**Propuesta Formal:**

***“Detector de Sonidos de Fallas Automotrices”***

Bonomi Ariel Matías

Marzo 2022

Universidad del Salvador

Ingeniería Informática

Proyecto final de Ingeniería Informática

# Capítulo 1. Introducción

Uno de los problemas más habituales para las personas que conducimos, es la detección

de los problemas que nuestros autos suelen tener a lo largo de su vida útil.

Todos conocemos ese típico “Ruido” que empiezan a aparecer con el uso cotidiano, los

cuales pueden indicar desde problemas leves, a problemas realmente urgentes que

deberían ser atendidos de inmediato.

Mi objetivo es poder llevarle a las personas que, como la gran mayoría no tienen un

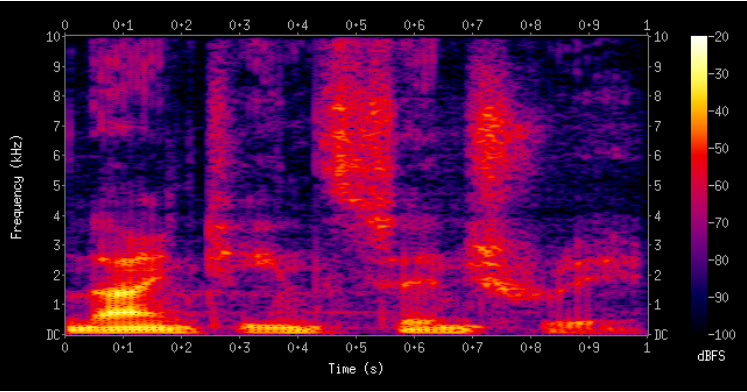
entendimiento profundo de los automóviles, una forma fácil de saber que le está pasando

su auto, porque existe este ruido, nivel de emergencia y sus posibles soluciones.

Todo esto a través de un detector de sonidos (Desde el propio teléfono), con el

cual poder grabar dicho sonido y conseguir el objetivo propuesto en el proyecto.

Se estudiaran los patrones de sonidos que puedan producir nuestros automóviles a través de un espectrograma y así poder encontrar similitudes con todos los sonidos que estan guardadas en nuestra base de datos.



Con este mismo principio trabaja la aplicación “Shazam” para detectar canciones con solo oír un fragmento de estas, de esta misma aplicación recolectare toda la información necesaria para el reconocimiento tan preciso de diferentes tipos de sonidos.

Y como funcionalidad extra, un lector de imágenes para que las personas, a través de fotos que le puedan

sacar al tablero del auto, sepa también el porqué de las luces que aparecen en dicho

tablero.

Se sabe que en el manual se brinda esta información, pero en esta era de digitalización las

personas buscan obtener respuesta lo más rápido posible, y el uso del manual en jóvenes

es cada vez menor. Por lo cual encontré una forma sencilla de poder realizar una consulta

sobre estas luces sencillamente a través de una fotografía.

Principalmente el detector de sonidos comenzara reconociendo aquellos sonidos que produzcan un sonido particular, entre los cuales podrían ser:

- Fuga en el caño de escape

- Pastillas de freno gastadas

- Problemas con la batería

- Problema con la correa

Esto debido a que algunos sonidos producidos pueden ser muy similares y representar muchos problemas diferentes, o bien un único problema que produzca una inmensa variedad de sonidos diferentes entre ellos.

Además de muchas otras barreras las cuales pueden ser:

**Barrera Cultural**: Se desconoce si las empresas automovilísticas quieren adaptar un software ajeno que apunte específicamente a los problemas del automotor, debido a que muchas agencias poseen personal especializado y herramientas especializadas para detectar estos problemas.

Al existir la posibilidad de que las personas conozcan lo que le esta sucediendo a su automóvil sin llevarlo a los lugares recomendados por el manual lo más probable es que estos intentarían solucionarlo por sus propios medios (Ya sea por comodidad o por temas económicos), perdiendo así garantía o la seguridad de arreglar los autos en los talleres oficiales.

**Barrera de Espacio**: Véase como espacio al almacenamiento necesario para todos los sonidos que tienen los automóviles, sin contar que cada modelo o marca tiene su propio sonido en cuanto a problemas.

**Barrera de Tiempo**: Esta barrera va de la mano con la anterior, la inmensidad de autos que existen hoy en día da como resultado un tiempo en la carga, de todos los sonidos que estos pueden hacer, muy grande, además del espacio que las bases de datos requerirán.

**Barrera Técnica**: La tecnología utilizada hoy en día para reconocer sonidos es escasa y más en el área del automovilismo donde algunos sonidos de un mismo problema pueden escucharse de maneras muy distintas, por lo cual una barrera importante será la del reconocimiento como tal de ciertos sonidos los cuales no son demasiados característicos, como un sonido metálico o de chapa.

