# Sistema Automatizado De Monitoreo De Pauta Publicitaria Con Base En Herramientas Speech To Text.

**Realizado por:**

**Luis Enrique García Moreno**
**Jorge Esteban Aguilar Chacón**



**UNIVERSIDAD SANTO TOMÁS**
**FACULTAD DE INGENIERÍA ELECTRÓNICA**
**BOGOTÁ D.C.**
**2022**

# Sistema Automatizado De Monitoreo De Pauta Publicitaria Con Base En Herramientas Speech To Text.

**Proyecto de grado presentado como requisito para optar al título de INGENIERO ELECTRÓNICO**

**Director: Ing. Dario Alejandro Segura.**

**UNIVERSIDAD SANTO TOMÁS DE AQUINO**
**FACULTAD DE INGENIERÍA ELECTRÓNICA**
**BOGOTÁ D.C.**
**2022**

**Dedicatoria**

Esta tesis está dedicada a nuestros padres quienes con su altruismo y esfuerzo nos impulsaron por el camino del saber para construir un mejor futuro personal y profesional.

A nosotros, como amigos y como compañeros durante toda nuestra etapa como estudiantes, pues sin el apoyo mutuo hubiese sido imposible.

Finalmente, a todas las personas involucradas en nuestro proceso como profesionales, que con una palabra, gesto, oración y acto de bondad nos acompañaron y nos impulsaron a continuar con nuestro sueño de ser profesionales.

El trabajo de grado titulado **Sistema Automatizado de Monitoreo de Pauta Publicitaria con Base en Herramientas Speech to Text**, realizado por los estudiantes **Jorge Esteban Aguilar Chacón y Luis Enrique García Moreno**, cumple con los requisitos exigidos por la **Universidad Santo Tomás** para optar por el título de **Ingeniero Electrónico**.

# Agradecimientos

Nuestro profundo agradecimiento a todo el personal educativo que hace parte de la Universidad Santo Tomás, por brindarnos consejo personal y profesional, por confiar en nosotros y abrirnos la puerta para culminar esta etapa educativa.

Del mismo modo, a la Facultad de Ingeniería Electrónica y a todos nuestros maestros, en especial al Ingeniero Dario Alejandro Segura, quien con sus valiosos conocimientos, dedicación, paciencia y profesionalismo nos guio durante todo nuestro proceso como estudiantes hasta la última instancia, siendo nuestro principal colaborador y asesor de tesis.

Finalmente, agradecer a nuestra familia y amigos, quienes nos ayudaron a crecer como personas inculcando los valores que hoy nos caracterizan.

# Tabla de contenido

# Tabla de figuras

# Tabla De Tablas

# Resumen

El objetivo de esta tesis es diseñar y evaluar un sistema de monitoreo de pauta publicitaria para detectar la emisión de un comercial dentro de la transmisión de una emisora de radio, aplicando herramientas Speech To Text (STT), teniendo en cuenta que estos sistemas tienen grandes alcances en diversos campos y que la publicidad en radio es un foco importante de dinero en Bogotá. Se ha desarrollado una investigación de cuáles serían los módulos necesarios para implementar un prototipo del sistema el cual sea capaz de determinar si una pauta publicitaria fue emitida. El sistema fue diseñado pensado en módulos por separado, en cascada, para los cuales se implementaron tres (3) algoritmos pertinentes individualmente con el fin de explorar las mejores soluciones para cada uno de ellos y fue evaluado basado en unos criterios de rendimiento definidos con cinco casos concretos, grabaciones en las cuales se encontraban las pautas publicitarias en cuestión.

# 1. Introducción

Partiendo del hecho de que la publicidad emitida en radio es un foco de dinero significativo en la ciudad de Bogotá, y entendiendo que en esta última década se ha visto en alta competencia con respecto a las redes sociales y la televisión, esta se sigue manteniendo en vigor. Los medios de comunicación ponen a disposición espacios en su programación con precios fluctuantes según diversos parámetros predeterminados. Las empresas y negocios que pagan por este servicio de publicidad en radio tienen la necesidad de verificar que efectivamente fue emitido al aire el servicio por el cual pagaron.

Durante el nuevo siglo los sistemas automatizados han generado gran desarrollo e impacto en la sociedad, esto debido a que se pueden generar soluciones a diversos problemas de la actualidad y con gran probabilidad, también para el futuro, con lo cual se piensa automatizar un sistema que permita realizar un monitoreo de una pauta publicitaria aplicando herramientas Speech To Text (STT) para detectar la emisión de un comercial dentro de la transmisión de una emisora de radio.

# 2. Problema

## 2.1 Formulación del problema

¿Qué componentes debe tener un sistema de monitoreo de pauta publicitaria basado en herramientas STT para que sea funcional?

## 2.2 Planteamiento del problema

En la actualidad existen diversos sistemas automatizados aplicados a la verificación, monitoreo y auditoría de pautas publicitarias en medios audiovisuales. Estos sistemas son utilizados en empresas instauradas a nivel mundial en temas de publicidad y marketing, para el caso latinoamericano algunas de estas empresas son Auditsa, Media5 Corporation, etc. A raíz de que son empresas comerciales, no se tiene conocimiento explícito de cómo operan dichos sistemas, con lo cual, en este proyecto, se aborda la problemática de identificar los componentes de un sistema que pueda reconocer un texto implícito en una señal audible basándose en herramientas Speech to text (STT) con el objetivo de monitorear una pauta publicitaria por la cual una empresa haya pagado, verificando de este modo que fue emitida al aire.

## 2.3 Delimitación del problema

Se busca investigar acerca de cómo utilizar herramientas STT para implementarlas en un sistema automatizado de monitoreo y certificación de pautas que sean descritas únicamente en el idioma español. Se realiza la captura de una señal audible, esta señal será obtenida a través de la salida de audio de un radio receptor y será enviada a un dispositivo encargado de realizar la digitalización de la señal. El sistema no contemplará una interfaz de gestión de la información debido a que el lineamiento del problema se ciñe a los componentes del sistema de monitoreo de una pauta publicitaria y no a la gestión de la información que se extrae de este. El estudio por realizar contempla la evaluación de al menos tres (3) algoritmos pertinentes por cada componente que lo requiera de acuerdo con la metodología, además de esto, para la evaluación del sistema, se decretan cuatro criterios específicos:

- Los falsos positivos: este criterio se compone del número de veces en las que la pauta publicitaria no ingresa al sistema y es verificada erróneamente.
- Los falsos negativos: este criterio se compone del número de veces en las que la pauta publicitaria ingresa al sistema y es rechazada erróneamente.
- Porcentaje de aciertos: este criterio se compone del número de veces en que al sistema ingresa la señal audible con la pauta publicitaria y es verificada correctamente.
- Velocidad de verificación: La velocidad de verificación será medida únicamente en las pautas publicitarias verificadas acertadamente, desde el momento en que la señal audible entra al sistema, hasta que se indica la verificación.

Un caso se compone de dos grabaciones de audio en las cuales se encuentra la misma pauta publicitaria, el sistema diseñado será evaluado con cinco (5) casos.

# 3. Justificación

La publicidad emitida en radio es una de las más pagadas en la ciudad de Bogotá, debido a que este medio sigue en pie, aunque tenga medios directos de competencia como las redes sociales o la televisión. Las ventajas que se evidencian de la publicidad de radio con respecto a otros tipos de publicidad son notables en ciertos aspectos, aunque principalmente en su omnipresencia, pues cualquier teléfono celular, equipo de sonido, computador y televisor tienen acceso a la radio. Partiendo de este punto, se origina el proceso de certificar los contenidos y/o características de un producto, generada principalmente por la desaparición de las relaciones directas entre el productor y el consumidor, las cuales constituían un factor de confianza para con el consumidor (Pons, Jean-Claude; Sirvardiere, Patrick. 2002).

Las cadenas radiales ofrecen espacios para publicidad a cualquiera que quiera difundir sus cuñas/pautas empresariales o políticas, y estos tienen el deber de seleccionar una cuota variable con base en unos parámetros predeterminados. Las empresas que pagan esta publicidad radial tienen la necesidad de verificar que efectivamente las pautas, por las cuales pagaron, fueron emitidas al aire las veces exactas en el tiempo correcto. A través del nuevo siglo los sistemas automatizados han generado gran desarrollo e impacto en la sociedad, esto debido a que se pueden generar soluciones a diversos problemas del presente y, con gran probabilidad, también para el futuro, con lo cual, se puede automatizar un sistema que permite realizar un monitoreo de una pauta publicitaria aplicando herramientas Speech To Text (STT) para detectar la emisión de un comercial dentro de la transmisión de una emisora de radio.

# 4. Objetivos

## 4.1 Objetivo general

Diseñar y evaluar un sistema de monitoreo de pauta publicitaria para detectar la emisión de un comercial dentro de la transmisión de una emisora de radio, aplicando herramientas Speech To Text (STT).

## 4.2 Objetivos específicos

- Identificar los componentes de un sistema de búsqueda de texto en una señal de audio para desarrollar un prototipo basado en herramientas STT.
- Identificar algunos de los diferentes algoritmos utilizados en cada uno de los componentes del sistema para seleccionar el más acorde a los criterios de rendimiento definidos.
- Evaluar el sistema con base en los criterios de rendimiento definidos para un sistema de monitoreo de pauta publicitaria.

# 5. Marco Referencial

En este capítulo se documentan conceptos importantes relacionados con el desarrollo del proyecto.

## 5.1 Antecedentes

### *5.1.1 Sistema De Conversión Speech To Text En Tiempo Real Utilizando Filtro De Kalman Bidireccional En Matlab.*

Se implementó un sistema de conversión de palabras/discurso a texto en tiempo real en un contexto exacto al que el locutor lo pronuncia. Se utilizó el diseño de un filtro de Kalman bidireccional no estacionario que incrementara la habilidad de este sistema para realizar la tare, ya que el filtro de Kalman ha sido evaluado como el mejor estimador de ruido en ambientes resonantes no estacionarios. Como propósito de este proyecto se tuvo el introducir un nuevo sistema de reconocimiento el cual fuese computacionalmente más simple y robusto contra el ruido que el sistema de reconocimiento de discurso basado en HMM. Para llevar a cabo este proyecto se implementó una base de datos propia creada por los autores para aumentar la flexibilidad del sistema, y adicionalmente una base de datos TIDIGIT para su precisión en la comparación con el sistema de reconocimiento de discurso HMM.

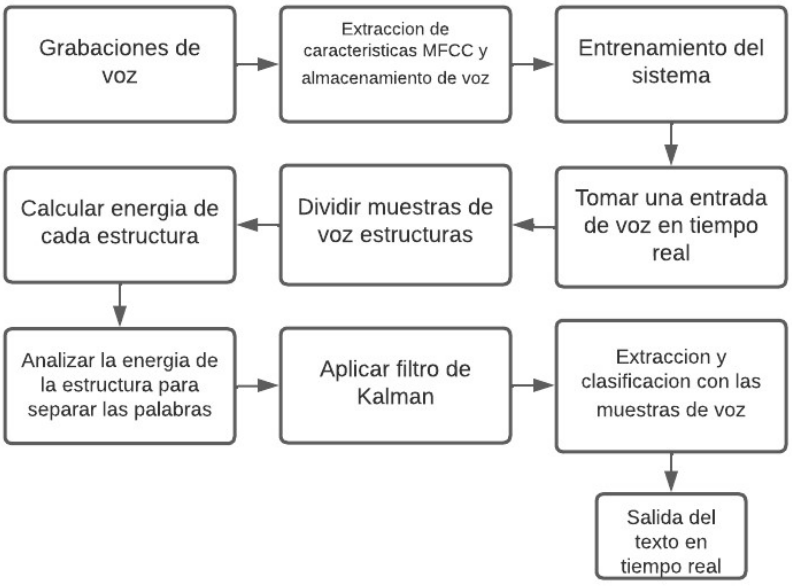


**Figura 1.** ***Proceso de reconocimiento de voz.***
**Tomado de: (Sharma, Neha. Sardana, Shripa, 2016. Jaipur, India).**

El proceso utilizado para la puesta en marcha del sistema se describe en la figura 1.

Para evaluar el sistema se realizaron diferentes pruebas con ambientes en condiciones de ruido diferentes. En primer lugar, el sistema fue puesto a prueba en un ambiente cerrado, con el ruido de un ventilador, como resultado se obtuvo un 100% de precisión. Posterior a esta prueba, se puso el sistema en un laboratorio de la universidad de Chandigarh con el mismo ruido del ventilador y música de fondo, en este escenario el sistema dio un resultado del 80%. Basados en esto, se determina que el sistema fue un 90% preciso. Esta precisión puede variar dependiendo de los escenarios en los que se encuentre el sistema (Sharma, Neha. Sardana, Shripa, 2016. Jaipur, India).

### *5.1.2 Convertidor Speech To Text basado en SVM (Support vector machine) para idioma turco*

Se ha desarrollado un SVM basado en el idioma turco con el fin de desarrollar un sistema Speech To Text. Con este sistema, MFCC (Mel Frequency Cepstral Coefficients) han sido aplicados con el fin de extraer características del turco. Debido a que el turco es un idioma basado en fonemas, su estructura morfológica fue tomada en consideración durante el desarrollo del sistema. A diferencia de los clasificadores de multiclase que son utilizados en sistemas de reconocimiento de voz basados en SVM-MFCC, se implementa un nuevo sistema clasificador SVM el cual utiliza menos clases en capas, incrementando de este modo el número de capas multiclase. De igual manera, se propuso un nuevo algoritmo de comparación del texto el cual utiliza únicamente secuencias de fonemas para medir similitudes. (Burak, TombaloĞlu. Hamit, Erdem. 2017, SIU).

### *5.1.3 Aplicación de reconocimiento Speech To Text y traducción computarizada para soportar comunicación multilenguaje en un proyecto de aprendizaje cultural.*

Se implementa un sistema de reconocimiento de voz Speech To Text y traducción computarizada (CAT) para soportar comunicaciones entre diferentes idiomas con el fin de participar en un proyecto de aprendizaje multi cultural, en el cual los participantes estaban comprometidos con compartir información en su lengua natal la cual fuese transmitida pasando por el sistema traduciéndola al idioma elegido.

Los participantes del proyecto hablan y el sistema de reconocimiento de voz genera un texto a partir de las voces de entrada. Inmediatamente, el sistema de traducción computarizada realiza la traducción al inglés, el cual una vez traducido se publica en plataformas de comunicación social. La herramienta Speech To Text implementada en este proyecto fue Google, obteniendo unos porcentajes de similitud de 98.15% para el español, 98.02% para el ruso y 97.95% para el francés, esto debido a que la base de datos de traducción computarizada es más amplia para dichos idiomas (Rustam Shadiev, Barry Lee Reynolds, Yueh-Min Huang, Narzikul Shadiev, Wei Wang, Rai Laxmisha, and Wanwisa Wannapipat. 2017, Tainan, Taiwan).

### *5.1.4 Monitoreo y auditoria de pautas publicitarias*

El monitoreo en servicios de Broadcast tiene como pionero en Latinoamérica a una empresa llamada Auditsa; es una empresa de monitoreo y auditoría de publicidad en medios a nivel

internacional, con más de 10 años de experiencia, la cual utiliza una tecnología de reconocimiento que trabaja con base a un algoritmo de reconocimiento de audio y un procesamiento digital de señales lo cual permite detectar y entregar información de las transmisiones en tiempo real monitoreando de esta manera dicha señal captada, la cual se ciñe a la condición de tener una duración de 6 segundos o más.

Para el monitoreo y auditoria de una pauta publicitaria se realiza proceso ilustrado en la figura 2:

**Figura 2.** ***Proceso de monitoreo y auditoria de la publicidad.***
**Tomado de: Auditsa Mexico. Auditsa Ad Tracking Presentación.**

Spot es como se le llama a la publicidad de un servicio o un producto durante una programación televisiva o radial. Sabiendo esto, el primer paso es cargar el spot a la base central y posteriormente se recibe la transmisión la cual se digitaliza, para luego ser analizada con el fin de generar una huella asociada a la señal, en dado caso que la huella generada coincida con la carga del material se detecta un HIT, que es como se le denomina a una verificación exitosa (Auditsa Mexico, 2016).

## 5.2 Marco teórico

### *5.2.1 Grabación de audio*

La cadena de audio, se refiere al proceso por el que pasa el sonido durante la grabación y la reproducción. Esto requiere ciertas técnicas para transformar la señal original en otra señal de otra naturaleza y calidad en varias etapas. Esto lo sitúa en apoyo a su conservación y posterior uso.

