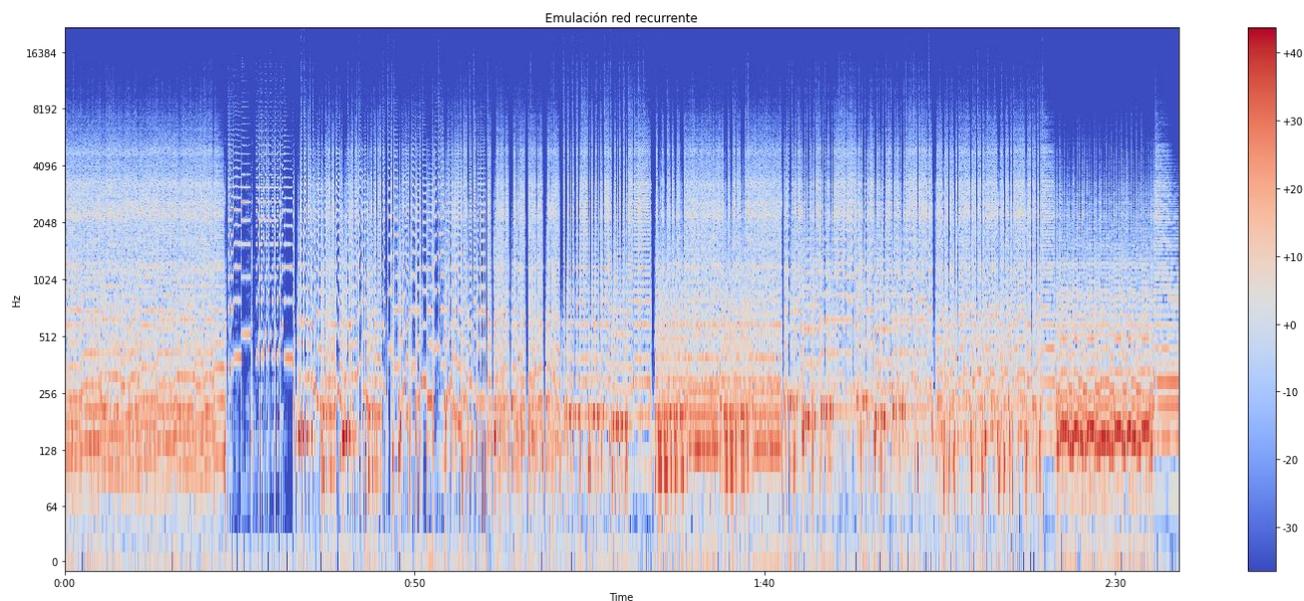


Figura 34

Espectrograma del modelo recurrente del amplificador



Pedal de distorsión Ibanez Tube Screamer

Como se puede observar en la Figura 35 la forma de onda en azul representa la señal del pedal real, esta presenta pequeñas diferencias en la amplitud con respecto a las señales de emulación convolucional representada en rojo y la señal de emulación recurrente mostrada en verde. Al igual que en el caso de las diferencias entre los modelos convolucional y recurrentes obtenidos, para la emulación del amplificador en el caso del pedal se puede apreciar que la amplitud de la emulación convolucional es ligeramente superior a la de la emulación recurrente lo que hace que la emulación convolucional se acerque más al pedal real en forma de onda.

En la Figura 36 se puede ver la comparación de las mismas tres señales, pero en el dominio de la frecuencia en la cual en la señal del pedal real se presenta un pico en la frecuencia más baja, la cual es emulada de la misma manera en su modelo convolucional, mas no en el modelo recurrente, sin embargo al ser una frecuencia muy baja no posee

características sonoras audibles, más bien se puede tratar de un armónico generado por la circuitería analógica del pedal de distorsión o incluso alguna especie de error en la grabación o en los convertidores de la interfaz de audio, en el resto de frecuencias ambos modelos se asemejan mucho al pedal real.

En las Figuras 37, 38 y 39 se presentan los espectrogramas del pedal real, de la emulación convolucional y de la emulación recurrente respectivamente las diferencias son casi imperceptibles por lo que los modelos se asemejan en gran medida al pedal real.