Además, muchos problemas que poseen los autos no tienen ningún ruido hasta que el auto se descompone por completo, como podría ser la perdida de agua refrigerante y la subida de temperatura del automóvil. Estos problemas pasaran desapercibidos por la aplicación debido a que no producen ruidos significativos, hasta que el auto se rompe.

El proyecto busca solucionar el problema más grande de las personas con los automóviles, facilitando de manera importante el entendimiento de problemas de estos. En una era de digitalización la adaptación del Software va a ser inmediata y evitara miles de complicaciones a largo plazo.

El estudio de los sonidos y la recolección de datos se tiene que hacer de una manera única, debido a la complejidad de estos sonidos, lo cual agrega una relevancia importante al proyecto. En el enfoque veremos que el sistema contara de dos fases importantes y la conexión entre estas.

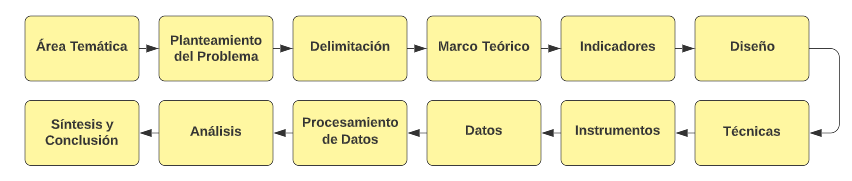
Si bien se utilizara como guía algunos softwares existentes, lo que se plantea en el proyecto llevara toda esta tecnología un paso más adelante, no solo comparando un único sonido, sino detectando miles de variantes de este mismo, ya que un mismo problema puede tener muchas variantes a la hora de ser detectados.

Hoy en día los automóviles de alta gama vienen con un sistema incorporado de detección de problemas, pero en Argentina, aproximadamente el 22% de la población puede acceder a estos vehículos. El otro 78% tiene autos de gama media o baja, por lo cual el proyecto adquiere una relevancia enorme para todo este gran sector de la sociedad.

# Capítulo 2. Metodología y procedimientos

Durante el proyecto se adopta una metodología de investigación con pasos a seguir para una correcta documentación de la misma, la cual facilita el entendimiento del problema y ayuda a tener una conclusión certera de todo lo planteado a continuación.

La metodología contara con doce pasos los cuales están representados en la figura 1, y además se acompaña con un diagrama de Gantt con el objetivo de marcar tiempos a cumplir y llevar un detalle con cada uno de los pasos de la metodología.

****

**Figura 1**

Área temática: El área temática son los problemas de los vehículos los cuales producen un sonido particular, esto acompañado con la tecnología de detección de audio para poder dar una explicación a estos ruidos que producen los problemas de los autos.

Planteamiento del problema: El problema a tratar esta enfocado en aquellas personas, mas bien jóvenes, que poseen autos usados o antiguos los cuales luego de un tiempo empiezan a tener fallas notorias, fallas las cuales produzcan un sonido particular.

Delimitación de la investigación: El proyecto tiene como delimitación principal a Argentina, debido a que los estudios realizados son sobre los vehículos en este país. Además, gran parte de la población argentina posee autos de gama media o baja, donde suceden estos sonidos con mucha más frecuencia.

Con la escalabilidad del proyecto en un futuro se podría pensar en expandir hacia Latinoamérica.

Marco teórico: Desarrollo de la información que respalda el proyecto tomando la bibliografía referenciada. Investigando se encuentra muchos estudios al respecto, no estrictamente en el área del automovilismo, o no al menos de la forma planteada en el proyecto, pero si con fines industriales y/o mecánicos.

Algunos documentos importantes encontrados podrían ser “Sistema inteligente para el tratamiento de ruidos”, investigación creada para la Universidad de Palermo o también “Diseño de sistemas de control pasivo de ruido en recintos industriales” de la asociación Argentina de Mecánica Computacional, entre muchos otros documentos.

Diseño concreto: El diseño concreto es llevar a cabo un dispositivo que registre sonidos para dar respuesta a problemas proveniente de vehículos. Llevando una nueva tecnología con base en la detección de audios.

Instrumentos: Para medir el nivel de éxito de la solución se hicieron exhaustivas pruebas hasta que el sistema dio resultados exitosos, luego se compartió con diferentes usuarios para que dieran su perspectiva sobre la solución y midan el nivel de éxito. Siendo exitoso cuando los problemas detectados sean efectivamente los correctos.

Datos: Se realiza el preprocesamiento de datos siguiendo los pasos siguientes pasos, seleccionados del Knowledge Discovery in DataBases (K.D.D), ya que este concepto comprende la limpieza, interpretación y descubrimiento de patrones en la información.   
- *Limpieza*: Consiste en remover toda la información no representativa del dataset seleccionado.

- *Selección*: Consiste en seleccionar las columnas del dataset que se apliquen al problema que se está tratando.