Refiriéndose a lo tecnológico, se pueden encontrar diversidad de sistemas tales como: captadores de sonido, amplificadores, procesadores, digitalizadores etc. Un factor importante para tener en cuenta es que se debe mantener la calidad de todos los elementos implicados en la cadena de audio, pues como ocurre en otros procesos, la calidad final de la señal es determinada por el elemento de peor calidad involucrado en la cadena. Debe recordar que hay una gran cantidad de elementos técnicos que se pueden utilizar dentro de su cadena de

sonido. Cada uno de ellos requiere un conocimiento profundo de su uso y sus limitaciones para poder superarlos. La insuficiencia puede conducir a efectos indeseables.

### *5.2.2 Grabación analógica y digital*

La grabación analógica del sonido es el proceso mediante el cual una onda continuamente variable es convertida en un patrón de modulación sobre un medio físico. Este patrón debe ser análogo a la señal original, y podría ser almacenado y reproducido posteriormente, de manera que, en el caso de usar un micrófono, para trasformar las variaciones de presión sonora en variaciones de voltaje, este voltaje variable se puede adaptar después a distintos medios: cambios del patrón de magnetización de la cinta, alteraciones de las zonas claras y oscuras en una banda sonora de cine, desviaciones variables en el surco, etc. El problema radica en que el sistema de reproducción es incapaz de distinguir entre señales deseadas y señales no deseadas, siendo estas el resultado del proceso de grabación que termina siendo imperfecto. Por ejemplo: la aguja de un tocadiscos no puede distinguir si el movimiento que sufre es debido a un arañazo en el disco o a una desviación importante como consecuencia de un transitorio fuerte en la música.

Por otro lado, la grabación digital transforma la onda generada por un micrófono en una serie de números, cada uno representando un instante específico en el tiempo. Estos números se almacenan de manera codificada, lo que permite que el sistema detecte si la señal reproducida es correcta o no. El audio digital es mucho más tolerante con los canales de grabación defectuosos que el audio analógico, y las distorsiones e imperfecciones durante el proceso de grabación no tienen por qué afectar necesariamente la calidad del sonido, tanto en grabación como en reproducción (Francis Rumsey. Tim Mc Cormick).

### *5.2.3 Clasificación de señales de audio y señales de voz*

Relacionado con temas de reconocimiento de audio y voz en una señal, existe una técnica que se basa en el uso de reconocimiento de patrones de flujo espectral (SF). La organización Internacional de Normalización junto con la Comision Electrotecnica Internacional generan modelos que estandarizan las transmisiones multimedia, lo cual es llevado a cabo específicamente por el grupo MPRG por sus siglas en inglés (Moving Picture Experts Group).

Para el reconocimiento preciso de un patrón de audio se utiliza el modelo de probabilidad de mezcla Gaussiana, que tratándolo de una manera simple se le puede denominar un tipo de algoritmo de agrupamiento. Como su nombre lo indica, cada grupo se modela de acuerdo con una distribución gaussiana diferente (Sangkil Lee, Jieun Kim, & Insung Lee, 2012). Este enfoque flexible y probabilístico para modelar los datos significa que, en lugar de tener asignaciones difíciles en grupos como en el algoritmo de Lloyd (k-means), se tienen asignaciones suaves. Esto significa que cada punto de datos podría haber sido generado por cualquiera de las distribuciones con una probabilidad correspondiente. En efecto, cada distribución tiene alguna "responsabilidad" para generar un punto de datos particular. Además de este algoritmo, se utiliza un teorema conocido como maximización de

expectativas por sus siglas en inglés (Expectation-Maximization). El algoritmo EM consta de dos pasos, un paso E o paso de Expectativa y un paso M o paso de Maximización. Se dice que se tienen algunas variables latentes $\gamma$ y puntos de datos X. El objetivo es maximizar la probabilidad marginal de X dados los parámetros (denotados por el vector θ). Esencialmente se puede encontrar la distribución marginal como la unión de X y Z y sumar sobre todas las Z (regla de suma de probabilidad) (Foley, Daniel. 2011).

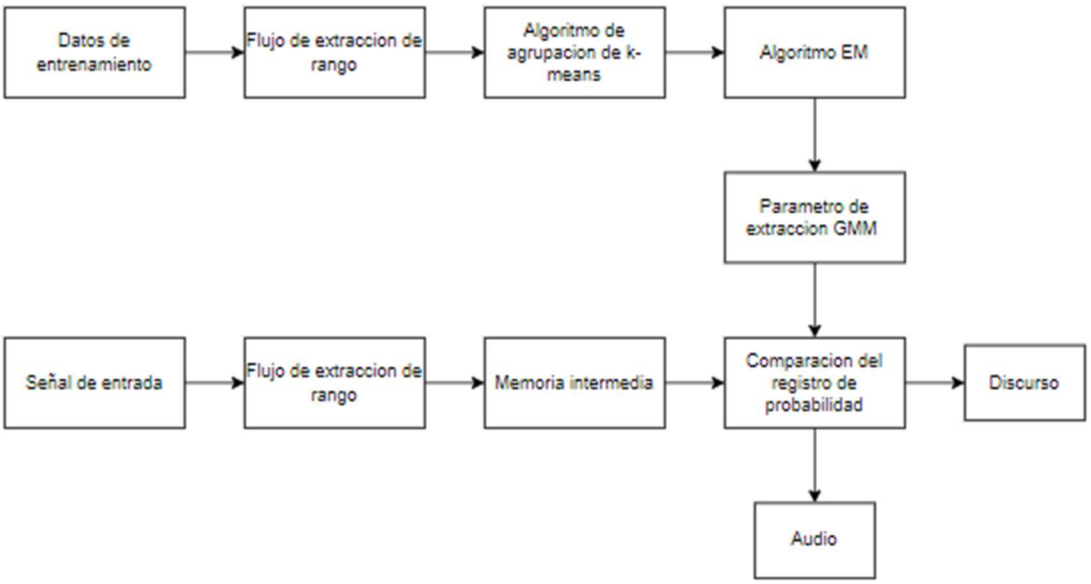


**Figura 3.** ***Diagrama de bloques propuesto para el algoritmo de clasificación de audio/voz***
**Fuente: (Sangkil Lee, Jieun Kim, & Insung Lee, Oct 2012)**

En la figura 3 se observa el diagrama de bloques propuesto para la ejecución de los algoritmos mencionados anteriormente, con el fin de clasificar la voz y el audio de una señal.

### *5.2.4 Digitalización de audio*

La mayoría de las señales prácticas son analógicas tales como: voz, biológica, sísmica, radar, sonar y varias señales de comunicación, como audio y video. Para poder procesar digitalmente una señal analógica, primero debe convertirse a formato digital. En otras palabras, se requiere convertirlo en una secuencia de precisión finita. (Proakis & Manolakis, 2007). La conversión A/D es un proceso de tres pasos, que son los siguientes:

Muestreo: Consiste en tomar *n* muestras (medidas) de valores de señal por un tiempo determinado, con *n* niveles de tensión por dicho tiempo. Por ejemplo, para realizar el muestreo de canales telefónicos de voz, ocho mil (8.000) muestras por segundo lo que es equivalente a una (1) muestra cada 125 microsegundos, entregaría la información necesaria para reconstruir la señal muestreada, pero en tiempo discreto.

Cuantificación: En la cuantificación una señal muestreada se transforma en una señal de valor discreto en puntos de tiempo discretos, es decir, una señal digital. El valor de cada muestra está dado un valor discreto relacionado con los niveles de tensión obtenidos en la etapa de muestreo. La diferencia entre la muestra no cuantificada $x(n)$ y la salida cuantificada $x_q(n)$ es el error de cuantificación, el cual se define en la ecuación 1 como la distancia entre la señal original y la señal cuantificada:

$$\boldsymbol{e[n] = y[n] - x[n] = Q[x(n)] - x[n]} \quad \textbf{(1)}$$

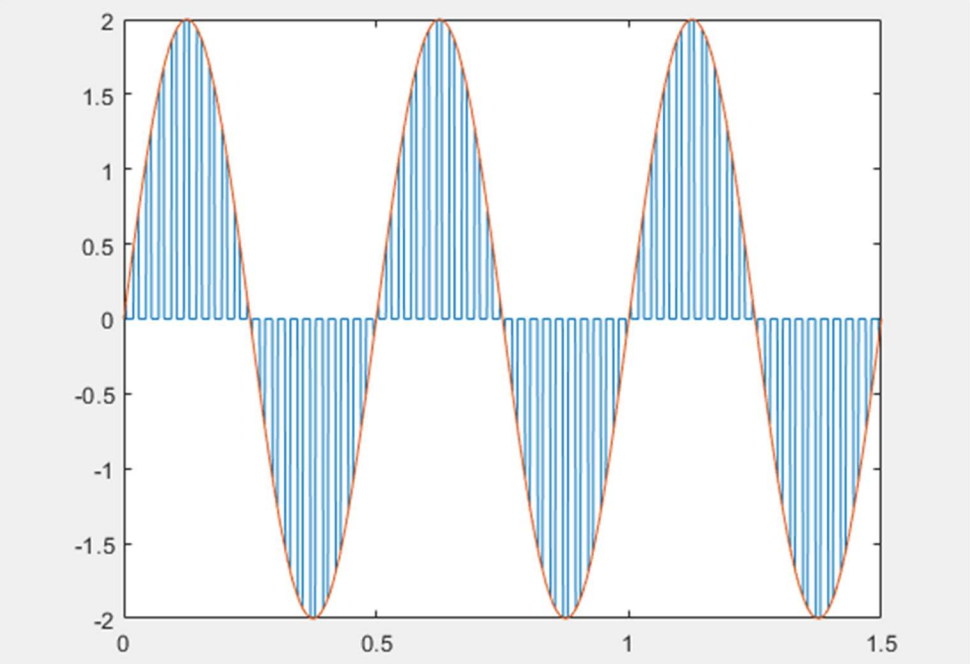


**Figura 4.** ***Representación de una señal analógica.***
**Tomado de: Autor.**

En la figura 4 se observa la representación de una señal analógica, con el color rojo. La señal de color azul, la cual es el resultado de la cuantificación sobre la señal roja. Para este paso se encuentran también algunos conceptos básicos, como:

Número de bits que especifican el número de estados de salida para el cuantificador. El nuevo valor que asume la señal cuantificada, el nivel de cuantificación dado por el número de bits en el cuantificador, y el rango dinámico que indica los valores mínimo y máximo de la señal cuantificada. (Marta Ruiz Costa-jussà & Helenca Duxans Barrobés).

Codificación: Los valores discretos $X_q(n)$ son representados por una secuencia de bits binaria. La conversión análoga a digital se realiza mediante un dispositivo que toma $X_a(T)$ y produce un numero codificado en binario (Proakis & Manolakis, 2007). El muestreo y la cuantificación son ejecutadas en dicho orden respectivamente. En lo que respecta al procesamiento de audio, las señales se pueden convertir de digital a analógica. El proceso de convertir una señal digital en una señal analógica se denomina D/A. Todos los convertidores D/A utilizan alguna forma de interpolación para conectar los puntos de una señal digital. Su precisión depende de la calidad del proceso de conversión.

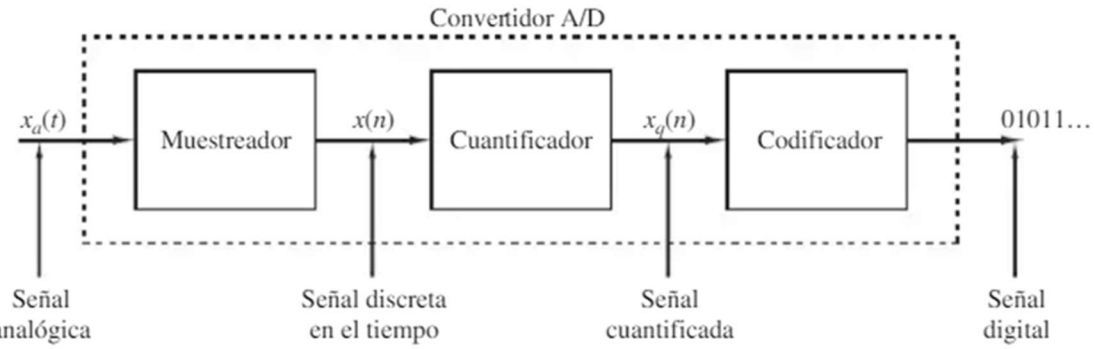


**Figura 5.** ***Partes básicas de un convertidor analógico-digital (A/D).***
**Fuente: (Proakis & Manolakis, 2007)**

En la figura 5 se observa un diagrama de bloques que describe las etapas de funcionamiento de un convertidor A/D, donde se incluye la señal de entrada pasando a través del muestreador, luego al cuantificador y finalizando en el codificador, para por último obtener la señal digital.

Específicamente, en el tratamiento digital de señales auditivas se utilizan tres técnicas en concreto, que son: Sobremuestreo, el dither y el filtrado noise-shaping, las cuales hacen parte del proceso de cuantificación, específicamente, de la conversión Análoga-Digital y Digital-Análoga de señales.

### *5.2.5 Sobremuestreo*

Con el fin de realizar el muestreo de una señal de banda limitada con una frecuencia máxima dada por $f_m$ la frecuencia de muestreo que se aplica debe cumplir la condición de $f_s > 2f_m$, esta se conoce como la frecuencia de Nyquist.

Desde este punto de vista, el sobremuestreo es una técnica utilizada para lograr una mejor calidad o evitar problemas técnicos, que consiste en utilizar una frecuencia de muestreo superior a la frecuencia de Nyquist. (Semeria, Marce2015)

### *5.2.6 Dithering*

El dither o interpolación puede definirse como la inclusión de algo de ruido blanco a una señal analógica destinada a ser digitalizada. Retrospectivamente, esta técnica se ha utilizado principalmente en el campo del audio para mejorar el sonido del audio digital. De manera similar, en el mundo de conversión analógico-digital, el difuminado se puede utilizar para mejorar el rango dinámico del convertidor de analógico-digital. El tramado tiene el efecto de difundir los contenidos falsos espectrales de la señal sobre su espectro. Esta propiedad se obtiene gracias a las características del tramado:

- No correlaciones en el tiempo.
- No correlación con la señal analógica.
- Constante.

El dither debe considerarse como un ruido aleatorio que tiene efectos predeterminados en la señal analógica que se quiere digitalizar. El nivel de este ruido blanco debe calcularse con respecto al nivel de ruido que se espera que disminuya el dither (Gamble, Andrew; Wright, Tony, 2019).

### *5.2.7 Noise Shaping*

Es la combinación de interpolación y ecualización, utilizada tanto para cubrir el ruido de cuantización como para impulsar cualquier ruido creado en áreas menos perceptibles del espectro de frecuencia. Con el uso de noise-shaping, se puede aplicar difuminado, mientras que, en teoría, se reduce el nivel de ruido general prevenible. Para explicar el concepto de noise-shaping se hace referencia a la siguiente ilustración:

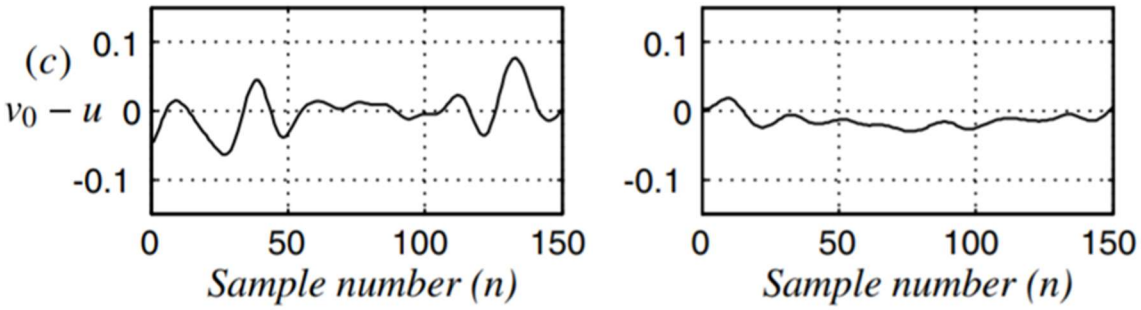


**Figura 6.** ***Efecto de noise-shaping en una señal analógica.***
**Tomado de: (Gamble, Andrew; Wright, Tony. (2019, enero).**

Observando la figura 6 es evidente que el error de cuantificación se suprime en gran medida por el efecto del Noise-shaping. (Gamble, Andrew; Wright, Tony. Enero 2019)

### *5.2.8 Formato de audio digitalizado WAV (wave form audio file format)*

El formato de archivo de audio de forma de onda, a menudo denominado WAV o WAVE, es un estándar de formato de archivo de audio para almacenar un flujo de bits de audio. Se basa en el formato contenedor de formato de archivo de intercambio de recursos (RIFF) y almacena datos en fragmentos, cada fragmento consta de datos con su identificador y longitud. WAV se usa típicamente para almacenar audio sin comprimir, por ejemplo, el formato de modulación de código de pulso lineal (LPCM). El contenido sin comprimir denota que los archivos suelen ser muy grandes, pero existe una limitación de tamaño de 4 GB de datos de audio por fragmento de datos. El formato fue desarrollado originalmente por Microsoft e IBM en 1991 y es compatible con la mayoría de los sistemas operativos más utilizados, incluidos Windows, Macintosh y Linux. La popularidad de WAV tiene mucho que ver con su familiaridad con los profesionales del audio y su estructura relativamente simple. El formato se puede utilizar para codificar audio digital, como, por ejemplo, con varios paquetes de grabación musical que permiten la creación de archivos WAV, además de ser un formato objetivo para actividades de digitalización. (WAV format preservation assessment. Agosto 26, 2022)

El formato WAV ofrece dispone de ciertas características:

- Un archivo WAV de 44.100 Hz y 16 bits contiene una respuesta en frecuencia por encima de 22KHz.
- Un archivo WAV no tiene perdidas ni tampoco es comprimido.
- Un archivo WAV puede contener LPCM, ADPCM o incluso datos cifrados MP3.