Figura 35

Comparación de la forma de onda del pedal real con los modelos emulados

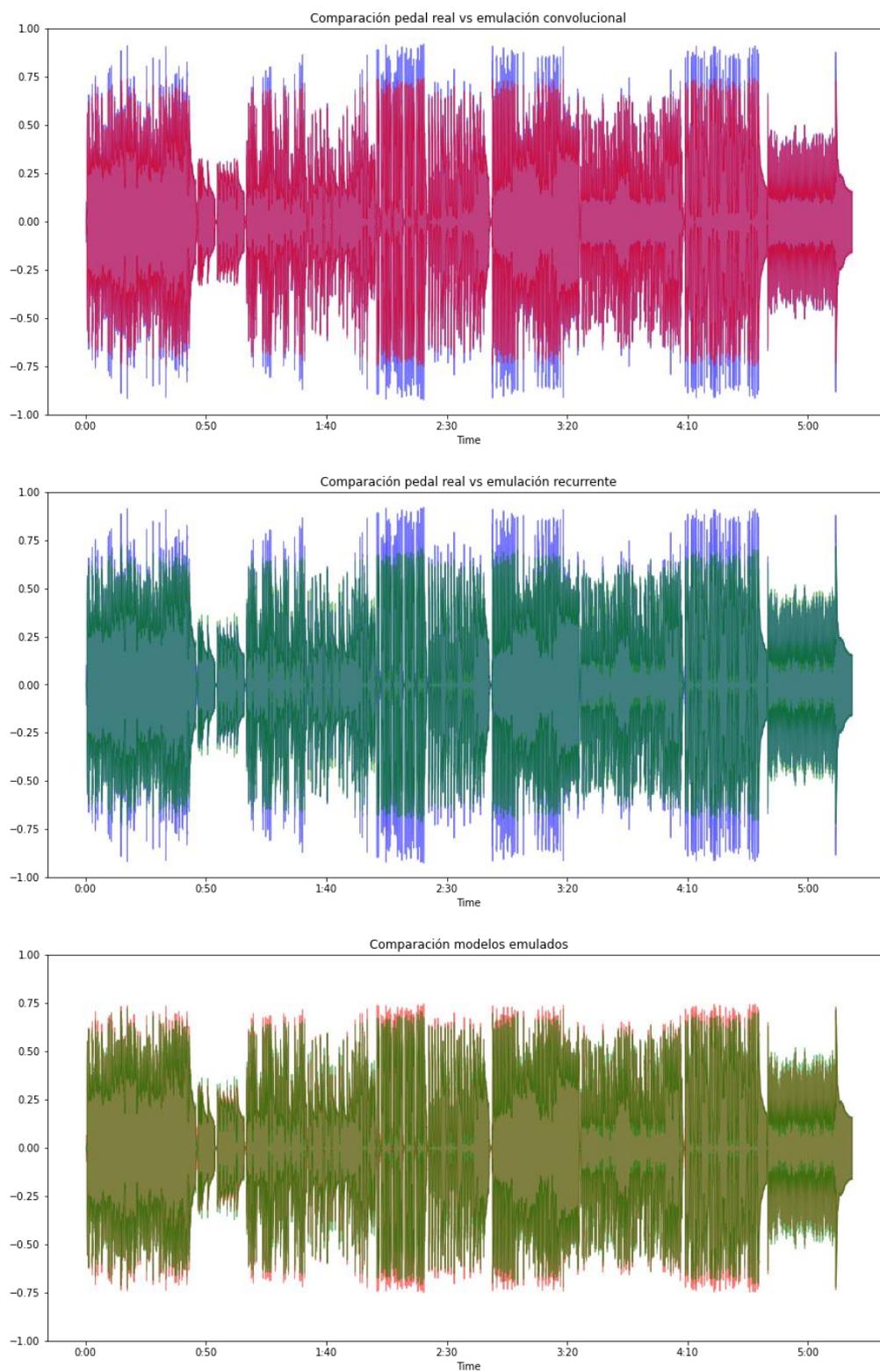


Figura 36

Comparación en el espectro de frecuencias del pedal real con los modelos emulados