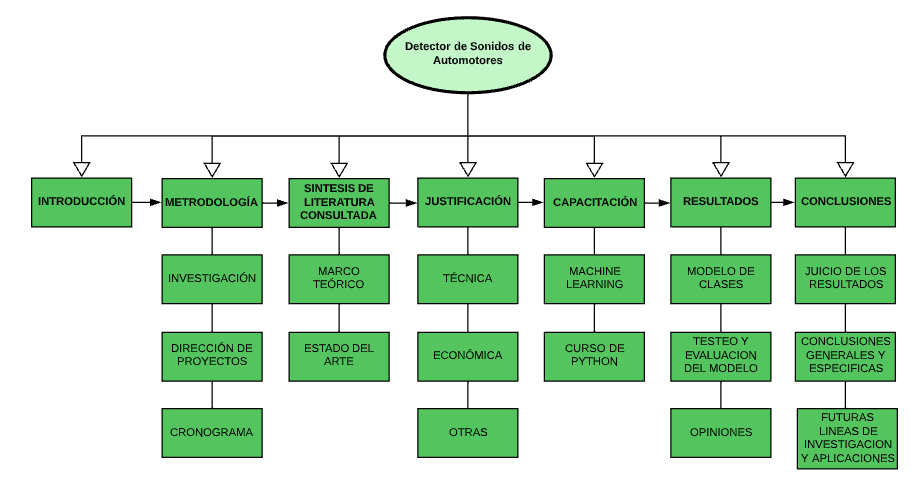
- *Transformación*: Consiste en realizar la transformación de los datos de modo tal, que puedan ser utilizados por los algoritmos y también para sacar mejor provecho de los modelos.

- *Minería de datos*: Consiste en seleccionar el algoritmo que mejor se adapte a nuestro problema y ejecutarlo para ver los resultados.

Análisis de datos obtenidos: Para el análisis se utilizó el paso de “Evaluación” del K.D.D que consiste en evaluar si la información y eficiencia obtenidas resultan relevantes para el problema planteado.

Conclusiones: Luego del análisis, se especifica los resultados obtenidos y se realiza una conclusión global en base a dichos resultados.

**La estructura de la metodología de dirección de proyectos planteada es la siguiente:**

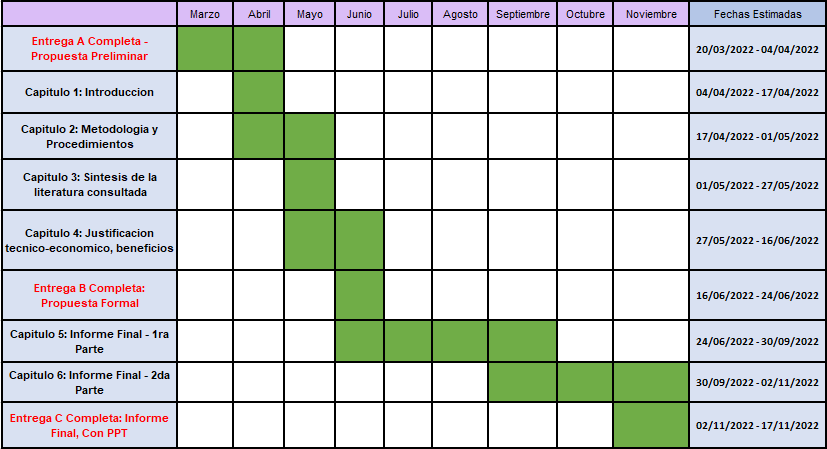


A modo de explicación, el grafico anterior puede ser leído de derecha a izquierda y de arriba abajo. Cabe remarcar, que el proyecto es desarrollado siguiendo la presente estructura en forma iterativa e incremental, sin necesidad de cumplir con un orden especifico. El orden de los elementos del gráfico no debe ser tomado en cuenta para su entendimiento, finalizando el proyecto cuando todas sus etapas estén completas.

**Cronograma**

El cronograma lo voy a presentar en forma de cuadro, explicando el paso a paso de la investigación que se ira realizando, acompañado con diagramas de Gantt para un entendimiento más grafico del tiempo y el proceso realizado.

Comenzando la primera semana de abril de 2022 y finalizando a mediados de noviembre de 2022, se cuenta aproximadamente con 8 meses para la finalización del proyecto.