La tabla 1 muestra las características generales del formato WAV, las cuales son relevantes en el módulo de adquisición y almacenamiento de la señal.

**Tabla 1.** ***Características del formato WAV.***
**Tomado de: British Libary, WAV format preservation assesment.**

| Formato | Muestreo | Velocidad | Calidad | Peso |
|---|---|---|---|---|
| Wav/Aiff | 8kHz – 16kHz | 8 bits | Muy baja | Pequeño |
| | 16kHz – 32kHz | 16 bits | Buena | Mediano |
| | 44kHz | 16 bits | Excelente | Grande |
| | 48kHz | 16 bits | Perfecto | Muy grande |

### *5.2.9 Los filtros digitales*

"Un filtro es un dispositivo que tiene una entrada y una salida, capaz de transmitir una banda de frecuencias limitada" (Gomez Gutierrez, 2010). Los filtros son utilizados en muchos ámbitos del procesamiento de señales. Un filtro es como un objeto que altera el espectro o el contenido frecuencial de una señal. Los filtros más comunes son los filtros de paso bajo, paso alto, paso de banda y de rechazo de banda, se puede observar en la figura 7 la respuesta en frecuencia de estos filtros mencionados.

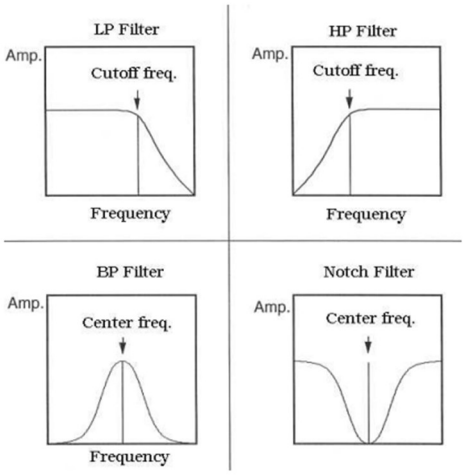


**Figura 7.** ***Tipos de filtros corrientes.***
**Tomado de: Introducción al filtrado digital. Escola superior de música de Catalunya. Departamento de Senología.**

Los filtros de paso bajo dejan pasar las frecuencias situadas por debajo de alguna frecuencia determinada, de igual manera, los filtros de paso alto permiten el paso de frecuencias situadas por encima de alguna frecuencia determinada. En cuanto a el filtro de pasa banda deja pasar todas las frecuencias situadas en una determinada banda de frecuencia y el filtro rechaza banda permite el paso de las frecuencias que no están situadas dentro de cierta banda. (Gutiérrez, 2012)

### *5.2.10 Procedimiento de diseño de filtros*

Para diseñar un filtro se requiere un circuito en el cual la respuesta frecuencial de su función de transferencia cumpla con una delimitación de frecuencia establecida. Con base en esto, se ejecuta el siguiente proceso:

- Determinar la función de transferencia:

$$\boldsymbol{G(s) = \frac{V_0}{V_i} = \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b1^s + b_0}{a_n s^n + a_{n-1} s^{n-1} + \cdots + a1^s + a_0}} \quad \textbf{(2)}$$

- Descomponer los polinomios resultantes en factores de segundo orden.

$$G(s) = \prod_{J=1}^{N} G_j(s) = \prod_{J=1}^{N} K_j \frac{s^2 + e_j s + f_j}{s^2 + c_j s + d_j} \quad (3)$$

- Estimar el valor de las variables para cada filtro según su función.

Con base en lo anterior, y con el fin de diseñar un filtro pasa banda en Matlab:

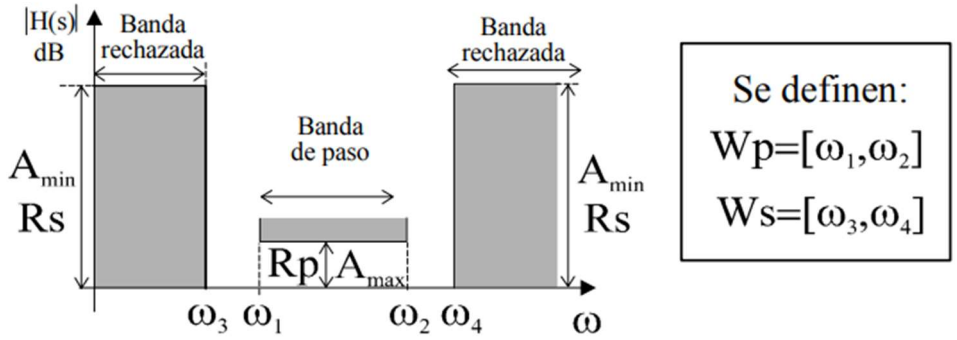


**Figura 8.** ***Variables para el diseño de un filtro pasa banda en Matlab.***
**Tomado de: Diseño de filtros con Matlab, Universidad de Cantabria, España.**

Es necesario definir las variables señaladas en la figura 8 para incluirlas en la siguiente función:

$$[N, Wnc] = Buttord(Wp, Ws, Rp, Rs, 's')$$

Dicha función también retorna $Wn = [Wn1, Wn2]$ que corresponden a las frecuencias naturales 1 y 2 del filtro.

De tal manera, si el filtro a diseñar es un filtro pasa alto:

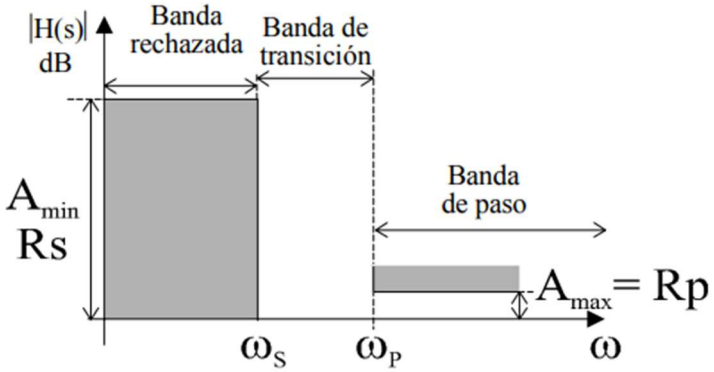


**Figura 9.** ***Variables para el diseño de un filtro pasa altos en Matlab.*** **Tomado de: Diseño de filtros con Matlab, Universidad de Cantabria, España.**

$$[N, Wn] = Buttord\ (Wp, Ws, Rp, Rs, 's')$$

Determinando las variables en la figura 9 se incluyen en la función para diseñar el filtro pasa alto.

### *5.2.11 Sistema Speech To Text*

Un sistema de conversión de voz a texto comprende como mínimo una señal de audio de entrada adquirida por el usuario, y un módulo o factor de procesamiento, en donde se reconoce el audio grabado y se genera el texto. (Bijl & Hyde-Thomson, 2001).

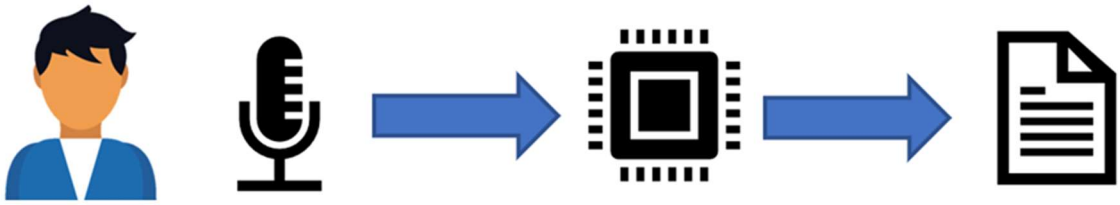

**Figura 10.** ***Flujo de un sistema STT.***
**Tomado de: Autor.**

La transcripción de documentos de voz como conferencias, charlas, presentaciones y debates, son una de las aplicaciones más importantes de los sistemas Speech To Text. La recuperación y obtención de la información que contiene una señal de audio no se obtiene de manera sencilla, es por ello por lo que se desarrollaron los sistemas Speech To Text, sistemas muy completos de diferentes marcas, en diferentes idiomas y alimentados con bases de datos que diariamente se entrenan con ayuda de inteligencia artificial (Furui, Kikuchi, Shinnaka, & Hori, 2004). Los sistemas Speech To Text son sistemas diseñados inicialmente para usarse de manera independiente en plataformas de escritorio, el flujo de un sistema STT se evidencia en figura 10, la adaptabilidad de estos sistemas va de acuerdo con el entorno en que se desempeñan, las condiciones exteriores obligarán a que los modelos implementados se reajusten. Después de la adaptación de reconocimiento automático de voz (de sus siglas en inglés ASR), la mayoría de los sistemas STT se rigen por dos métricas de evaluación (precisión del texto resultante y tiempo requerido para producirlo) (Bijl & Hyde-Thomson, 2002).

En el campo de la tecnología existe una gran variedad de servicios Speech To Tex los más usados comercialmente son desarrollados por compañías de software reconocidas. Algunos de estos servicios se conocen como “Google web Speech API” creado por Google, “Watson” creado por IBM y “Azure Speech Services” creado por Microsoft. Estas APIs proveen transcripción en tiempo real, permitiendo a los sistemas implementar herramientas como comandos de voz, conversaciones para adaptar diferentes acentos de voz o patrones de voz sin cambiar el texto resultante de una conversación. “Google web Speech API” y “Google Cloud Speech” son APIs implementadas por páginas web, domóticas y otros dispositivos que utilizan reconocimiento de voz. La precisión de las APIs incrementa proporcionalmente a el tamaño de la red neural que la soporta, las mejoras en cada versión y las actualizaciones para

cada idioma y región. Python posee una librería llamada “SpeechRecognition”, esta librería funciona con diferentes APIs para reconocimiento de voz y audio para conversión a texto, incluyendo IBM, Microsoft Services y Google. Herramientas como “Google Web Speech API” o “Google Cloud Speech” cuentan con restricciones de uso, bien sea de uso diario limitado o pruebas gratis por cierto número de usos. (Aguilar-Chacón, J. E., & Segura-Torres, D. A. 2020).

“Azure Speech services” es una API de Microsoft que permite la transcripción de voz, que, a comparación de otras herramientas mencionadas anteriormente, permite métodos de integración con leguajes de programación como C# O JAVA.

### *5.2.12 Procesamiento de texto*

En el procesamiento de texto digital es necesario definir los caracteres, palabras y frases que lo conforman. Definir estas unidades representa diferentes retos dependiendo del lenguaje que está siendo procesado y del origen de los archivos, lo cual no es una tarea fácil, considerando la variedad de idiomas y sistemas de escritura. El lenguaje natural contiene ambigüedades y los sistemas de transcripción amplifican dichas ambigüedades y generan algunas nuevas, con lo cual muchos de los retos del procesamiento natural del lenguaje involucran resolver estas. El trabajo inicial del procesamiento del lenguaje ha sido enfocado en un número reducido de idiomas, no obstante avances significativos han sido alcanzados usando fuentes con estructuras de alto rango, incluyendo una vasta fuente dinámica de suministro de texto proveniente de Internet. El procesamiento del texto puede dividirse en dos ramas: clasificación del texto y segmentación del texto.

La clasificación del texto consiste en la conversión de un conjunto de archivos digitales en documentos de texto bien definidos. Para las estructuras iniciales, este proceso era lento, compuesto por como máximo, unos pocos millones de palabras. En contraste, las estructuras actuales cultivadas por el Internet pueden abarcar billones de palabras al día, lo cual requiere una automatización completa del proceso de clasificación del texto. El proceso puede incluir diversos pasos, dependiendo del origen del archivo que va a ser procesado. En primer lugar, para hacer que cualquier tipo de documento digital, sin importar su idioma ni su comprensibilidad para la máquina, sus características deben ser presentadas en una codificación de caracteres. Existe otro subproceso llamado Identificación de Carácter Encriptado, el cual determina la encriptación (o encriptaciones) de cualquier archivo y lo convierte entre codificaciones si se requiere. Además, selecciona el algoritmo para el idioma específico en el que se encuentra el documento. La sectorización del texto identifica el contenido actual dentro del archivo mientras descarta elementos indeseados, como imágenes, tablas, encabezados, enlaces y HTML. El producto de la etapa de clasificación es una estructura de texto organizada por idioma, adecuada para la segmentación y análisis futuro.

La segmentación del texto es el proceso de convertir una estructura de texto en sus palabras y frases, dicha segmentación divide la secuencia de caracteres en un texto localizando las fronteras entre palabras, los puntos en donde una palabra termina y otra nueva empieza. Para términos computacionales, las palabras identificadas de este modo se denominan tokens, esa

segmentación es conocida también como tokenización. La tokenización está vinculada a la codificación de caracteres subyacente del texto que es procesado y la identificación de los caracteres es siempre el primer paso. La tokenización está establecida de una manera concreta para lenguajes artificiales como los de programación. Sin embargo, dichos lenguajes artificiales pueden ser estrictamente definidos para eliminar ambigüedades en estructura o léxico, lo cual no es posible con idiomas naturales, donde el mismo carácter puede tener significados de acuerdo con el contexto dado que su sintaxis no está definida estrictamente. Existen diversos métodos en los cuales una palabra puede ser ubicada, y cada uno de estos métodos existe para idiomas segmentados (inglés, español, francés) o sin segmentación (Japones, chino, Thai). La morfología de las palabras en un idioma puede ser aislante, donde las palabras no dividen en unidades más pequeñas; aglutinador donde las palabras dividen en unidades más pequeñas (morfemas) con fronteras claras entre morfemas, o flexivo donde las fronteras entre morfemas no están claras y donde los componentes de los morfemas pueden expresar más de un significado. (Indurkhya. Damerau. 2010, Handbook Natural Lenguange Processing).

### *5.2.13 Tokenización de idiomas segmentados.*

En muchos de los idiomas que se componen del alfabeto latino, sus palabras se separan por espacios en blanco. Aun así, en estructuras de frases bien conformadas, existen muchos problemas que resolver con respecto a la tokenización. Muchos de los problemas de la tokenización existen alrededor de los signos de puntuación (puntos, comas, signos de interrogación o exclamación, etc) esto debido a que la puntuación puede actuar de diferente manera dependiendo del contexto y ubicación de este. Una manera lógica de tokenizar estos idiomas es considerar como un token a cada secuencia de caracteres precedida y/o seguida por un espacio, lo que tokeniza satisfactoriamente palabras en una secuencia de caracteres alfabéticos, pero no toma en cuenta la puntuación. (Indurkhya. Damerau. 2010, Handbook Natural Lenguange Processing).

# 6. Diseño Metodológico

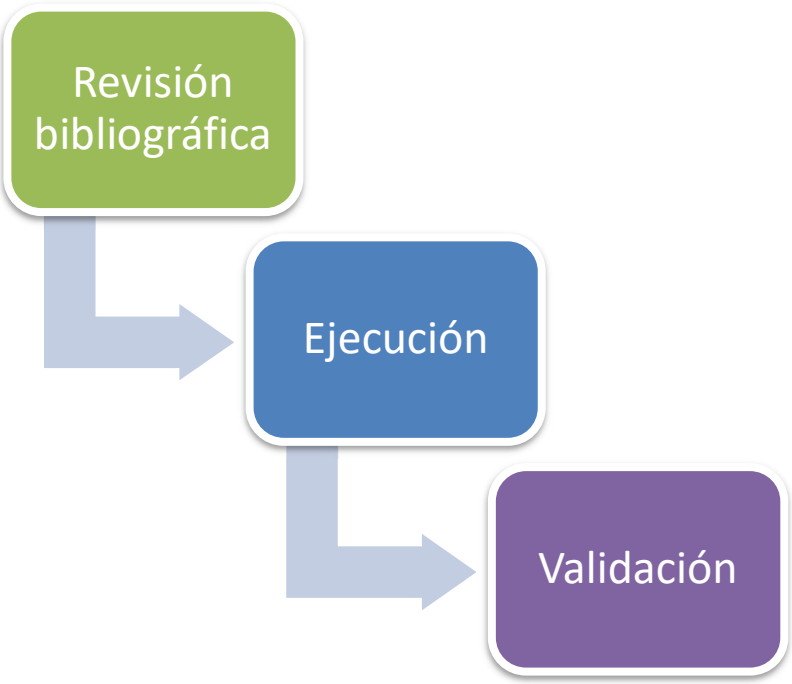


**Figura 11.** ***Flujo de metodología planteada.***
**Tomado de: Autor.**

En la figura 11 se detallan las tres etapas para el desarrollo de la metodología, iniciando por una revisión bibliográfica, en la cual se realizó una investigación del temario propio a resolver para dar paso a la ejecución. En la ejecución se plantearon sistemas que cumplieran con los objetivos específicos de cada módulo. En la etapa de validación, se realizaron verificaciones de que los sistemas, algoritmos o desarrollos implementados en la etapa anterior en realidad funcionaban.

