



UNIVERSIDAD DE ESPECIALIDADES ESPÍRITU SANTO

Facultad de Ingeniería

Escuela de Computación y Telecomunicaciones

**REVISIÓN ANALÍTICA DE 3 ALGORITMOS DE RECONOCIMIENTO DE
CONTEXTO DE RUIDO PARA LA DEFINICIÓN DE UNA PROPUESTA DE UN
ALGORITMO EMPÍRICO.**

Trabajo de Titulación que se presenta como requisito para el título de Ingeniero en
Sistemas

Autor: Iván Andrés López Cazón

Tutor: Washington Antonio Caraguay Ambuludi

Samborondón, Septiembre, 2020



APROBACIÓN DEL TUTOR

En mi calidad de Tutor del estudiante Iván Andrés López Cazón, que cursa estudios en el programa de TERCER nivel: Ingeniería en Sistemas, dictado en la Facultad de Sistemas, Telecomunicaciones y Electrónica de la UEES, en modalidad presencial.

CERTIFICO

Que he revisado el Trabajo de Titulación denominado: “Revisión Analítica de 3 algoritmos de reconocimiento de contexto de ruido para la definición de una propuesta de un algoritmo empírico.”, presentado por el estudiante Iván Andrés López Cazón, como requisito previo para optar por el Grado Académico de Ingeniero en Sistemas CERTIFICO que el Trabajo de Titulación ha sido analizado y reúne todos los requisitos para ser presentado y sometido a los procesos de revisión estipulados por la Facultad.

Atte.

Washington Antonio Caraguay Ambuludi
0703527465

Revisión Analítica de 3 algoritmos de reconocimiento de contexto de ruido para la definición de una propuesta de un algoritmo empírico.

Analytical review of 3 noise context recognition algorithms for the definition of a proposal for an empirical algorithm.

Iván Andrés López Cazón¹

Resumen

El presente trabajo de titulación tiene como objetivo principal de proponer un algoritmo de reconocimiento de contexto para ruidos de fondo mediante un análisis comparativo de algoritmos desarrollados por otros investigadores, donde se verá reflejado las técnicas de reconocimiento de sonidos no-habla, desde un punto de vista práctico y funcional. La metodología utilizada para el desarrollo de la presente investigación es de carácter exploratorio, cualitativo y aplicado, para lo cual fue necesario aplicar una estrategia con base en la recopilación de información necesaria para el desarrollo de un algoritmo. Se revisaron cuatro técnicas de reducción de ruido de fondo: el filtro adaptable RACE (siglas en inglés de realzado adaptable por correlación), la sustracción espectral, el filtrado por medio de GSVD (siglas en inglés de descomposición en valores singulares generalizada) y el análisis de componentes independientes. Lo primero que se realizó fue descomponer la señal de audio en un solo canal, posteriormente mediante funciones se procedió a la detección del ruido de la señal mediante el uso de la transformada rápida de Fourier, los filtros para DSP (siglas en inglés de procesamiento digital de señales) y finalmente se logró la separación del ruido de fondo.

Palabras clave:

Algoritmo, desarrollo, auditivo, sonido, discapacidad auditiva, tecnología asistiva, análisis de ruido, filtrado de sonido, Python, análisis comparativo, audición.

Abstract

The main objective of the present degree work is to propose a context recognition algorithm for background noises through a comparative analysis of algorithms developed by other researchers, where non-speech sound recognition techniques will be reflected from a point of view. practical and functional sight. The methodology used for the development of this research is exploratory, qualitative and applied, for which it was necessary to apply a strategy based on the collection of information necessary for the development of an algorithm. Four background noise reduction techniques were reviewed: the adaptive RACE (correlation adaptive enhancement) filter, spectral subtraction, GSVD (generalized singular value decomposition) filtering, and generalized singular value decomposition. of independent components. The first thing that was done was to decompose the audio signal into a single channel, then through functions, the noise of the signal was detected by using the fast Fourier transform, the filters for DSP (digital signal processing) and finally the separation of the background noise was achieved.

Keywords

Algorithm, development, auditory, sound, hearing impairment, assistive technology, noise analysis, sound filtering, Python, comparative analysis, hearing.

¹ Estudiante en Facultad de Ingeniería y Escuela de Sistemas y Telecomunicaciones, Universidad Espíritu Santo – Ecuador. E-mail ivalopez@uees.edu.ec.

Introducción

El oído es uno de los sentidos más importantes del ser humano. Después de la vista, que es el sentido que más información aporta sobre el medio en el que nos movemos, podemos considerar el oído como la herramienta perceptiva humana con más trascendencia en cualquier actividad diaria. El oído humano es capaz de distinguir un gran número de sonidos ambientales con tan solo prestar un poco de atención a nuestro entorno. Estos sonidos pueden ser entendidos y manejados como una gran cantidad de datos de información que son transmitidos por un medio, forman parte del proceso de comunicación entre el entorno y las personas que reciben los mensajes (Levratto, 2017).

Pero... ¿Qué sucede cuando esta capacidad de comunicación no existe o está muy limitada? Las personas con discapacidad auditiva son un claro ejemplo de ello, sufren cada día los inconvenientes que la imposibilidad de detectar e identificar sonidos conlleva. Diferenciar sonidos del entorno es una capacidad inherente de las personas, y carecer de ella puede suponer muchos problemas en la cotidianidad de la vida. Una llamada telefónica o una alarma de incendios a horas intempestivas avisando de un peligro inminente, pueden ser concebidas como sonidos de un orden prioritario a detectar. El no hacerlo, no solo podría poner en riesgo a la persona en cuestión, sino también puede significar un permanente estado de preocupación e inseguridad, lo cual también es perjudicial para la salud del individuo (Olaya, 2018).

Desarrollar un sistema capaz de actuar con la suficiente inteligencia que permita percibir y tomar las decisiones adecuadas como si de un ser humano se tratase, ha sido y sigue siendo en la actualidad una de las metas más grandes para los investigadores (Pico & Lindao, 2019). En un sistema de este tipo, la tecnología de base son las técnicas de reconocimiento de eventos acústicos, sin embargo, los esfuerzos actuales de investigación en detección e identificación de señales sonoras han estado en su gran mayoría en el campo del reconocimiento del habla, dejando a un lado todo lo referente a otros sonidos producidos en el entorno. No obstante, la concienciación social y la búsqueda de la aplicabilidad, cada vez más

exigida en la investigación, hacen que esta tendencia este cambiando constantemente, surgiendo nuevos proyectos orientados al reconocimiento de este tipo de sonidos, pero siempre alejados del mercado al cual estamos aplicando y de un producto robusto y consolidado que este orientado a la comunidad de personas con discapacidad auditiva (Rasgado & Mereles, 2017).

Es por esto que el desarrollo de un sistema capaz de detectar y clasificar de forma automática los diversos sonidos de interés que puedan surgir en una vivienda, en la oficina, en el centro de estudio, en la cotidianidad de la vida, etc., se concibe como una necesidad primordial para el colectivo de personas con disfunciones sensoriales auditivas que contribuirá a la mejora de su calidad de vida (Carranza & Alzamora, 2017).

La finalidad del trabajo de titulación es desarrollar un algoritmo de reconocimiento de contexto para ruidos del fondo, basado en los diferentes estudios realizados por otros investigadores contribuyendo con el estado del arte del desarrollo de tecnologías asistivas con miras a mejorar la calidad de vida del colectivo de personas con discapacidad auditiva, como el desarrollo de una interfaz visual con detalles de los ruidos identificados, mediante el estudio de las técnicas de reconocimiento de sonidos no-habla, desde un punto de vista práctico y funcional. El lenguaje a utilizar para el desarrollo del algoritmo es Python.

MARCO TEÓRICO

Existe mucha información acerca de artículos de reconocimiento de ruido ambiental y acústico, sobre todo existen herramientas de análisis y reconocimiento, pero no ofrecen el conocimiento sobre los algoritmos que están contruidos.

Por lo tanto, se documenta sobre la tecnología del habla y las últimas técnicas utilizadas en el procesamiento y reconocimiento de voz.

Sonidos no-habla

Desde un punto de vista físico, el sonido es un fenómeno vibratorio transmitido en forma de ondas. Las vibraciones pueden ser transmitidas a través de diferentes medios elásticos. Entre los más comunes se encuentran el aire y el agua (Bermúdez, 2016).

Cuando se hace mención a sonidos no-habla, se está refiriendo a aquellos sonidos audibles, no generados por el aparato fonador de un ser humano, que no tengan una construcción lógica; como palabras o frases. Uno de los mayores problemas de este tipo de sonidos es que no están acotados frecuentemente, como lo está la voz o la música en general, este tipo de sonidos no solo abarcan el espacio disponible de la voz y de la música sino también el restante hasta alcanzar el umbral de audibilidad.

Existen pocos intentos para definir el término: sonido no-habla. VanDerveer (1979), define cuatro puntos generales que pueden ayudar a identificarlos:

- a) Son producidos por eventos reales.
- b) Tienen un significado en virtud de hechos casuales.
- c) Son más complejos en su composición que los sonidos generados en laboratorios como tonos puros.
- d) No son parte de un sistema de comunicación como el habla.

