Redes Neuronales (8654)
Uso del perceptrón multicapa para distorsionar audio de guitarra eléctrica

10 de febrero de 2016

Matías H. Senger (95589, SengerEfectos@hotmail.com)

Facultad de Ingeniería de la Universidad de Buenos Aires

Resumen

En el presente trabajo se desarrolló una serie de perceptrones multicapa al los cuales se les enseñó a distorsionar audio de guitarra eléctrica con el fin de obtener aplicaciones en el ámbito de la música moderna. Los perceptrones reciben una señal de audio de una guitarra eléctrica y en sus salidas se obtienen las señales distorsionadas en forma similar a la que lo haría un circuito analógico de transistores.

Índice

1. Introducción

En géneros musicales modernos como el rock, la distorsión es un *efecto de sonido*[1] muy común que se aplica por lo general a las guitarras eléctricas, aunque también se suele aplicar a bajos y con menos frecuencia a teclados y otros instrumentos. El efecto de la distorsión consiste justamente en distorsionar la señal de entrada con el fin de obtener una nueva composición armónica y darle más riqueza al sonido del instrumento. En la actualidad existen sistemas distorsionadores digitales y analógicos, nosotros nos limitaremos a los analógicos (que son los originales) ya que los digitales intentan, de alguna manera, emular a los analógicos.

[1] Un *efecto de sonido* es, desde un punto de vista matemático, una función que se aplica a la señal de audio original con el fin de obtener una versión modificada de la misma que tenga otras cualidades sonoras. En el ámbito de la guitarra eléctrica es muy común la presencia de los denominados *pedales de efectos* que son literalmente cajas metálicas con conectores de entrada y de salida que poseen en su interior un circuito electrónico encargado de generar algún efecto. Los efectos más comunes en pedales de guitarra son distorsión, *overdrive*, *chorus*, *delay*, *wah wah*, etc.

Los distorsionadores analógicos son amplificadores que, de alguna forma, se salen de su régimen lineal. Esto hace que la señal de salida gane determinados armónicos que no están presentes en la señal de entrada y así le confieren un nuevo color al sonido. Debido a que el amplificador trabaja en una región no lineal, la relación entre la señal de entrada y la de salida también es no lineal. Esta relación no lineal hace que muchas veces sea muy complicado obtener la transferencia del sistema lo cual dificulta la imitación de un distorsionador analógico mediante técnicas digitales.

Dado que los perceptrones multicapa son capaces de implementar funciones no lineales sin necesidad de conocer en forma explícita su formulación matemática, si no observando cómo actúa sobre distintas muestras, esto los convierte en buenos candidatos a la hora de recrear sistemas no lineales como los distorsionadores de guitarra eléctrica.

El objetivo fundamental de este trabajo es el de obtener perceptrones capaces de aprender cómo distorsionar una señal de audio genérica de forma similar a la que lo haría un circuito de distorsión analógico.

2. Desarrollo del trabajo

Con el fin de llevar a cabo el proyecto, lo primero que se realizó fue codificar un *script* de *Octave*[2] que permitiera implementar perceptrones con un número arbitrario de capas y de neuronas por capa. Para el entrenamiento de estos perceptrones se utilizó el algoritmo *back propagation* con la variante de agregar un parámetro de *momentum*, tal como detalla [1] en la página 123, con el fin de acelerar el entrenamiento del perceptrón. Una vez implementado el algoritmo completo, se lo puso a prueba enseñándole a perceptrones de diversas estructuras distintas funciones sencillas, que nada tienen que ver con el audio de una guitarra eléctrica, únicamente con el fin de verificar el correcto funcionamiento del algoritmo escrito. Luego se procedió luego a implementar la aplicación de audio.

2.1. Arquitectura utilizada

Dado que el perceptrón se implementó en una computadora, el mismo debe procesar señales grabadas digitalmente. El formato utilizado para almacenar las muestras de entrenamiento fue el formato WAV, con una frecuencia de muestreo de 42,1 kHz que corresponde al estándar utilizado en los discos compactos de música.

Una forma posible de implementar el perceptrón hubiera sido utilizar N neuronas de entrada y N neuronas de salida, con N igual a la cantidad de muestras de la señal a procesar, y así obtener una señal de igual longitud en las N neuronas de salida. Este enfoque es poco práctico en términos computacionales debido a que para procesar una señal de audio de un segundo, muestreada a 42,1 kHz, se requerirían 42100 neuronas de entrada y otras tantas de salida, más las de las capas ocultas. Esto es imposible de implementar en forma práctica en una computadora estándar actual. Por otro lado también limita la longitud de los audios a procesar a una longitud fija determinada por la cantidad de neuronas de las capas de entrada y salida del perceptrón.