Cada una de estas etapas se realizó de manera individual, para cada una de las fases de elaboración del proyecto, fases descritas en la figura 12.


**Figura 12.** ***Fases del proyecto.***
**Tomado de: Autor.**

## 6.1 Módulos

Los módulos del sistema son aquellas etapas necesarias para la construcción del prototipo para una posterior validación, la adquisición de la señal, las etapas de preprocesamiento, procesamiento y posprocesamiento son algunos de los componentes implementados. La aplicación de la metodología para los componentes se desarrolló de la siguiente manera:

### *6.1.1 Revisión bibliográfica*

En esta etapa de la metodología se realizó una investigación bibliográfica de los sistemas de identificación y extracción de texto existentes, a partir de ello se determinaron los módulos que debía tener el sistema a diseñar.

### *6.1.2 Ejecución*

En esta etapa de la metodología se planteó el diseño del sistema, donde se abordó la estructura que debía tener este.

### *6.1.3 Validación*

En esta etapa de la metodología se realizó una implementación básica con un algoritmo para verificar el funcionamiento del modelo en cada uno de los componentes identificados.

## 6.2 Algoritmos

La aplicación de la metodología para los algoritmos de desarrolló de la siguiente manera:

### *6.2.1 Revisión bibliográfica*

En esta etapa de la metodología se hizo una investigación bibliográfica para buscar algunos de los algoritmos existentes y se decidió cuáles de estos eran acordes al objetivo de las etapas del modelo diseñado.

### *6.2.2 Ejecución*
En esta etapa de la metodología se probaron tres de los algoritmos seleccionados durante la etapa anterior.

### *6.2.3 Validación*
En esta etapa de la metodología se realizó la medición de los criterios de evaluación que fueron seleccionados inicialmente.

**Tabla 2.** ***Fases de elaboración con metodología planteada.***
**Tomado de: Autor.**

| | **Módulos** | **Algoritmos** |
|---|---|---|
| **Revisión Bibliográfica** | Investigación | Investigación |
| **Ejecución** | Diseño del sistema | Desarrollo e implementación del sistema |
| **Validación** | Implementación básica | Medición de los criterios |
| **Documentación** | Descripción escrita de lo realizado | Descripción escrita de lo realizado |

En la tabla 2 se puede observar la matriz asociada a la metodología propuesta para el cumplimiento de los objetivos del proyecto, y se desarrolló aplicando cada etapa a cada fase sin ninguna excepción.

# 7. Identificación de los módulos del sistema

Para llevar a cabo el diseño de los módulos de un sistema capaz de identificar una pauta publicitaria basado en herramientas Speech To Text, fue necesario realizar un análisis de las necesidades y así plantear los diferentes módulos.

Se determinaron cinco módulos para el sistema con la metodología en cascada, la cual es bastante implementada en el diseño de algoritmos y software, fue popularizada por Winston Royce en 1970 (Argomedo Pflücker & Cóndor Ruiz, 2015). Las principales características de esta metodología son: flujo lineal, los módulos están relacionados secuencialmente y cada módulo tiene una entrada y una salida, el nombre y el posicionamiento de cada uno de ellos se evidencia en la figura 13.

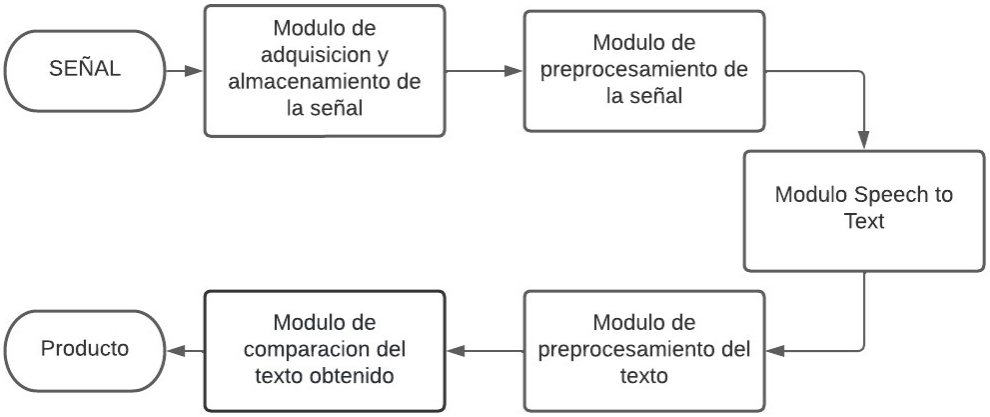


**Figura 13.** ***Componentes del sistema.***
**Tomado de: Autor.**

## 7.1 Módulo de adquisición y almacenamiento de la señal

Primeramente, como entrada del sistema se tiene una señal de audio analógica la cual debe ser convertida en una señal digital con el fin de analizar su información. Para ello se requiere captar dicha señal y almacenarla en la máquina. El producto obtenido como salida del módulo de adquisición y almacenamiento de señal es un archivo de audio digital en formato WAV.

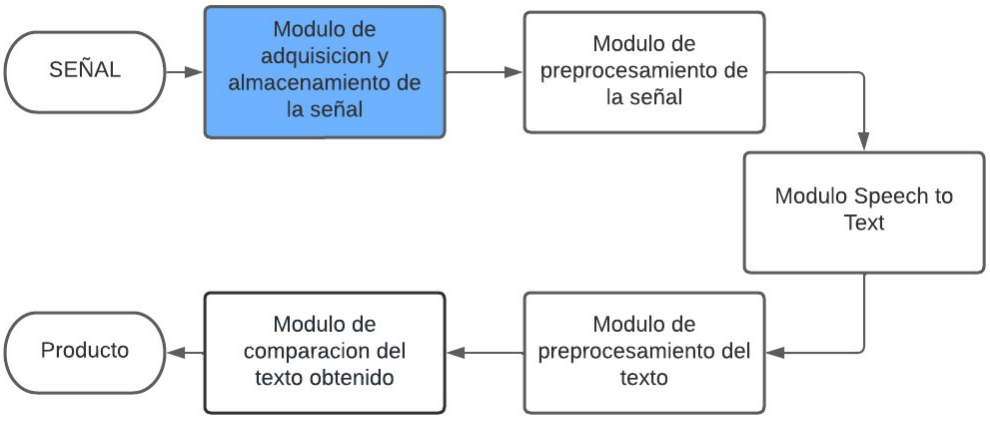


**Figura 14.** ***Módulo de adquisición y almacenamiento de la señal*****.**
**Tomado de: Autor.**

## 7.2 Módulo de preprocesamiento de la señal

La calidad de los archivos de audio puede variar según su formato, su codificación, su tasa de compresión, y sus respectivos bits por muestra. Aunque un archivo de audio tratado digitalmente puede guardarse, copiarse y reproducirse infinitamente sin perder su calidad, estas características fueron analizadas y definidas en el momento del almacenamiento de la señal.

Para este caso fue necesario un acondicionamiento de la señal para la reducción de características no deseadas. Es por ello por lo que se implementó un módulo de preprocesamiento de la señal, el cual como entrada tiene el archivo de audio y a la salida se obtiene un archivo con reducciones en las características no deseadas, aun así, manteniendo el formato inicial del archivo de audio.

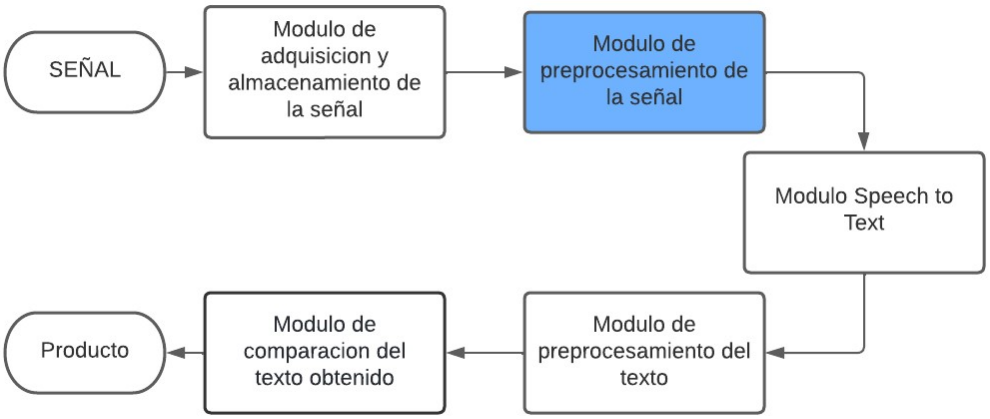


**Figura 15.** ***Módulo de preprocesamiento de la señal.***
**Tomado de: Autor.**

## 7.3 Módulo Speech To Text

El módulo Speech To Text, el cual no fue implementado en el proyecto por definición inicial del proyecto, en donde se pacta la utilización de una herramienta de uso libre, para este caso específico, la de Azure. A la entrada del módulo Speech To Text se tiene un archivo de audio, luego a la salida se obtiene la transcripción en forma de cadena del audio a texto. Se realiza lectura del archivo y del evento de reconocimiento continuo, el módulo reconoce siempre y cuando tenga un archivo sobre el cual realizar el análisis.

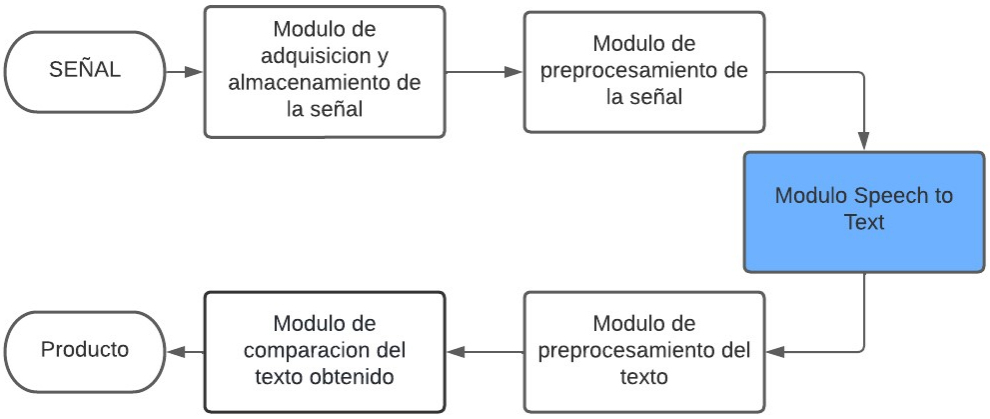


**Figura 16.** ***Modulo Speech To Text.***
**Tomado de: Autor.**

## 7.4 Módulo de preprocesamiento del texto

Dada la experiencia en trabajos previos con herramientas STT, se tiene la certeza que la precisión en las palabras obtenidas no es del ciento por ciento, es decir que al implementar la herramienta STT esta puede entregar palabras erróneas haciendo referencia a falsos positivos en donde se reconoce una palabra que no hace parte de la señal de audio, o falsos negativos en donde se omite una palabra que ciertamente hace parte de la señal. Por ello se implementó un módulo de preprocesamiento de texto, en donde a la entrada, llegan todas las palabras provenientes de una cadena de texto y a la salida del módulo se tiene esa cadena de texto con correcciones realizadas.

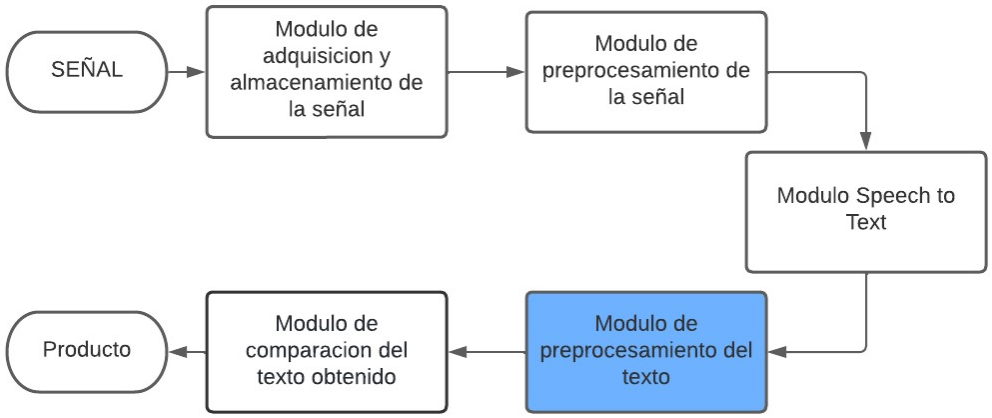


**Figura 17.** ***Módulo de preprocesamiento del texto.***
**Tomado de: Autor.**

## 7.5 Módulo de comparación del texto obtenido

Una vez se tiene el texto corregido, este es evaluado con respecto al texto definido para la pauta publicitaria en cuestión. El módulo de comparación del texto se encarga de confrontar la información del audio transcrita a la entrada del sistema con la resultante luego del procesamiento del texto.

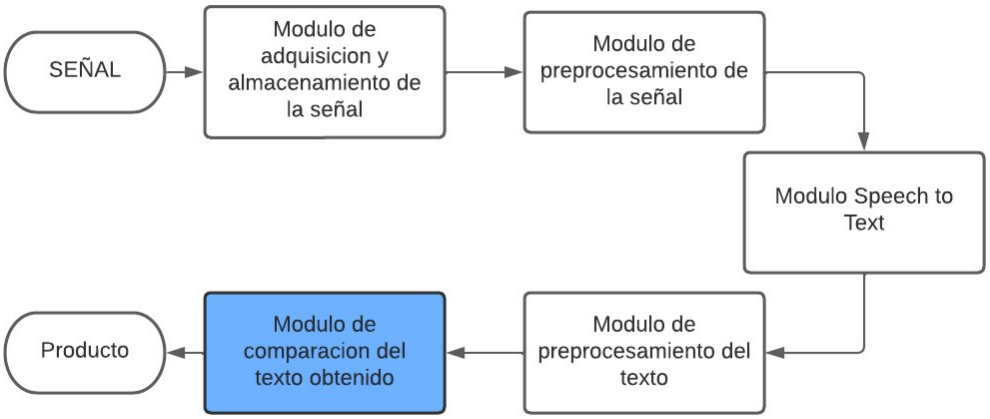


**Figura 18.** ***Módulo de comparación del texto obtenido.***
**Tomado de: Autor.**

# 8. Identificación de los algoritmos

## 8.1 Módulo de adquisición y almacenamiento de la señal.

El primer módulo cuenta con un algoritmo implementado en C# con el cual se realizó el proceso de captación y almacenamiento de la señal.

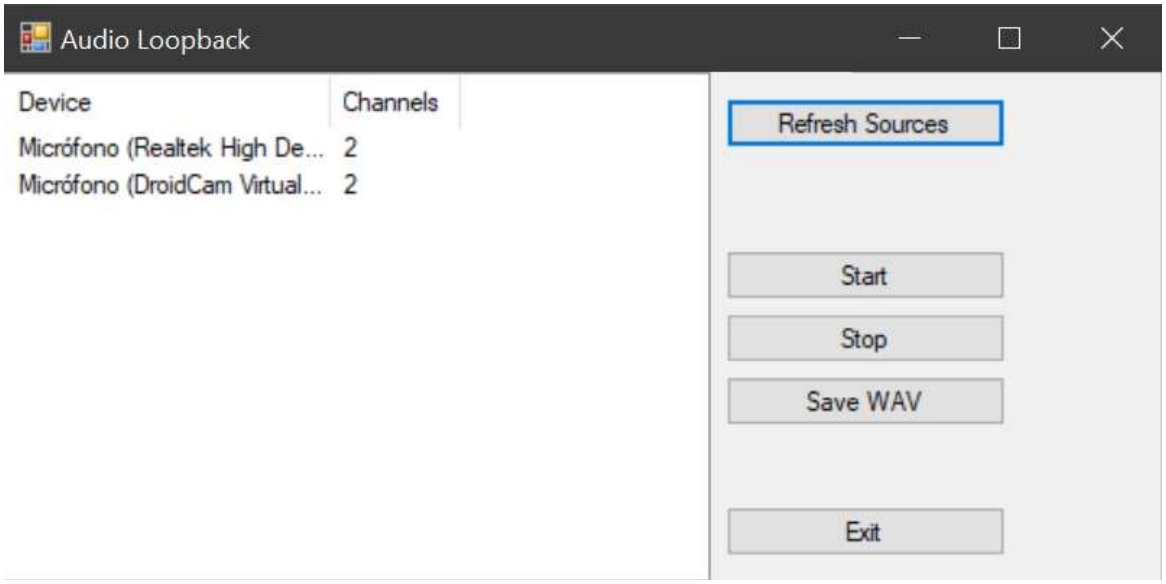


**Figura 19.** ***Audio Loopback ventana de texto.***
**Tomado de: Autor.**

En la figura 19 se observa la ventana de un formulario creado con *Windows Forms* el cual permite visualizar un listado de fuentes disponibles de las cuales se puede obtener el audio.