Sin embargo, estos puntos excluyen sonidos como: timbres de puerta, teléfonos, etc., y se centra más en sonidos puramente ambientales.

Diferentes estudios se contradicen en la forma en la que el ser humano es capaz de percibir los sonidos no-habla. VanDerveer (1979), indican como los seres humanos perciben los sonidos no-habla de forma equivalente al habla. (p. 36).

Estudios posteriores en psico acústica, afirman que mientras el hemisferio izquierdo (oído derecho) es el encargado de procesamiento de sonidos del habla, es el hemisferio derecho (oído izquierdo), el que controla el procesamiento de sonidos no-habla. Esto perfectamente consiste con el talento del lado derecho del cerebro en el reconocimiento de patrones, en asociar inmediatamente un nuevo estímulo a una experiencia previa almacenada en memoria, sin primero tener que descifrar el contexto asociado a ese sonido. Un sonido no-habla es un evento y el lado derecho del cerebro lo reconoce como tal, muy diferente a una sílaba al azar en una secuencia determinada de sonidos del habla, que es un mero significativo, en relación a otros significantes (Kane, 2004).

Además, si el conjunto a reconocer, no está bien acotado, el reconocimiento de estos sonidos, puede ser en muchos casos complicado a realizar. Investigaciones sobre percepción sonora apuntan que al igual que con el reconocimiento automático del habla, existen sonidos homónimos (knigth y nighth para el caso del habla) que no pueden ser determinados sin el contexto de otros sonidos (Cowling, 2004). Si, por ejemplo, se produce un sonido metálico precedido de un chirrido, la interpretación semántica podría ser la de un choque de coche, entendiéndose ese chirrido como el sonido de los neumáticos al derrapar por la carretera. Por otra parte, si el mismo sonido metálico se combina con goteo de agua y ráfagas de aire, la interpretación podría ser la del ruido de una maquina en una fábrica.

A pesar de que no existe ningún tipo de semántica que ayude a la clasificación de estos sonidos, estudios realizados en el ámbito de reconocimiento demuestran que, trabajando con un conjunto limitado de clases, ciertos parámetros o características de una señal dan información suficiente para que esta pueda ser clasificada de una manera correcta, incluso cuando la fuente del sonido que se produzca, pueda estar en movimiento (Covreur, 1997). Para continuar con la revisión bibliográfica en contexto, es necesario conocer algunos parámetros relevantes en lo que concierne a señales de audio. A continuación, su descripción:

Sonido

Sensación percibida por el oído producida por rápidas fluctuaciones de la presión del aire. (Blanco Á. , 2017)

Propiedades:

- **Velocidad del sonido:** La velocidad del sonido indica la distancia viajada por las ondas sonoras en unidad de tiempo. (Linares, 2017).
- **Presión acústica:** La presión acústica (p) es la raíz media cuadrática de la variación periódica de la presión en el medio en que se propaga la onda sonora. Aproximadamente, según la sensibilidad de las personas, su presión acústica varía entre 2×10^{-5} y; 2×10^4 [Pascal] (Dolores Zuñiga, 2004).

Fuente: (Gil, 2018)

- **Frecuencia:** La frecuencia de un sonido u onda sonora expresa el número de vibraciones por segundo medidos en Herzt (Hz) (Ortiz, 2016).
- **Longitud de onda:** La longitud de onda describe las características de una onda en el espacio. Es la distancia entre dos puntos análogos de una onda a una frecuencia (Sepúlveda, 2019).
- **Decibel:** Es la unidad que refleja la presión y la intensidad acústicas, y como herramienta matemática simplifica los valores que tomen éstas en la escala logarítmica del sonido (Hernández, 2019).

Caracterización de Ruido

Según la Organización Mundial de la Salud (O.M.S) es todo sonido indeseable, generalmente de carácter aleatorio, que no presenta componentes de frecuencia bien definidos. (Muñoz, 2016). El ruido afecta al sonido produciendo una serie de perturbaciones como el siseo o hiss (ruido impulsivo o click), producido por las ralladuras de un disco CD (Gil, 2018).

- **Ruido de banda ancha (hiss)**

El hiss o siseo aparece por la fricción de dos superficies como la aguja de un tocadiscos y el disco de vinilo. Se asemeja mucho al ruido blanco por la repartición equitativa de la energía en todas las frecuencias audibles como se muestra en la figura 1, por lo que la gráfica de su densidad espectral de potencia es prácticamente plana en todos los instantes de tiempo, aunque disminuye ligeramente en las frecuencias altas (Gil, 2018). En la figura 1 se expone a continuación, la estimación de la densidad espectral de potencia del ruido producido por el siseo.

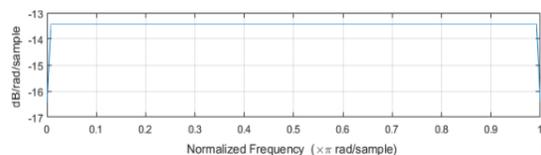


Figura 1. Estimación de la densidad espectral de potencia del ruido producido por el siseo

- **Ruido impulsivo (click y dust)**

Aparecen en la reproducción en los discos reproductores físicos y causan el efecto click en la lectura (Cuevas, 2017).

La densidad espectral de potencia del ruido producido por el efecto click, guarda un parecido con la densidad espectral de potencia de una delta como se muestra en la figura 2:

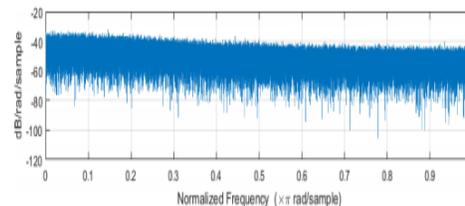


Figura 2. Densidad espectral de potencia cuando se produce un click de máxima amplitud
Fuente: (Cuevas, 2017).

- **Ruido de baja frecuencia (hum)**

El ruido hum, a diferencia de los anteriores, es un ruido dominante en baja frecuencia, información que se puede observar en una gráfica como lo muestra la figura 3 (González, 2019):

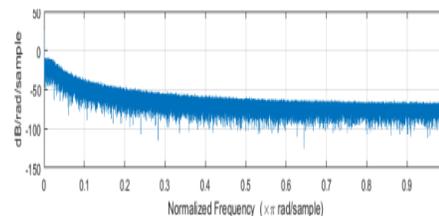


Figura 3. Estimación de la densidad espectral de potencia del ruido hum
Fuente: (González, 2019)

Puerta del ruido en el dominio del tiempo

Este efecto es una puerta de ruido simple: las muestras de una señal de audio cuya amplitud no supere al del umbral se atenúan o se eliminan. A pesar de lo sencillo de construir o programarlo hay dos problemas fundamentales:

- No elimina el ruido en toda la grabación, sino solo en aquellas partes que apenas en audible, como silencios, descansos de una canción.
- No hay diferencia de ruido y partes de la señal de audio con amplitud baja

$$E[k] = |X[k]|^2 \quad (3)$$

(García L., 2019). A continuación se muestra en la figura 4 la onda de reacción en lo que respecta al Fragmento de una canción donde en varias partes suenan los hi-hat a bajo volumen.

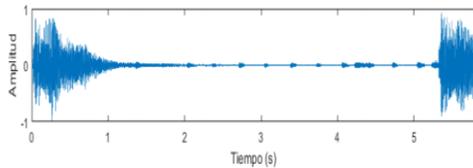


Figura 4. Fragmento de una canción donde en varias partes suenan los hi-hat a bajo volumen
Fuente: (García L., 2019).

Puerta del ruido en el dominio de la frecuencia

La diferencia con el dominio del tiempo es que puede detectar con precisión aquellas partes de la canción donde exista ruido y se puede distinguir de otros tipos de sonidos, ya que realiza un análisis por bandas de frecuencia (Perez, 2019). Este ejemplo se ilustra en la figura 5, que se expone a continuación.

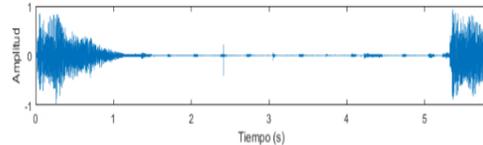


Figura 5. Mismo fragmento de audio que el anterior pero aplicado la reducción de ruido por bandas de frecuencia
Fuente: (Perez, 2019).

Enventanado

Es una técnica que se utiliza en procesamiento de señales para evitar discontinuidades en los bloques analizados.

- Se denomina ventana a un tipo de funciones reales positivas que toman valor nulo fuera del intervalo elegido y que comparten una serie de características comunes (Blasina, 2019).

