



Eliminación de ruido espectral basado en redes neuronales

Autor: Ignazio F. Finazzi

Tutora: Patricia Jiménez Fernández

25 de julio de 2020

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Introducción

Definición y objetivos

Introducción al Procesamiento de señal

Redes LSTM

Estado del arte

Obtención, procesado y almacenamiento de datos

Datos de Audio

Datos de Ruido

Análisis exploratorio de datos

Análisis de integridad

Análisis de los datos de audio

Diseño e implementación de modelos

Pre-procesado de datos

Modelo de capas LSTM

Análisis de los resultados obtenidos

Conclusiones y propuestas de mejora

Introducción

●○○○

Obtención,
procesado y
almacenamiento
de datos
○○

Análisis
exploratorio de
datos
○○

Diseño e
implementación
de modelos
○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Definición y objetivos

El desarrollo del trabajo propuesto consiste en el diseño y prototipado de un sistema la eliminación de ruido en conversaciones habladas.

- ▶ Velocidad
- ▶ Posición
- ▶ Actitud

Tipos de navegación:

- ▶ Autónoma → no depende de medidas externas ni comunicación con el exterior.
Ejemplo: Sistema de Navegación Inercial (INS)
- ▶ No autónoma → depende de medidas exteriores y/o comunicación con el exterior.
Ejemplo: Sistemas de navegación por satélite (GNSS), radio-ayudas o métodos basados en imágenes, entre otros

Introducción

○●○○

Obtención,
procesado y
almacenamiento
de datos
○○

Análisis
exploratorio de
datos
○○

Diseño e
implementación
de modelos
○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Digitalización

Proceso mediante el cual, una señal analógica $x_a(t)$ continua en tiempo y valores pasa a una señal digital $x_d(t)$ discreta en tiempo y valores.

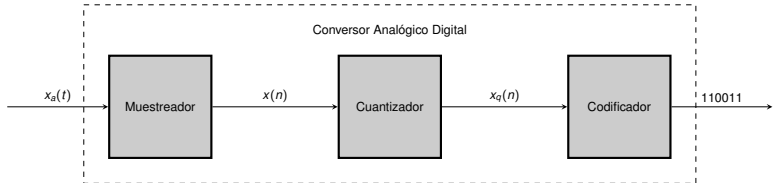


Figura: Esquema de la conversión analógico a digital

Introducción al procesamiento de señal.

Muestreo I

UEMC

4

Introducción

●●○○

Obtención,
procesado y
almacenamiento
de datos
○○

Análisis
exploratorio de
datos
○○

Diseño e
implementación
de modelos
○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Digitalización

Proceso mediante el cual, una señal analógica $x_a(t)$ continua en tiempo y valores pasa a una señal $x(n)$ discreta en tiempo y continua en valores.

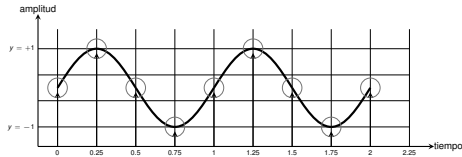


Figura: Esquema del muestreo de una señal de 1Hz muestreada a 4 muestras por segundo

UEMC

Introducción al procesamiento de señal.

Muestreo II

UEMC 5

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

UEMC

Teorema de muestreo de Nyquist-Shannon

Si la frecuencia más alta contenida en una señal analógica $x_a(t)$ es $F_{max} = B$ y la señal se muestrea a una tasa $F_s > 2F_{max} \equiv 2B$, entonces $x_a(t)$ se puede recuperar totalmente a partir de sus muestras mediante la siguiente función de interpolación

$$g(t) = \frac{\sin 2\pi Bt}{2\pi Bt} \quad (1)$$

Así, $x_a(t)$ se puede expresar como:

$$x_a(t) = \sum_{n=-\infty}^{\infty} x_a\left(\frac{n}{F_s}\right) g\left(t - \frac{n}{F_s}\right)$$

donde $x_a\left(\frac{n}{F_s}\right) = x_a(nT) \equiv x(n)$ son las muestras de $x_a(t)$