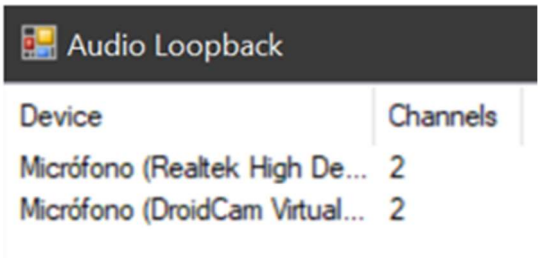


**Figura 20.** ***Listado de fuentes de audio.***
**Tomado de: Autor.**

En este caso, el audio se debe capturar desde la tarjeta de sonido del computador en donde se conecta el radio. Las fuentes disponibles se actualizarán cada vez que se dé clic en el botón *"Refresh Sources"*.

Luego de seleccionar la fuente, hay tres posibles opciones a elegir.

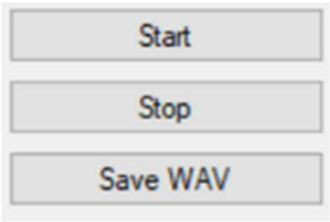


**Figura 21.** ***Botones de acción durante el proceso de grabación.***
**Tomado de: Autor.**

El botón "*Start*" contiene una serie de declaraciones y métodos que se ejecutan de manera secuencial, lo primero es comprobar que exista una fuente seleccionada, de no ser así el método no realizará ninguna acción, sí de lo contrario, existe algún dispositivo elegido, el método continuará con su ciclo, lo primero es instanciar un objeto de entrada de tipo *"WaveIn"*, este objeto pertenece a una biblioteca de código abierta llamada *"nAudio",* de esta biblioteca se extraen la mayoría de métodos y propiedades que permiten el correcto funcionamiento de este módulo. El objeto de tipo *"WaveIn"* es donde se almacena en memoria la señal de audio, este mismo permite modificar propiedades como la frecuencia de muestreo, el número de bits y el número de canales, también se crea un objeto *"WaveOut"* de tipo *"DirectSoundOut"*, a este objeto se le envía por parámetro el elemento *"WaveIn"* para que mediante un método seleccionado pueda reproducir por el altavoz del dispositivo lo que a la entrada está llegando.

El botón *"Save WAV"* tiene un proceso similar al flujo descrito anteriormente, solo que ahora la señal no se reproduce sino se almacena en una variable cada vez que un evento detecte que hay información a la entrada, de igual forma, luego de ajustar las propiedades del objeto *"WaveIn"* se instancia un objeto *"waveWriter"* de tipo *"WaveFileWriter"* que será capaz de almacenar el objeto de entrada en un archivo nuevo en el disco del computador. El botón *"Stop"* permite detener cualquiera de los dos procesos seleccionados y de esa forma liberar el software para volver a elegir alguna opción, por último, el botón *"Exit"* cerrará la ventana y terminará por completo la ejecución del algoritmo.

### *8.1.1 Método de adquisición de la señal número uno: Jack 3.5mm*

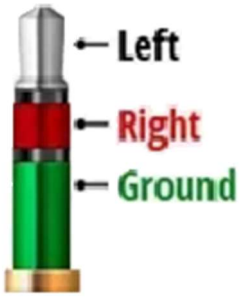


**Figura 22.** ***Configuración de conexión de plug.***
**Tomado de: Autor.**

Con la adquisición del audio mediante este método, el objetivo fue que esta señal llegase al computador para que posteriormente fuera procesada, por lo tanto, no se usó micrófono. En la figura 23 se evidencia el diagrama de conexión utilizado para dar cumplimiento al objetivo anteriormente mencionado, los colores enmarcados en el diagrama de la figura 23 hacen referencia a las conexiones denotadas en la conexión de terminal de audio universal de 3.5mm de la figura 22.


**Figura 23.** ***Diagrama de conexión Radio-PC.***
**Tomado de: Autor.**

### *8.1.2 Método de adquisición número dos: Software Audacity.*

La captación de audio con este método se compuso de dos partes, la señal de radio se obtuvo directamente del sitio web de emisoras radiales, y el software Audacity que es un software

de uso libre, el cual tiene diferentes funcionalidades con respecto al tratamiento y grabación de audio.

De tal manera, se obtuvieron diferentes archivos de audio WAV con las siguientes configuraciones:

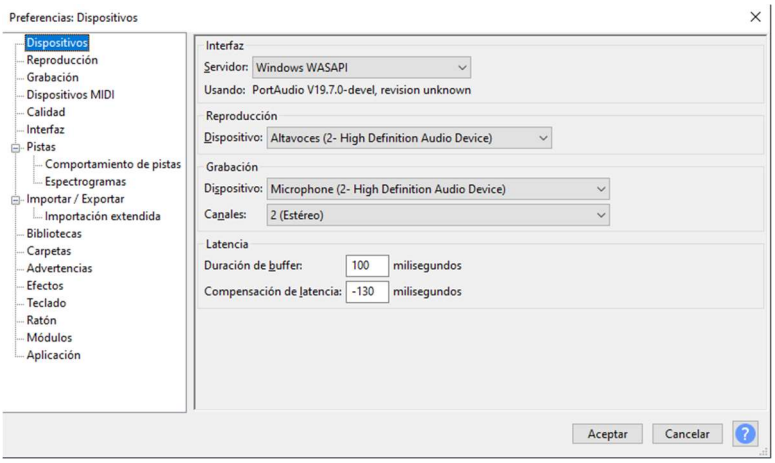


**Figura 24.** ***Ventana de preferencia de dispositivos en el software Audacity.***
**Tomado de: Autor.**

En la figura 24 se evidencia la configuración del servidor para Windows Audio Session API (WASAPI) es crucial, pues de esta manera Audacity graba directamente el sonido de la tarjeta de audio del computador. Además de eso, se tiene una latencia predeterminada para esta fase.

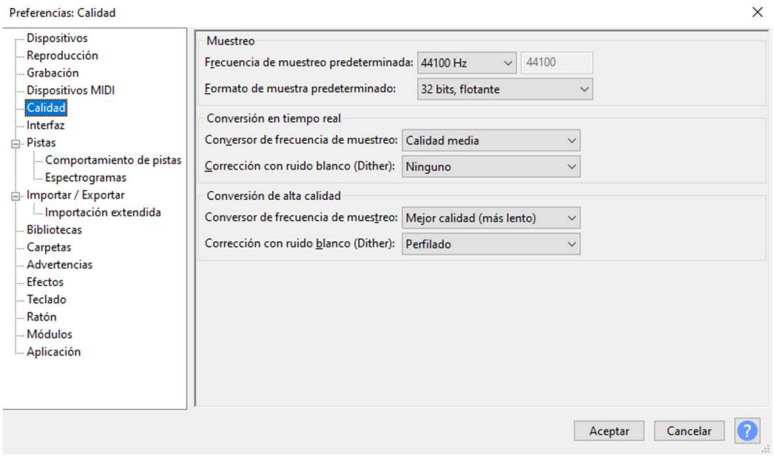


**Figura 25.** ***Ventana de preferencia de calidad en el software Audacity.***
**Tomado de: Autor.**

En la figura 25 se evidencia la frecuencia de muestreo para este módulo fue de 44100Hz con un formato de 32 bits.

### *8.1.3 Método de adquisición número tres: Tarjeta de sonido USB externa.*

Como tercer método se utilizó una tarjeta de sonido Chanel 7.1 Externa USB 2.0 para PC 3.5 mm, el cual es un adaptador de audio externo, es un periférico bidireccional, lo que significa que es una herramienta que se conecta externamente a un computador y tiene la capacidad de enviar y recibir datos. La mayor ventaja obtenida de este método es que se independiza el proceso del computador, ya que la tarjeta tiene como objetivo exclusivo procesar el audio.


**Figura 26.** ***Tarjeta de sonido Chanel 7.1 externa USB 2.0.***
**Tomado de: Manual de usuario Tarjeta Chanel 7.1**

Este método funciona en conjunto con el algoritmo diseñado específicamente para la captación del audio en formato WAV, el cual fue mencionado anteriormente.

## 8.2 Módulo de preprocesamiento de la señal

Para este módulo se tomaron diferentes casos, tal como fue indicado en la delimitación del problema. Estos casos fueron organizados en diferentes escenarios, como se observa en la tabla 3:

**Tabla 3.** ***Tabla de escenarios de prueba.***
**Tomado de: Autor.**

| | **Escenario** |
|---|---|
| **1** | Voz, pauta, música |
| **2** | Voz, pauta, Voz |
| **3** | Música, Pauta, música |
| **4** | Música, pauta, Voz |

### *8.2.1 Método número uno: Filtros digitales.*

El rango de frecuencias en la voz humana ronda entre los 50 Hz y 350 Hz (Antón, E. R), esto dependiendo del género y el tono de quien produce la señal. A partir de ello se realizó un análisis y se propuso implementar filtros digitales, que una vez aplicados a la señal, fueron evaluados utilizando el módulo de Speech To Text y el módulo de comparación del texto. Para el primer método se decide realizar unas pruebas de dichos filtros en Matlab.

### *8.2.2 Método numero dos: Cambio de frecuencias de muestreo*

Para el segundo método, se realizó una serie de pruebas con los archivos de audio. Mas específicamente se cambió la frecuencia de muestreo. Dicha frecuencia se cambió entre 11.025Hz, 22050Hz, 32000Hz y 44010Hz.

### *8.2.3 Método número tres: Cambio de bits de resolución*

Para el tercer método, se cambió la cantidad de bits de resolución por muestra, para 8 bits y 16 bits.

## 8.3 Módulo de preprocesamiento de texto.

Para realizar análisis de texto, generalmente se requiere diversas fases, tales como la identificación del carácter, identificación del lenguaje y segmentación del texto. Ya que el sistema está diseñado únicamente para el idioma español y el texto proviene de una herramienta Speech To Text, se evaluó solamente la segmentación del texto. Para ello los tres algoritmos implementados para este módulo siguieron el flujo descrito en la figura 27.

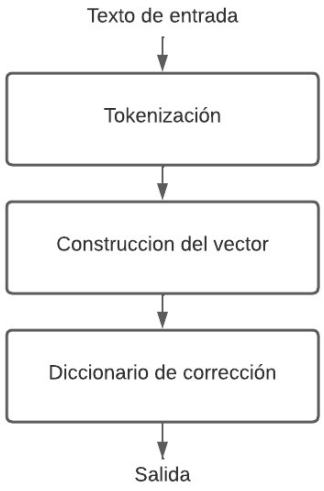


**Figura 27.** ***Diagrama de flujo para el algoritmo de preprocesamiento del texto basado en tokenización.*** **Tomado de: Autor.**

### *8.3.1 Tokenización excluyendo puntuación con diccionario cerrado.*

Se realiza la tokenización de la frase obtenida del módulo de Speech To Text, la cual es transformada en un vector excluyendo la puntuación y se vería de la siguiente manera:

**Tabla 4.** ***Ejemplo de elementos de un vector luego del proceso de tokenización excluyendo puntuación.***
**Tomado de: Autor.**

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| me | encanta | que | las | cosas | buenas | se | vuelvan |

La pauta se encuentra fraccionada en elementos de un vector, es comparada con un diccionario cerrado en el cual se encuentran todas las palabras que incluye la pauta originalmente, permitiendo de este modo remplazar aquellas palabras incorrectas que fueron tokenizadas y hacen parte del diccionario como una llave.

### *8.3.2 Tokenización incluyendo puntuación con diccionario cerrado*

Se realiza la tokenización de la frase obtenida del módulo de Speech To Text, la cual es transformada en un vector incluyendo la puntuación resultando de la siguiente manera:

**Tabla 5.** ***Ejemplo de elementos de un vector luego del proceso de tokenización incluyendo puntuación.***
**Tomado de: Autor.**

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| trocipollo, | nada | mejor | fue | un | paquete | de | sabor |

Como se observa, la coma hace parte del token 0, y de esa manera a lo largo de toda la pauta, los signos de puntuación se conservan en el token más cercano. La pauta se encuentra ahora en un elemento del vector, es comparada con un diccionario cerrado en el cual se encuentran todas las palabras que incluye la pauta originalmente, permitiendo de este modo remplazar aquellas palabras incorrectas que fueron tokenizadas y hacen parte del diccionario como una llave.

### *8.3.3 Diccionario de la pauta*

Se tienen dos cadenas de texto para este algoritmo, La primera es la obtenida de la grabación de la pauta luego de ser procesada por el sistema Speech To Text y la segunda es la cadena objetivo, o la cual se pretende encontrar. Se convierte la cadena objetivo en un arreglo unidimensional en donde cada elemento del arreglo es equivalente a una palabra de la cadena. Posteriormente se realiza una tokenización excluyendo puntuación a la cadena número uno.

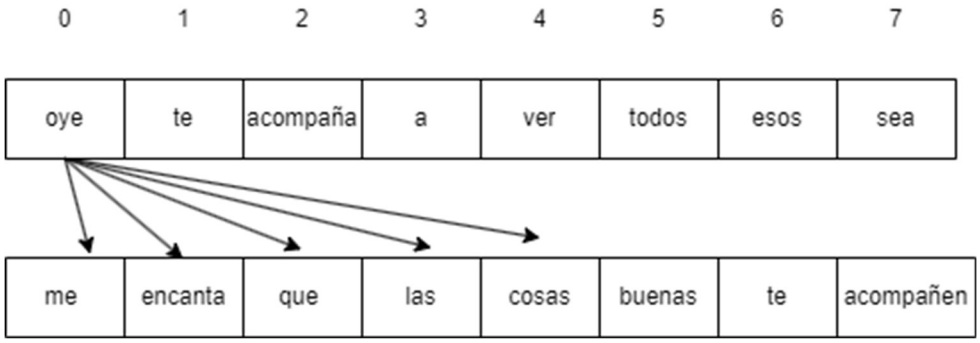


**Figura 28. Algoritmo de diccionario de la pauta.**
**Tomado de: Autor.**

Como se observa en la figura 28, el elemento 0 de la cadena tokenizada, es decir la superior, recorre todos los elementos de la cadena inferior en busca de elementos iguales, en caso de encontrarlas, no elimina el elemento de la cadena, por el contrario, si no existe ninguna igualdad, el elemento es eliminado depurando así la información no relevante para la identificación, dejando solamente las palabras iguales.

## 8.4 Módulo de comparación del texto obtenido.

### *8.4.1 Algoritmo número uno: Distancia de Levenshtein*

La distancia de Levenshtein es una medida de similitud entre dos frases (fuente *f* y objetivo *o*). La distancia es el número de supresiones, inserciones o sustituciones requeridas para transformar *f* en *o.* Entre más grande sea la distancia de Levenshtein, más grande es la diferencia entre las frases. En este caso la frase fuente es la pauta publicitaria tomada de la radio y el objetivo es el texto de dicha pauta tomado del diccionario especificado (Haldar, R., & Mukhopadhyay, D. 2011). El algoritmo se desarrolla en tres pasos.

Paso 1: Inicialización

- Se configura la longitud de *n* para que sea la longitud de *s* (Source) y se configura *m* para que sea la longitud de *t* (Target).
- Se construye una matriz que contiene columnas y filas de 0 hasta *m.*
- Se inicializa la primera fila de 0 hasta *n.*
- Se inicializa la primera columna de 0 hasta *m*.

Paso 2: Procesamiento

- Analizar s (desde la i a la l para *n*)
- Analizar t (desde la i a la l para *m*)
- Si $S_i$ (Source inicial) resulta igual a $t_j$ (Target inicial), el costo es 0.
- Si $S_i$ es diferente de $t_j$, el costo es 1.