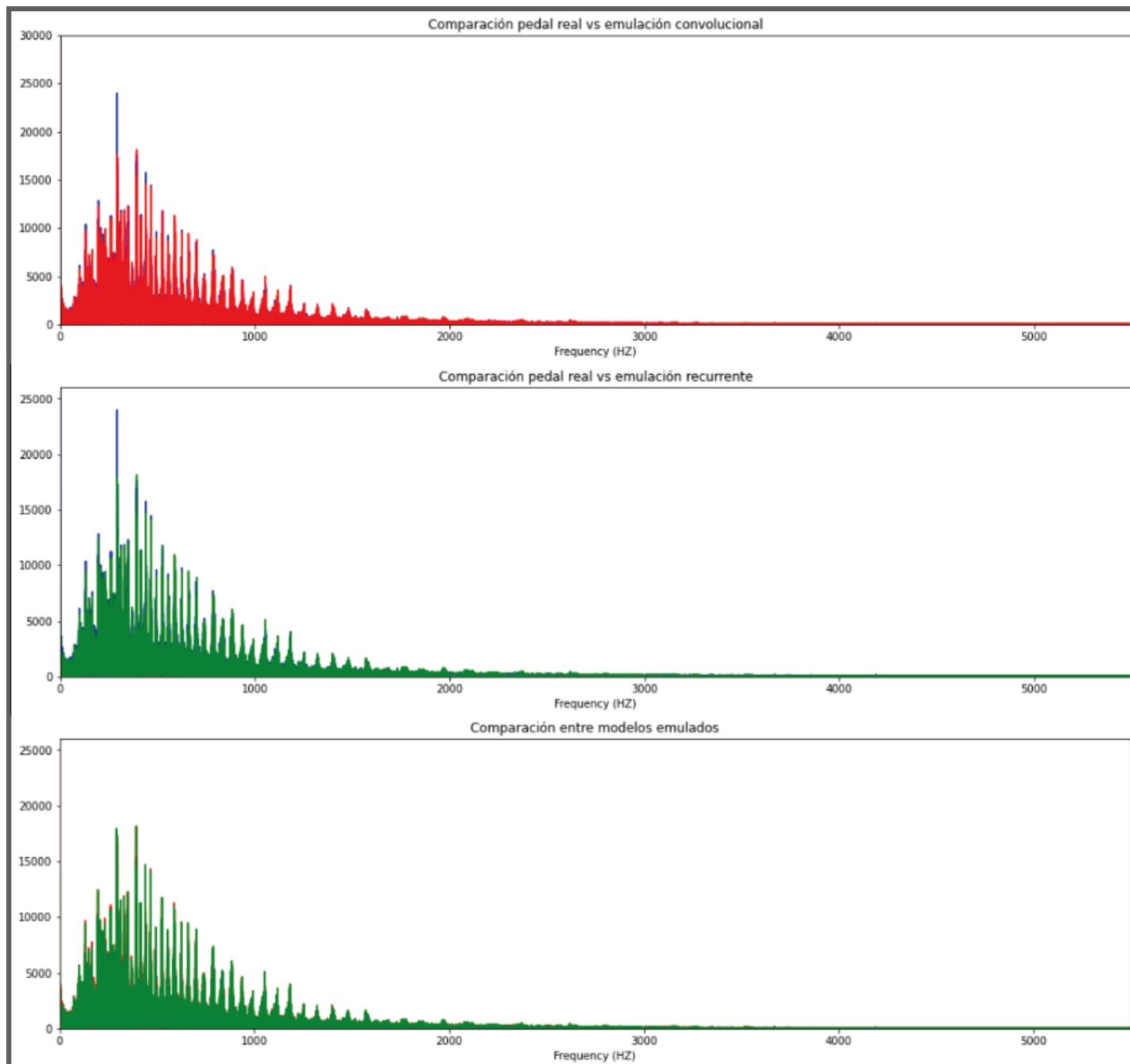
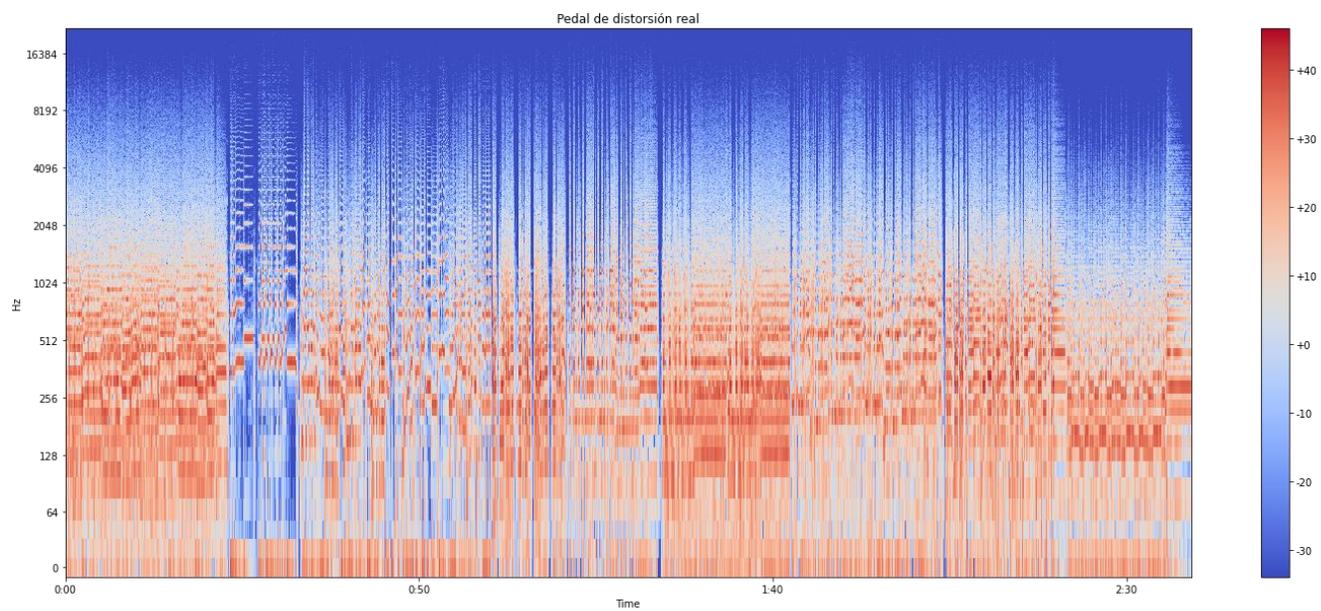