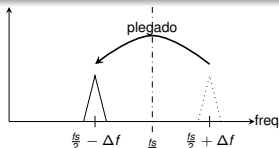
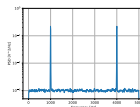
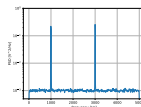


Figura: Esquema del plegado de una señal con aliasing



(a) Densidad espectral de potencia para la suma dos senos de 1kHz y 4kHz muestreados a 10kps



(b) Densidad espectral de potencia para la suma dos senos de 1kHz y 7kHz submuestreados a 10kps

Introducción al procesamiento de señal.

Cuantización

UEMC 6

Introducción

○●○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Digitalización

Proceso mediante el cual, una señal $x(n)$ discreta en tiempo y continua en valores pasa a una señal digital $x_q(n)$ discreta en tiempo y en valores.

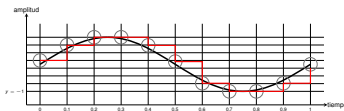


Figura: Esquema de la cuantización con 4 bits a 10 muestras por segundo, i.e., 8 valores posibles

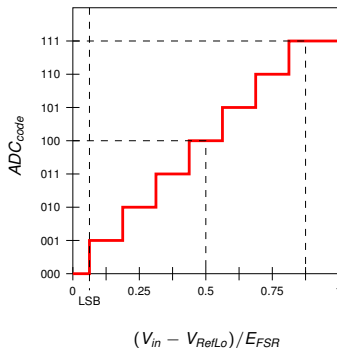


Figura: Resolución del ADC

Redes recurrentes. Celdas LSTM

Introducción

○○●○

Obtención,
procesado y
almacenamiento
de datos
○○

Análisis
exploratorio de
datos
○○

Diseño e
implementación
de modelos
○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

- Redes **recurrentes** → Las redes neuronales recurrentes son aquellas que retienen información de sus estados anteriores y ante los mismos estímulos de entrada no siempre producen las mismas salidas, debido al estado de la celda.

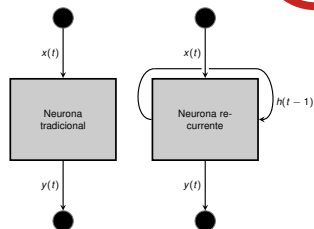


Figura: Comparación de neuronas tradicionales con neuronas recurrentes

- Celdas LSTM (Long Short-Term Memory)
 - Propuestas por Sepp Hochreiter y Jürgen Schmidhuber en 1997.
 - Evitan el problema de las dependencias a largo plazo.

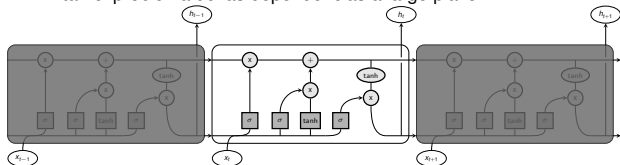


Figura: Celdas LSTM

Introducción

○○●

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

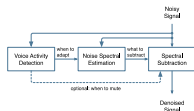
Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

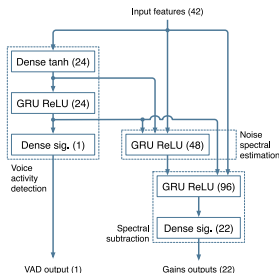
Algoritmos dominio público



(a) Estructura del algoritmo presentado por Jean-Marc Valin



(b) Diagrama de bloques del eliminación de ruido en el dominio de la frecuencia



(c) Red neuronal de RNNNoise

Algoritmos privativos

► RTX Voice

- Producto de NVIDIA
- Gratuito
- Plugins para algunas aplicaciones de chat por voz
- Necesita una gráfica NVIDIA RTX