- **Ventana de Hamming**

Se trata de una versión optimizada de la ventana de Hann, de modo que se minimice el nivel del primer lóbulo lateral (Donoso, 2018):

$$w(n) = \begin{cases} 0.54 - 0.46 \cos \left(\frac{2\pi n}{N-1} \right); & 0 \leq n < N \\ 0; & \text{resto} \end{cases} \quad (1)$$

Donde:

W = hamming (L) devuelve una ventana Hamming simétrica de puntos. L

W = hamming (L, sflag) devuelve una ventana Hamming utilizando el muestreo de ventana especificado por .sflag

Transformada de Fourier de tiempo reducido (STFT)

Se define como la transformada de Fourier de la señal enventanada. Se trata de una secuencia de Transformada de Fourier Discreta (DFT, siglas en inglés) para diferentes posiciones de la ventana $w[n - m]$ (Gilibert, 2019).

$$S[k] = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}} x[n] w[n - m] W_N^{nk}, k = 0, 1, \dots, N - 1 \quad (2)$$

Donde es el instante de tiempo en que se analiza $x[n]$, $w[n - m]$ es la ventana y N es el tamaño de la FFT y de la ventana.

El análisis localizado permite extraer características dinámicas de la señal de forma eficaz, ya que analiza sus partes de forma independiente.

Reducción de ruido

La reducción de ruido se realiza comparando la energía de la señal del audio con la del ruido en cada banda. La energía de una banda de frecuencia se calcula como (Ylurdoz, 2016):

Donde $X[k]$ es la transformada de Fourier Discreta de la señal $x[n]$ en la banda k .

El análisis sobre el ruido se ha realizado de igual manera que el análisis de la señal hasta este punto: enventanado, seguido del cálculo del algoritmo Transformada rápida de Fourier

(FFT), y almacenamiento del valor de la energía (Torres, 2018).

- Los resultados para cada tipo de ruido se encuentran guardados en el servidor y no es necesario analizarlo cada vez que se aplica la reducción (Guerrón, 2018).

El proceso de reducción de ruido para cada banda de frecuencia de señal se muestra en la figura 6

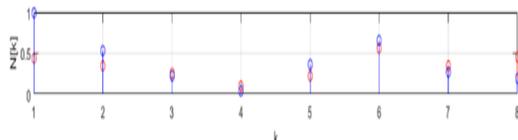


Figura 6. DFT de una ventana de tamaño 8
Fuente: (Guerrón, 2018).

El proceso de reducción de ruido para cada banda de frecuencia de ruido se muestra en la figura 7 donde se compara su energía (en términos computacionales y se compara el nivel de señal en la figura 8:

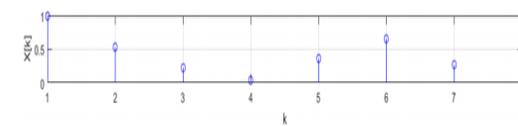
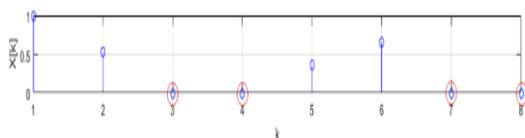


Figura 7. En señal de color rojo es ruido



Fuente: (Guerrón, 2018).

Figura 8. Señal tras aplicar eliminación de ruido
Fuente: (Guerrón, 2018).

Algoritmos para reducción de ruido de fondo

Quispe (2018), hacen hincapié en la falta de un alfabeto fonético de sonidos no-habla. Concluye que esto es debido a que el habla es producida por un conjunto limitado de diferentes acciones por los mecanismos de fonación de los seres humanos, y en cambio, los sonidos no-habla pueden ser producidos desde un amplio rango de fuentes, un sonido no-habla, puede ser producido por el agua de la lluvia al caer sobre un tejado o por el sonido producido por un taladro mecánico en contacto con el asfalto de la carretera. Es claro por tanto el gran abanico de posibilidades que

esto otorga y la complejidad que implica para el área de reconocimiento. (p. 91-114)

La gran cantidad de sonidos no-habla, hace necesario acotar el problema a un conjunto más limitado. Este es el motivo por el cual usualmente los autores desarrollan diferentes taxonomías de sonidos, estructurando los sonidos en niveles semánticos. El desarrollo de una taxonomía de sonidos, ayuda a entender mejor el dominio de los datos (Paredes, 2012), e incrementa la precisión y la velocidad de la clasificación. Diferentes artículos aplican una jerarquía básica de clasificación entre sonidos no-habla o entre sonidos habla, no-habla y música, pero la aplicación de una jerarquía dentro del conjunto no-habla, no esta tan extendida por la dificultad que eso entraña (Lozano, 2016).

Pérez (2003), hace una primera aproximación común a la descripción jerárquica de tres niveles: “En el primer nivel se encuentran las clases materiales y las interacciones que pueden causar al sonar. En un segundo nivel los eventos se dividen en tres categorías generales: aquellas que involucran sonidos vibrátiles, sonidos aerodinámicos y sonidos líquidos. Finalmente se muestran los eventos producidos a bajo nivel, definidos por la simple interacción que pueden causar los sólidos, gases y líquidos al sonar”. (p.36)

Bajo este concepto es cómo funcionan los sistemas actuales, lo que nos lleva a que, centrarse en el colectivo de personas con discapacidad auditiva hace que el abanico de sonidos a identificar se reduzca sustancialmente, sin embargo, no parece tan evidente que sea una necesidad primordial para ella (Zubillaga, 2010).

La innovación de un conjunto de base de datos ayudara de manera permanente los segmentos que contengan sonidos para entretener a los usuarios.

Base de datos

El actual desarrollo en los sistemas e información conllevan a una tendencia hacia el uso de técnicas probabilísticas, las cuales ayudan en las organizaciones mediante objetivos estratégicos, en la implementación, de un sistema adecuado para proteger las herramientas. Estas técnicas se basan en el entrenamiento del sistema mediante bases de

datos de sonidos. Se requiere, por tanto, el uso de bases de datos que contengan suficiente cantidad de muestras de sonidos para que los modelos que se obtengan, tengan información suficiente para discernir entre todas las clases del conjunto de sonidos a revisar (Muñoz, 2018).

Durante los aspectos evaluados es correspondiente la clasificación de eventos acústicos. En 2004 tuvo lugar la primera evaluación CHIL "Computer in the Human Interaction Loop". El objetivo de la clasificación de eventos acústicos en este trabajo, fue adquirir conocimiento del entorno monitorizando el sonido ambiente. En esta evaluación, y en otras posteriores varias bases de datos de sonidos no-habla fueron creadas y puestas a distribución (Vázquez & Quizhpi, 2016). El entorno de las grabaciones consistió en una sala de reuniones donde fueron grabados sonidos tales como apertura y cierre de puertas, teléfonos, aplausos, movimiento de sillas, etc.

Posteriormente centrado en sonidos de oficina, en 2013 tuvo lugar una evaluación IEEE AASP, donde se incluyó también reconocimiento de escena. Sin embargo, posiblemente por su más reciente creación, no ha tenido muchas referencias y comparaciones en artículos publicados. Es por eso que en este proyecto se evalúan las dos bases anteriores RWCP y CHIL (López, 2013).

Tanto el conjunto de sonido como la metodología de grabación son totalmente distintos entre las bases de datos de RWCP y CHIL. Cabe destacar que en RWCP los sonidos fueron grabados independientemente en una cámara insonorizada, mientras que en CHIL, los sonidos están grabados dentro de sesiones continuas (con la posibilidad de analizar su detección) y en un entorno real como es la sala de reuniones (Lozano, 2016). Estos métodos sirven de ayuda para la señal que se produce a causa de un dominio, ya sea de tiempo o frecuencia, con una función compleja.

Aunque el uso de bases de datos es de gran ayuda a la hora de analizar y mejorar algoritmos, estas solo expresan un subconjunto de posibles estudios de comportamientos. Esto implica que, en pruebas finales, cuando se trabaja en entornos no controlados se producen caídas

considerables en el rendimiento del sistema. El funcionamiento de los sistemas de reconocimiento sufre degradaciones importantes cuando se incorpora ruido de fondo a las grabaciones o cuando las condiciones acústicas de los datos de entrenamiento y los datos de test son muy diferentes (Robles, 2018).

Lo descrito por Robles, es muy habitual en los sistemas que funcionan en explicaciones de tiempo real en las que, por una parte, la fuente de sonido suele estar contaminada por la presencia de diversos ruidos de fondo, separación de los micrófonos, reverberaciones, movimiento, etc., y, por otra parte, resulta imposible disponer de datos de entrenamiento grabados en todas las condiciones acústicas posibles. Estas circunstancias han dado lugar, en los últimos años a una extensa investigación sobre diversas técnicas cuyo objetivo primordial es el de proporcionar una gran robustez a los sistemas de reconocimiento inmersos en un entorno acústico diferente al existente en la fase de entrenamiento. Para tal fin, se pueden encontrar artículos que comparan la robustez de sus algoritmos mezclando los sonidos a detectar con ruido de fondo de diferentes entornos, como, por ejemplo: aire acondicionado (Gallardo, 2002). El ruido artificial generado, otras técnicas existentes se basan en adaptar el sistema "online" al ruido de fondo que se está detectando, reentrenando los modelos dinámicamente

Extracción de características acústicas

En lo que respecta a la extracción de características acústicas. Una sola característica difícilmente será capaz de dar información para clasificar todas las clases de sonidos, y cuando el conjunto obtenido es amplio, se hace necesaria la aplicación de métodos de validación basados en selección y búsqueda de características que requieren la realización de varias pruebas con los datos (Quispe Y. , 2017).