Para solventar estas dificultades, se optó por implementar un perceptrón con un menor número de neuronas que realice el procesamiento de la señal en forma progresiva avanzando desde el principio hasta el final del archivo de audio. De esta forma se obtiene una red de un tamaño mucho menor que en el otro caso y la longitud de las señales a procesar es arbitraria, logrando además una mayor flexibilidad.

Se optó por trabajar en el dominio del tiempo y no en el de la frecuencia ya que los distorsionadores de guitarra eléctrica son no lineales por lo que se consideró, a priori, que no habría beneficio alguno al pasar al dominio de la frecuencia. Por otro lado se evitan efectos de ventaneo al pasar al dominio de la frecuencia, los cuales podrían ser no despreciables si la cantidad de neuronas de la capa de entrada es pequeña.

En la figura 1 se muestra la arquitectura utilizada. El perceptrón posee un número arbitrario de neuronas de entrada las cuales van recorriendo en forma progresiva la señal a procesar, avanzando de a una muestra por vez, y una única neurona de salida en la cual se obtiene la señal de salida.

No se consideraron más neuronas en la capa de salida ya que hubiera sido más complicada la reconstrucción de la señal debido a la superposición de las ventanas de la señal en las neuronas de salida. En cambio, al utilizar una única neurona, cada muestra se genera en forma individual y no es necesario ningún proceso de reconstrucción.

Como se mencionó anteriormente, la estructura interna del perceptrón de la figura 1 es completamente arbitraria en lo que respecta a la cantidad de capas y la cantidad de neuronas por capa. La nomenclatura utilizada para identificar a las estructuras en el presente informe será colocar en forma de vector la cantidad de neuronas que tiene cada capa del perceptrón, por ejemplo [4,2,3,1] identifica a un perceptrón de cuatro capas, con cuatro neuronas en la capa de entrada, una neurona en la capa de salida y dos y tres neuronas en las capas ocultas respectivamente, tal como se ilustra en la figura 2.

[2]Octave es un software de código abierto similar a Matlab.

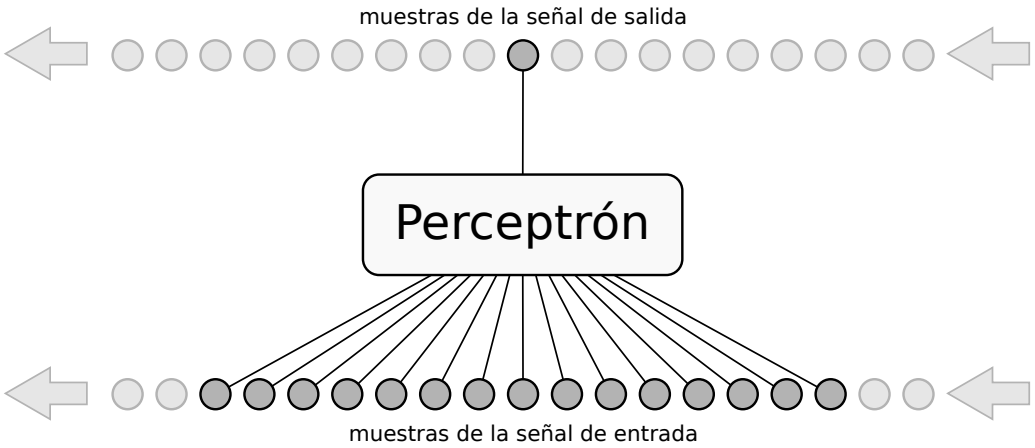

Figura 1: Arquitectura utilizada para procesar las señales de audio con el perceptrón desarrollado.

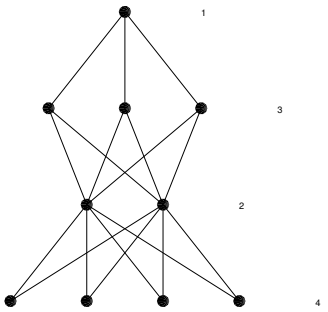

Figura 2: Estructura de un perceptrón de ejemplo implementado mediante el algoritmo desarrollado. La estructura de este perceptrón se describe mediante el vector [4,2,3,1]. La capa inferior es la capa de entrada y la capa superior es la capa de salida.