► Krisp

- Producto de Krisp
- De pago (gratuito hasta 120 $\frac{\text{min}}{\text{semana}}$)
- Funciona con el dispositivo de audio del sistema operativo, luego funciona con todo
- No necesita hardware dedicado

Obtención, procesamiento y almacenamiento de datos. Pistas de Audio I

UEMC

9

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

●○

Análisis
exploratorio de
datos

○○




Diseño e
implementación
de modelos

○○






Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Extracción del audios de videos de Youtube

- ▶  Gran cantidad fuente de información
- ▶  Vídeos de todo tipo que pueden contener ruido
- ▶  Una descarga masiva requiere una revisión de todos los audios obtenidos

Generación sintética

- ▶ A partir de audios existentes
 - ▶  En audio no tiene sentido. Se usa en entorno de imágenes donde éstas se giran o se ordenan las columnas en orden inverso (espejo)
- ▶ A partir de texto (Text to speech)
 - ▶  Fuente infinita
 - ▶  Conocimiento de lo que se dice (se genera a partir de texto)
 - ▶  Sintetizadores tradicionales generan voces metálicas
 - ▶  Sintetizadores basados en NN costosos de entrenar (Real Time Voice Cloning)

UEMC

Obtención, procesamiento y almacenamiento de datos. Pistas de Audio II

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

●○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Descarga masiva de audiolibros

Técnica de web scrapping:

- Exploración visual del sitio web.
- Análisis de la URL. En muchas ocasiones, parte de la información que el sitio web devuelve viene filtrada mediante una serie de filtros que se definen en la URL.
- Análisis del sitio web mediante herramientas de desarrollador. Consiste analizar el sitio web para encontrar dónde está la información que se quiere extraer y cómo viene en el código de la página. Para esto se usa un navegador web y se analiza el código mediante el inspector del navegador. Esta herramienta resalta cada una de las partes del código de la página web que se corresponden con las partes visuales de la misma.
- Extracción y almacenamiento de la información.

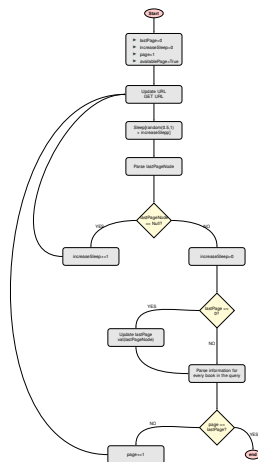


Figura: Diagrama de flujo para el scraper de LibriVox

Obtención, procesamiento y almacenamiento de datos. Pistas de Ruido

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○●

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Descarga de videos con ruido a partir de YouTube y extracción de audio

- ▶ Ruido de ciudad
- ▶ Ruido de lluvia
- ▶ Ruido de murmullo
- ▶ Librería *youtube_dl*
- ▶ Almacenamiento en base de datos de toda la información de las pistas

Análisis exploratorio de datos. Integridad

UEMC 12

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

●○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Nombre	Tipo de dato	Descripción
id	int	Identificador de la pista de audio
book_name_dummy	text	Nombre del libro con caracteres ASCII reducidos
book_name	text	Nombre completo del libro
book_author	text	Autor del libro
book_url	text	URL de descarga del libro
book_language	text	Lenguaje del libro
book_path	text	Dirección de almacenamiento del libro comprimido
book_n_tracks	int	Número de pistas del libro
track_name	text	Nombre de la pista de audio
track_path	text	Directorio de almacenamiento de la pista de audio
track_channels	int	Número de canales de la pista de audio
track_sample_rate	int	Tasa de muestreo de la pista de audio
track_duration	real	Duración de la pista de audio
track_status	text	Estado del archivo (OK → descargado, DELETED → eliminado)
track_insert_datetime	int	Fecha y hora de inserción del registro

Tipo de pista	Duración [horas]	Tasas de muestreo [muestras segundo]
Audiolibros	707.70 (555.73)	[22050]
Ruido	52.91	[44100, 48000]