A continuación, en los siguientes apartados se explica los puntos más críticos tales como son: la segmentación de la señal en ventanas, continuando con una revisión de características acústicas más importantes y finalizando con las técnicas utilizadas para la reducción de variables.

Enventanado

La división de la señal en pequeños trozos (conocido como “enventanado” o segmentación), es un proceso necesario siempre que se trabaje en tiempo real. Además, este proceso también es necesario debido a que, en general, un sonido no-habla, no es una señal estocástica estacionaria. Sin embargo, si esta se divide en porciones pequeñas, puede ser considerada, al dividir la señal en partes más pequeñas. Habitualmente su tamaño es fijo, de entre 10 a 100 milisegundos. Existen estudios como Ding (2009), que trabajan con ventanas variables estimadas mediante técnicas de lógica difusa.

Se introduce la idea de los Descriptores de Unidad Acústica (AUD) donde la división de la señal se hace en base a unidades atómicas, extraídas de forma no supervisada. Los autores asumen la posibilidad de poder definir la mayoría de los sonidos con un conjunto limitado de estas unidades, capaz de representarlos a todos ellos de una manera precisa (Osses & Espinoza, 2010).

Características acústicas

Una vez que se tiene dividida la señal en ventanas, será de estas secciones de audio de las que se obtenga la información a enviar al modelo de clasificación. Esta información son los parámetros o características acústicas descriptivas del audio. La pregunta que surge a los investigadores es: ¿Qué característica hace diferente un sonido, del resto de sonidos que se puedan dar en un entorno dado? Como, por ejemplo; sonidos de pasos. Según Benítez (2018), la extracción de características puede ser dividida en dos tipos: estacionarias (basadas en frecuencia) y no estacionarias (basadas en tiempo-frecuencia). Las características estacionarias producen un resultado general detallando las frecuencias contenidas en la señal completa. Con la extracción de características estacionarias no hay distinción de donde estas frecuencias ocurren dentro de la señal.

En el reporte técnico de Fuertes (2018), divide las características en dos categorías: perceptuales y físicas. Las características perceptuales están basadas en la forma en que los humanos escuchan un sonido. Ejemplos de características conceptuales son:

pitch, timbre y ritmo. Las características físicas están basadas en propiedades estadísticas y matemáticas de las señales. Ejemplos de características físicas son; la frecuencia fundamental, Zero Crossing Rate y Energía. Se indica como algunas características perceptuales están relacionadas con características físicas, tal y como la frecuencia fundamental está relacionada con el pitch, y el timbre con el contenido espectral. (p. 69-71).

Al margen de las clasificaciones, a continuación, se presentan los parámetros acústicos más utilizados en el reconocimiento de sonidos ambientales, algunos de los cuales son comúnmente utilizados en el reconocimiento de instrumentos musicales y la gran mayoría de ellos son utilizados en el reconocimiento del habla. La definición de los parámetros acústicos se presenta en la tabla 1.

Tabla 1. Definición de los parámetros acústicos. Adaptado y modificado de (Solís Estrella, 2006)

Parámetros acústicos	Definición
Short Time Energy (STE)	Es la energía total comprendida en una ventana de audio
Zero Crossing Rate (ZTR)	Calcula las veces que pasa por cero la señal en un cierto intervalo de tiempo
Root Mean Square Level (RMS)	Medida que cuantifica la intensidad de señal. Su valor es calculado obteniendo la raíz cuadrada de la media de las muestras al cuadrado (Breebaat, 2003), lo que viene a ser la raíz cuadrada de Short Time Energy.
Band Energy Ratio	Establece la relación entre la energía de las bajas y de las altas frecuencias de la señal
Fundamental Frequency	Es la frecuencia fundamental de la señal. Aplicable cuando los sonidos a reconocer son periódicos
Spectral Centroid	Calcula la frecuencia en la que reside el centro de la energía espectral de la señal analizada. Perceptualmente mide como de brillante es el sonido que se escucha.
Roll-off Point (RF)	Calcula el ancho de banda de frecuencias en la que se concentra el 85% o 95% (este valor cambia dependiendo de la implementación) de la energía del espectro
Spectral Bandwidth	Una medida de la extensión del espectro alrededor del Spectral Centroid. Se calcula a través de la media ponderada de las distancias entre los componentes espectrales y del Spectral Centroid
Mel Frequency Cepstral Coefficients (MFCCs)	Se computan los Cepstral Coefficient utilizando una escala logarítmica tipo Mel. Algunos autores analizan diferentes modificaciones de estos

Linear Frequency Cepstral Coefficients (LFCCs):	parámetros teniendo en cuenta el espectro de los sonidos a clasificar. A diferencia que los MFCCs, estos coeficientes utilizan una escala lineal para crear su banco de filtros
Linear Prediction Coefficients (LPCs):	Utilizados frecuentemente en sistemas de comunicación para codificar y decodificar señales de voz el tracto vocal es modelado mediante un filtro todo polo de respuesta infinita impulsiva. Los coeficientes LPC son coeficientes de la combinación lineal de este filtro, siendo coeficientes que representan la envolvente espectral de la señal
Wavelet Coefficients	Obtenidos de la transformada Wavelet capaz de trabajar en el dominio secuencial con varias resoluciones. Indican que tan parecida es la señal analizar de una función madre, siendo los portadores del detalle de la señal.

Fuente: (Essid, 2006).

De acuerdo con Pérez M. (2018), introducen el concepto de: El uso de características de fluctuación espectro-temporales basadas en el algoritmo Harmonic Percussive Sound Separation (HPSS) donde extrae tres espectrogramas que contienen los componentes estacionarios, transitorios e intermedios de la señal (p.46).

Otros estudios aplican conocimientos del campo de la percepción auditiva humana, usando características acústicas extraídas de un análisis del espectro denominado "cocleograma", que imita la respuesta de la membrana basilar (Miró, 2013).

Según el autor Lengua (2014) son guiados también por estudios de psico acústica, se usan los llamados Amplitude Modulation Spectrograms (AMS, siglas al inglés) motivados por la importancia de modulaciones temporales para el reconocimiento de objetos acústicos. AMS representa una descomposición de la señal a través de las dimensiones de frecuencia, modulación y tiempo, a su vez estas con computadas por una descomposición espectral de sub-bandas. Esta técnica es aplicada para detectar sonidos que no han sido previamente recogidos en la base de datos, aunque los experimentos presentados tan solo han tratado con un corpus de cuatro clases de sonidos. (p.1658-1662).

Burak y Buke (2012) utilizan una característica para medir el rango de variación de la frecuencia fundamental postulando la hipótesis de que en muchos sonidos no-habla este cambia abruptamente sin permanecer estacionario (Burak & Buke, 2012). En este caso, estos parámetros obtienen resultados muy similares a los MFCCs, añadiendo información adicional si se complementan. Otros trabajos como los de Chu, Narayanan, & Kuo (2008), utilizan características basadas en la técnica Matching Pursuit (MP) aplicando funciones de Gabor, pero en este caso, la aplicación es más dirigida al reconocimiento de entornos que al de sonidos aislados. Sin embargo, el cálculo computacional que estas técnicas requieren es elevado. La implementación de estos algoritmos en dispositivos móviles en tiempo real requeriría unas prestaciones de hardware alejadas de las actuales disponibles en el mercado. El procesamiento de todos los filtros y técnicas aplicadas incrementan considerablemente el tiempo de respuesta del sistema, haciendo inviable su uso.

En los últimos años, las técnicas aplicadas en el reconocimiento de canciones basados en las teorías de Wang (2003), han tenido gran impacto en la literatura científica. Estas técnicas buscan encontrar puntos clave dentro del espectrograma que definan un patrón en dos dimensiones del sonido (tiempo – frecuencia). No obstante, la base de datos con la que se trabaja contiene un número muy limitado de clases. Además, estas técnicas están basadas en la realización de conexiones lo largo del tiempo y esto implica que la duración del sonido no puede ser muy corta para su correcto reconocimiento. (p. 713-718).

Reducción de variables

La reducción de variables tiene como objetivos disminuir la información, cuando el número de parámetros utilizados es muy elevado, los clasificadores sufren problemas de sobredimensionamiento. Para mitigar este efecto contraproducente es importante hacer una reducción previa de variables para evitar parámetros redundantes y una alta complejidad en el clasificador.

En los sistemas de clasificación, se utilizan mecanismos de reducción de parámetros que permiten separar aquellas características que realmente aportan información descriptiva, de

las que únicamente añaden ruido a la parametrización (Vivar, 2017). Con ello también se consigue reducir la dimensión de la parametrización, lo que reduce la carga computacional. El detectar los parámetros verdaderamente discriminantes puede permitir además comprender mejor la relación que sucede entre los sonidos ambientales y las características acústicas de la señal de audio. Existen dos aproximaciones: Construcción de Características y Selección de Conjuntos de Características (García V. , 2019). A continuación, un breve resumen de las mismas.