2.2. Entrenamiento del perceptrón

Para entrenar al perceptrón se utilizó el algoritmo de *back propagation* tal cual lo detalla [1] en la sección correspondiente, con el agregado del parámetro de *momentum* desarrollado en la misma obra. En cada paso que se aplicó el algoritmo de *back propagation* se procedió de la siguiente forma:

1. Se aplicó a las N neuronas de entrada las muestras $[k, k+N-1]$ de la señal de entrada y se propagó el resultado hasta obtener la muestra de salida k.
2. Se aplicó el algoritmo *back propagation* con la consecuente corrección de pesos de interconexión entre neuronas.
3. Se avanzó la ventana de entrada a las muestras $[k+1, k+N]$ y se volvió al paso 1.

Este procedimiento se repitió hasta haber recorrido por completo la señal a procesar.

2.2.1. Obtención de las muestras de entrenamiento

Dado que el objetivo del perceptrón es, en el presente trabajo, distorsionar señales provenientes de una guitarra eléctrica, lo que se hizo fue grabar con la computadora distintos *riffs*[3] utilizando una guitarra eléctrica sin distorsión. Luego se conectó la salida de audio de la computadora a la entrada de un pedal analógico de distorsión, en particular el modelo *Rat* de la compañía *ProCo,* y la salida del mismo a la entrada de audio de la computadora. En la figura 3 se muestra un esquema de este procedimiento.

De esta forma se grabó una versión distorsionada de cada uno de los *riffs* grabados previamente, obteniéndose así dos audios para cada *riff*

- clean*i*.wav → versión limpia[4] del *riff* i-ésimo grabado.

[3] *Riff* es un término empleado en el ámbito del rock que hace referencia a un fraseo melódico tocado con una guitarra eléctrica. Un ejemplo de un *riff* es la melodía con la que empieza la canción *Day Tripper* de *The Beatles*.

[4] Sin distorsión.

Figura 3: Esquema de obtención de las señales con y sin distorsión utilizadas para entrenar al perceptrón.

- dist$i$.wav → versión distorsionada del *riff* i-ésimo grabado.

Estos archivos (cinco en total), cuya extensión temporal oscila entre 3 y 5 segundos, fueron utilizados como señal de entrada (clean$i$.wav) y señal de salida (dist$i$.wav) para entrenar al perceptrón. En la figura 4 se muestra un ejemplo de estas señales.

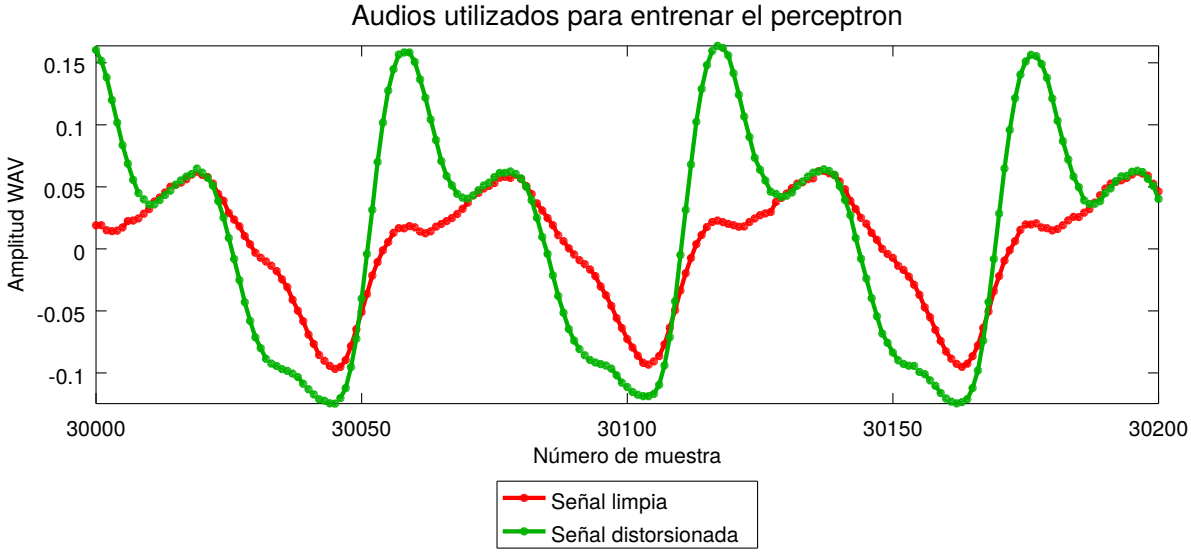

Figura 4: Ejemplo de un fragmento correspondiente a un archivo clean$i$.wav y dist$i$.wav utilizado para entrenar al perceptrón.