Figura 37

Espectrograma del pedal de distorsión real

**Figura 38**

Espectrograma de la emulación convolucional del pedal de distorsión

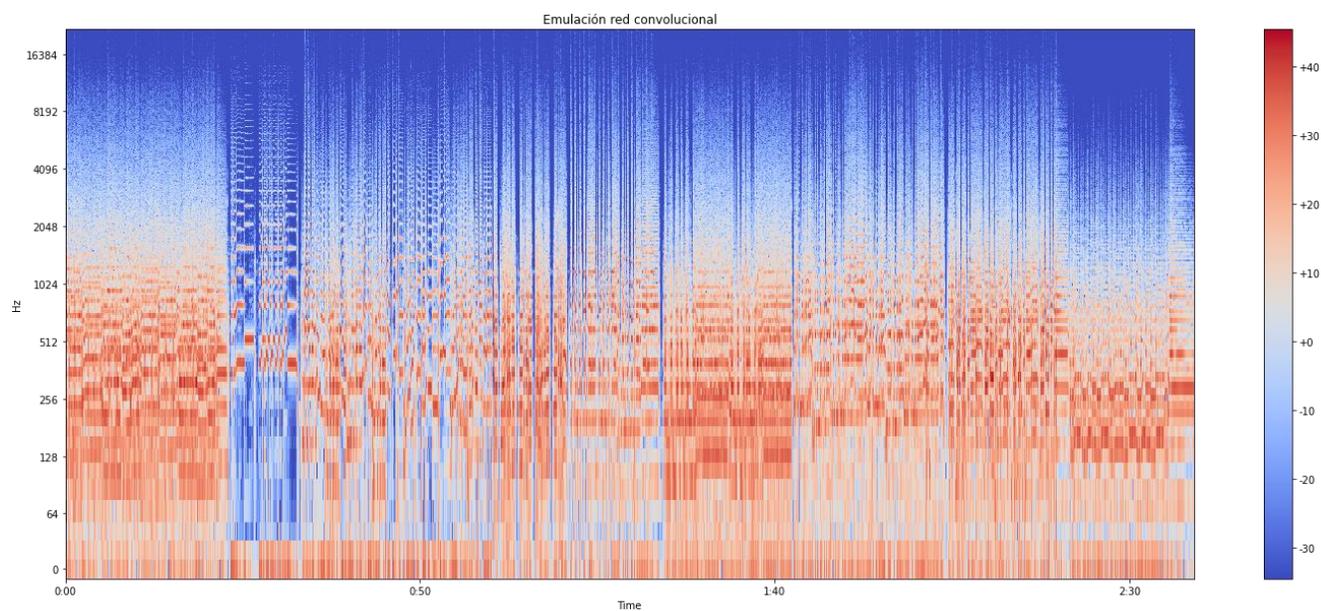
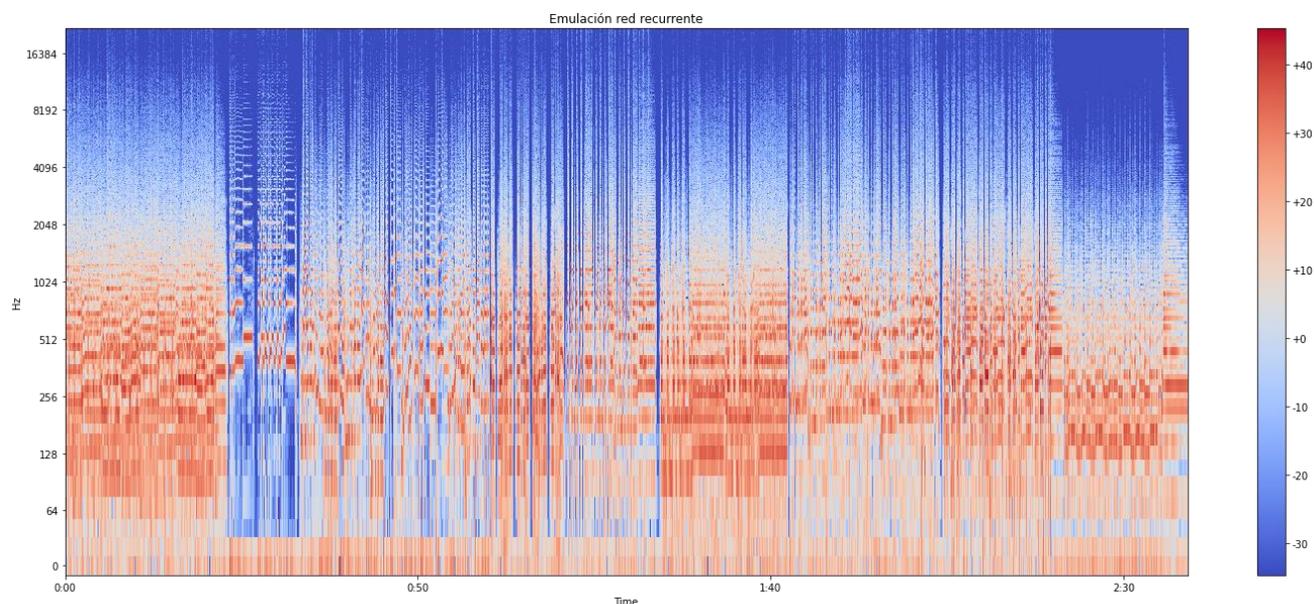


Figura 39

Espectrograma de la emulación recurrente del pedal de distorsión



Resumen objetivo de los modelos

El resumen de los datos obtenidos en el análisis matemático de las señales se puede ver en la Tabla 8, donde se aprecia que el modelo convolucional que emula al pedal de distorsión *Ibanez Tube Screamer* es el que más se asemeja al dispositivo real con un ESR de 0.0057 el cual es realmente bajo. De la misma manera se puede apreciar que en general los modelos convolucionales presentan un error más bajo en relación con los modelos recurrentes que emulan el mismo dispositivo, todos los modelos pueden ser ejecutados en tiempo real lo que demuestra un costo computacional bajo para su ejecución.

Tabla 8

Tabla de resumen de resultados objetivos de análisis.

	ESR	Error ABS pico	Error ABS medio	Ejecución tiempo real
CNN amplificador	0.0451	0.175	0.075	Si
RNN amplificador	0.1097	0.200	0.100	Si

	ESR	Error ABS pico	Error ABS medio	Ejecución en tiempo real
CNN pedal	0.0057	0.300	0.050	Si
RNN pedal	0.0220	0.320	0.100	Si

Análisis subjetivo

Se realizó una encuesta a 30 personas con el fin de verificar si de alguna manera pueden identificar guiándose en el sonido, cual podría ser el equipo analógico real y cuales las emulaciones tanto del amplificador *Mesa Boogie* así como del pedal de distorsión *Ibanez Tube Screamer*, y así establecer si existe una diferencia notoria entre el audio real y las emulaciones y finalmente tener una idea de la calidad de audio percibida por los encuestados. Para ello se realizaron cinco preguntas, las cuales se basan en escuchar seis audios que han sido etiquetados con números para que el test se base totalmente en la percepción del oído de cada encuestado, mas no en la etiqueta del nombre de cada uno. El contenido de cada audio se describe en la Tabla 9.