**Cronograma general**:

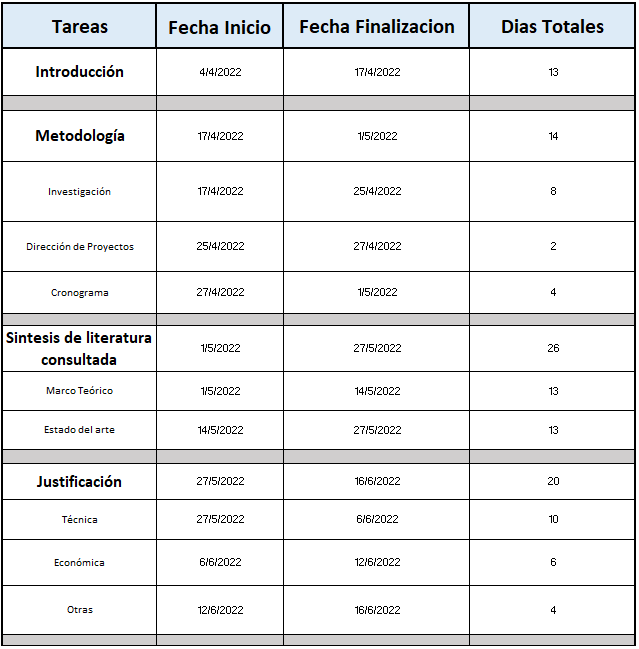
**Cronograma especifico:**

En el cronograma especifico se hizo un cronograma donde expliqué cada una de las tareas mostradas en la estructura de la metodología de dirección del proyecto.

Comenzando con la introducción y finalizando con las conclusiones del mismo.

Sera dividido en dos partes, la primera haciendo hincapié en entregar una propuesta formal del proyecto, y la segunda, en terminar con todas las tareas del informe final del mismo, concluyendo en la presentación final.

Primera parte:



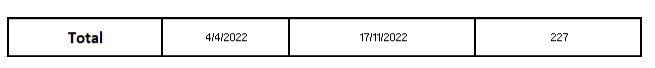
En esta primera sección del proyecto se realizó un enfoque de una entrega completa del informe formal en la fecha 16 de junio, con un total de 60 días pasados desde el inicio de la propuesta preliminar.

Segunda parte:



En la segunda sección del proyecto tal como se especificó anteriormente, comenzara con 4 meses de capacitación para poder llevar a cabo un prototipo del proyecto, para luego pasar por los resultados del mismo y obtener de esta manera conclusiones sobre el proyecto.

Una vez finalizado se hizo una presentación final del proyecto.

Total:

# Capítulo 3. Síntesis de literatura consultada

**Estado del Arte**

En esta sección del capítulo se hace un resumen de algunas investigaciones encontradas acerca de los temas a tratar en el proyecto.  
Revisar todo el conjunto de saberes o el desarrollo que se ha conseguido en el área en la cual se investiga es clave para lograr el objetivo propuesto para toda esta investigación.

El estado del arte se divide en dos importantes pilares, la detección de audio y una aplicación la cual aplica esta detección de audio.

**Detección de Audio**:

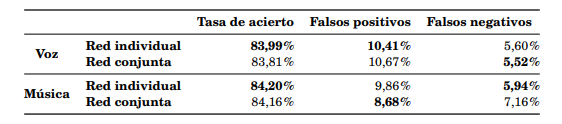
Primer Investigación:

* Título: “*Detección de voz y música en un corpus a gran escala de eventos de audio*”.
* Objetivos:   
    
  1. Estudio, comprensión y descripción de la base de datos AudioSet de Google.   
    
  2. Consulta del estado del arte del área, experimentos relacionados y resultados previos.   
    
  3. Diseño de modelos de clasificación basados en diferentes arquitecturas de redes neuronales.   
    
  4. Diseño y ejecución de experimentos para la evaluación del rendimiento de los modelos de clasificación propuestos.   
    
  5. Interpretación y discusión de los resultados obtenidos.   
    
  6. Generación de código orientado al manejo de la base de datos AudioSet y su uso en futuras investigaciones.
* Muestra: 77.396 segmentos de audio de 10 segundos (216 horas) pertenecientes a vídeos de YouTube, extraídos de la reciente base de datos AudioSet de Google.

Evaluados con arquitecturas de redes neuronales, basadas en DNN, CNN y LSTM.

* Palabras Claves: Aprendizaje automático, deep learning, redes neuronales, redes convolucionales, LSTM, audio, voz, música, reconocimiento de eventos de audio.
* Resultados:

Las redes entrenadas para las tareas de detección de voz (4.1.1), detección de música (4.1.2) y detección conjunta de voz y de música (4.1.3) alcanzan resultados muy similares en cuanto a tasas de acierto, que se recogen en la tabla 4.4. Las tasas de falsos positivos y falsos negativos se expresan como porcentaje sobre el total de segmentos clasificados. En ambos casos (música y voz), la precisión es levemente superior con las redes individuales, pero no lo suficiente como para considerarlo una diferencia significativa.



* Conclusión

Han sido evaluadas diferentes arquitecturas de redes neuronales, basadas en DNN, CNN y LSTM. Los resultados indican un rendimiento muy destacable de las arquitecturas basadas en redes convolucionales, que logran rendimientos muy similares en voz y en música, con tasas de acierto en torno al 84%. También se ha destacado el rendimiento de la red conjunta, que alcanza resultados semejantes a los de las dos redes individuales combinadas, pero contando aproximadamente el mismo número de parámetros entrañables que sólo una de ellas.