Adicionalmente se generó una serie de archivos de audio con señales senoidales puras y se las distorsionó con el pedal de distorsión de la misma forma que se explicó anteriormente.

2.3. Medición del error

En el caso de señales de audio como las utilizadas en el presente trabajo, resultan de escasa utilidad los estimadores de error tradicionales, como el error cuadrático medio, ya que no se puede dar una relación certera entre la calidad del audio obtenida y un determinado valor del error cuadrático medio. Puede ocurrir que, a pesar de que el perceptrón no produzca la misma distorsión exacta que el sistema analógico original, el audio obtenido en la salida sea de buenas cualidades sonoras, lo cual resulta muy complicado de medir en forma matemática. Es por ello que en el análisis de resultados no se utilizó un estimador matemático, como es común, si no que se juzgaron los resultados de forma visual mediante una superposición gráfica de las señales y se realizó una comparación auditiva reproduciendo las señales de distorsión originales y las producidas por los distintos perceptrones. Esta comparación auditiva fue la más importante a la hora de juzgar los resultados obtenidos.

3. Resultados obtenidos

Debido a que no se conoce una forma analítica que permita determinar la estructura óptima de un perceptrón (número de capas y neuronas por capas) para la tarea a realizar, se procedió de forma empírica con una guía previa en base a la experiencia del autor de [2]. Se realizaron numerosas pruebas variando tanto la estructura

del perceptrón (capas y número de neuronas) como así también los parámetros de aprendizaje del algoritmo *back propagation*.

3.1. Resultados exitosos

Redes de más capas demostraron ser mejores para la tarea de distorsionar audios de guitarra eléctrica. Los mejores resultados que se obtuvieron fueron con redes de estructuras [32,16,8,8,1] y [32,16,8,4,2,1]. En estos casos el sonido obtenido se asemeja en gran medida al original y además el ruido de fondo, aún presente, es amplificado en menor medida que con las demás topologías de redes. Por otro lado estas redes demostraron ser capaces de aprender a distorsionar mucho más rápido que otras redes.

Perceptrón [32,16,8,8,1] Se obtuvieron resultados satisfactorios con una red de estructura [32,16,8,8,1], tal como se muestra en la figura 5. En la misma se puede ver un fragmento de la señal con la distorsión original y el mismo fragmento correspondiente a la distorsión generada por el perceptrón. El perceptrón fue entrenado con los audios 1, 2 y 3 (cada audio en su extensión completa) en dicho orden, con parámetros $\eta = 0{,}01$ y momentum $= 0{,}01$. En la figura mencionada se encuentra un fragmento del audio número 1 luego de que al perceptrón se le enseñasen los tres audios, lo cual demuestra que el mismo ha sido capaz no solo de aprender el audio número 1 si no también de "recordarlo" luego de que se le hayan enseñado los otros dos audios. A pesar de que se observan diferencias (en la figura), ambas señales presentan una gran similitud y, más importante aún, auditivamente son similares. El error cuadrático medio cometido al procesar el audio de la figura 5 (completo) fue de 721×10^{-6}. La distorsión generada por este perceptrón tiene levemente más componentes armónicas de alta frecuencia pero no deja de tener un sonido agradable y con el mismo carácter que la distorsión original con lo cual su utilización como efecto distorsionador sería perfectamente posible.

