Luis Felipe Flores Sánchez
Luis Darío Hinojosa Quijada

A01274880
A01028822

Análisis de Sistemas de Señales

Dr. José Luis Cuevas Ruiz

Santa Fe, 9 de septiembre de 2021

A01022687@itesm.mx
A01274880@itesm.mx

Repositorio de Github:

El código fuente de la solución puede ser encontrado en el siguiente repositorio:
https://github.com/LuisDarioHinojosa/VoiceRecognition

Preámbulo:

El presente escrito tiene como fin presentar el diseño, resultados, y conclusiones finales de nuestra propuesta de solución a la situación problema. Esta consiste, a grandes rasgos, en un sistema de registro y reconocimiento de voz.

Introducción:
En la actualidad la voz ya es la puerta para muchos campos en la sociedad, en el ámbito de la tecnología muchos servicios y aplicaciones que se usan en el día a día ocupan el aparato vocal como medio de identificación, la biometría de voz es un sistema que identifica la autenticidad de cada individuo a través de patrones de voz, donde la tecnología ha aprovechado para implementarlo como un tipo NIP único para cada persona y así convertirse inmune a imitaciones. La huella vocal toma importancia para el uso de aplicaciones donde con nuestra voz podamos interactuar como un medio de identificación

Para la implementación de un sistema de registro se debe abordar diferentes cosas de la voz, ya sea por sus propiedades físicas o sobre qué datos se pueden llegar a analizar con una simple muestra, ya que la voz en conjunto es una muestra de datos que se puede desplegar con diferentes técnicas, ya sea con la estadística descriptiva o con el uso de Software que puede brindar información para hacer distintas referencias y hacer una diferencias sobre las muestras presentadas, donde esto se vuelve tendencia en relación al tema de la biometría de voz, además de considerar aspectos técnicos y la recolección de datos que se pueden dar.

Desde hace tiempo la tecnología ha formalizado el uso de la voz en interfaces para experiencia de usuario, donde se tienen los famosos asistentes de voz, por mencionar algunos como Alexa, Google Assistant, Bixby, entre otros que destacan por su funcionalidad, donde estos tienen la capacidad de análisis en un sistema que permite identificar señales sonoras, donde el trabajo es identificar frases o palabras del lenguaje natural y este pueda convertirse en un formato legible que pueda entender la máquina, ya dependiendo de la programación, se tiene el software de rendimiento indumentario que pueden llegar a ser muy complejo que otros y que tienen un mejor análisis del lenguaje natural.

El desarrollo de este escrito consiste en presentar una propuesta de solución para identificar muestras de voz a través del uso del Software Matlab presentando una demostración de la diferencia de voces, a partir de técnicas de la estadística donde se menciona que la voz es un conjunto de datos valioso que se puede estudiar para fines determinados, donde en ello la tecnología es una ventaja para mejorar la observación que se desea obtener. Dado ello se busca examinar y dar planteamiento a una problemática que aborda la voz para permitir identificar personas.

Diagrama de bloques (versión 2.0):

Durante la etapa 1 del reto, se presentó un primer diagrama de bloques basado en nuestra investigación preliminar:

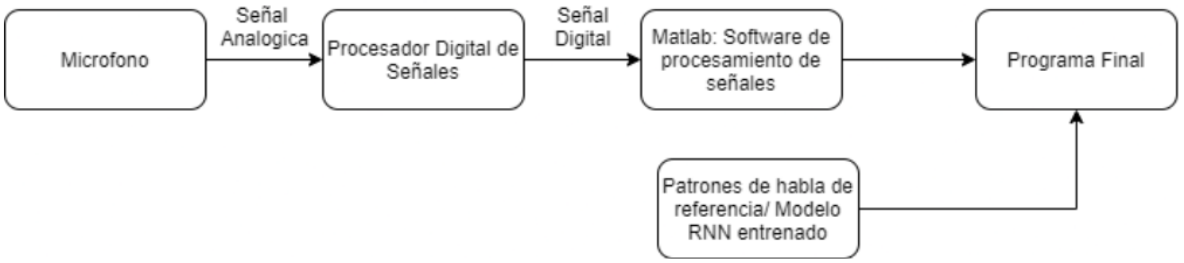


Figura 1 Diagrama de Bloques Versión Preliminar

Este primer diagrama se bosquejo a partir de un sistema industrial actual. Algunos aspectos que incorporaba estaban fuera de alcance, porque el lapso del proyecto (3 semanas) imposibilitaba buscar un dataset lo suficientemente grande como para entrenar un modelo de RNN en Python. Entonces, decidimos simplificar el diagrama de bloques. Para esto, se eliminaron los elementos de IA, y se obviaron los elementos de hardware (micrófono y microprocesador) pues ya están implícitos dentro de nuestros ordenadores personales. Nos centramos en el aspecto del reconocimiento (**¿a que miembro del equipo pertenece una muestra de audio determinada?**) y validación (**¿que tan seguros podemos estar de que dicha muestra pertenece a un individuo en particular?**).

El diagrama de bloques final se presenta a continuación:

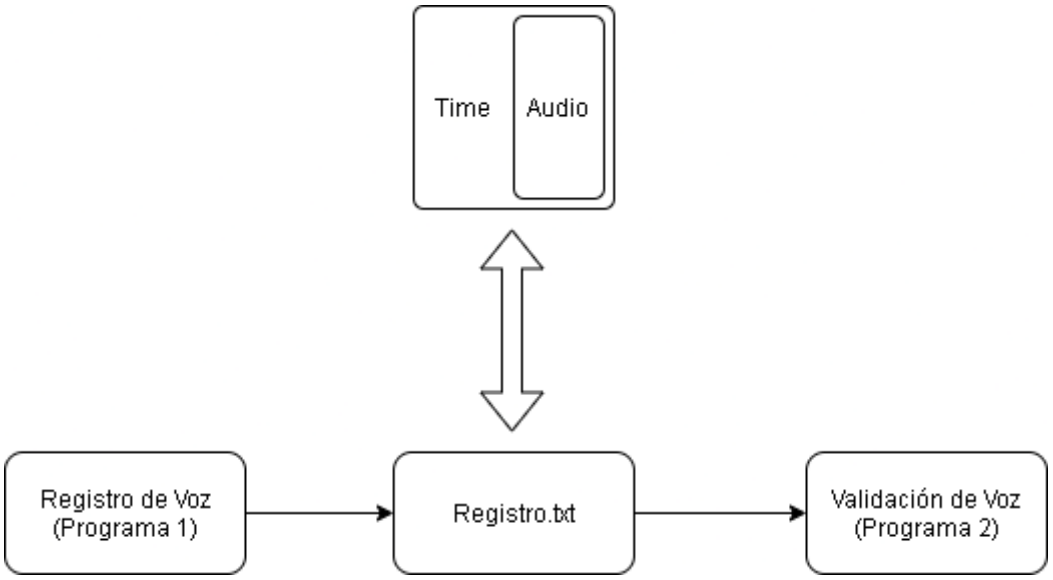


Figura 2 Diagrama de Bloques

- Registro de señales: Este consiste en un script de matlab que se encarga solamente de dar de alta un usuario, y de registrar una muestra de audio que servirá como referencia para el reconocimiento de voz. Es decir, para hacer reconocimiento de voz, se requiere que el usuario que quiere revalidar su voz proporcione una muestra de audio para ser comparada con una referencia que ya está almacenada en el

sistema. El script de registro de señales se encarga de registrar en el sistema las referencias a manera de archivos de texto.

- Reconocimiento de voz: En este script se realiza la tarea de validación y reconocimiento de voz. A grandes rasgos, se extrae la información de las referencias (almacenadas en documentos de texto), se graba una muestra de voz (de un usuario que quiere revalidar su voz), se hacen los cálculos pertinentes (ver sección: Metodología), y se identifica a la persona con base en su voz.

Metodología:

A continuación, se explicará a detalle la manera en la que se logró la funcionalidad de cada script.