Construcción de Características

Las técnicas de construcción de características buscan las posibles relaciones entre las variables y devuelven un nuevo conjunto de variables, combinando las originales.

Las Técnicas de análisis de componentes, por ejemplo; Principal Component Analysis (PCA), son un caso de amplio uso, en el que el espacio transformado se forma de elementos ortogonales y aunque no se considera ningún modelo de aleatoriedad, en la estimación de las respectivas matrices de covarianza, es conocido que la efectividad de la técnica es más alta entre más se ajuste la condición de Gaussividad para las distribuciones de características (Camacho, 2019).

Selección de conjuntos de características:

Los algoritmos de selección de parámetros se suelen dividir en dos grupos, los wrappers, que es una subrutina en una biblioteca de software o un programa de computadora cuyo propósito principal es llamar a una segunda subrutina o una llamada al sistema con poco o ningún cálculo adicional. A continuación, se presenta una breve descripción de cada uno de ellos para facilitar la comprensión de los mismos.

Los métodos wrapper, utilizan un clasificador como mecanismo para estimar la precisión que puede alcanzar un cierto conjunto de parámetros. Esta estimación se obtiene entrenando y evaluando el clasificador con cada conjunto de parámetros considerado, generalmente mediante algún tipo de validación cruzada. Una vez estimado el error

de clasificación cometido por cada conjunto de parámetros, se selecciona aquel que proporciona mayor tasa de aciertos (Picó & Silva, 2018). En este tipo de métodos, existen dos aproximaciones básicas:

- Selección Forward: se empieza sin ninguna variable y se van añadiendo una a una. En cada paso se añade la que más hace decrecer el error hasta que no se encuentren mejoras significativas.
- Selección Backward: se empieza con todas las variables y se van eliminando una a una. En cada paso se elimina la que más hace decrecer el error hasta que no se encuentren mejoras significativas.

Este tipo de métodos proporcionan conjuntos de parámetros que alcanzan una gran precisión para el clasificador considerado. Sin embargo, los resultados no siempre son generalizables a otros clasificadores. Además, son algoritmos de gran carga computacional, debido a la necesidad de entrenar el clasificador por cada conjunto de parámetros analizado (Espinoza, 2019).

Los métodos filtrados de señales seleccionan los parámetros en función de su distribución y su relación con el resto de parámetros y con la clase objetivo. Como resultado, el conjunto seleccionado no está optimizado para ningún clasificador, y generalmente proporciona una precisión comparable con diferentes clasificadores. También requieren una carga computacional significativamente menor que el método wrapper, al no tener que entrenar un clasificador en cada etapa. Estos métodos emplean medidas de ganancia de información, distancia o consistencia, entre el parámetro y la clase (Leite & Bradzil, 2002); sin embargo, debido a que miden la importancia de cada parámetro en forma aislada, no pueden detectar si existen parámetros redundantes, y tampoco son capaces de determinar si la combinación de dos o más parámetros, aparentemente irrelevantes en forma aislada, se pueden transformar en relevantes (Mejía & Frausto, España). Algunas aproximaciones de este problema son planteadas en el Performance of Basic Spectral Descriptions and MRMR Algorithm to the Detection of Acoustic Events, en donde utilizan el algoritmo denominado Minimum Redundancy Maximum Relevance (Vidal, 2013).

Clasificación de sonidos

Las técnicas de clasificación de audio, permiten decidir a qué clase de entre un conjunto de clases previamente definido pertenece cada observación (conjunto de características acústicas). Si en el proceso de reducción de variables se ha utilizado un método wrapper la técnica de clasificación está ligada a dicho proceso. Aunque existen estudios que aplican algoritmos semi-supervisados, o técnicas deterministas, la gran mayoría de ellos aplican algoritmos supervisados en los que es necesario que los sonidos sean etiquetados previamente (Garces, 2008).

Técnicas de Clasificación de Sonidos

Clarkson, Sawhney, & Pentlant (1998), diseñaron un sistema para reconocer diferentes tipos de sonidos tales como habla, sonidos de vehículos, etc., usando modelos ocultos de Markov (HMM) como método de clasificación.

En la siguiente tabla se listan los artículos más relevantes desde 2010, e indicando el o los algoritmos de clasificación utilizados. Los artículos y clasificaciones más relevantes se observan en la tabla 2.

Tabla 2. Artículos y clasificaciones más relevantes

Artículo	Clasificador
Real-time digital speech processing strategies for the hearing impaired.	FILTRO ADAPTABLE RACE Este filtro propone una solución implementada en un DSP (procesador digital de señales) a la que llaman realizado adaptable por correlación en tiempo real o RACE, por sus siglas en inglés (Bermúdez, 2016).
Quantile based noise estimation for spectral subtraction and Wiener filtering	SUSTRACCIÓN ESPECTRAL La idea principal de la sustracción espectral es estimar el ruido ambiental y posteriormente sustraer dicho estimado espectral de la señal ruidosa para así mejorar la componente de voz (Chu, Narayanan, & Kuo, 2008)
GSVD-Based Optimal Filtering for Single and Multimicrophone Speech Enhancement	FILTRADO POR GSVD Usa el criterio de error de mínimos cuadrados promedio (MSE) (Hernández, 2019).

Combined approach of array processing and independent component analysis for blind separation of acoustic signals

Análisis de componentes independientes
Esto se puede expresar con las ecuaciones lineales la voz y el ruido (García V. , 2019).

Fuente: (Guilles, 1994)

Los detectores de actividad de voz (VAD), son fundamentales en el uso eficiente de ancho de banda [30], por lo cual resulta útil en señales HS y LS. (Mayorga, Chahez, Arguellez, & Druzgalski, 2017, pág. 15).

En muchas aplicaciones de procesamiento de señales de voz, la detección de actividad de voz (VAD) juega un papel esencial para separar un flujo de audio en intervalos de tiempo que contienen actividad de voz e intervalos de tiempo en los que no hay voz. En la literatura se introdujeron muchas características que reflejan la presencia del habla.

METODOLOGÍA

El enfoque metodológico de la presente investigación es de tipo exploratorio y descriptiva porque consiste en la recopilación de información desarrollada por otros investigadores que servirán como base para el desarrollo del algoritmo. Para lo cual se hace uso de base de datos como SCOPUS, Google Scholar, de donde se extrajeron los artículos utilizados en la redacción de la sección del Marco Teórico. Es cualitativa y aplicada porque hace uso de técnicas y parámetros utilizados por otros investigadores para el desarrollo del algoritmo, para lo cual se revisan cuatro técnicas de reducción de ruido ambiental. (Bermúdez, 2016)

1. El filtro adaptable RACE (realizado adaptable por correlación),
2. La sustracción espectral,
3. El filtrado por medio de GSVD (descomposición en valores singulares generalizada).
4. El análisis de componentes independientes.

El proceso de análisis de señales, mediante las etapas de procesamiento se expone en la figura 9.

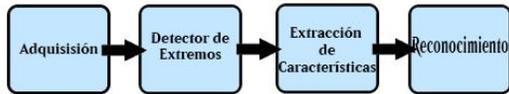


Figura 9. Etapas del procesamiento (Lozano, 2016)

Considerando las técnicas expuestas, se desarrollará el algoritmo de reconocimiento de contexto utilizando Python, considerando los siguientes pasos: para el procesamiento de señales, el primer paso es leer el archivo de audio con ruido de fondo para convertirlo en un arreglo de datos analógicos. Con ese arreglo se trazan puntos discretos con el cual se muestran las gráficas desde matplotlib (una librería de Python para gráficos). Con el uso de la teoría de procesamiento digital de señales, se usa el concepto de frecuencia máxima dividido por dos (llamado la frecuencia de Nyquist). Por consiguiente, se aplica la transformada rápida de Fourier (FFT) y podemos ver la señal en el espectro de las frecuencias. Finalmente, para detectar o reconocer el ruido, en el mapa espectral aparecen señales de forma aleatoria que se puede suprimir mediante un filtro pasa alto o paso bajo.

El ruido de fondo produce una simultánea a la realización de una medida acústica, la cual afecta al resultado, por ende, el resultado debe ser favorable evitando errores de forma simultánea sobre la fuente sonora.

DESARROLLO DEL APORTE PRÁCTICO

En lo que respecta a este punto, en el, se procederá a describir todo el proceso de diseño del algoritmo para el reconocimiento del contexto de ruido de fondo.

Según Chu, Narayanan, & Kuo, (2008) para diseñar técnicas de reducción de ruido se consideran los ruidos blancos, que se los puede reducir con un filtrado común en la señal contaminada, pero en otros casos, se debe eliminar el ruido por otros filtros de banda.