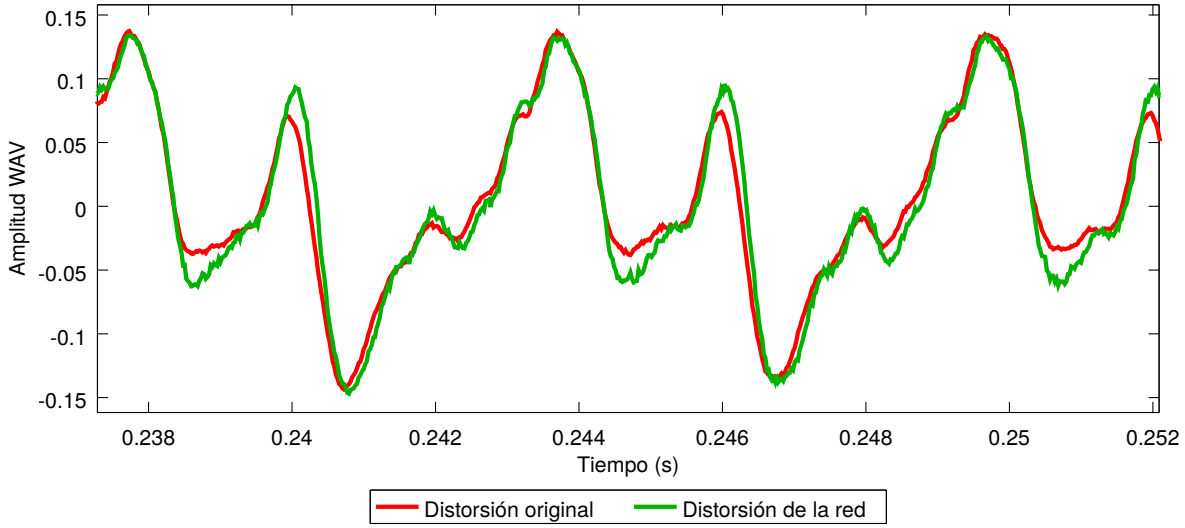

Figura 5: Superposición de un pequeño fragmento de las señales con distorsión original y distorsionada por un perceptrón de estructura [32,16,8,8,1]. El error cuadrático medio cometido al procesar este audio completo fue de 721×10^{-6}.

Una prueba que se realizó con este mismo perceptrón de estructura [32,16,8,8,1], en el mismo estado de aprendizaje[5] que al momento de procesar la señal de la figura 5, fue la de procesar el archivo de audio número 5 el cual *jamás había sido presentado anteriormente al perceptrón*. El resultado obtenido nuevamente fue satisfactorio ya que el audio resultante es prácticamente idéntico al original producido por la distorsión analógica, con un error cuadrático medio de 682×10^{-6}. En la figura 6 se puede ver un fragmento de este resultado. Nuevamente se observan algunas diferencias gráficas, sin embargo a nivel auditivo se puede considerar que el sonido original y el generado por el perceptrón son prácticamente idénticos. Esta prueba demuestra que *el perceptrón fue capaz de distorsionar un audio completamente desconocido y hacerlo de la misma forma en que lo hizo la distorsión analógica.*

Perceptrón [32,16,8,4,2,1] El perceptrón de estructura [32,16,8,4,2,1] demostró ser el de mejores cualidades entre los ensayados. El mismo aprendió a distorsionar únicamente mediante la enseñanza de dos audios, más rápido que todos los demás ensayados. Además presentó la mejor relación señal-ruido y la mejor fidelidad a la distorsión original. En la figura 7 se muestra un fragmento de una señal generada por este perceptrón. Como

[5] Es decir que no se le enseñó ningún otro audio nuevo.