Cuadro: Duración total de todas las pistas

Cuadro: Columnas de la tabla pista de los audiolibros

id	book_name	track_name
1351	Antología de Cuentos Fantásticos	antologiacuentosfantasticos_01_various_64kb.mp3
3410	Coppelius (1ra Parte) (in Antología de Cuentos Fantásticos)	antologiacuentosfantasticos_01_various_64kb.mp3
3935	De lo que aconteció a un déan de Santiago con don Illan el mágico, que moraba en Toledo (in Antología de Cuentos Fantásticos)	antologiacuentosfantasticos_01_various_64kb.mp3

Cuadro: Ejemplo de registros repetidos

UEMC

Análisis exploratorio de datos. Audio.

Dominio de la Frecuencia

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○●

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

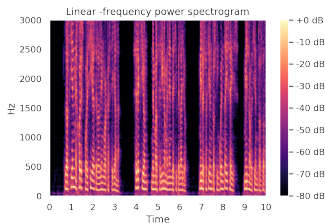


Figura: Espectrograma de la voz humana

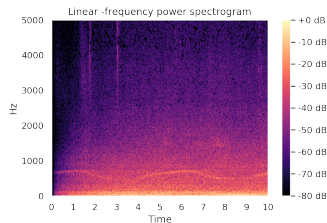


Figura: Espectrograma del ruido

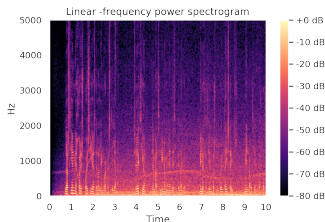


Figura: Espectrograma de las pistas combinadas

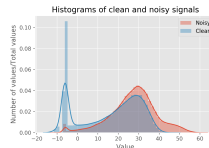


Figura: Espectrograma de las pistas combinadas

Análisis exploratorio de datos. Audio.

Dominio del Tiempo

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

●●

Diseño e
implementación
de modelos

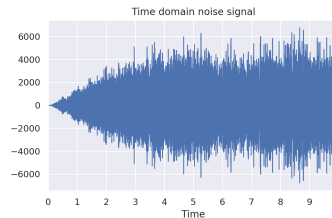
○○

Análisis de los
resultados
obtenidos

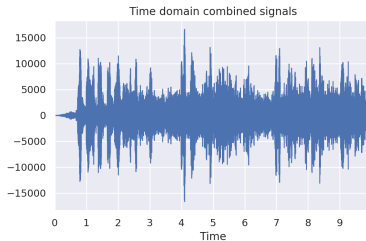
Conclusiones y
propuestas de
mejora



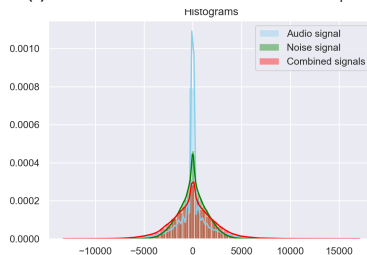
(a) Forma de la señal de audio en el dominio del tiempo



(b) Forma de la señal de ruido en el dominio del tiempo



(a) Forma de la señal combinada de audio y ruido en el dominio del tiempo



(b) Histogramas superpuestos de las tres señales

Análisis exploratorio de datos. Audio.

Conclusiones

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

●●

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

- ▶ **La voz humana no es continua en el tiempo** (la persona se para para respirar mientras habla) lo que crea un espectro seccionado en el tiempo. Por el contrario el **ruido sí es continuo en el tiempo**.
- ▶ El **espectro de la voz humana está acotado en frecuencia** mientras que el espectro del ruido, en general, no. Como implicación directa tiene que para analizar el audio en el dominio de la frecuencia se puede acotar la tasa de muestreo a la banda donde la voz humana se encuentra, ahorrando mucho procesado.
- ▶ La **voz humana** se caracteriza por tener unas **frecuencias características** que varían en el tiempo y están relacionadas entre sí (tono fundamental y armónicos). Por el contrario el ruido, no presenta relación entre sus componentes espectrales.
- ▶ Tras la combinación de los audios, si las amplitudes del ruido no son lo suficientemente grandes, las componentes frecuenciales de la voz siguen siendo predominantes.