- Se configura una celda $d_{(i,j)}$ de la matriz igual al mínimo de:
  - La celda ubicada abajo más 1: $d_{(i-1,j)} + 1$
  - La celda a la izquierda más 1: $d_{(i,j-1)} + 1$
  - La celda en diagonal abajo y a la izquierda más el costo: $d_{(i-1,j-1)} + \text{costo}$

Paso 3: Resultados. El procedimiento descrito en el paso dos es realizado hasta que el valor de $d_{(n,m)}$ sea encontrado.

### *8.4.2 Algoritmo numero dos: Similitud de Jaro Winkler*

La similitud de Jaro Winkler es usada ampliamente para medir la semejanza de frases. Fue desarrollada para la detección de nombres de personas duplicadas en una base de datos de nombres. Comparado con otras medidas esta provee buenos resultados en computación, sin embargo, el cálculo secuencial de la similitud para la búsqueda de similitudes en largos conjuntos de frases es prolongada.

$$\boldsymbol{Jaro(s_1, s_2)} = \begin{cases} \boldsymbol{\frac{1}{3} * \left(\frac{m}{|s_1|}\frac{m}{|s_2|}\frac{m-t}{m}\right) : m > 0} \\ \boldsymbol{0 : de\ lo\ contrario} \end{cases} \quad \textbf{(4)}$$

Donde $|s_1|, |s_2|$ son las longitudes de ambas frases, *m* es el número de caracteres que coinciden, y *t* es el número de transposiciones. Los caracteres coincidentes son caracteres que ambas frases tienen en común con un máximo de distancia de:

$$\boldsymbol{w = \frac{\max(|s_1||s_2|)}{2} - 1} \quad \textbf{(5)}$$

El número de transposiciones *t* es la mitad del número de posiciones desiguales en la frase concatenada de todos los caracteres coincidentes en orden de ocurrencia (T. Grust et al. (Hrsg), 2019).

### *8.4.3 Algoritmo número tres: Similitud del coseno*

Similitud del coseno es un algoritmo de medida implementado ampliamente en extracción de información. Este algoritmo modela un documento de texto en un vector de términos, de tal manera, la similitud entre dos documentos puede ser derivada mediante el cálculo del valor del coseno entre estos. La implementación de esta métrica puede ser aplicada a cualquier texto, como frases, párrafos o documentos completos (Lahitani, A. R., Permanasari, A. E., & Setiawan, N. A, 2016). Una frase puede ser representada como un vector de la siguiente manera:

$$\boldsymbol{\vec{d} = (w_{do,} w_{d1} \dots w_{dk})} \quad \textbf{(6)}$$

De la misma manera, se representa el vector a comparar de la siguiente manera:

$$\vec{q} = (w_{q0,} w_{q1} \dots w_{qk}) \quad \textbf{(7)}$$

En donde $w_{di}$ y $w_{qi}(0 \leq i \leq k)$ son números flotantes indicando la frecuencia de cada carácter dentro de la frase, mientras que la dimensión del vector corresponde a un carácter disponible en la frase. Con base en la similitud del vector, la similitud entre los dos vectores puede ser definida por:

$$Sim(\vec{q}, \vec{d}) = \frac{\vec{q} * \vec{d}}{|\vec{q}| * |\vec{d}|} = \frac{\sum_{k=1}^{t} w_{qk} * w_{dk}}{\sqrt{\sum_{k=1}^{t} (w_{qk})^2} * \sqrt{\sum_{k=1}^{t} (w_{dk})^2}} \quad \textbf{(8)}$$

# 9. Resultados

En este capítulo se hace un análisis de resultados de cada uno de los módulos definidos para el sistema los cuales fueron señalados en el capítulo anterior, y que basados en la experimentación, se definieron los más adecuados para el propósito del sistema.

## 9.1 Descripción de los casos de prueba

Un caso se compone de dos grabaciones de audio en las cuales se encuentra la misma pauta publicitaria, el sistema diseñado fue evaluado con cinco casos. Inicialmente fueron construidos diferentes escenarios para cuatro pautas publicitarias en concreto y posterior a estas pruebas iniciales, se da paso a grabar escenarios reales, que son aquellos tomados de franjas de radio en vivo y con los cuales se realizaron las pruebas del sistema definido. Los escenarios construidos pueden ser escuchados en el enlace relacionado al anexo 1, donde se encuentran los audios para cada pauta.

Caso 1. Pauta publicitaria Café Aguila Roja.
La pauta publicitaria es: *"felicidad es todo aquello que se brinda sin reservas una flor un beso la ternura del amor la navidad es todo aquello que nos hace recordar que la vida es bella que diciembre es amor navidad aguila roja navidad aguila roja en esta navidad café aguila roja te acompaña con cariño"*

Caso 2. Pauta publicitaria Trocipollo
La pauta publicitaria es: *"me encanta que las cosas buenas se vuelvan requete buenas cómo tener una cita con tu crush o que te regalen unos trocipollo eso es más que bueno y ahora más que trocipollo tienen nuevos sabores requeté bueno son los dos nuevos sabores de trocipollo pollo brasa picante y pollo california cuál es tu favorito"*

Caso 3 Pauta publicitaria Smirnoff
La pauta publicitaria es: *"para los que aman la fiesta a mí y a mi también eso para todos ustedes que disfrutan el sabor que no necesita explicación de smirnoff x 1 ahora tienen que probar el nuevo smiroff x 1 lulo sin azúcar vive la fiesta diferente sin explicaciones tomalo frio y en shots nuevo smirnoff lulo sin azúcar sin explicaciones que no te invita a disfrutar con responsabilidad el exceso de alcohol es perjudicial para la salud prohibase el expendio este de bebidas embriagantes a menores de edad 25% volumen de alcohol"*

Caso 4 Pauta publicitaria Fibra ETB
La pauta publicitaria es: *"Emocionarte con la misma velocidad de subida y bajada está en ti compra 200 megas de internet fibra óptica etb desde 37450 pesos mensuales por cuatro meses llama ya al 6013714000 etb"*

Caso 5 Pauta publicitaria Aguila
La pauta publicitaria es: *"una fría para este calor suave una fría para la fiesta suave una fría un viernes suave una fría siempre aguanta y más si es con la suavidad de águila por solo*

*2000 pesitos precio sugerido al público para cerveza águila light y águila original en botella retornable"*

## 9.2 Hardware y desarrollo de algoritmos

Para este trabajo se creó un proyecto principal en donde se instancian los demás proyectos utilizados, se implementó un proyecto (aplicación de consola) para cada módulo, y estos son utilizados en el proyecto principal que hace el flujo de funcionamiento ilustrado en la figura 29. Para el desarrollo fue utilizado .NET framework en lenguaje C#.

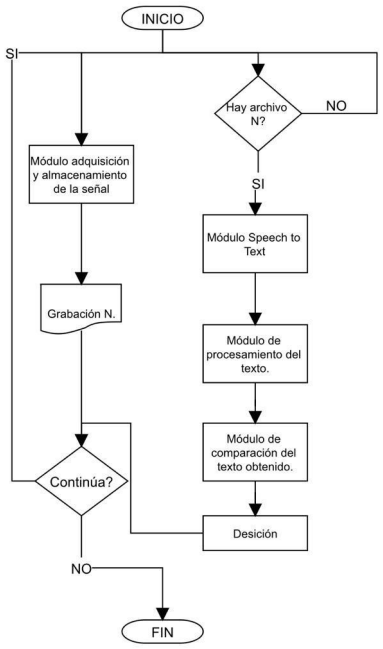


**Figura 29. Diagrama de flujo del sistema general.**
**Tomado de: Autor.**

En el diagrama de flujo existen dos hilos, un hilo solamente para la captación y almacenamiento de la señal y otro para realizar todo el proceso de manipulación de la señal almacenada y toma de decisión acerca de la emisión o no de la pauta publicitaria.

Para la medición del tiempo del sistema en general o de cuanto tarda cada uno de los módulos, se implementa un objeto de tipo TimeSpan que sirve para realizar estas medidas. El objeto DateTime.Now.TimeOfDay se implementa en un intervalo de inicio y fin y se realiza la correspondiente operación matemática para obtener la diferencia de tiempo (Microsoft, recuperado 28 septiembre 2022).

Las especificaciones técnicas de la máquina en la que se realizaron los procesos son las siguientes:

- Procesador Intel(R) Core (TM) i7-6500U CPU @ 2.50GHz 2.59 GHz
- RAM 8,00 GB (7,86 GB usable)

- Sistema operativo de 64 bits, procesador basado en x64

El código de encuentra en el enlace adjunto al anexo número 5.

## 9.3 Módulo de adquisición y almacenamiento

En este módulo se tuvieron en cuenta diferentes factores para la determinación de cuál sería el método ideal de captación de la señal. Por un lado, mantener la calidad de grabación y almacenamiento fue un punto determinante. De igual manera fue necesario pensar en la sostenibilidad del sistema en términos de automatización, hacerlo independiente del internet o de un software externo, obteniendo una señal de alta calidad, con un método robusto y fácil de integrar al sistema.

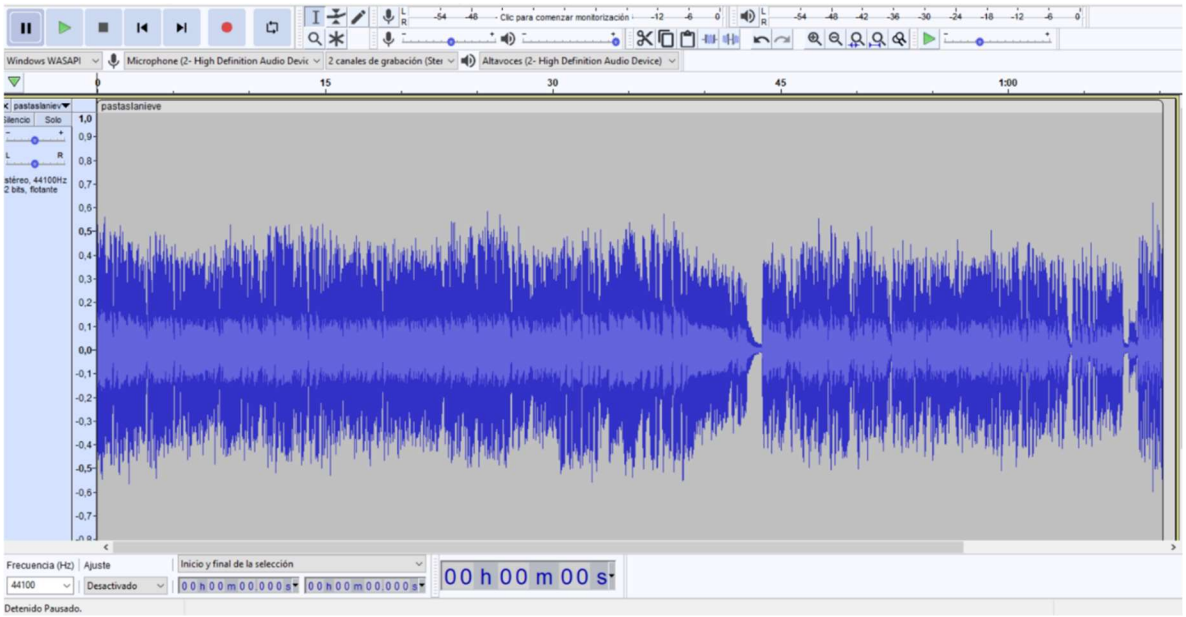


**Figura 30.** ***Audio obtenido mediante el algoritmo desarrollado en C# con el método de adquisición número uno*****.**
**Tomado de: Autor**

En la figura 30 se puede observar la grabación obtenida del primer método de adquisición, el Jack 3,5mm y utilizando el algoritmo desarrollado para la obtención en formato WAV.

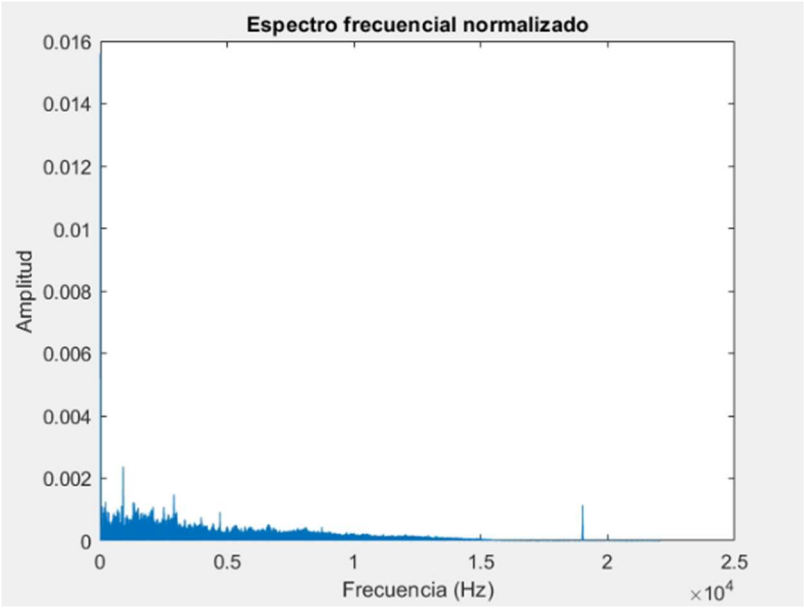


**Figura 31.** ***Espectro frecuencial de audio obtenido mediante el algoritmo desarrollado en C# con el método de adquisición número uno.***
**Tomado de: Autor**

En la figura 31 se evidencia el espectro de frecuencia de audio obtenido con el primer método.

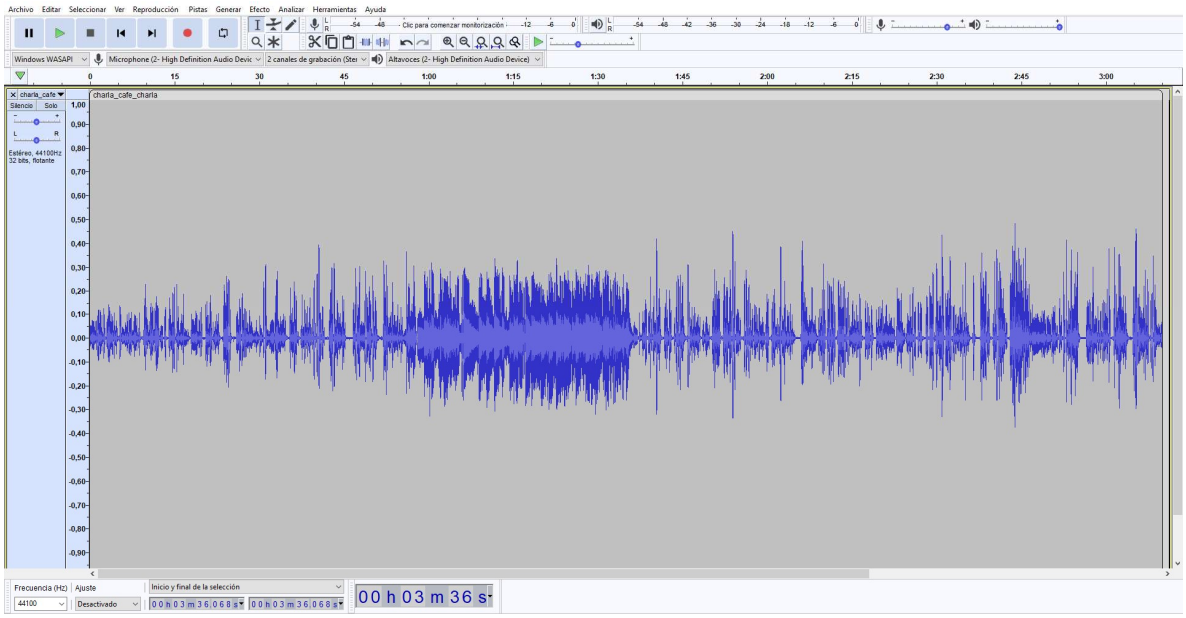


**Figura 32.** ***Audio obtenido con el software Audacity.***
**Tomado de: Autor**

En la Figura 32 se puede observar la señal de audio que contiene una pauta publicitaria grabada directamente con el software Audacity.