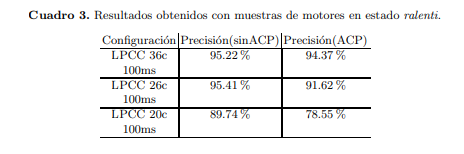
Segunda Investigación:

* Título: “*Análisis y Clasificación de Sonidos de Motores de Carros, para Diagnosticar Fallas, con Ayuda de Redes Neuronales*”.
* Objetivo:   
    
  El objetivo de la investigación es reconocer problemas en el motor a traves de su sonido, utilizando las vibraciones que este realiza cuando existe algún tipo de falla. Con ayuda también de redes neuronales.
* Muestra:

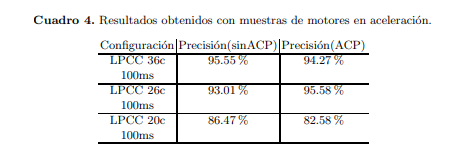
Las muestras de los sonidos fueron sacadas de tres talleres se especializan en motores de autos. Cada muestra grabada fue obtenida con ayuda de una grabadora digital ZicPlay y los archivos de salida fueron guardados en formato WAV, con las siguientes características: 8 khz, 16 bits y canal monoaural. Cada uno de estos archivos tiene un periodo de tiempo de 3 a 5 segundos, y fueron grabados en total 26 diferentes máquinas. Estas muestras fueron clasificadas en 4 categorías: motores en buenas condiciones (clase 1), problemas relacionados con el cigüeñal (clase 2), con pistones (clase 3) y con problemas en válvulas (clase 4)

* Resultados:

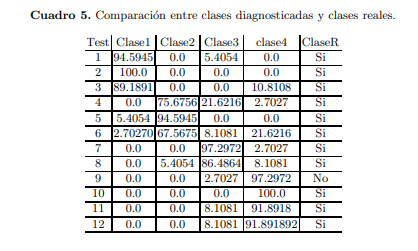
En el Cuadro 3, se pueden ver los resultados obtenidos usando muestras de máquinas en estado ralenti, el Cuadro 4, muestra los resultados obtenidos de muestras grabadas de máquinas en aceleración. Cada resultado, es el obtenido de la media aritmética de 10 experimentos previos cada uno. En ambos cuadros, la primera columna muestra el número de características extraídas por cada segmento de tiempo (ms).



Los resultados en ambas tablas muestran que la mejor precisión fue dada cuando usamos 36 y 26 LPCC por cada 100 ms, cuando usamos muestras de motores con aceleración; podemos decir que con 36 coeficientes por segmento de tiempo, el lapso de entrenamiento fue mayor y no había mucha diferencia en los resultados de reconocimiento, mostrando con esto, que 26 LPCC dan mejores resultados globales.



Tomando esto en cuenta, se seleccionaron al azar doce muestras de sonido en formato WAV, provenientes de ruido no usado en la etapa de entrenamiento, y separado para este fin; tomando una RNA previamente entrenada, se alimentaron estas muestras para demostrar la precisión que logra alcanzar dicha red. Los resultados obtenidos se muestran en el Cuadro 5.



* Conclusiones

Los resultados muestran que en un motor en malas condiciones existen diferencias acústicas, ´ comparándolo con un motor afinado o en buenas condiciones. Al menos estas diferencias entre sonidos pueden ser reconocidas por una Red Neuronal con Retraso en el Tiempo, claro está, dependiendo del tipo de falla que esté afectando la eficiencia del motor. Esta información puede ayudar al especialista del motor a dar diagnósticos rápidos, precios estimados por la reparación y permitirle trabajar directamente con el problema del motor. También podemos concluir que es viable trabajar en este problema e implementar un sistema barato que pueda ser usado por fabricantes de motores o dueños de talleres.

**Aplicaciones Similares**: ***Shazam***

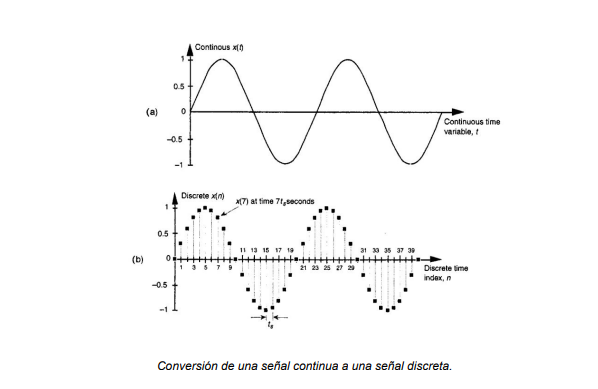
Titulo: “*Shazam: Cómo funciona el algoritmo de reconocimiento de canciones de la popular aplicación”*

Introducción: Shazam es una aplicación móvil que permite la identificación de canciones. La aplicación hace uso del micrófono que llevan incorporados la mayoría de smartphones para poder grabar una muestra de música de 10 segundos que se esté reproduciendo en el ambiente. A partir de esta breve muestra de audio, la aplicación es capaz de reconocer exactamente qué canción está sonando, brindándole al usuario información útil y precisa.