De tal forma surgen principios generales de reducción de ruido, uno de ellos el filtrado fijo en el cual se eliminan las bandas no críticas para la voz, esto es muy básico porque ni la voz ni tipos comunes de ruido son estacionarios o siempre tienen el mismo contenido en frecuencia. (Blasina, 2019)

Una evolución natural son los filtros adaptables, los cuales modifican su respuesta en función de las características de las señales. Asimismo, otro principio que se puede usar es el filtrado espacial aprovechando “observaciones” de dos o más transductores colocados de manera específica para obtener más información sobre las señales de voz y ruido (Valhondo, 2003)

Para el desarrollo del presente trabajo se revisaron cuatro técnicas de reducción de ruido ambiental: el filtro adaptable RACE (realizado adaptable por correlación), la sustracción espectral, el filtrado por medio de GSVD (descomposición en valores singulares generalizada) y el análisis de componentes independientes.

$$\hat{R}_{xx}(n, l) = \beta R_{xx}(n-1, l) + (1-\beta)x(n)x(n+l)$$

Las primeras dos son técnicas clásicas de reducción de ruido implementadas en prototipos de auxiliares auditivos, el filtrado por GSVD es una propuesta reciente la cual está implementada en tiempo real en una computadora.

El análisis de componentes independientes es una técnica con un enfoque distinto al del filtrado adaptable (en el que se basan de manera general las otras tres técnicas), forma parte de una familia de técnicas utilizadas para la separación a ciegas de fuentes y no ha sido implementado en auxiliares auditivos, pero ha mostrado resultados interesantes en otras áreas.

Método de Filtro adaptable RACE

RACE es básicamente un filtro adaptable de respuesta al impulso finita (FIR) que actualiza sus coeficientes de acuerdo con la estimación recursiva de la correlación de la señal de entrada según la expresión (Solís, 2006):

Donde β es una constante de suavizado tomando valores de 0 y 1, $R_{xx}(n, 1)$ y $R_{xx}(n-1, 1)$ son los estimados de una autocorrelación de la señal de entrada $x[n]$ (Blanco M., 2008).

Método de supresión de ruido por sustracción espectral

El método utilizado para realizar la reducción de ruido por sustracción espectral es el siguiente: 1. Obtener la FFT por segmentos de la señal contaminada. 2. Estimar el espectro

de potencia del ruido en periodos de inactividad de voz y ponderarlo. 3. Sustraer el estimado de ruido del espectro de potencia de la señal contaminada.

4. Realizar la rectificación de media onda: las partes del espectro que resulten negativas después de la sustracción son reemplazadas por una versión muy atenuada de la señal o con cero. 5. La fase de la señal ruidosa se combina con el resultado de la rectificación de media onda y se aplica la IFFT (Solís, 2006).

Método de FILTRADO POR GSVD

Usando el criterio de error de mínimos cuadrados promedio (MSE) usados por (Muñoz, Giraldo, & Bonilla, 2016), se tiene que el filtro se expresa como:

W_{WF} se deriva de la descomposición en valores singulares de las matrices Yk y Vk mencionadas en la ecuación.

Método de componentes independientes

Bajo ciertas condiciones, se puede considerar como una suma ponderada de las señales emitidas por las fuentes de sonido (la voz y el ruido). Esto se puede expresar con las ecuaciones lineales (Solís, 2006):

$$\begin{aligned}x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) \\x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t)\end{aligned}\quad (4)$$

Donde a_{ij} son parámetros que dependen de las distancias de los transductores a las fuentes.

Conjunto de datos experimentales

El conjunto de registros de voz que se utilizó para realizar la evaluación consta de grabaciones de voz de siete locutores diferentes, este registro funciona como base para ser diferenciados de los ruidos no-habla. Quince de las grabaciones provienen de un locutor de género masculino y los quince restantes de un locutor de género femenino. Son archivos en formato WAV (Waveform Audio Format) con codificación de Modulación por impulsos codificados conocido por sus siglas en inglés como PCM lineal, una frecuencia de muestreo de 11025 Hz y 16 bits por muestra. Los archivos consisten en una canción de lenguaje español.

Los ruidos grabados son ruido de oficina, ruido de conversaciones indistintas en un lugar concurrido y ruido de un medio de transporte. El ruido de oficina consta del sonido de equipo

de cómputo funcionando y de escritura en teclados y no incluye voces de personas.

Detección de actividad de voz

Debido a que los detectores de actividad de voz más comunes no son confiables en condiciones donde la SNR (Relación de señal y ruido) es cercana a 0 dB, y el objetivo del presente trabajo es desarrollar un algoritmo para reconocimiento de un ruido, para eso se necesitan técnicas que requieren de un detector de actividad de voz, la señal de fue calculada a partir de los registros limpios de voz.

Metodología de evaluación en entorno de simulación

Se aplica esta metodología de desempeño a las técnicas de reducción de ruido porque no todos los algoritmos mencionados nos sirven para todos los escenarios, unas técnicas son dedicadas cuando hay una segunda voz en un audio, otros son usados para ambientes de ruido en segundo plano, por esta razón es necesario indicar la mejor relación de señal – ruido presente en el algoritmo. Para revisar el desempeño de las técnicas de reducción de ruido se utilizó la siguiente metodología (basado en los algoritmos de detección de señal y filtrado (Blanco M. , 2008):

1. Se carga de un archivo la señal limpia de voz y se calcula el nivel en decibeles (dB).
2. Se detecta la actividad de voz.
3. Se genera o se carga de un archivo con la señal de ruido.
4. Se ajusta el nivel del ruido con la expresión, de forma que se obtenga una relación señal/ruido de -9 dB
-6 dB, -3 dB, 0dB, 3 dB y 6 dB en referencia al nivel de la señal de voz calculado previamente.

5. Se hacen las mezclas aditivas de voz con ruido con las SNR previamente mencionadas, la mezcla se hace de la forma 0.5 (voz + ruido). En la mezcla se toma la media aritmética en vez de la suma para evitar que la señal salga del rango [-1,1]. En el caso del filtrado por GSVD y de ICA, que requieren dos canales, el segundo canal está compuesto por una mezcla en la cual el ruido tiene 1 dB más que en el paso anterior y la voz se atenúa 1 dB, simulando que un micrófono está más orientado hacia la voz y otro más hacia el ruido.

6. Se obtiene una señal mejorada con las técnicas utilizando los archivos de mezcla de voz con ruido.

7. Se calcula la SNR de la señal mejorada de acuerdo con la expresión tomando en cuenta que, si se asume que en la señal mejorada se recuperó la componente de voz de la mezcla, se debe ajustar el nivel de la energía al nivel de ésta. Así pues, la señal s en la componente de voz de la mezcla, la voz y la señal de ruido residual r es por lo tanto la diferencia entre la señal mejorada y la componente de voz en la mezcla.

8. Se calcula la SNR ponderada

9. La mejora a la SNR que proporciona la técnica se obtiene como la diferencia entre la SNR de la señal mejorada y la SNR de la mezcla de voz con ruido (Durán, 2013).

Desarrollo del algoritmo

Para el desarrollo del algoritmo se debe considerar los parámetros de canal, lectura de un archivo de audio, señal de muestreo (para pasar de una señal analógica a una digital) y la transformada rápida de Fourier (Blanco M. , 2008).

Los pasos utilizados para reducir el ruido de fondo son los siguientes:

- Recogemos la data del audio con la función de `waves.read`
- Se procesa el archivo de audio en Python, se coloca el audio en un solo canal.
- Se calcula una FFT sobre el clip de audio de ruido
- Las estadísticas se calculan sobre FFT del ruido (en frecuencia)
- Se calcula un umbral basado en las estadísticas del ruido (y la sensibilidad deseada del algoritmo)
- Se calcula una FFT sobre la señal
- Una máscara se determina comparando la señal FFT con el umbral
- La máscara se suaviza con un filtro sobre frecuencia y tiempo.
- La máscara se aplica a la FFT de la señal y se invierte

El script se muestra en la figura 10 es básicamente la formulación del algoritmo realizado.

```
import IPython
from scipy.io import wavfile
import scipy.signal
import numpy as np
import matplotlib.pyplot as plt
import librosa
import scipy.io.wavfile as waves
import time
from datetime import timedelta as td

# archivo = input('archivo de sonido:')
archivo = 'star_is_born.wav'
muestreo, sonido = waves.read(archivo)

# canales: monofónico o estéreo
tamano = np.shape(sonido)
muestras = tamano[0]
m = len(tamano)
canales = 1 # monofónico
if (m>1): # estéreo
    canales = tamano[1]
# experimento con un canal
if (canales>1):
    canal = 0
    uncanal = sonido[:,canal]
else:
    uncanal = sonido

# rango de observación en segundos
inicia = 1.000
termina = 2.002
# observación en número de muestra
a = int(inicia*muestreo)
b = int(termina*muestreo)
parte = uncanal[a:b]

# Salida # Archivo de audio.wav
print('archivo de parte[] grabado...')
waves.write('parte01.wav', muestreo, parte)

- .
noise_len = 25 # seconds
noise = band_limited_noise(min_freq=4000,
                           max_freq = 12000,
                           samples=len(archivo),
                           samplerate=muestreo)*100
noise_clip = noise[:muestreo*noise_len]
audio_clip_band_limited = noise+noise

IPython.display.Audio(data=audio_clip_band_limited,
                      rate=audio_clip_band_limited)

#ruido removido
output = removeNoise(audio_clip=audio_clip_band_limited,
                    noise_clip=noise_clip,
                    [verbose=True,visual=True])

# play back a sample of the song
IPython.display.Audio(data=output, rate=44100)
```

Figura 10. Formulación del algoritmo

Comprobación y Resultados del Algoritmo.