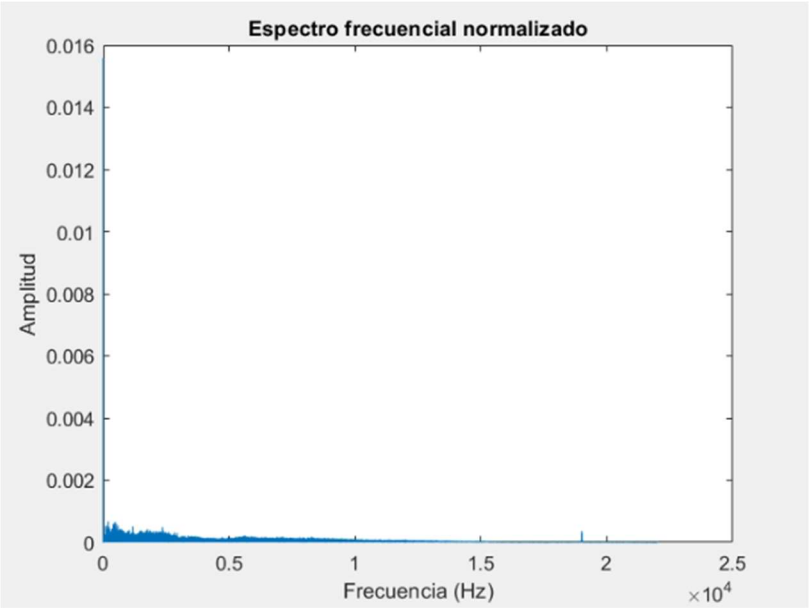


**Figura 33.** ***Espectro frecuencial de audio obtenido mediante el algoritmo desarrollado en C# con el método de adquisición número dos.***
**Tomado de: Autor**

La calidad de este audio es mayor que en el método anterior, como se evidencia en la figura 33, pero integrar este sistema se presentaban complicaciones como la dependencia a internet y el desarrollo de una interfaz de integración.

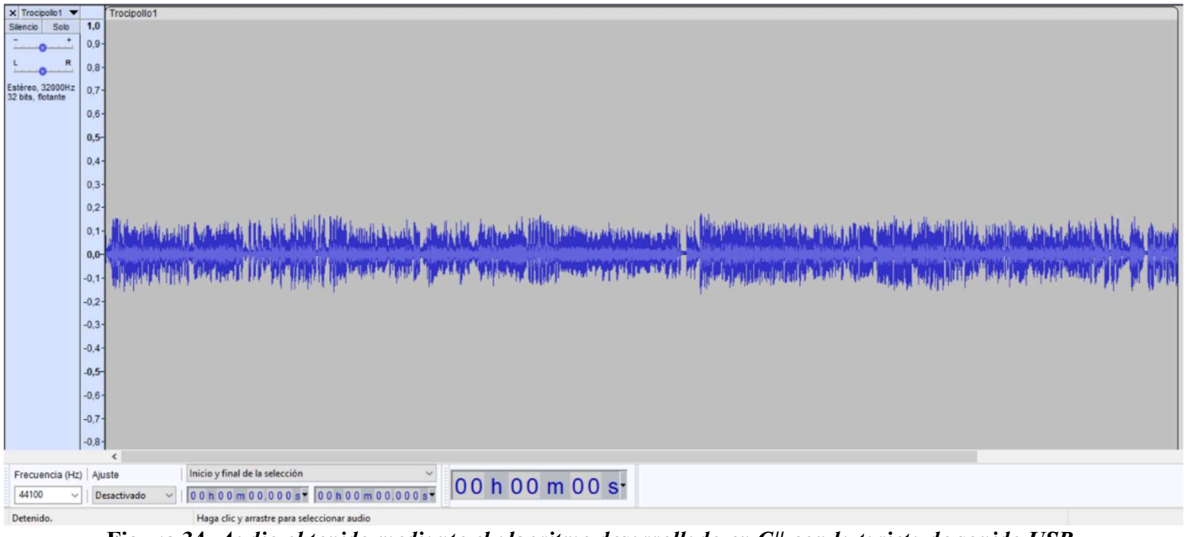


**Figura 34.** ***Audio obtenido mediante el algoritmo desarrollado en C# con la tarjeta de sonido USB.***
**Tomado de: Autor**

Por último, en la figura 34 se observa la señal utilizando la tarjeta de audio USB.

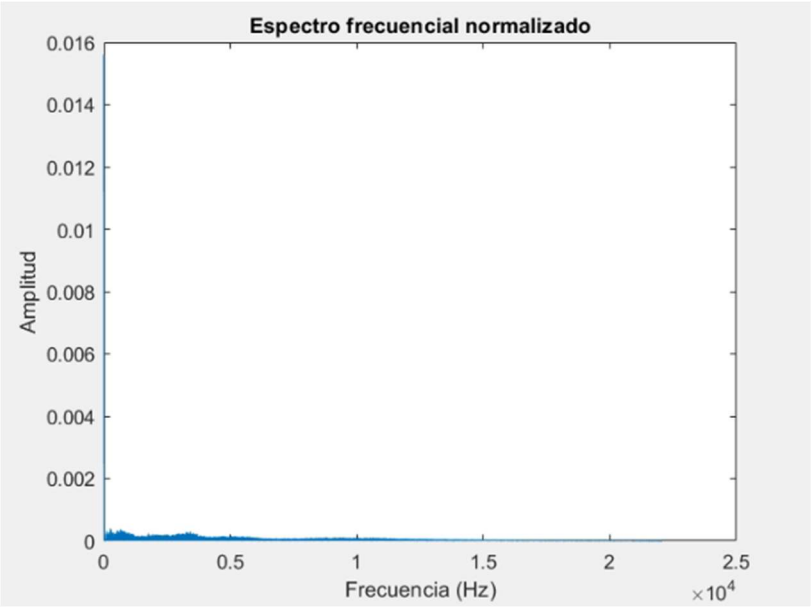


**Figura 35.** ***Espectro frecuencial de audio obtenido mediante el algoritmo desarrollado en C# con el método de adquisición número tres.***
**Tomado de: Autor**

En la figura 35 se observa el espectro de frecuencia de la señal. Se determinó que el método de adquisición número tres: Tarjeta de sonido USB externa era el más adecuado entre los métodos evaluados, ya permitió obtener el audio para que fuera grabado y almacenado por el algoritmo con una mejor calidad, directamente en el formato requerido para el prototipo (WAV) evitando la necesidad de integrar un software externo, aportando de esa manera flexibilidad al sistema. Utilizando este método se realizó la grabación y almacenamiento de las pautas publicitarias que compondrían los casos de estudio para los siguientes módulos.

## 9.4 Módulo de preprocesamiento de la señal

Para el método de preprocesamiento de señal, se tienen los archivos WAV de las captaciones almacenadas para ser procesadas. Un factor determinante para este algoritmo es que se realizó una evaluación de similitud con y sin los diferentes procesamientos de la señal, esto con el fin de determinar qué tan influyente era el preprocesamiento para el sistema.

Los casos estructurados de los que se habla en la delimitación del problema fueron evaluados junto con el módulo Speech To Text, el cual es fijo, y se realizó una comparación de los resultados obtenidos frente al texto de la pauta publicitaria. Los audios con los que se realizaron estas pruebas se encuentran en el enlace adjunto en el anexo 2.

**Tabla 6.** ***Tabla de similitudes de pautas publicitarias sin preprocesamiento de señal.***
**Tomado de: Autor**

| **16 bits de resolución - Frecuencia de muestreo 44.1 kHz** | | |
|---|---|---|
| | Escenario 1 sin procesamiento de señal | Escenario 2 sin procesamiento de señal |
| Pauta 1 | 77,90% | 68,50% |
| Pauta 2 | 91,20% | 92,18% |
| Pauta 3 | 83,03% | 78,64% |
| Pauta 4 | 88,27% | 84,36% |
| Pauta 5 | 91,39% | 91,39% |
| Pauta 6 | 97,96% | 94,65% |
| Pauta 7 | 98,76% | 98,02% |
| Pauta 8 | 74,24% | 76,13% |

En la tabla 6 se observan los resultados adquiridos para las nueve pautas con las que se realizó la evaluación de este módulo, evidenciando aquí dos escenarios y su porcentaje de similitud sin recibir ningún procesamiento de señal.

### *9.4.1 Filtros digitales.*

Para realizar la prueba con los filtros digitales en MATLAB se utilizó la función *audioread* que permite importar, crear, procesar y guardar archivos en formato WAV. Se aplicaron tres filtros digitales, con los que la señal pudiese mejorar y así contribuir a que su similitud fuera mayor. Luego de realizar el procesamiento de las señales con los filtros propuestos, se evaluó el porcentaje de similitud obtenido para determinar si existe mejora en el sistema. Los audios con los que se realizaron estas pruebas se encuentran en el enlace adjunto en el anexo 3.

**Tabla 7.** ***Tabla de similitudes de pautas publicitaria aplicando filtros digitales.***
**Tomado de: Autor**

| **16 bits de resolución - Frecuencia de muestreo 44.1 kHz** | | | | | | |
|---|---|---|---|---|---|---|
| | Filtro pasa banda Butterworth | | Filtro pasa bajos Butterworth | | Filtro pasa banda Chebyshev | |
| | Escenario 1 | Escenario 2 | Escenario 1 | Escenario 2 | Escenario 1 | Escenario 2 |
| Pauta 1 | 77,90% | **80,66%** | 73,48% | **76,79%** | 76,24% | 73,48% |
| Pauta 2 | 88,27% | **89,90%** | 91,20% | **92,83%** | 87,84% | **87,94%** |
| Pauta 3 | 91,18% | **93,21%** | **93,01%** | 90,61% | **88,42%** | 79,44% |
| Pauta 4 | 85,34% | 83,71% | 86,31% | **89,25%** | **86,31%** | 86,31% |
| Pauta 5 | 90,06% | **91,39%** | 76,82% | **94,04%** | 81,45% | **94,03%** |
| Pauta 6 | 98,47% | 94,14% | **97,20%** | 94,14% | **95,41%** | 94,14% |
| Pauta 7 | 98,76% | 98,02% | **98,49%** | 98,24% | **98,76%** | 98,02% |
| Pauta 8 | 79,54% | 77,13% | 71,21% | 75,37% | **75,37%** | 73,48% |

En la tabla 7 se pueden evidenciar los resultados obtenidos. Al comparar la tabla 7 con la tabla 6 es posible demostrar que para algunas pautas se ve una mejoría notable, pero para muchas otras es lo contrario.

### *9.4.2 Cambio en bits de resolución y frecuencia de muestreo.*

Al cambiar los bits de resolución y la frecuencia de muestreo se quiso verificar que el sistema incrementaba su desempeño proporcionalmente a la calidad de señal de audio que analizaba.

**Tabla 8.** ***Tabla de similitudes de pautas publicitarias cambiando los bits de resolución (16 bits) y frecuencias de muestreo.***
**Tomado de: Autor**

| **Bits de resolución: 16 bits** | | | | | | |
|---|---|---|---|---|---|---|
| **Porcentaje de similitud para las diferentes frecuencias** | | | | | | |
| | Frecuencia 11.025 Hz | | Frecuencia 22.050 Hz | | Frecuencia 32.000 Hz | |
| Pauta 1 | 77,90% | 79,55% | **78,45%** | 72,92% | 77,90% | 60,77% |
| Pauta 2 | **89,90%** | 87,62% | **92,83%** | 85,66% | **92,83%** | 83,71% |
| Pauta 3 | **87,02%** | 84,83% | **84,63%** | 79,44% | **82,23%** | 75,04% |
| Pauta 4 | 82,73% | **88,92%** | 86,31% | **86,64%** | **85,99%** | 85,99% |
| Pauta 5 | 80,13% | **86,75%** | **91,39%** | 91,39% | **91,39%** | 91,39% |
| Pauta 6 | **98,21%** | 94,14% | **99,49%** | 94,65% | **100%** | 94,65% |
| Pauta 7 | 98,76% | **99,50%** | **98,76%** | 98,02% | **98,76%** | 98,02% |
| Pauta 8 | 75,75% | **77,27%** | **78,78%** | 74,62% | 73,48% | **76,89%** |

En la tabla 8, se cambia la frecuencia de muestreo de la señal obtenida con 16 bits de resolución, la cual, al compararla con la tabla 8 se determina que la aplicación del proceso tiene un impacto imparcial para la mayoría de ellas, pues en algunos escenarios se obtiene mejor resultado, pero en muchos otros no.

**Tabla 9.** ***Tabla de similitudes de pautas publicitarias cambiando los bits de resolución (8 bits) y frecuencias de muestreo.***
**Tomado de: Autor**

| **Bits de resolución: 8 bits** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Frecuencia 11.025 Hz | | Frecuencia 22.050 Hz | | Frecuencia 32.000 Hz | | Frecuencia 44.010 Hz | |
| Pauta 1 | 75,00% | **78,03%** | 68,93% | 76,13% | 76,89% | **77,65%** | **78,40%** | 76,13% |
| Pauta 2 | **92,50%** | 86,97% | **92,18%** | 89,25% | **92,18%** | 88,92% | **89,25%** | 88,92% |
| Pauta 3 | 77,04% | **83,43%** | 73,85% | **84,23%** | **83,03%** | 81,83% | **82,23%** | 74,45% |
| Pauta 4 | 88,27% | **88,92%** | **88,59%** | 82,08% | **86,31%** | 86,31 | 86,31% | **86,64%** |
| Pauta 5 | 79,47% | **93,37%** | 88,74% | **91,39%** | **92,71%** | 91,39% | **92,71%** | 91,39% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Pauta 6 | **95,92%** | 94,14% | **98,47%** | 94,14% | **97,70%** | 94,65% | **98,47%** | 94,14% |
| Pauta 7 | 98,76% | **99,50%** | 97,28% | **98,02%** | **98,76%** | 98,02% | **98,76%** | 98,02% |
| Pauta 8 | 75% | **78,03%** | **73,48%** | 71,59% | 77,65% | **78,40%** | 77,65% | **78,03%** |

La tabla 9 muestra de igual manera el mismo comportamiento de poca fiabilidad y contundencia, pues los porcentajes se comportan de manera positiva solamente en ciertos casos.

Luego de realizar estas pruebas, no fue posible determinar un método sobresaliente sobre los otros, con lo cual se decide proceder con la calidad más alta de audio (16 bits a 44kHz), que, aunque con falta de contundencia presentó porcentajes de similitud superiores, comparada con otras.

## 9.5 Módulo de preprocesamiento del texto

El módulo de preprocesamiento del texto fue evaluado con los tres algoritmos señalados en el capítulo anterior para cada una de las pautas publicitarias (casos) en cuestión.

### *9.5.1 Caso uno: Pauta publicitaria de trocipollo con el algoritmo de tokenización excluyendo signos de puntuación con diccionario cerrado.*

Texto proveniente del módulo Speech To Text: *"Oye te acompaña a reconocer esos momentos que marcan la diferencia a escuchar esa voz que te invita a descubrir lo que te apasiona en la vida. Quiere decirte algo, emprender, disfrutar, ser feliz está en ti una invitación de ETB y me encanta que las cosas buenas se vuelvan. Requeté buenas. Cómo tener una cita con tu Crush o que te regalen nosotros y pollo, eso es más que bueno y ahora más que otras y pollo tienen nuevos sabores.Requeté, bueno, son los dos nuevos sabores de trozzi, pollo, pollo, brasa picante y pollo, California. ¿Cuál es tu favorito?Escuchas tropicana, la más bacana bacanísima si el dolor de cabeza te deja sin libertad y hacer lo que quieres liberarte con calmidol, liberate para que disfrutes tu mundo. Liberate para reír sin parar. Liberate para vivir a tu manera. Libérate del dolor de cabeza activa tu mente y no pares calmidol tan bueno para los cólicos como para el dolor de cabeza. Es un medicamento no exceder su consumo si los síntomas persisten, consultar al médico soy David luna, candidato al Senado por cambio radical."*

Texto proveniente del módulo Speech To Text con procesamiento: "*oye te acompaña a reconocer esos momentos que marcan la diferencia a escuchar esa voz que te invita a descubrir lo que te apasiona en la vida quiere decirte algo emprender disfrutar ser feliz está en ti una invitación de etb y me encanta que las cosas buenas se vuelvan requeté buenas cómo tener una cita con tu crush o que te regalen unos trocipollo pollo eso es más que bueno y ahora más que otras y pollo tienen nuevos sabores requeté bueno son los dos nuevos sabores de trocipollo pollo brasa picante y pollo california cuál es tu favorito escuchas tropicana la más bacana bacanísima si el dolor de cabeza te deja sin libertad y hacer lo que quieres liberarte con calmidol liberate para que disfrutes tu mundo liberate para reír sin parar liberate para vivir a tu manera libérate del dolor de cabeza activa tu mente y unos*

*trocipollo pares calmidol tan bueno para los cólicos como para el dolor de cabeza es un medicamento unos trocipollo exceder su consumo si los síntomas persisten consultar al médico soy david luna candidato al senado por cambio radical."*

Pauta publicitaria precisada: *"me encanta que las cosas buenas se vuelvan requete buenas cómo tener una cita con tu crush o que te regalen unos trocipollo eso es más que bueno y ahora más que trocipollo tienen nuevos sabores requeté bueno son los dos nuevos sabores de trocipollo pollo brasa picante y pollo california cuál es tu favorito."*

Pauta publicitaria obtenida: *"me encanta que las cosas buenas se vuelvan. Requeté buenas. Cómo tener una cita con tu Crush o que te regalen nosotros y pollo, eso es más que bueno y ahora más que otras y pollo tienen nuevos sabores.Requeté, bueno, son los dos nuevos sabores de trozzi, pollo, pollo, brasa picante y pollo, California. ¿Cuál es tu favorito?"*

Pauta publicitaria obtenida y procesada: *"me encanta que las cosas buenas se vuelvan requeté buenas cómo tener una cita con tu crush o que te regalen unos trocipollo pollo eso es más que bueno y ahora más que trocipollo pollo tienen nuevos sabores requeté bueno son los dos nuevos sabores de trocipollo pollo brasa picante y pollo california cuál es tu favorito"*

## 9.6 Módulo de comparación del texto obtenido

El módulo de comparación del texto obtenido fue evaluado con los tres algoritmos señalados en el capítulo anterior para cada una de las pautas publicitarias (casos) en cuestión, las cuales se evidencian entre los puntos 9.4.1.1 a 9.4.1.3. Los audios con los que se realizaron estas pruebas se encuentran en el enlace adjunto en el anexo 4.