Registro de voces:

En general, nos basamos en el programa de muestra proporcionado para todo lo que conllevase grabar audio (ver Anexo 1).
Nuestro programa de registro de voces es prácticamente el código fuente de muestra con ligeras modificaciones. Primero, se ideó un código básico para saber a qué miembro del equipo corresponde la referencia que será grabada: se pide al usuario que entre su matrícula del tec. Por cuestiones de tiempo y simplicidad, sólo son válidas las matrículas de los miembros del equipo:

```
4   p1 = "Enter you university id: A######## \n";
5   pe = "User is not registered. Enter another one: \n";
6
7   mat = input(p1,'s')
8   % user validation pending
9   while strcmp(mat,"A01028822") == 0 | strcmp(mat,"A01028822") == 0
10      mat = input(pe,'s')
11  end
12  fprintf("Identification passed.\nPlease say in loud voice: Hola mundo! Esto es una prueba.\n ");
13
```

Figura 3 Registro de identidad

Nótese que en caso de que el usuario entre cualquier otra cosa que no sean nuestras matrículas, el sistema seguirá pidiendo un dato válido, y no se dejará avanzar al usuario. Evidentemente, el sistema no es confiable en términos de escalabilidad y ciberseguridad. Sin embargo, dado el lapso de tiempo para la elaboración del proyecto, consideramos que cumple a la perfección su propósito demostrativo.
Para la grabación de la referencia de audio, se usó el bloque de código muestra proporcionado en canvas (ver Anexo 1). Únicamente se cambió el tiempo de grabación de 5 segundos a 3 segundos.
Finalmente, los datos son exportados a archivos de texto con el formato “Matricula.txt”. De igual manera, se implementa lógica básica para determinar el nombre del archivo de texto:

```
31  if strcmp(mat,"A01028822") == 1
32      dlmwrite('A01028822.txt',tab,'delimiter','\t','newline','pc')
33      fprintf("Data written in A01028822.txt \n")
34  else
35      dlmwrite('A01274880.txt',tab,'delimiter','\t','newline','pc')
36      fprintf("Data written in A01274880.txt \n")
37  end
```

Figura 4 Lógica de exportación.

En concreto, los datos que son exportados son:

- Timesteps de la señal de audio.
- Valores de la señal de audio en el dominio del tiempo.

En matlab, estos valores son representados por arreglos. Para su exportación, se acomodan en una matriz, de modo que cada arreglo forme una columna, y las filas conformen cada pareja (tiempo,valor):

```
5422   0.67763 -0.21875
5423   0.67775 -0.19531
5424   0.67788 -0.14844
5425   0.678   -0.023438
5426   0.67812 0.13281
5427   0.67825 -0.36719
5428   0.67838 -0.0078125
5429   0.6785  0.20312
5430   0.67863 0
5431   0.67875 -0.0078125
5432   0.67888 -0.17188
```

Figura 5 Ejemplo de datos recopilados

Reconocimiento de voz:

En cuanto al script de reconocimiento de voz, se comienza leyendo y extrayendo la información de los archivos de texto. Para cada referencia, se lee la matriz del archivo de texto, y se reordena en 2 arreglos que contiene la información de los valores de la señal de audio en el dominio del tiempo y los timesteps respectivamente:

```
fprintf("Retrieving Data... \n");
ref1 = readmatrix("A01028822.txt");
ref2 = readmatrix("A01274880.txt");
fprintf("Data Retreived \n");

% extract arrays from first reference
tRef1 = ref1(:,1);
rRef1 = ref1(:,2);

% extract arrays from second reference
tRef2 = ref2(:,1);
rRef2 = ref2(:,2);

```

Figura 6 Lectura y extracción de señales de audio

Para grabar la muestra de audio (la cual debe ser comparada con las referencias extraídas de los archivos de texto para el reconocimiento de voz) se usa nuevamente el código de muestra de canvas con el tiempo de grabación ajustado a 3 segundos.