Tabla 9

Etiquetas de los audios de la encuesta

Audio 1	Emulación del amplificador RNN
Audio 2	Emulación del amplificador CNN
Audio 3	Emulación del pedal de distorsión RNN
Audio 4	Emulación del pedal de distorsión CNN
Audio 5	Amplificador real
Audio 6	Pedal de distorsión real

Figura 40

Pregunta 1 de la encuesta

En los siguientes audios se encuentra 1 guitarra grabada con un amplificador real y 2 emulaciones hechas con redes neuronales. Seleccione la opción que usted crea es el amplificador real.

30 respuestas



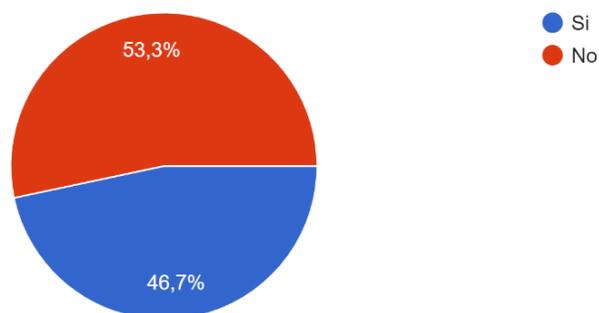
La figura 40 muestra la primera pregunta de la encuesta la cual nos muestra que un 56.7% respondieron a la pregunta de manera correcta ya que en este caso el Audio 5 corresponde al amplificador real, otro 33.3% creyó que la emulación convolucional del amplificador se trataba del amplificador real y por último apenas un 10% creyó que la emulación recurrente se trataba del amplificador real, lo cual muestra en este caso que más de la mitad de los encuestados pudieron reconocer cual es el amplificador real, mientras que otro gran porcentaje creyó que la emulación del modelo convolucional era el real, lo cual demuestra de manera subjetiva que el modelo convolucional obtenido presenta mejores resultados que el modelo recurrente.

Figura 41

Pregunta 2 de la encuesta

¿Puede percibir alguna diferencia notable entre los 3 audios anteriores correspondientes al amplificador?

30 respuestas



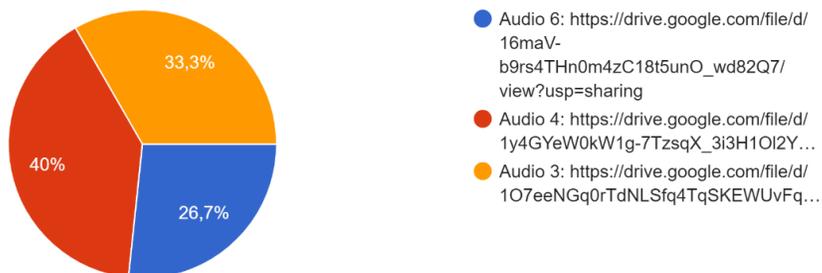
La Figura 41 muestra que un 53.3% de los encuestados no logra percibir alguna diferencia notable entre los audios correspondientes al amplificador real y a sus emulaciones convolucional como recurrente lo cual sugiere que la diferencia en el sonido no es demasiado grande para ser percibida por todos sin embargo esa diferencia es notable.

Figura 42

Pregunta 3 de la encuesta

En los siguientes audios se encuentra 1 guitarra grabada con un pedal de distorsión real y 2 emulaciones hechas con redes neuronales. Seleccione...ón que usted crea es el pedal de distorsión real.

30 respuestas



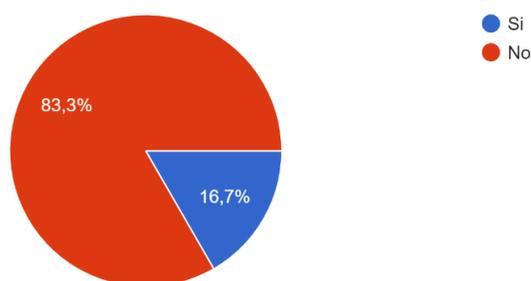
La Figura 42 nos muestra que apenas un 26.7% lograron responder de manera correcta ya que el Audio 6 se trata del pedal de distorsión real mientras que un 40% de los encuestados creyeron que la emulación correspondiente a la emulación del modelo convolucional se trataba del pedal real y un 33.3% creyeron que la emulación recurrente se trataba del pedal real, estos resultados sugieren que los modelos de emulación del pedal son muy cercanos al real ya que el porcentaje de personas que acertaron es el más bajo.

Figura 43

Pregunta 4 de la encuesta

¿Puede percibir alguna diferencia notable entre los 3 audios anteriores correspondientes al pedal de distorsión?

30 respuestas

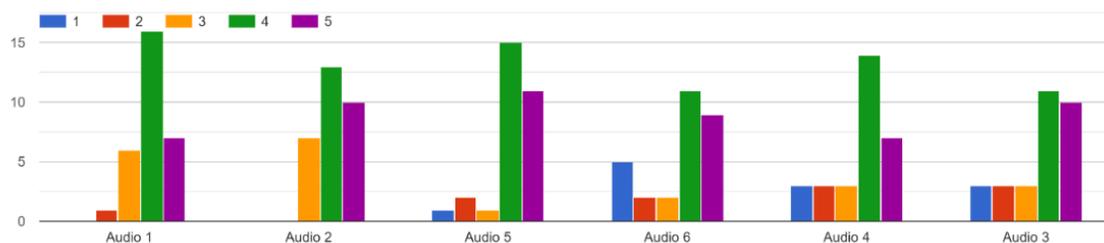


La Figura 43 nos muestra que un 83.3% de los encuestados no perciben una diferencia notable entre los audios del pedal de distorsión real y los modelos convolucional y recurrente que lo emulan, esto sugiere que la diferencia en este caso es muy sutil y de hecho no se puede determinar con exactitud cuál es el equipo real de cuál es el equipo emulado.