### *9.6.1 Caso uno: Pauta publicitaria de Trocipollo con los tres algoritmos sin procesamiento de texto.*

Pauta publicitaria precisada: *"me encanta que las cosas buenas se vuelvan requete buenas cómo tener una cita con tu crush o que te regalen unos trocipollo eso es más que bueno y ahora más que trocipollo tienen nuevos sabores requeté bueno son los dos nuevos sabores de trocipollo pollo brasa picante y pollo california cuál es tu favorito"*

#### 9.6.1.1 Distancia de Levenshtein:

"me encanta que las cosas buenas se vuelvan requeté buenas cómo tener una cita con tu crush o que te regalen unos trocipollo pollo eso es más que bueno y ahora más que trocipollo pollo tienen nuevos sabores requeté bueno son los dos nuevos sabores de trocipollo pollo brasa picante y pollo california cuál es."

#### 9.6.1.2 Similitud de Jaro Winkler:

"me encanta que las cosas buenas se vuelvan requeté buenas cómo tener una cita con tu crush o que te regalen unos trocipollo pollo eso es más que bueno y ahora más que trocipollo pollo

tienen nuevos sabores requeté bueno son los dos nuevos sabores de trocipollo pollo brasa picante y pollo california cuál es”

#### 9.6.1.3 Similitud del coseno:

“ncanta que las cosas buenas se vuelvan requeté buenas cómo tener una cita con tu crush o que te regalen unos trocipollo pollo eso es más que bueno y ahora más que trocipollo pollo tienen nuevos sabores requeté bueno son los dos nuevos sabores de trocipollo pollo brasa picante y pollo california cuál es tu”

**Tabla 10.** ***Resultados de similitud y tiempo de proceso para el caso uno con cada algoritmo propuesto para el módulo de comparación del texto obtenido.***
**Tomado de: Autor**

| Algoritmo | Tiempo de proceso | Porcentaje de similitud (texto sin procesar) | Porcentaje de similitud (texto procesado) |
|---|---|---|---|
| **Levenshtein** | 5.38s | 91.20 | 91.85 |
| **Jaro Winkler** | 0.32s | 89.56 | 90.64 |
| **Coseno** | **0.14s** | 93.01 | **97.35** |

De la misma manera, se realizó la evaluación para las diferentes pautas publicitarias y los resultados obtenidos fueron los mostrados en la tabla 10, en donde el mejor desempeño con respecto al tiempo de proceso se ve en la similitud del coseno, y en conjunto con el texto procesado, alcanza un porcentaje de similitud con la pauta emitida de 97.35%

**Tabla 11.** ***Resultados de similitud y tiempo de proceso para el caso dos con cada algoritmo propuesto para el módulo de comparación del texto obtenido.***
**Tomado de: Autor.**

| Algoritmo | Tiempo de proceso | Porcentaje de similitud (texto sin procesar) | Porcentaje de similitud (texto procesado) |
|---|---|---|---|
| **Levenshtein** | 8.84s | 83.06 | 94.62 |
| **Jaro Winkler** | 0.52s | 85.48 | 89.59 |
| **Coseno** | **0.19s** | 90.91 | **98.13** |

En la tabla 11 se observa que nuevamente con respecto al tiempo de proceso, la similitud del coseno es la que tiene un desempeño mayor, con 0.19 segundos de tiempo de proceso antes de entregar el porcentaje de similitud con la pauta principal, y logrando un total de 98.13% para el texto procesado.

**Tabla 12.** ***Resultados de similitud y tiempo de proceso para el caso tres con cada algoritmo propuesto para el módulo de comparación del texto obtenido.***
**Tomado de: Autor.**

| Algoritmo | Tiempo de proceso | Porcentaje de similitud (texto sin procesar) | Porcentaje de similitud (texto procesado) |
|---|---|---|---|
| **Levenshtein** | 6.79s | 88.56 | 97.38 |
| **Jaro Winkler** | **0.30s** | 91.76 | 93.11 |
| **Coseno** | 0.37s | 90.36 | **98.59** |

En la tabla 12 se observa que el algoritmo de similitud de Jaro Winkler entregó la respuesta en 0.30 segundos, esta vez 7 segundos más rápido que la similitud del coseno, aunque con un porcentaje de similitud inferior en 5.59 % cuando se compara con el algoritmo del coseno.

**Tabla 13.** ***Resultados de similitud y tiempo de proceso para el caso cuatro con cada algoritmo propuesto para el módulo de comparación del texto obtenido.***
**Tomado de: Autor.**

| Algoritmo | Tiempo de proceso | Porcentaje de similitud (texto sin procesar) | Porcentaje de similitud (texto procesado) |
|---|---|---|---|
| **Levenshtein** | 9.33s | 64.98 | 97.47 |
| **Jaro Winkler** | **0.308s** | 87.67 | 93.35 |
| **Coseno** | 1.5s | 86.12 | **98.94** |

En la tabla 13 se observa nuevamente que Jaro Winkler tiene un desempeño notablemente superior en cuanto al tiempo de proceso con 0.308 segundos, y similar a con el caso anterior, el porcentaje de similitud para el coseno es superior a los otros dos algoritmos con 98.94 %

**Tabla 14.** ***Resultados de similitud y tiempo de proceso para el caso cinco con cada algoritmo propuesto para el módulo de comparación del texto obtenido.***
**Tomado de: Autor.**

| Algoritmo | Tiempo de proceso | Porcentaje de similitud (texto sin procesar) | Porcentaje de similitud (texto procesado) |
|---|---|---|---|
| **Levenshtein** | 3.5s | 74.24 | 82.19 |
| **Jaro Winkler** | 0.76s | 95.42 | **98.58** |
| **Coseno** | **0.11s** | 89.96 | 93.77 |

En la tabla 14 se puede evidenciar que el tiempo de proceso del algoritmo del coseno es notablemente menor contra los demás algoritmos, pero reduciendo el porcentaje de similitud, pues el mayor porcentaje de similitud se observa para Jaro Winkler con 98.58%.

**Tabla 15. Resultados del sistema evaluado con respecto a los criterios de rendimiento definido. Tomado de: Autor**

| **Tabla de evaluación del sistema** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Pauta Nº** | % Similitud | | Falsos positivos | Falsos negativos | Porcentaje aciertos | Tiempo de proceso de verificación | |
| | Esc1 | Esc2 | | | | Esc1 | Esc2 |
| **1** | 96.98% | 97.30% | 0 | 0 | 100% | 29.941seg | 21.317seg |
| **2** | 98.13% | 90.98% | 0 | 1 | 50% | 28.312seg | 38.309seg |
| **3** | 98.59% | 93.16% | 0 | 1 | 50% | 33.166seg | 35.366seg |
| **4** | 98.94% | 94.59% | 0 | 1 | 50% | 92.830seg | 82.603seg |
| **5** | 93.77% | 91.47% | 0 | 2 | 0% | 28.94seg | 43.572seg |

En la tabla 15 se observan los resultados obtenidos de las pruebas para el prototipo del sistema implementado, en el cual se evaluaron todos los criterios de rendimiento establecidos al principio de este trabajo. Se realizó una prueba por cada escenario. El porcentaje de acierto determinado para la emisión de una pauta publicitaria fue del 95%. Teniendo en cuenta el criterio de evaluación de falsos positivos, se puede demostrar que el sistema no entregará emisiones erróneas de la pauta, sin embargo, en un porcentaje de fallo del 5%, es probable que sean rechazadas algunas pautas que fueron emitidas correctamente. Los porcentajes de similitud fueron en su mayoría por encima del umbral determinado.

# 10. Conclusiones

Durante la investigación del proyecto se plantearon diferentes componentes para el sistema con el fin de obtener resultados positivos de los cuales, posterior a diversas pruebas se descartaron algunos de ellos, entendiendo que, para los sistemas de reconocimiento y transcripción de texto, algunos de los procesamientos de señal no son benignos y por el contrario pueden afectar el rendimiento del sistema. Con lo cual se concluye que, para el prototipo implementado en este trabajo, los componentes finales otorgan una serie de resultados positivos para la mayoría de las pautas analizadas. Haber determinado los criterios de rendimiento con los cuales el sistema iba a ser evaluado desde un principio, guio la investigación hacia el rendimiento esperado.

El porcentaje de similitud ajustado para la aprobación de la emisión de la pauta con el que se realizaron estas pruebas fue del 95%, siendo un valor alto y exigente para evaluar el sistema. Cabe mencionar que con más del 90% se puede asegurar la detección de la emisión de la pauta, si se coloca este valor de acuerdo con los resultados anteriores se puede determinar que todas las pautas probadas en los diferentes escenarios plasmados fueron emitidas durante la transmisión radial.

Existen diversas extensiones, librerías y herramientas de acceso libre las cuales son constantemente mejoradas y probadas por sus respectivos desarrolladores, en distintos escenarios y con QA automatizados que fueron de ayuda para la contextualización bibliográfica y el desarrollo de los algoritmos planteados en este trabajo, ya que ofrecen soluciones a la vanguardia y por lo tanto permiten orientar las soluciones a diversos objetivos, con lo cual fue posible tomar la teoría aplicada a este trabajo obteniendo conjuntos de resultados que encaminaron a obtener un prototipo optimizado.

# Referencias bibliográficas


Furui, S., Kikuchi, T., Shinnaka, Y., & Hori, C. (2004). Speech-to-text and speech-to-speech summarization of spontaneous speech. IEEE transactions on speech and audio processing: a publication of the IEEE Signal Processing Society, 12(4), 401–408. https://doi.org/10.1109/tsa.2004.828699

Yang, L., & Yan, Z. (2011). Study on audio signal's classification based on BP neural network. 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 5153–5155.

Pavan, S., Schreier, R., & Temes, G. C. (2017). Sampling, oversampling, and noise-shaping. En Understanding Delta-Sigma Data Converters (pp. 27–61). John Wiley & Sons, Inc.

La inversión publicitaria cae un 2,2% en el tercer trimestre del año. (2019, octubre 21). PR Noticias. https://prnoticias.com/2019/10/21/inversion-publicidad-tercer-trimestre-prensa-radio-television/

Lee, S., Kim, J., & Lee, I. (2012). Speech/Audio Signal Classification Using Spectral Flux Pattern Recognition. 2012 IEEE Workshop on Signal Processing Systems, 232–236.

Gómez, E., & Anàlisi, G. (s/f). Introducción al filtrado digital. Wordpress.com. Recuperado el 26 de agosto de 2022, de https://processamentdelso.files.wordpress.com/2012/02/tema7-filtrosdigitales.pdf

Baxter, P. (1985). The role of the British library R & D department in supporting library and information research in the United Kingdom. Journal of the American Society for Information Science. American Society for Information Science, 36(4), 275–277. https://doi.org/10.1002/asi.4630360411

Shadiev, R., Reynolds, B. L., Huang, Y.-M., Shadiev, N., Wang, W., Laxmisha, R., & Wannapipat, W. (2017). Applying speech-to-text recognition and computer-aided translation for supporting multi-lingual communications in cross-cultural learning project. 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), 182–183.

Aguilar-Chacon, J. E., & Segura-Torres, D. A. (2020). Evaluation methodology for Speech To Text Services similarity and speed characteristics focused on small size computers. IOP conference series. Materials science and engineering, 844(1), 012039. https://doi.org/10.1088/1757-899x/844/1/012039

(N.d.). Edu.Ar. Retrieved August 26, 2022, from http://carina.fcaglp.unlp.edu.ar/senales/apuntes/frec_Nyquist.pdf

Antón, E. R. (n.d.). El tono de la voz masculina y femenina en los informativos radiofónicos:

un análisis comparativo. Ubi.Pt. Retrieved August 26, 2022, from http://www.bocc.ubi.pt/pag/rodero-emma-tono-voz-femenina.pdf

Haldar, R., & Mukhopadhyay, D. (2011). Levenshtein distance technique in dictionary lookup methods: An improved approach. In arXiv [cs.IT]. http://arxiv.org/abs/1101.1232

T. Grust et al. (Hrsg.): Datenbanksysteme für Business, Technologie und Web (BTW 2019), Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 205

Lahitani, A. R., Permanasari, A. E., & Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment. 2016 4th International Conference on Cyber and IT Service Management

Pons, J., & Sirvardiere, P. (2002). Certificación de calidad de los alimentos orientada a sellos de atributos de valor en países de américa latina.

Proakis, J. G., & Manolakis, D. G. (2007). Tratamiento digital de señales. (4th ed.). Madrid, España.: PEARSON EDUCATION SA.

Sangkil Lee, Jieun Kim, & Insung Lee. (Oct 2012). Speech/audio signal classification using spectral flux pattern recognition. Paper presented at the 232-236. doi:10.1109/SiPS.2012.36 Retrieved from https://ieeexplore.ieee.org/document/6363260

Inversión publicitaria. (2019). Retrieved from https://prnoticias.com/marketing/inversion-publicitaria/20175821-inversion-publicidad-tercer-trimestre-prensa-radio-television

Digital preservation team preservation assessment: WAV format preservation assessment. (n.d.). Dpconline.org. Retrieved August 26, 2022, from https://wiki.dpconline.org/images/4/46/WAV_Assessment_v1.0.pdf

N. Sharma and S. Sardana, "A real time Speech To Text conversion system using bidirectional Kalman filter in Matlab," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016, pp. 2353-2357, doi: 10.1109/ICACCI.2016.7732406.

B. TombaloĞlu and H. Erdem, "A SVM based Speech To Text converter for Turkish language," 2017 25th Signal Processing and Communications Applications Conference (SIU), 2017, pp. 1-4, doi: 10.1109/SIU.2017.7960486.

R. Shadiev et al., "Applying Speech-to-Text Recognition and Computer-Aided Translation for Supporting Multi-lingual Communications in Cross-Cultural Learning Project," 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), 2017, pp. 182-183, doi: 10.1109/ICALT.2017.20.

Rumnsey. Mc Cormick, F. T. (2014). *Introduccion al Sonido y la Grabacion* (7.ª ed.). IORTV.

Indurkhya. Damerau., N. I. (2010). *Handbook of Natural Lenguage Processing* (2.ª ed.). CRC.

Semeria, Marcelo (2015): Los tres teoremas: Fourier - Nyquist - Shannon, Serie Documentos de Trabajo, No. 582, Universidad del Centro de Estudios Macroeconómicos de Argentina (UCEMA), Buenos Aires

dotnet-bot. (s/f). *TimeSpan struct*. Microsoft.com. Recuperado el 28 de septiembre de 2022, de https://learn.microsoft.com/en-us/dotnet/api/system.timespan?view=net-7.0

## ANEXOS

Anexo 1: Casos estructurados manualmente.

https://drive.google.com/drive/u/3/folders/1JT3hkZobHHXsp27U0jxFgZubJ7b8RrgE

Anexo 2: Casos preprocesados con los filtros digitales.

https://drive.google.com/drive/u/3/folders/1caoY2k5q1l7WxSm_M4i6lQhtMRXT-D5T

Anexo 3: Casos con diferente frecuencia de muestreo y bits de resolución

https://drive.google.com/drive/u/3/folders/1axTr1RB1UbRz3zf9_FERivOlOFJsbVKr

Anexo 4: Pautas preprocesadas utilizadas para el módulo de comparación de texto.

https://drive.google.com/drive/u/3/folders/1z6QnwHMMgIaVQZXGJpptXBPkW4xtTXmG

Anexo 5: Algoritmo del prototipo implementado.
https://drive.google.com/drive/u/3/folders/10YSzy3lwWIaMbW9-V3WbLrKz9P7oebtN