Nota:

La razón por la que el tiempo de grabación se estableció en 3 segundos, es por que nuestra frase (“Hola mundo! Esto es una prueba”) toma de 2 a 3 segundos en recitarse. Si se usaran los 5 segundos, todos los valores en cero repercuten significativamente en las estadísticas requeridas para la creación de las distribuciones de probabilidad.

Ahora, el procesamiento de reconocimiento de voz, se dividió en 2 criterios: la función de densidad de probabilidad, y el coeficiente de correlación.

Función de densidad de probabilidad (PDF):

Esta es básicamente una función que indica, con base en una distribución, cuál es la probabilidad de que una variable aleatoria tome un valor determinado. Es decir, nos da una idea de que valores en concreto toma la señal de audio y la frecuencia con lo que lo hacen (resume de manera compacta el comportamiento de la señal de audio).

Nuestra propuesta de soluciones se basa en que si la muestra de voz corresponde a la misma voz que la referencia entonces la diferencia entre la distribución de probabilidad será mínima. Con esto en mente, decidimos que un criterio para identificar la voz es que si la raíz del error medio cuadrado es despreciable, entonces las distribuciones de probabilidad de muestra y referencia son prácticamente idénticas, y la voz pertenece a la misma persona.

La ecuación del error medio cuadrado se muestra a continuación:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (valorReferencia - valorMuestra)^2}{N}}$$

Para computar el error, se suman las diferencias elevadas al cuadrado (para eliminar sesgos por la sumatoria de números negativos) de cada uno de los valores de la PDF de la

referencia y la PDF de la muestra. Toda la sumatoria se divide entre el número de muestras (ambos arreglos tienen la misma longitud), y calcula la raíz cuadrada.

Coeficiente de correlación:

Nuestro segundo criterio fue el coeficiente de correlación lineal entre la muestra de audio con cada una de las referencias. A grandes rasgos, esta medición nos indica el grado de proporcionalidad directa o proporcionalidad inversa entre dos variables. La idea es que dependiendo del valor del coeficiente de correlación podemos determinar cómo se mueven las señales:

R^2	Significado
1	Proporcional directa perfecta
0	No hay relación
-1	Proporcional directa perfecta

Figura 7 Lectura de coeficiente de correlación

Entonces, lo que se realizó fue computar una matriz de correlación, y extraer los coeficientes de la muestra con cada una de las referencias. Después eso, los coeficientes se normalizaron a 1 para obtener una distribución de probabilidad una función de activación:

$$A^{[n]} = \frac{e^{m^{[n]}}}{\sum_{i=1}^{N} e^{m[i]}}$$

Esta operación se le conoce como Softmax y sirve para tareas de clasificación (como determinar si una voz pertenece a A01028822 o a A01274880). Los coeficientes de correlación fueron ajustados de modo que su suma sea igual a 1. Si la voz corresponde, por ejemplo, a A01028822, el coeficiente de correlación de la muestra con esa referencia va a ser mayor que el otro. Entonces, a la hora de normalizarse, este tendrá una probabilidad mayor.

Resultados:

Recopilación de pruebas:

Con el fin de determinar apropiadamente las condiciones de operación, se simularon diferentes escenarios para probar el desempeño del sistema ante condiciones adversas. En el siguiente video se pueden encontrar algunos de los experimentos realizados (por motivos de tiempo y practicidad, solo se seleccionó un caso de cada experimento):

- Demos del sistema (Recopilación):

  https://www.youtube.com/watch?v=AV4_rfetyAs

Es importante recalcar que se realizaron 5 pruebas en cada escenario (excepto para el experimento con el filtro de voz). Dichos escenarios se listan a continuación:

1. Pruebas en ambiente controlado (en un espacio cerrado sin ruidos de fondo y con un micrófono bluetooth).
2. Pruebas de reconocimiento de una voz femenina.
3. Pruebas de reconocimiento de voz en un entorno urbano (con ruidos de ciudad de fondo).
4. Pruebas de reconocimiento de voz durante una tormenta (con ruidos de lluvia de fondo).
5. Pruebas de reconocimiento de filtros de voz.
6. Pruebas de reconocimiento ante animales (con ladridos de perros de fondo).