Diseño e implementación de modelos.

Preparación de los datos I

UEMC 16

Introducción
○○○○

Obtención,
procesado y
almacenamiento
de datos
○○

Análisis
exploratorio de
datos
○○

Diseño e
implementación
de modelos
●○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

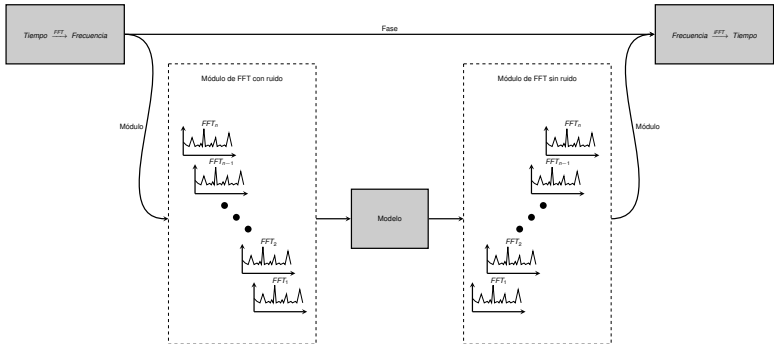


Figura: Esquema de las entradas y salidas de datos del modelo

Diseño e implementación de modelos.

Preparación de los datos II

UEMC 17

Introducción
○○○○

Obtención,
procesado y
almacenamiento
de datos
○○

Análisis
exploratorio de
datos
○○

Diseño e
implementación
de modelos
●○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

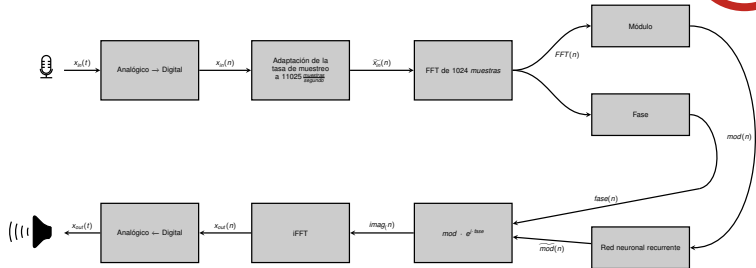


Figura: Cadena completa de procesamiento de señal

- Cálculo de las FFTs:
 - Tasa de muestreo: 11025 sps
 - Tamaño FFT: 1024 muestras (513 bandas de 10.74Hz)
 - Overlap: 50 %



- Enventanado Hann
- Almacenado en HDF5

UEMC

Diseño e implementación de modelos.

Preparación de los datos III

UEMC 18

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

●○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

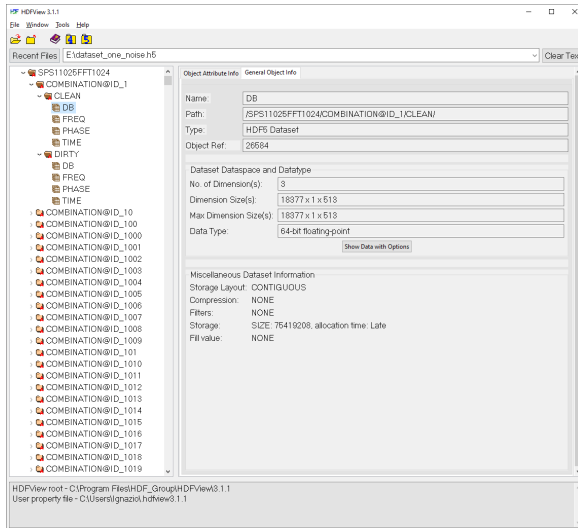


Figura: Estructura de los datos almacenados en el archivo HDF5

UEMC

Diseño e implementación de modelos.