Figura 44

Pregunta 5 de la encuesta

En una escala del 1 al 5 donde 1 es malo y 5 excelente, califique la calidad de audio de los audios escuchados previamente.



La Figura 44 nos sirve para determinar la calidad de audio percibido en una escala del 1 al 5, en donde se tiene que la mayoría de encuestados califican a los audios con un puntaje de 4 seguidos por la calificación de 5 lo cual sugiere que la calidad de audio percibida es muy buena, en la Tabla 10 se presenta un promedio de la calificación de la calidad de cada audio.

Tabla 10

Calidad subjetiva de los audios

	Calidad subjetiva
Amplificador real	4.10
Emulación CNN del amplificador	3.73
Emulación RNN del amplificador	3.63
Pedal Real	3.80
Emulación CNN del pedal	4.10
Emulación RNN del pedal	3.97

Conclusiones

Las redes neuronales recurrentes, así como las convolucionales pueden tener una gran utilidad en el campo del procesamiento digital de señales. En el caso específico de este trabajo, en señales de audio, ya que como se pudo observar en el apartado de análisis objetivo como subjetivo de los modelos presentados, ambas redes neuronales cumplen la función de una emulación bastante cercana a los dispositivos reales ya que los errores presentados por ambos tipos de modelos no superan el 0.1097 llegando incluso a ESR de 0.0057.

Existe mayor diferencia de las emulaciones respecto de los dispositivos originales cuando la señal de entrada tiene mayor amplitud, esto se debe al comportamiento no lineal de los dispositivos analógicos. Esa no linealidad representa una gran dificultad en el campo del modelado, de todas maneras, el error nunca es apreciablemente grande lo que demuestra que las emulaciones con redes neuronales pueden emular de manera bastante cercana las características no lineales de los dispositivos.

Los modelos con mayor error son los que pretenden emular al amplificador *Mesa Boogie Dual Rectifier* con ESR de 0.0451 para el modelo convolucional y un 0.1097 para el modelo recurrente, esto se debe a que la emulación de este dispositivo tiene mayor complejidad ya que es un modelo compuesto que pretende emular tanto el amplificador mencionado, conectado a un Gabinete y su señal es grabada a través de un micrófono lo cual hace que también sea susceptible al comportamiento de estos dispositivos.

Los modelos con menor error fueron los que pretenden emular el pedal de distorsión *Ibanez Tube Screamer* con errores para el modelo convolucional de 0.0057 y para el modelo recurrente de 0.022 son realmente valores muy bajos. Esto se debe principalmente a que la complejidad en la emulación de este dispositivo es menor que la del amplificador, ya que este pedal con los parámetros con los que fue configurado no realiza una distorsión tan marcada

como es el caso del amplificador, además que en este caso para emular este pedal la señal va directamente del pedal hacia la interfaz sin pasar por algún otro dispositivo intermedio que permita captar sus características.

El costo computacional requerido para ejecutar estos modelos es relativamente bajo para los CPU actuales, ya que permite correr todos los modelos en tiempo real sin ningún tipo de problema.

La mayoría de diferencias de los modelos emulados con los dispositivos reales se presentan en las frecuencias bajas, esto se debe propiamente a la señal de entrada, ya que la señal de entrada original de la guitarra ya presenta varias componentes en las frecuencias graves y al aplicar efectos sobre la misma estos efectos actúan también en esa sección del espectro sin embargo algunas de esas frecuencias bajas no son audibles y no representan las características audibles que realmente le dan carácter a estos dispositivos, ya que las frecuencias más importantes se encuentran en la parte media del espectro donde se pudo verificar que el comportamiento de las emulaciones funciona de manera muy cercana. Por lo tanto, este inconveniente en las señales puede ser resuelto con la aplicación de filtros que eliminen las frecuencias no audibles pudiendo así obtener modelos con un menor error objetivo.

En el campo del sonido, sobre todo el sonido orientado a la música y al arte en general, lo que realmente importa es lo que nuestros oídos pueden percibir. Del análisis subjetivo de estas emulaciones presentadas en el capítulo IV, se puede concluir en concordancia con el análisis objetivo, que la emulación convolucional del pedal de distorsión tuvo un puntaje de 4.1/5 en calidad percibida el cual es el puntaje mayor de las emulaciones, el resto de audios de las emulaciones oscilan entre 3.63/5 y 4.1/5 en promedio. Lo cual indica que la calidad de audio que se puede percibir de las emulaciones es bastante buena llegando casi a excelente,

además se puede ver que, para las emulaciones del pedal de distorsión, los encuestados en general no pueden percibir diferencias notables en los audios de las emulaciones, respecto del dispositivo real. Mientras que, para el caso de las emulaciones del amplificador, alrededor de la mitad de encuestados pudo percibir una diferencia notable. Esto muestra nuevamente en concordancia con la conclusión de la parte subjetiva que se podría deber a la complejidad en la emulación de cada dispositivo.

Finalmente se puede concluir que el audio digital puede resultar una manera muy conveniente de procesar las señales, ya que ofrecen muchas ventajas a comparación de los equipos analógicos como podrían ser el tema de costos, mantenimiento, espacio, versatilidad, entre otros. Además de ofrecer resultados bastante similares a los del hardware real, sin que la diferencia sea demasiado notoria.

Recomendaciones

Para trabajar con redes neuronales es muy importante el *set de datos* que se utiliza para el entrenamiento tanto como el modelo y otros parámetros. Por lo tanto, se deben obtener los datos más representativos para los modelos que se quieran entrenar. Esto aplica no solo para las aplicaciones de procesamiento de señales de audio, sino para cualquier tipo de aplicación de redes neuronales ya sea esta de modelos regresivos, de clasificación o de cualquier tipo.