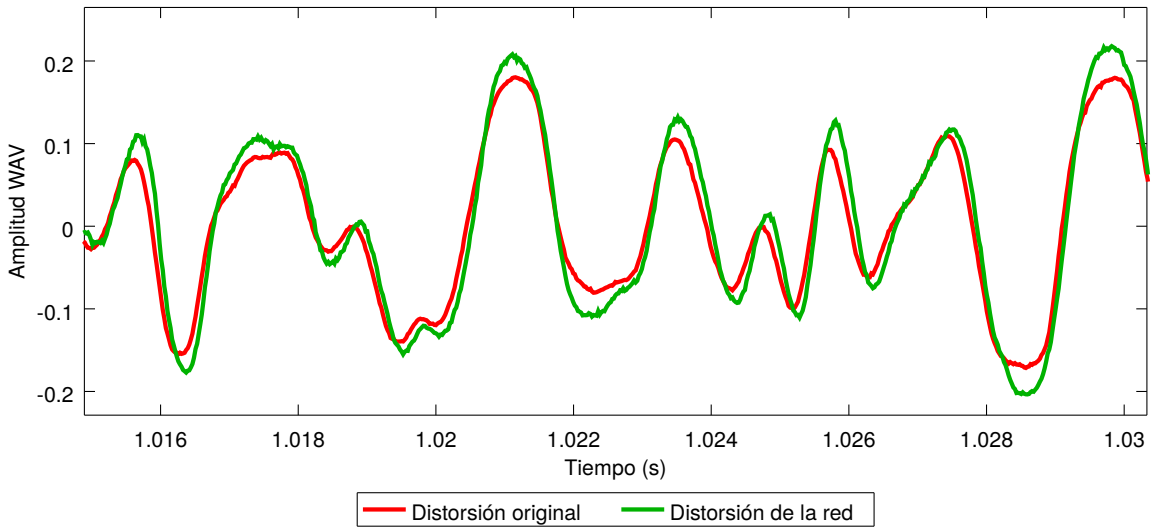

Figura 6: Resultado obtenido al procesar el audio número 5 mediante el perceptrón de estructura [32,16,8,8,1] cuando solo se le habían enseñado los audios 1, 2 y 3 (es decir, *el audio 5 jamás había sido presentado anteriormente*). Se observa que, a pesar de que el audio es completamente nuevo para el perceptrón, ambas señales poseen una alta correlación, *lo cual indica que el mismo ha aprendido a distorsionar audios que nunca se le presentaron anteriormente de una forma similar a la que lo haría la distorsión analógica original.* El error cuadrático medio cometido a lo largo de todo este audio fue de 682×10^{-6}.

puede verse la misma posee una alta similitud con la original. En este caso el error cuadrático medio obtenido fue de 948×10^{-6} que si bien es mayor que en el caso del perceptrón de estructura [32,16,8,8,1] presentado anteriormente, como se dijo anteriormente no refleja en forma exacta la calidad del audio obtenido.

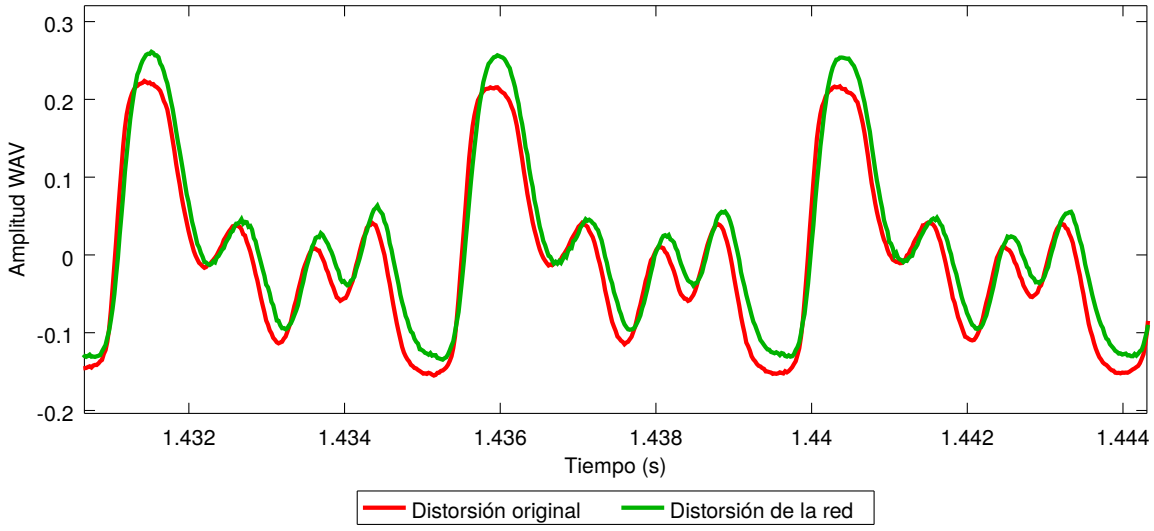

Figura 7: Fragmento de un audio (audio 1) procesado por el perceptrón de estructura [32,16,8,4,2,1]. Al momento de generar esta señal el perceptrón fue entrenado con los audios 1 y 2. El error cuadrático medio cometido en este caso fue de 948×10^{-6}.

3.2. Resultados no exitosos

Las estructuras de tres capas, e.g. [20,10,1], demostraron no ser capaces de dar resultados satisfactorios. Se probaron diversas estructuras de tres capas como por ejemplo [10,20,1], [32,16,1], [64,64,1], etc. En todos los casos los fraseos melódicos de la guitarra fueron perfectamente reconocibles, pero las cualidades del sonido obtenido estaban muy lejos de la distorsión original y además la relación señal-ruido se empobreció considerablemente. Los errores cuadráticos medios obtenidos fueron de $9{,}81 \times 10^{-3}$ para la red [10,20,1], $2{,}73 \times 10^{-3}$ para la red [32,16,1] y $28{,}8 \times 10^{-3}$ para la red [64,64,1], todas habiendo procesado el audio número 1.