Observaciones y resultados experimentales:

A decir verdad, a pesar de que hubo errores ocasionales el sistema se desempeñó mejor de lo esperado. También, se confirmó la hipótesis planteada en el primer experimento en condiciones controladas: el reconocimiento de voz mediante el error medio cuadrado sobre las funciones de densidad de probabilidad es más confiable que la función softmax.

Resumen de desempeño:

Las siguientes gráficas resumen el desempeño del sistema en cada uno de los escenarios:

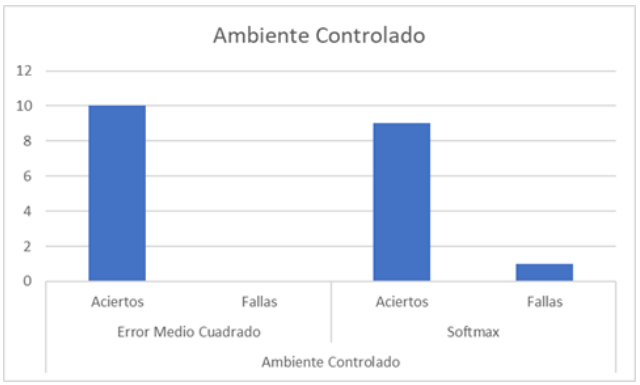


Figura 7 Desempeño en ambiente controlado

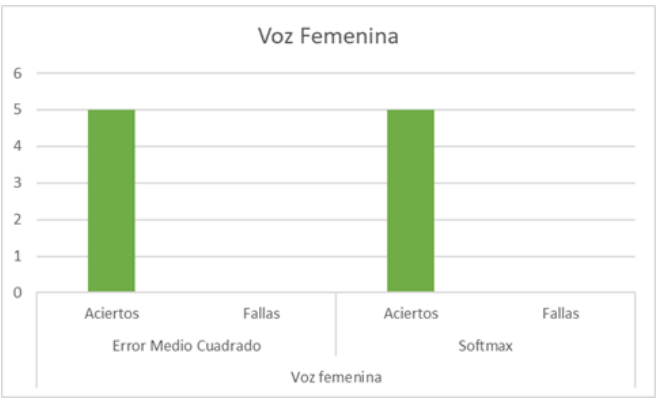
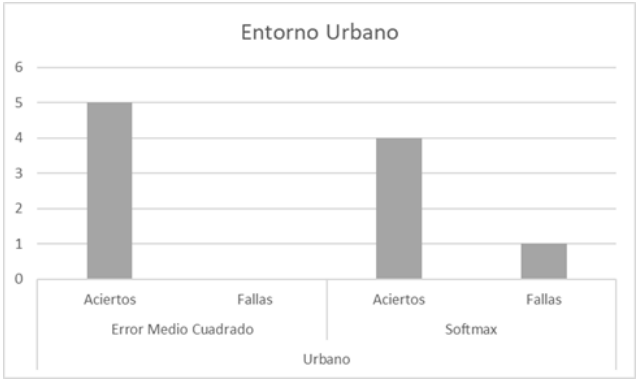