Funcionamiento: La aplicación se compone de un amplio catálogo de “huellas digitales” de audio que son almacenadas y catalogadas en una base de datos. Cuando un usuario graba con su teléfono una canción durante algunos segundos, esa grabación es procesada inmediatamente por la aplicación, creando una huella acústica (más adelante profundizaremos sobre este concepto) y enviándola a los servidores de Shazam a través de Internet. Una vez allí, se ejecuta una búsqueda de coincidencias en una base de datos que contiene millones de canciones (entre 8 y 11 millones, según algunas fuentes). Cuando existe una coincidencia, el usuario recibe en pocos segundos una respuesta con información de la pista original como el título de la canción, artista, álbum, e información adicional.

¿Como se logra este funcionamiento?

El micrófono del teléfono convierte la señal de presión de aire continúa producida por la misma canción, en una señal de voltaje analógica. De aquí se toma un muestreo que luego será analizado.



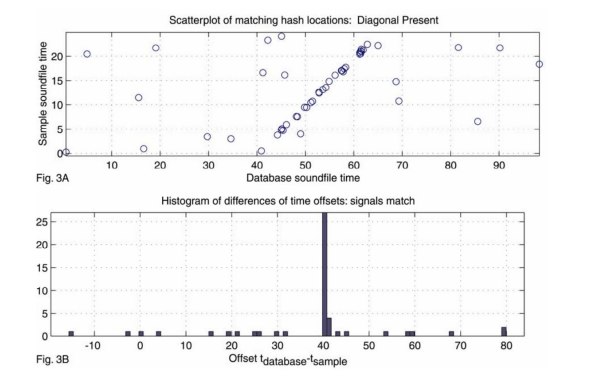
Si utilizamos esa frecuencia de muestreo, podremos reproducir la señal posteriormente a partir de las muestras tomadas.

Gracias a estas muestras se consigue a través de un largo procedimiento las denominadas “**Huellas Acústicas**” las cuales son un “resumen” digital del archivo, generado a partir de la señal de audio original. Gracias a este procedimiento, es posible identificar o localizar unívocamente un archivo de audio en una base de datos, sin necesidad de contar con información adicional.

Luego de analizar estas huellas acústicas, se pasa a la **detección de coincidencias**:

Ahora que hemos creado la huella acústica de nuestro archivo de audio, la información codificada en un conjunto de hashes con cada uno de los pares de puntos de anclaje de la grabación del usuario se envía a la base de datos de Shazam. Esta búsqueda devolverá las huellas acústicas de todas las canciones que contengan coincidencias de hash, es decir, que tengan los mismos pares de puntos de anclaje.

Si visualizamos este proceso en un diagrama de dispersión donde el eje y representa el momento en que se produce el hash en la grabación y el eje x representa el tiempo en que se produce el hash en el archivo de audio de la base de datos, podemos ver cómo se formarán los puntos coincidentes en la línea diagonal:



El desplazamiento en el tiempo se calcula restando el tiempo de ocurrencia del par de puntos de anclaje en la señal de entrada del tiempo de hash correspondiente en el archivo de audio de la base de datos. Si una cantidad significativa de hashes coincidentes tienen el mismo coeficiente de desplazamiento, se determina que esa canción es una coincidencia.

Este método de búsqueda de audio es lo suficientemente preciso para encontrar coincidencias incluso cuando la señal de entrada contiene ruido, como personas hablando, ruido de calle o incluso otras canciones sonando en simultáneo. Debido a que el número de hashes de puntos de anclaje creados por una huella acústica es mucho mayor que la cantidad de coincidencias de puntos de anclaje requeridas para devolver un resultado de búsqueda positivo, los hashes de puntos de anclaje que están enmascarados por el ruido externo no son suficientes para evitar que Shazam siga buscando coincidencias recursivamente en otros fragmentos de la señal.

**Marco Teórico**

En esta sección esta la recopilación de antecedentes y consideraciones teóricas en las que se sustenta este proyecto de investigación.

A continuación, se definen algunos conceptos importantes para el entendimiento del mismo.

# Mecánica

# Comportamiento del sonido y espectrogramas

Denominamos sonido a cualquier fenómeno que involucre la propagación de ondas mecánicas a través de un medio que este generando el movimiento vibratorio de un cuerpo.

La física del sonido es estudiada por la acústica, que trata tanto de la propagación de las ondas sonoras en los deferentes tipos de medios continuos como la interacción de estas ondas sonoras con los cuerpos físicos.

Estos a su vez son ondas muy complicadas por tener muchos componentes de diferente índole (armónicos, parciales y ruidos) que evolucionan muy rápidamente en el tiempo por lo que su análisis y estudio resulta bastante moroso.