Es importante el preprocesamiento de las señales a ingresar en las redes. En el caso de señales de audio, dependiendo si es un problema de clasificación o regresión se puede utilizar múltiples técnicas, por ejemplo, en problemas de clasificación resulta conveniente preprocesar las señales para convertirlas en espectrogramas y analizar su información en base a las imágenes correspondientes, mientras que en muchos casos de regresión de audio puede

resultar conveniente trabajar directamente con las señales de audio en pequeñas muestras de tiempo de algunos ms.

Para el entrenamiento, así como para la validación, es importante utilizar métricas que estén acorde con el problema que se pretende solventar. En caso de audio resulta conveniente utilizar filtros de pre-énfasis en las frecuencias medias y agudas para los cálculos del error en la etapa de validación debido a que la cantidad de frecuencias graves pueden enmascarar algunas frecuencias del espectro que contienen información importante sobre la señal, otro método sería filtrar las frecuencias graves.

Para elegir un tipo de red neuronal es importante conocer el tipo de problema que se pretende resolver, por lo general pueden ser problemas de regresión y clasificación para lo cual existen varios tipos de redes neuronales con distintas arquitecturas que pueden ser más eficientes para ciertas labores, además de eso hay que tener en cuenta el tipo de datos que van a ingresar a la red para elegir de manera correcta la red a utilizar.

Es importante realizar test con diferentes parámetros para determinar cuáles van a ser los parámetros que mejor se adapten a las necesidades del problema a resolver, teniendo en cuenta además factores como la velocidad de procesamiento, eficiencia y otras limitantes según sea el caso.

Trabajos Futuros

La realización de esta investigación puede servir como fundamento para entrar más en profundidad en los hiper-parámetros de entrenamiento de las redes con el fin de optimizar los mismos para entregar las emulaciones más similares a las reales posibles. Además, se pueden poner a prueba otros modelos de redes neuronales para la emulación de equipo analógico en el ámbito de sonido e incluso se pueden poner a prueba redes similares para otros usos

prácticos en el sonido por ejemplo la eliminación del ruido de fondo, extraer pistas de instrumentos de una canción, entre otros.

En este trabajo se presentaron modelos de redes neuronales los cuales trabajan sobre la señal de audio como tal. En trabajos futuros se pueden poner a prueba redes neuronales que trabajen en otros dominios por ejemplo el dominio de la frecuencia es decir que los datos de entrada para la red sean los valores de la transformada de Fourier FFT o incluso el uso de espectrogramas para el entrenamiento de las redes.

Bibliografía

- Burgos, J., & Vargas, J. (2016). *Análisis de sistemas procesadores de señales de guitarras*. Ingenium. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=5762963>
- Calvo, D. (20 de Julio de 2017). *Diego Calvo*. Obtenido de <https://www.diegocalvo.es/red-neuronal-convolucional/>
- Cardona, J., Leal, J., & Ustariz, J. (2020). *Modelado matemático de caja blanca y negra en educación en ingeniería*. Form. Univ. vol.13 no.6 La Serena. Obtenido de https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-50062020000600105#:~:text=Es%20una%20metodología%20mediante%20la,requieren%20tipos%20de%20análisis%20particular.
- Cayetano Matamoros, A. J. (2017). *Amplificador de válvulas para guitarra. Implementación, propuestas de mejora y diseño de un método de análisis de comportamiento objetivo y subjetivo*. Madrid: E.T.S.I. y Sistemas de Telecomunicación (UPM). Obtenido de <https://oa.upm.es/52498/>
- Cobas, L. (2018). *Diseño y desarrollo de un pedal de efecto para la guitarra eléctrica*. Master en Ingeniería industrial - Universidad Pontificia de Comillas. Obtenido de <https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/32602/TFM-CobasAranguren%2CLuis.pdf?sequence=1>
- Dogorov, M. (2019). *Deep Learning para modelado de amplificadores valvulares*. Buenos Aires: Universidad de Buenos Aires. Obtenido de https://www.researchgate.net/publication/334289316_Deep_Learning_para_modelado_de_amplificadores_valvulares

Eero-Pekka, D., Lauri, J., & Vesa, V. (s.f.). *Real-Time Modeling of Audio Distortion Circuits with Deep Learning*. Acoustics Lab, Department of Signal Processing and Acoustics.

Obtenido de https://www.smc2019.uma.es/articles/S5/S5_02_SMC2019_paper.pdf

Eero-Pekka, D., Lauri, J., Etienne, T., & Vesa, V. (2019). *Deep Learning for tube amplifier emulation*. Finland: Acoustics Lab, Dept. of Signal Processing and Acoustics. Obtenido de <https://arxiv.org/abs/1811.00334>

Evangelista, G. (2010). Physically inspired playable models of guitar, a tutorial. *4th International Symposium on Communications, Control and Signal Processing (ISCCSP)* (págs. 1-4). Limassol, Cyprus: IEEE.

Fermatta. (6 de 12 de 2019). *Fermatta*. Obtenido de <https://www.fermatta.edu.mx/2019/12/06/que-es-un-plugin/>

Gavilán, I. (2020). *Catálogo de componentes ed redes neuronales (III): funciones de pérdida*. Obtenido de <https://ignaciogavilan.com/catalogo-de-componentes-de-redes-neuronales-iii-funciones-de-perdida/>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Obtenido de <http://www.deeplearningbook.org>

Henshall, M. (2011). *Sound Matters*. Obtenido de <https://www.yoursoundmatters.com/history-development-magnetic-pickups/>

Hunter, D. (2002). *Guitar: A Complete Guide for the Player*. Thunder Bay Pr.