Figura 8 Desempeño con voz femenina



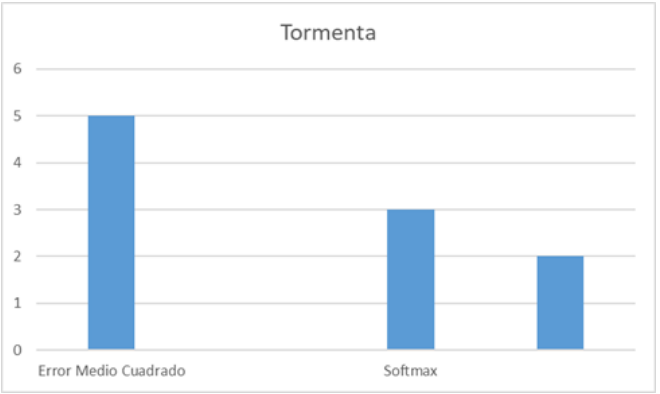


Figura 9 Desempeño en entorno urbano

Figura 10 Desempeño en lluvia

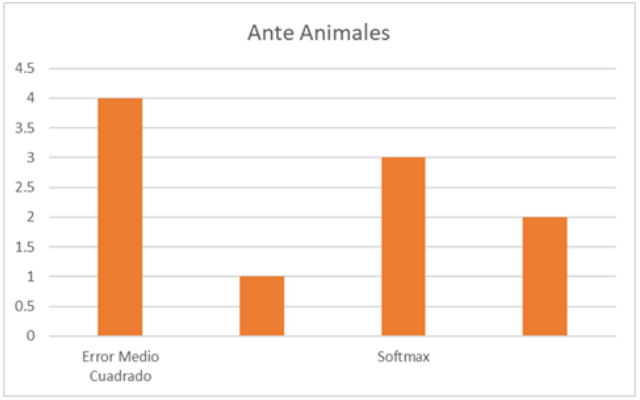


Figura 11 Desempeño ante animales

Como se pudo observar, la racha de errores del error medio cuadrado es mucho menor al de la normalización softmax. En un total de 31 pruebas, el error medio cuadrado solo se equivocó una vez, mientras que la función softmax se equivocó 6 veces.
En términos de probabilidad experimental, nuestro estimado de error es del 3.2% y del 19.35% respectivamente. Basado en este estimado, en el peor caso de desempeño el sistema tendría una precisión de prácticamente el 80%. Tomando en cuenta que el sistema se desarrolló con conceptos fundamentales y un programa relativamente básico, la precisión que se obtuvo es bastante aceptable.

A pesar de que la predicción por normalización tiene un 80% de precisión, la predicción por error medio cuadrado tiene un 96.8% de precisión en el peor de los casos. Es decir, podemos confiar en la predicción por error medio cuadrado siempre que el sistema determine predicciones diferentes para una misma muestra.

Una vez dicho esto, es importante recalcar que solo se hicieron 5 repeticiones para cada escenario, y en probabilidad experimental la precisión del estimado con el número de repeticiones. Para cuestiones de escalabilidad, valdría la pena hacer muchos más experimentos. Se sugeriría buscar un dataset de muestras de audio en línea para probar su efectividad con diferentes voces. Estos datasets son muy comunes, pues se utilizan para el desarrollo de redes neuronales recursivas, y se encuentran en el dominio del tiempo. Si se obtienen porcentajes similares con el dataset, se podría asumir con más seguridad que los porcentajes de error son adecuados. Por cuestiones de tiempo, este experimento expandido no se pudo llevar a cabo.

Condiciones de operación:

Con base en los resultados del experimento podemos concluir que el sistema funciona con mayor precisión en condiciones controladas; es decir, en espacios cerrados, sin ruido de fondo, y con un buen micrófono. Además, parece ser que no tiene problemas en distinguir voces de hombres y mujeres. Nuestra hipótesis era que el sistema podría llegar a tener un menor desempeño con mujeres, porque fue debutado con voces masculinas. Sin embargo, parece que el reconocimiento de voces femeninas no representa ningún problema. En principio, el sistema es ideal para su utilización en espacios cerrados.

En cuanto a otras condiciones, el sistema se desempeña correctamente en un entorno urbano. Además de una pequeña disminución en la precisión de la predicción de la función softmax con respecto al entorno controlado, no hay fallas considerables en el desempeño. De este modo, el sistema podría ser utilizado en espacios abiertos, siempre y cuando la predicción principal sea la del error medio cuadrado.

Ocurre algo similar en caso de lluvia. Hay que basarse en el error medio cuadrado para la predicción. La función softmax tiene menos precisión en este caso que con el entorno urbano.

Finalmente, con el ruido de perros de fondo, tanto la predicción por error medio cuadrado como la predicción por softmax disminuyeron su desempeño. Visto que en esta situación incluso el error medio cuadrado perdió confiabilidad, no creemos que sea recomendable que el sistema opere en entornos cercanos a perros o animales en general.