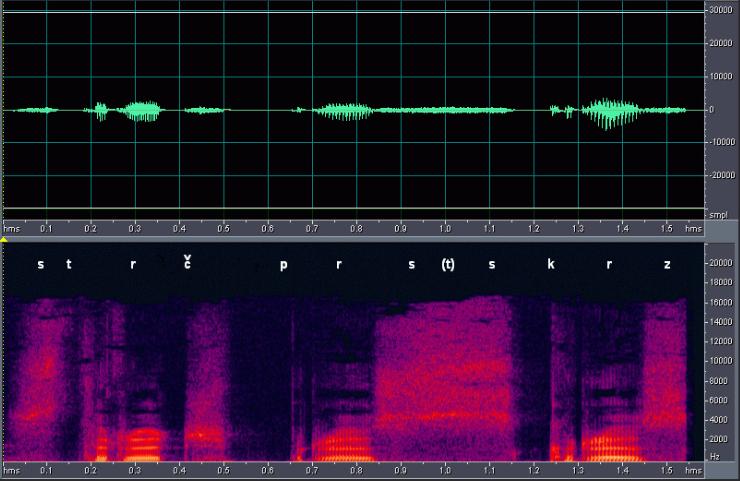
Cada sonido posee una propagación, resonancia, magnitud y velocidad única las cuales unidas generan un espectro de frecuencia, completamente diferentes uno al otro, y en consecuencia de esto nos resulta sencillo diferencias un sonido de otro, o reconocer de donde provienen estos.

¿Pero cómo puede un sistema reconocer la diferencia entre un sonido y otro?

Y aquí es donde entran los **Espectrogramas**.

Un Espectrograma es resultado de calcular el espectro de una señal por ventanas de tiempo de la misma. Resulta una gráfica tridimensional que representa la energía del contenido frecuencial de la señal según va variando esta a lo largo del tiempo.

Ejemplo de espectrograma



Estos se crean a partir del análisis por tramas, que es el primer paso a seguir, debido a la naturaleza variante de una señal de audio, se aplica una ventana para escoger un número limitado de muestras a procesar.

La elección de la ventana es un compromiso entre el tamaño de la trama, el procesado y el análisis de la señal.

Existen muchos tipos de ventanas, con más o menos resolución y esta elección va a depender del uso que se le quiera dar al espectrograma.

Algunas aplicaciones de estos espectrogramas son:

* Los espectrogramas son útiles para ayudar en la superación de defectos del habla y en la formación del habla para la parte la población que es profundamente sorda.
* Los estudios de la fonética y síntesis de voz se han facilitado a través del uso de espectrograma.
* Al invertir el proceso de elaboración de un espectrograma, es posible crear una señal donde el espectrograma es una imagen arbitraria. Esta técnica se puede utilizar para ocultar una imagen en un pedazo de audio y ha sido empleado por varios artistas de la música electrónica.
* Algunos tipos de música moderna se crean utilizando un espectrograma como medio intermedio, para cambiar la intensidad de diferentes frecuencias a través del tiempo, o incluso para crear otros nuevos, mediante la elaboración y después la transformación inversa.
* El Espectrograma se puede utilizar para analizar los resultados al pasar una señal de prueba a través de un procesador de señales, como un filtro para comprobar su rendimiento.
* Análisis frecuencial de cualquier señal audiovisual
* Visualización de señales eléctricas y de comunicaciones.
* Análisis musical.
* Aplicaciones acústicas.
* Aplicaciones sismológicas.
* Análisis de señales biomédicas.
* Extracción de características.

# Detección de sonidos

La herramienta principal que se utiliza para la detección de sonidos son los micrófonos, este es un elemento que transforma una señal acústica de un audio en una señal eléctrica.

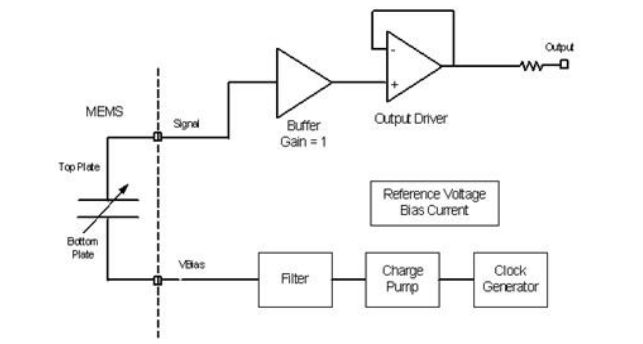
Existen muchos tipos diferentes de micrófonos que utilizan diferentes técnicas para realizar la transformación de la señal acústica en la señal eléctrica.

Todos los micrófonos se basan en la misma propiedad de transducir una variación de presión a una variación de tensión y varían en el concepto físico del procedimiento.