Lo primero que se realizó fue descomponer la señal de audio en un solo canal para que el sonido salga por ese canal como se muestra en la figura 12, 13 y 14. Las funciones que se realizaron fueron para detectar el ruido de la señal, por medio de la transformada rápida de Fourier, los filtros para DSP (procesamiento digital de señales).

En la figura 11 se mostrará datos del sonido con ruido.

```
===== RESTART:
frecuencia de muestreo: 44100
dimensiones de matriz: (1102500, 2)
datos de sonido:
[[ 0 0]
 [ 0 0]
 [ 0 0]
 ...
 [ 1086 -1160]
 [ 496 -1103]
 [ 215 -480]]
```

Figura 11. Datos del sonido con ruido.

Luego se llama al archivo en la función noise, que sirve para detectar el ruido de la señal a una máxima frecuencia de 12KHz y el dato del muestreo sacado del archivo original del audio, este proceso se expone en la figura 12. Al final sale el ruido removido.

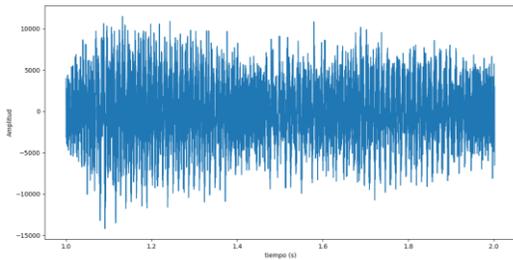


Figura 12. Gráfica del audio con ruido

Devoluciones: gráfica: la señal recuperada con ruido sustraído mostrado en la figura 13 y 14.

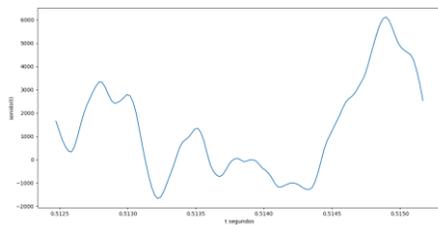


Figura 13. Gráfica del audio con ruido renderizado en el tiempo de forma analógica

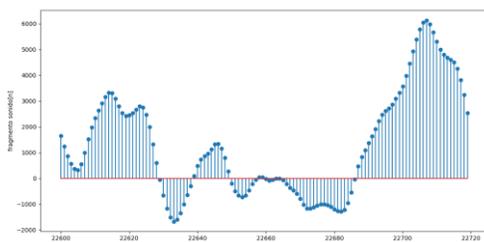


Figura 14. Gráfica del audio con ruido renderizado en el tiempo de forma digital

Finalmente, en la figura 15 se expone la gráfica del audio sin ruido renderizado en el tiempo de forma digital aplicando FFT

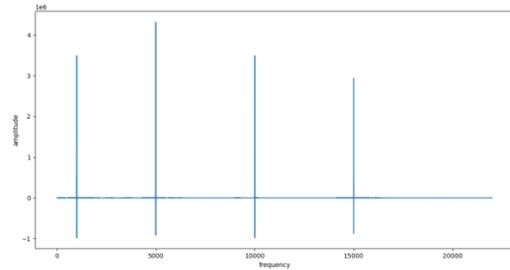


Figura 15. Gráfica del audio sin ruido renderizado en el tiempo de forma digital aplicando FFT

CONCLUSIONES

Luego de caracterización de los sistemas de reconocimiento de sonidos por diversas etapas de algunos sistemas de reconocimiento de sonidos no-habla dentro del ámbito de la investigación. Se observó la tendencia a la división de estos en dos grandes bloques: detección y clasificación, dentro de los cuales se encuentra implícita la extracción de características acústicas de la señal para el entrenamiento y testeo de los modelos usados para el reconocimiento.

El desarrollo del algoritmo permite un mayor entendimiento de los diversos métodos que existen para reducir el ruido y posteriormente analizarlo para obtener información que puede ser utilizada para asistir al desarrollo de tecnología que ayude a las personas con audición limitada.

En base al desarrollo del algoritmo propuesto en esta documentación, se concluye que mediante las gráficas y las salidas de archivos de audio obtenidos por la aplicación de las técnicas de procesamiento digital de señales como es la transformada rápida de Fourier y la función de remover el ruido mediante un filtro basado en los valores pasados y actuales de la señal - filtro IIR (Blanco M. , 2008) se logra la detección del ruido no-habla, permitiendo de esta forma un avance en el posible desarrollo de tecnología asistiva para personas con discapacidad auditiva, y de esta forma la creación de nuevas herramientas que mejorarían su calidad de vida.

Referencias

- Benítez, M. (2018). *Estudio y análisis de métodos para la extracción de características y clasificación de emociones basados en EEG*. Madrid: UAM. Departamento de Ingeniería Informática.
- Bermúdez, S. (2016). *La caracterización del sonido como onda mecánica: una propuesta para la enseñanza en la escuela*. Bogota: Repositorio Institucional UPN.
- Blanco, Á. (2017). *DESARROLLO DE UN SOFTWARE DE SIMULACIÓN ACÚSTICA PARA RECINTOS CERRADOS*. Cantabria: UNIVERSIDAD DE CANTABRIA.
- Blanco, M. (2008). *Tratamiento Digital de Señales*. España: Universidad de Alcalá .
- Blasina, F. (2019). *Procesamiento de señales acústicas aplicado al monitoreo de procesos*. Uruguay: Universidad de la República (Uruguay). Facultad de Ingeniería.
- Burak, U., & Buke, t. B. (2012). NON-SPEECH ENVIRONMENTAL SOUND CLASSIFICATION USING. *International Journal of Innovative*, 3511–3524.
- Camacho, J. (2019). *licación del análisis en componentes independientes (ICA) a la clasificación de imágenes*. Sevilla: Universidad de Sevilla.
- Carranza, L., & Alzamora, K. (2017). *Limitaciones en el desarrollo cognitivo a consecuencia de la escasa inclusión educativa para niños con discapacidad auditiva de 6 a 10 años en Lima Metropolitana*. Lima: Universidad San Ignacio de Loyola.
- Chu, S., Narayanan, S., & Kuo, C. C. (2008). Environmental sound recognition using MP based features. *2008 IEEE Int. Conf. Acoust. Speech Signal Process*, (págs. 1-4).
- Clarkson, B., Sawhney, N., & Pentlant, A. (1998). Auditory Context Awareness via wearable computing. *Work percept user interfaces*.
- Covreur, C. (1997). Environmental sound recognition: a statistical approach. *PhD Thesis*. Fac. Polytech. Mons.
- Cowling, M. (2004). Non-speech environmental sound classification system for autonomus surveillance. Griffith Univ.
- Cuevas, F. (2017). *VAMOS AL CINE: EL FENÓMENO*. Mexico : Universidad México.
- Dàvila, P., & Cahueñas, R. (2018). *Diseño de una cabina insonorizada para un laboratorio de calibración de equipos de monitoreo de ruido*. Quito : Universidad Internacional SEK.
- Dolores Zuñiga, A. B. (2004). *Menos ruido, más vida* . Andalucía, España: EDUCAM: Materiales de Educación Ambiental.
- Donoso, R. (2018). *Diseño e implementación de un sistema de reconocimiento de hablantes*. España : Universidad Carlos III de Madrid.
- Durán, M. (2013). *La simulación clínica como metodología de aprendizaje y adquisición de competencias en*