Modelo de capas LSTM I

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○●

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Tipos de arquitectura

- ▶ **One to one** → clasificación de imágenes, una entrada (matriz de la imagen), una salida, la categoría.
- ▶ **One to many** → generación de títulos para imágenes, una entrada (matriz de la imagen), varias salidas, las diferentes palabras.
- ▶ **Many to one** → análisis de sentimiento, entran varias palabras, se clasifica como positiva o negativa.
- ▶ **Many to many** → traducción de texto, varias palabras en un idioma a la entrada y varias a la salida en otro idioma.
- ▶ **Many to many** → entradas y salidas sincronizadas, clasificación de vídeo.

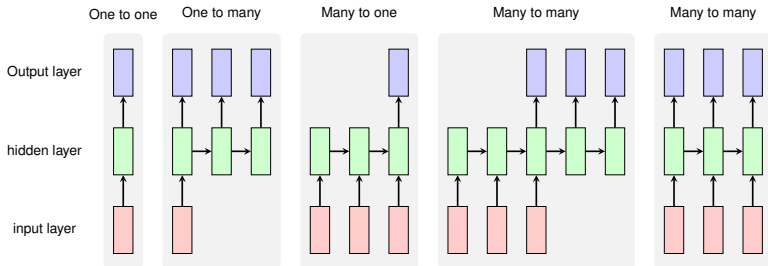


Figura: Arquitecturas de modelo

Diseño e implementación de modelos.

Modelo de capas LSTM II

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○●

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 1, 512)	1562624
lstm_1 (LSTM)	(None, 1, 512)	2099200
lstm_2 (LSTM)	(None, 512)	2099200
dense (Dense)	(None, 250)	128250

Total params: 5,889,274
Trainable params: 5,889,274
Non-trainable params: 0

Listing 1: Resumen del modelo

Diseño e implementación de modelos.

Modelo de capas LSTM III

UEMC 21

Introducción
○○○

Obtención,
procesado y
almacenamiento
de datos
○○

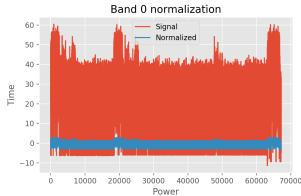
Análisis
exploratorio de
datos
○○

Diseño e
implementación
de modelos
○●

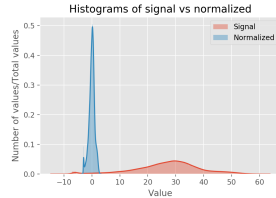
Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

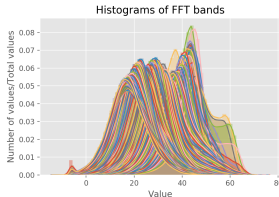
Debido al gran número de datos, **294 Gbytes**, se debe entrenar con un generador de secuencias, además los datos van a ser normalizados



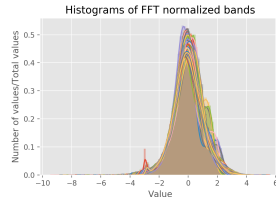
(a) Comparación de la señal con su normalización en tiempo para la banda cero de un mismo audio con ruido



(b) Comparación de los histogramas de la señal y su normalización para la banda cero de un mismo audio con ruido



(a) Histograma de todas las bandas



(b) Histograma de todas las bandas normalizadas

Análisis de los resultados obtenidos I

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

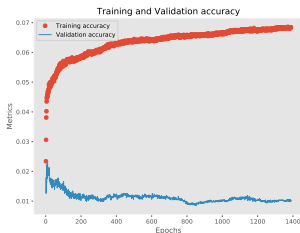
Diseño e
implementación
de modelos

○○

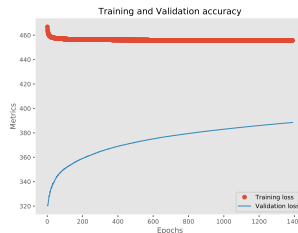
Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

- **Tamaño de batch** → 32
- **Número de FFT**
 - Entrenamiento → 63132
 - Validación → 4068
- **Número de épocas** → 2000¹
- **Tasa de aprendizaje** → Decreciente en el tiempo.