Perceptrón [5,5,5,1] En la figura 8 se muestran superpuestas la señal de distorsión original en rojo y la señal obtenida con un perceptrón de estructura [5,5,5,1] en verde. Al momento de generar dicha señal de distorsión la red había sido entrenada con los audios 1, 2 y 3 en su totalidad en dicho orden. El audio procesado al cual corresponden las señales de la figura 8 es el número 1. Si bien la distorsión obtenida puede llegar a tener aplicaciones musicales, en este caso el resultado no se lo consideró satisfactorio ya que la la misma no es similar a la original. Esta diferencia se puede apreciar tanto en el gráfico de la figura citada como también en forma auditiva. La distorsión obtenida a partir del perceptrón posee más componentes de alta frecuencia que la distorsión original y su sonoridad es considerablemente distinta. El error cuadrático medio obtenido en este caso fue de $6{,}31 \times 10^{-3}$.

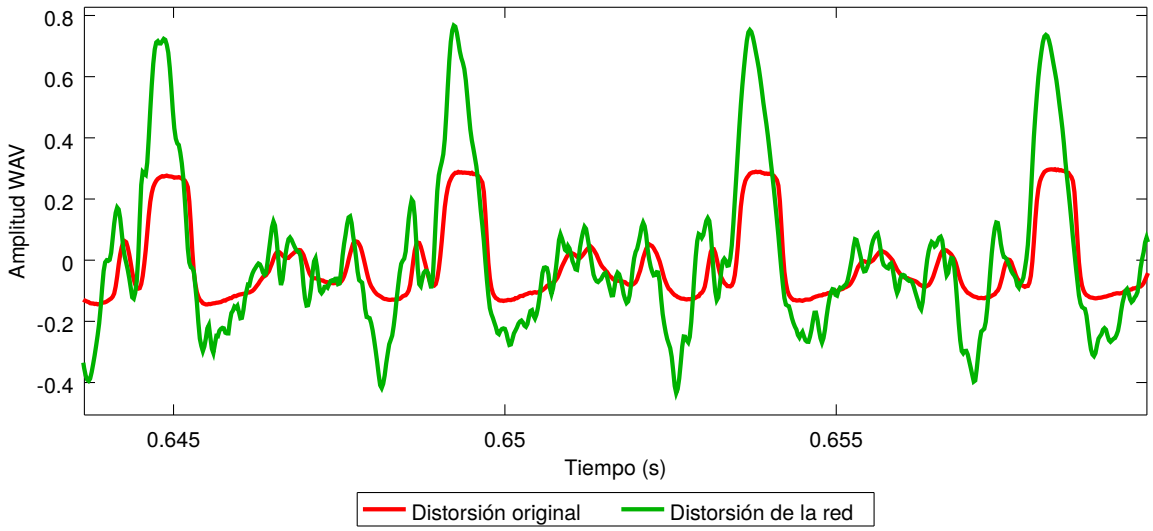

Figura 8: Superposición de un pequeño fragmento de las señales con distorsión original y distorsionada por un perceptrón de estructura [5,5,5,1]. El error cuadrático medio obtenido al procesar este audio fue de $6{,}31 \times 10^{-3}$.

3.3. Ruido de fondo

Como efecto adverso se puede mencionar que los audios procesados por todos los perceptrones ensayados poseen un mayor nivel de ruido de fondo que los audios originales. Según el perceptrón y el estado de aprendizaje del mismo este ruido varía en intensidad y en ancho de banda. En el caso del audio obtenido con el perceptrón de estructura [32,16,8,8,1] es este ruido de fondo el único parámetro que permite diferenciar (auditivamente) el audio original y el obtenido mediante el perceptrón. En la figura 9 se muestra un análisis en el dominio de la frecuencia del audio obtenido con el perceptrón de estructura [32,16,8,8,1]. En la misma se observa el mayor nivel de ruido de fondo para el perceptrón, presente en la banda de 5000 a 15000 Hz. También se ha graficado una FFT del audio limpio (sin distorsión). Puede observarse que la distorsión analógica introdujo ruido de fondo en la banda de 0 a 8 kHz aproximadamente. Este ruido se observa levemente disminuido en el caso del perceptrón pero, como se dijo anteriormente, se incorporó un ruido de mayor intensidad en otra banda de frecuencias donde antes no había haciendo que en forma global el perceptrón genere más ruido que la distorsión original.