Ibanez. (2022). *Ibanez*. Obtenido de https://www.ibanez.com/usa/products/detail/ts808_99.html

IBM. (2021). *IBM*. Obtenido de <https://www.ibm.com/docs/es/spss-modeler/saas?topic=networks-neural-model>

- Izaurieta, F., & Carlos, S. (2000). *Redes Neuronales Artificiales*. Departamento de Física, Universidad de Concepción, Concepción, Chile. Obtenido de https://www.academia.edu/750527/Redes_Neuronales_Artificiales
- Jakub, L. (2021). *Learning the Musical Instrument amplifier model with Neural Networks*. Prague: Faculty of Electrical Engineering. Obtenido de https://dspace.cvut.cz/bitstream/handle/10467/96736/F3-BP-2021-Lukes-Jakub-Learning_the_Musical_Instrument_Amplifier_Model_with_Neural_Networks.pdf
- Kingma, D., & Lei Ba, J. (2015). *Adam: A Method For Stochastic Optimization*. ICLR 2015. Obtenido de <https://arxiv.org/abs/1412.6980>
- Lähdevaara, J. (2012). *The science of electric guitars and guitar electronics*. Oulu.
- Larochelle, H., Bengio, Y., Louradour, J., & Lamblin, P. (2009). *Exploring Strategies for Training Deep Neural Networks*. Journal of Machine Learning Research. Obtenido de <https://jmlr.org/papers/volume10/larochelle09a/larochelle09a.pdf>
- López, A. (2019). *Reducción de ruido en señales de audio basada en una Red Neuronal Convolutiva*. Catalunya: UOC. Obtenido de <https://openaccess.uoc.edu/bitstream/10609/97227/8/alopezmoraTFG0619memoria.pdf>
- MathWorks. (2021). Measuring Signal Similarities. Obtenido de <https://es.mathworks.com/help/signal/ug/measuring-signal-similarities.html>
- Mesa Boogie. (2022). *Mesa Boogie*. Obtenido de <https://www.mesaboogie.com/en-US/Amp/2.DR.1.B.LC>
- Oord, A. v., & Sander, D. (2016). *WaveNet: A Generative Model for Raw Audio*. Deepmind. Obtenido de <https://arxiv.org/abs/1609.03499>

- Ordiales, H., & Barrera, G. (2019). *Transferencia de estilo entre audios mediante redes neuronales*. Buenos Aires: Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas. Obtenido de <https://reddi.unlam.edu.ar/index.php/ReDDi/article/view/83/171>
- Perez, J. L. (2012). *Distorsion de sonidos con redes neuronales*. Buenos Aires: Universidad de Belgrano. Obtenido de <http://repositorio.ub.edu.ar/handle/123456789/890>
- Pontificia Universidad Católica del Ecuador. (2022). *PUCE*. Obtenido de <https://puceapex.puce.edu.ec/web/cev/tipos-de-redes-neuronales-y-framework-de-desarrollo/>
- Rico, L. Á. (2015). *Turbo manual de guitarra eléctrica*.
- Roascio, M. (2007). *El rincón del guitarrista*. Obtenido de <http://www.guitarraonline.com.ar/ibanez5.htm>
- Sanchez, D. (2005). *Emulating electric guitar effects with neural networks*. Universitat Pompeu Fabra. Obtenido de <http://theguitar-blog.com/wp-content/uploads/2012/12/emulatin>
- Torres, J. (2019). *Jordi Torres AI*. Obtenido de <https://torres.ai/deep-learning-inteligencia-artificial-keras/>
- Universidad De Sevilla. (2022). *ISI*. Obtenido de <http://www.lsi.us.es/docencia/get.php?id=361>
- Valenzuela, D. (2013). *El universo sonoro de la guitarra amplificada*. Catalunya: esmuc. Obtenido de <https://www.recercat.cat/handle/2072/228072>
- Velasco, L. (2020). *Optimizadores en redes neuronales profundas: un enfoque práctico*. Obtenido de <https://velascoluis.medium.com/optimizadores-en-redes-neuronales-profundas-un-enfoque-práctico-819b39a3eb5#:~:text=Save->

,Optimizadores%20en%20redes%20neuronales%20profundas%3A%20un%20enfoco
%20práctico,requiere%20de%20bastante%20bagaje%20matemático.

Velten, K. (2009). *Mathematical Modeling and Simulation: Introduction for Scientists and Engineers*. Verlag GmbH & Co. KGaA.

Wonho, Y., Dixon, M., & Robert, Y. (1997). A modified bark spectral distortion measure which uses noise masking threshold. En *Workshop on Speech Coding for Telecommunicatinos Proceedings. Back to basics: Attacking Fundamental Problems in Speech Coding* (págs. 55-56). IEEE.

Wright, A., Eero-Pekka, D., Lauri, J., & Vesa, V. (2019). *Real-Time Guitar Amplifier Emulation with Deep Learning*. applied sciences. Obtenido de <https://www.mdpi.com/2076-3417/10/3/766>

Wright, Damskägg, & Välimäki. (2019). *Real-time black-box modelling with recurrent neural networks*. International Conference on Digital Audio Effects. Obtenido de https://dafx.de/paper-archive/2019/DAFx2019_paper_43.pdf

Zeigler, B., Muzy, A., & Kofman, E. (2019). *Theory of Modeling and Simulation*. San Diego: Academic Press.