(a) Precisión en entrenamiento y validación



(b) Pérdidas en entrenamiento y validación

¹ El entrenamiento se paró en 1390 al comprobar que los resultados empeoraban

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

```
-----  
Layer (type)                Output Shape  
Param #  
-----  
1stm (LSTM)                 (None, 2048)  
18833408  
-----  
dense (Dense)               (None, 250)  
512250  
-----  
Total params: 19,345,658  
Trainable params: 19,345,658  
Non-trainable params: 0  
-----
```

Listing 2: Resumen del modelo mono-capa

- ▶ **Tamaño de batch** → 128
- ▶ **Número de FFT**
 - ▶ Entrenamiento → 55657
 - ▶ Validación → 11543
- ▶ **Número de épocas** → 300
- ▶ **Tasa de aprendizaje** → Decreciente en el tiempo.

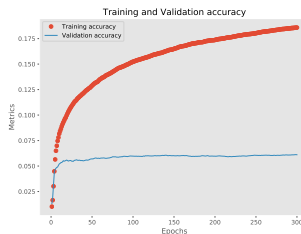


Figura: Precisión en entrenamiento y validación

Conclusiones y propuestas de mejora.

Conclusiones

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

- ▶ Hardware AMD Radeon empleado.
 - ▶ ✓ Framework de código abierto y mantenido por la comunidad.
 - ▶ ✓ AMD Radeon 570x al 100 % durante más de 48 horas seguidas.
 - ▶ ✗ Desempeño por debajo de NVIDIA.
- ▶ Métodos de scraper
 - ▶ ✓ Muy potentes y ahorran mucho tiempo.
 - ▶ ✗ Ética cuestionable y potencialmente peligrosos.
- ▶ Estructuración de los metadatos en base de datos.
 - ▶ ✓ Permite localizar fallos complejos de encontrar.
 - ▶ ✓ Ahorra tiempo en explotación.
 - ▶ ✗ Requiere tiempo y planificación del modelo de datos.
- ▶ Análisis en el dominio de la frecuencia y con redes LSTM
 - ▶ ✓ Resultados esperanzadores en la red monocapa.
 - ▶ ✗ Problema complejo.

Conclusiones y propuestas de mejora.

Propuestas de mejora

Introducción

○○○○

Obtención,
procesado y
almacenamiento
de datos

○○

Análisis
exploratorio de
datos

○○

Diseño e
implementación
de modelos

○○

Análisis de los
resultados
obtenidos

Conclusiones y
propuestas de
mejora

- ▶ **Crear un mecanismo de detección activa de voz.** Con este mecanismo automáticamente se elimina todo el espectro, o se le aplica una ganancia negativa muy alta, a las partes en las que no se detecte voz. De esta manera se elimina mucho ruido sin necesidad de llegar a la red neuronal.
- ▶ **Detección de la frecuencia principal y sus armónicos (pitch frequency).** Detectar la frecuencia fundamental y sus armónicos y pre-limpiar el espectro a la entrada del algoritmo.
- ▶ **Trabajar sobre la normalización de los datos.** Este es uno de los puntos con mayor influencia sobre los resultados. Los resultados de la red variaban mucho con datos normalizados o sin normalización. Se debe plantear una normalización que sea posible aplicarla en tiempo real, es decir sólo se conoce el valor de las muestras actuales y las pasadas, no las futuras como en entrenamiento.
- ▶ **Entrenar el último modelo con mayor número de datos.** Dados los resultados obtenidos, el primer punto a tratar sería entrenar el modelo de mayor precisión con más datos ayudándose del generador de secuencias.



GRACIAS POR SU
ATENCIÓN