- enfermería*. Madrid: Universidad Complutense de Madrid.
- Espinoza, J. (2019). *Desarrollo de un software para la gestión de proyectos de desarrollo PM4R que permita automatizar la etapa de planificación para estimar tiempos a través de algoritmos de aprendizaje automático. Tesis Para optar por el Título de Ingeniero Informático que*. Lima: PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ.
- Essid. (2006). Hierarchical classification of musical instruments on solo recordings. *vol 5*, 817-820. ICASSP06.
- Fuertes, P. (2018). *Características vocales de la voz cantada de un grupo de estudiantes de canto de música popular contemporánea de la ciudad de Lima*. Peru: PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ.
- Gallardo, A. (2002). *RECONOCIMIENTO DE HABLA ROBUSTO FRENTE A CONDICIONES DE RUIDO ADITIVO Y CONVOLUTIVO*. Madrid,: ESCUELA TÉCNICA SUPERIOR DE INGENIEROS .
- Garces, M. (2008). *Reducción de espacios dinámicos de representación*. Pereira: Universidad Tecnológica de Pereira.
- García, L. (2019). *Postproducción de Audio Digital: Edición, Mezcla y Masterización*. Alemania : Book demand .
- García, V. (2019). *Diseño de un sistema de reconocimiento automático de vehículos mediante el uso de redes neuronales profundas (DNN)*. Imbabura: Repositorio Digital Universidad Técnica del Norte .
- Gil, F. (2018). *Sistema de medida de tensión arterial en dedos con fotoplethismografía: sistema PPG*. Madrid: Repositorio de la Universidad Pontificia Comillas.
- Gilibert, J. (2019). *Implementación de dispositivo de procesamiento de señal de audio y generación de excitación a LEDs individualmente direccionables*. Sevilla: Universidad de Sevilla.
- González, Á. (2019). *Estimación del ritmo cardíaco mediante análisis de secuencias de vídeo*. Sevilla: Universidad de Sevilla.
- Guerrón, A. (2018). *Alternativa tecnológica con GEO posicionamiento y rastreo satelital para reducir la contaminación acústica producida por parte de los distribuidores de gas doméstico*. Ibarra : UNIVERSIDAD REGIONAL AUTÓNOMA DE LOS ANDES.
- Guilles, D. (1994). *LÓGICA DEL SENTIDO*. Escuela de Filosofía Universidad ARCIS. Obtenido de www.philosophia.cl : www.philosophia.cl
- Hernández, G. (2019). *Monografía "Audio y electroacústica"*. Tolima: Universidad de Ibagué.
- Kane, j. (2004). Poetry as Righthemispheric Language. *j. Conscious. Stud. Vol 11*, (págs. 21-59).
- Leite, R., & Bradzil, P. (2002). Decision tree-based attribute selection via sub sampling. En P. B. R. Leite, *Decision tree-based attribute*

- selection via sub sampling* (págs. 1157-1182). Sevilla.
- Lengua, I. (2014). *DESARROLLO DE UN SISTEMA DE ENTRENAMIENTO ACÚSTICO VIRTUAL PARA LOCALIZACIÓN DE SONIDOS VIRTUALES PARA PERSONAS INVIDENTES*. Valencia : Universitat Politècnica de València.
- Levratto, V. (2017). *Encuentro entre lectura en papel y lectura digital: hacia una gramática de lectura en los entornos virtuales*. España : Universidad Nacional Educación a Distancia (UNED).
- Linares, N. (2017). *Verificación del cumplimiento normativo de los niveles de presión sonora previstos en la Resolución 627 para el subsector hospitalario en la ciudad de Bogotá*. Bogotá: Universidad de La Salle. Facultad de Ingeniería. Ingeniería Ambiental y Sanitaria.
- López, J. (2013). *Detección de aglomeraciones estática*. Castellón: Universidad Jaume .
- Lozano, H. (2016). *Detección e identificación de señales sonoras en entornos asistivos*. Bilbao: Universidad del País Vasco .
- Luna, K. (2018). *Guía de Personas con Robots Móviles Basado en la Interacción Humano-Robot*. Mexico: Universidad Autónoma de San Luis Potosí.
- Mayorga, Chahez, Arguellez, & Druzgalski. (2017). Detección y extracción automática de eventos S1, S2, S3 y S4 en sonidos del corazón. *Res. Comput. Sci.*, 142, 9-20. Obtenido de https://www.rcs.cic.ipn.mx/2017_142/RCS_142_2017.pdf#page=9
- Mejía, M., & Frausto, J. (España). *Selección de Atributos en una Base de Datos de Facturación Eléctrica aplicando Programación Cóncava* . 2004: Univesidad de Garnada .
- Miró, M. (2013). *Una aportación al procesado neuromórfico de audio basado en modelos pulsantes. Desde la cóclea a la percepción auditiva*. Sevilla: Universidad de Sevilla.
- Muñoz, P., Giraldo, E., & Bonilla, J. (2016). Identificación y control de un Vehículo Aéreo no Tripulado tipo Quadcopter. *Ingenierías USBMed*, 5-10.
- Olaya, C. (2018). *Propuesta de intervención para el acceso de personas con discapacidad auditiva al servicio de emergencias navarro*. Pamplona : Universidad Pública de Navarra.
- Ortiz, A. (2016). *Monitoreo de vibraciones, ondas acústicas y determinación de la onda de soplo por efecto de voladura, en el túnel de aducción de la central hidroeléctrica chaglla*. Departamento de Huánuco. Huánuco: Facultad de Geología, Geofísica y Minas .
- Osses, A., & Espinoza, V. (2010). *PONDERACIÓN PSICOACÚSTICA EN FRECUENCIA Y AMPLITUD PARA SEÑALES DE AUDIO DIGITAL* . Santiago de Chile : Universidad Tecnológica de Chile INACAP.
- Paredes, F. (2012). *Propuesta para la enseñanza de la comunicación*

- umentativa y alternativa en la población con trastorno del espectro autista*. Bogota: Universidad Pedagógica Nacional.
- Perez, D. (2019). *Analisis de señales de tos para detección temprana de enfermedades respiratorias*. Valladolid : Universidad de Valladolid .
- Pérez, M. (2003). *Evaluación de la comprensión lectora en alumnos de doce años*. Madrid: Universidad Complutense de Madrid, Facultad de Educación, Departamento de Métodos de Investigación y Diagnóstico en Educación.
- Pérez, M. (2018). *Desarrollo de Karaoke (Minus One) basado en algoritmos de alineamiento y separación de señales musicales*. Linares: Repositorio de Trabajos Académicos de la Universidad de Jaén.
- Pico, E., & Lindao, G. (2019). *La inteligencia emocional para mejorar el desempeño laboral de los trabajadores de la empresa EMASA-EP, provincia de Santa Elena, año 2018*. La Libertad: Universidad Estatal Península de Santa Elena.
- Picó, M., & Silva, M. (2018). *Aprendizaje computacional para la generación automática de programas*. Montevideo : Universidad de la República (Uruguay). Facultad de Ingeniería.
- Quispe, C. (2018). *Retraso del lenguaje oral en niños/as de 2 a 4 años : (Estudio en el CENTRO INFANTIL RECREANDO, Ciudad de El Alto)*. Ciudad de La Paz: Universidad Mayor de San Andrés.
- Quispe, Y. (2017). *“Identificación de Personas a Partir de un Sistema Biométrico Bimodal audio-visual”*. Arequipa: Escuela Profesional de Ingeniería de Telecomunicaciones.
- Radostina, D. (2017). *Desarrollo y evaluación de métodos de selección de características para la predicción de eventos adversos en pacientes polimedicados*. Navarra: E.T.S. de Ingeniería Industrial,.
- Rasgado, D., & Mereles, C. (2017). *Aplicación para el reconocimiento automático de huellas digitales*. Catamarca: UNIVERSIDAD NACIONAL DE CATAMARCA .
- Robles, A. (2018). *Diseño óptimo de redes de distribución de agua mediante algoritmos de búsqueda tabú*. Sevilla: Universidad de Sevilla.
- Sepúlveda, F. (2019). *Radiación de Hawking en un fluido transónico - medición de propiedades termodinámicas de un sistema análogo*. Santiago de Chile: Universidad de Chile.
- Solís. (2006). Estudio comparativo de técnicas digitales de reducción de ruido para su posible uso en auxiliares auditivos. *Revista Mexicana de Ingeniería Biomédica*.
- Torres, C. (2018). *Estudio de factibilidad de un sistema de transferencia inalámbrica de energía para la alimentación de vehículos eléctricos utilizando*

- Beamforming*. Santiago de Chile: UNIVERSIDAD DE CHILE.
- Valhondo, D. (2003). *Gestión del conocimiento: del mito a la realidad*. Madrid: Ediciones Diaz de Santo.
- VanDerveer, N. (1979). *Ecikigical acoustics: Human Perception of Enviromental Sounds*. Cornell University .
- Vázquez, D., & Quizhpi, L. (2016). *Diseño de una aplicación multimedia para mejorar la comunicación de los niños con parálisis cerebral que sufren trastornos del habla y del lenguaje*. Azuay: Universidad del Azuay.
- Vidal, P. (2013). *Técnicas secuenciales y paralelas para la resolución de problemas computacionales en electromagnetismo*. Valencia: Universitat Politècnica de València.
- Vivar, A. (2017). *Desarrollo de sistema biométrico de obtención y clasificación de datos de miembro inferior derecho*. Santiago de Chile: Universidad Andres Bello.
- Wang, A. (2003). An industrial Strength Audio Search Algorithm. *Int. Conf. Music Inf. Retr.*, (págs. 713-718).
- Ylurdoz, A. M. (2016). *Universidad Politécnica de Madrid*. Obtenido de UPM:
http://oa.upm.es/41530/1/PFC_ANGEL_MUNOZ_SAGASETA.pdf
- Zubillaga, A. (2010). *La accesibilidad como elemento del proceso educativo : análisis del modelo de accesibilidad de la Universidad*
- Complutense de Madrid para atender las necesidades educativas de los estudiantes con discapacidad*. Madrid: Universidad Complutense de Madrid .