Los micrófonos se dividen en una gran clasificación, pudiendo ser dinámicos, de carbono, piezoeléctrico, entre muchos otros, pero todos ellos a su vez se agrupan en dos tipos principales, los **analógicos** y los **digitales**.

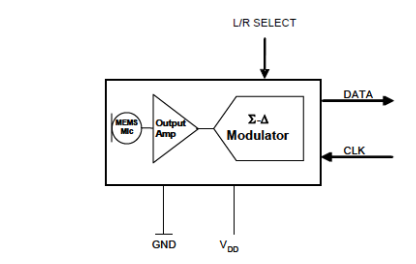
**Los micrófonos analógicos** son del tipo silicio, también llamados “MEMS”. Suelen estar construidos en encapsulados para montaje superficial para mayor comodidad a la hora de integrarlos en un circuito impreso para alguna aplicación. Su reducido tamaño permite poder colocarlos en el mejor lugar acústico para la aplicación. Las dimensiones mínimas rondan los 3mm de largo, 2 mm de ancho y 0.88 mm de altura.

Estructura de micrófono analógico



Y luego están los **micrófonos digitales**, que, a diferencia de los micrófonos analógicos, tienen la característica de llevar un conversor analógico-digital en el propio dispositivo de transducción para tener en la salida una señal digitalizada y en formato PWM. Esta mejora de prestación en la estructura de los micrófonos de silicio permite reducir el ruido del sistema considerablemente ya que no hay ni cables ni conexiones entre el codificador y el micrófono y permite eliminar directamente dicho componente lo cual nos deja la oportunidad de crear dispositivos más pequeños y con menos consumo.

Estructura de micrófono digital



Por último, es necesario nombrar a los **Códecs Audio** que es una abreviatura de codificador-decodificador.

El códec es un dispositivo electrónico que une las funciones de codificador y decodificador. Las funciones que realizan son muy simples. Convierten una señal analógica de audio o vídeo en una señal de audio o vídeo digital. Transforman la señal en primer lugar y luego la envían a través de un canal de comunicaciones y en un formato específico para que sea procesado en un dispositivo electrónico digital como por ejemplo un DSP.

Lo que importa de estos “Códecs” son sus aplicaciones, donde las mas comunes suelen ser en cámaras de video, cámaras fotográficas, tabletas digitales, teléfonos, ordenadores, y todos aquellos dispositivos que requieran un detector de sonido. En todos estos sistemas existe algún micrófono o sensor de fuente analógica que ha de ser transformada para su almacenamiento, tratamiento o transmisión.

Un teléfono por ejemplo recoge una señal de voz hablada, que ha de ser transformada a una señal digital para poder ser procesada y trasmitirá correctamente.

# Almacenamiento de Sonidos (AudioSet)

Uno de los problemas mas grandes a la hora de trabajar con audio/sonido siempre fue la inmensa cantidad de espacio que se requiere para almacenar cada uno de estos dentro de una base de datos. Sabiendo que un audio de 60 segundos pesa alrededor de 138KB (0,138MB) el solo hecho de pensar en crear aplicaciones donde se tenga que mantener un registro de estos hacia muy complicado llevar a cabo cualquier proyecto relacionado. Es por esto mismo que Google se vio impulsado a publicar lo que denominaron como “AudioSet”, basado en los conocidos Dataset.

AudioSet es una base de datos publicada por Google en 2017 para impulsar el problema de reconocimiento de eventos acústicos. Además de tener una gran variedad de sonidos para reconocer, con mas de 500 clases, el numero de ejemplos es mucho mayor que en los otros casos, con mas de dos millones de ejemplos.

Esta gran cantidad de ejemplos ha sido recopilada a partir de videos de YouTube, de los cuales se han extraído 10 segundos de audio, lo que hace que el total AudioSet cuente con mas de 5000 horas de audio. Al provenir de videos que no han sido grabados para este propósito, los sonidos no están aislados y pueden solaparse. Las 527 clases están organizadas bajo una estructura jerárquica con profundidad máxima 6, siendo los nodos más altos categorías de sonidos y los más bajos sonidos específicos. Cada ejemplo puede estar etiquetado como una o más clases. Los datos están disponibles en dos formatos. El primero contiene los enlaces a los videos de forma que se puede acceder directamente al audio para tratarlo con libertad, mientras que el segundo contiene las salidas de una red de embedding con la que Google ha tratado previamente el audio, de modo que se puedan entrenar modelos sin necesidad de procesar señales de audio

# Data Engineering, Data Science y Data Analytics

Se conoce como Data Engineering

**Capítulo 4. Justificación técnico-económica**

**BIBLIOGRAFIA**

“Shazam: Cómo funciona el algoritmo de reconocimiento de canciones de la popular aplicación” – Guido Sirna – ACADEMIA.

PFC\_AdquisicionAudioCodecsDSP\_MGH\_2011 (uam.es) - UNIVERSIDAD AUTÓNOMA DE MADRID ESCUELA POLITÉCNICA SUPERIOR