En el caso de la red de estructura [32,16,8,4,2,1] se observó que la misma generó audios con un ruido de fondo de menor intensidad, pero así y todo mayor que la distorsión original.

4. Conclusión

Se logró desarrollar exitosamente un perceptrón de estructura completamente arbitraria en lo que respecta a cantidad de capas y cantidad de neuronas por capa, y se obtuvieron ciertas estructuras capaces de aprender de forma bastante precisa la forma de generar una distorsión particular sobre un audio de guitarra eléctrica limpio, i.e. sin distorsión.

En todos los casos ensayados se observó que los perceptrones generan un mayor nivel de ruido de fondo y que el mismo se concentra en distintas bandas en función de la estructura de la red y del estado de entrenamiento de la misma.

La tecnología actual hace que el distorsionado de audio mediante redes neuronales, de la forma en que se lo implementó en este trabajo, sea lo muy lenta como para que no pueda ser implementada para procesado en tiempo real. Con el avance de la tecnología quizás algún día se puedan implementar redes neuronales a nivel de

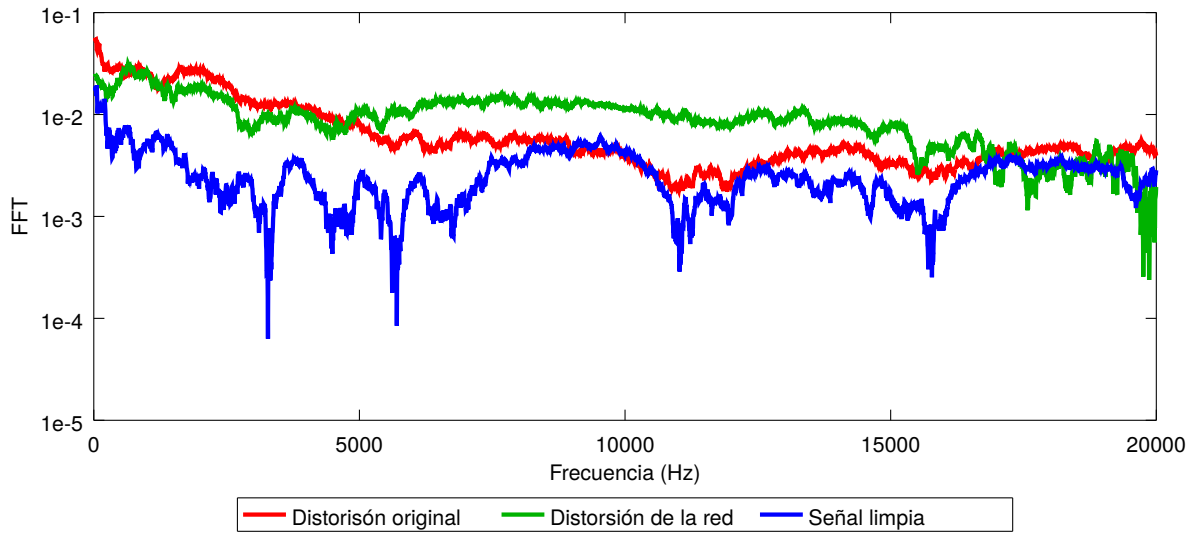

Figura 9: Gráfico en el dominio de la frecuencia del audio (completo) que se muestra en la figura 6. Se observa que el audio generado por la red neuronal posee un mayor nivel de ruido de fondo principalmente en la banda de 5000 a 15000 Hz. Para obtener estos gráficos se aplicó un promediado sobre las FFT calculadas con el fin de quitar ruido gráfico y apreciar el fenómeno que se menciona. Es por esto que no se aprecian los picos de frecuencias correspondientes a las notas musicales ejecutadas en el audio.

hardware y entonces sí se las podría utilizar para procesar audio en tiempo real aprovechando las estructuras aquí desarrolladas.

Referencias

[1] *Introduction to the theory of neural computation*, J. Hertz & A. Krogh & R. G. Palmer.

[2] *Emulating electric guitar effects with neural networks*, David Sanchez Mendoza, Universitat Pompeu Fabra, 2005.