Con el fin de maximizar las probabilidades de un buen desempeño, creemos que el sistema (en su estado actual) podría servir en:

1. Espacios cerrados no concurridos (ej. casas, garajes y todo tipo de bodegas).
2. Lugares cerrados y abiertos concurridos (ej. oficinas, escuelas, y accesos principales).

Nota: Los entornos de implementación son preliminares, ya que **el proyecto se encuentra en estado de prototipo, y aún no podría ser desplegado en entornos reales** (ver sección: Propuestas de mejora a largo plazo).

Desventajas del sistema:

Ahora, se comentarán algunas desventajas y limitaciones del sistema con base en nuestros experimentos.
Primero, es evidente que el sistema podría llegar a resultar confuso si se tienen dos predicciones diferentes. Es decir, en principio es ambiguo que mientras que el error medio cuadrático prediga a una persona, la función sigmoid prediga a otra persona.
Esto se debe principalmente, a que nuestro sistema está pensado para cubrir dos tareas. De manera análoga al reconocimiento facial y la validación facial, existe el reconocimiento por voz y la validación por voz.
A grandes rasgos, el validación responde a la pregunta: ¿esta persona es tal? (si o no).
Por otro lado, el reconocimiento responde a la pregunta: de todas las personas dadas de alta en el sistema: ¿quién es esta persona?
De acuerdo a nuestra investigación, uno de los métodos más efectivos para resolver problemas de clasificación (reconocimiento) es mediante la función softmax. Sin embargo, estos métodos funcionan mejor mientras más datos se tengan. En nuestro caso, solo dimos de alta dos voces. De momento, la falta de datos representa una desventaja que repercute en el desempeño del sistema, pero si este se escala a una base de datos completa el desempeño mejoraría considerablemente. Resulta normal que el sistema tenga errores cuando tiene que normalizar solo dos datos. Es importante recalcar que solo se tiene una muestra de 3 segundos como criterio de comparación y depende en gran medida de la distancia entre los coeficientes de correlación.

Otra desventaja potencial puede apreciarse de los minutos 7:00 a 7:15 de la compilación de demostraciones (ver sección: Demos del sistema (Recopilación)).
En aquel video, se puede observar que el sistema toma una voz como válida aunque el usuario se equivocó y no quería revalidar su voz. En principio, esto es un error del usuario, pero el sistema debería de ser capaz de responder apropiadamente. Sobre todo, esto tendría implicaciones importantes en cuestión de ciberseguridad. En principio, un sistema con tal deficiencia sería blanco fácil de ataques de ingeniería social (ver sección Propuestas de mejora a largo plazo para una propuesta de solución).

Valoración del proyecto:

El proyecto es un prototipo preliminar para un sistema más completo, que, evidentemente, no cuenta con los medios para competir con sistemas de seguridad industrial actuales (muchos de ellos se complementan con otras tecnologías como escaneos biométricos y reconocimiento facial).
Sin embargo, dado que nuestro time to market fue de apenas 3 semanas, que no contábamos con el tiempo para la implementación de un sistema de machine learning más avanzado, y que el desarrollo se limitó a la aplicación de conceptos fundamentales en matlab, consideramos que el proyecto fue aceptable.
En general, el proyecto cumple los requerimientos básicos de funcionalidad de un sistema biométrico de voz. Además, se presentó la oportunidad de probar el modelo en diferentes escenarios. Esto no estaba contemplado originalmente. Sin embargo, el sistema tuvo un buen desempeño en situaciones adversas.
Descubrimos que los métodos en el dominio del tiempo pueden competir con los métodos de análisis de voz en frecuencia, e incluso podrían presentar ciertas ventajas si son empleados adecuadamente.
Por un lado, los métodos de análisis estadístico en el dominio del tiempo pueden lograr una precisión similar a los tiempos en el dominio de la frecuencia, y generalmente ocupan menos recursos computacionales.
La transformada de fourier es una sumatoria de armónicos, operación relativamente más compleja.
El cálculo del error medio cuadrado tiene una complejidad de O(n), y el algoritmo FFT (Fast Fourier Transform) tiene una complejidad O(Nlog(N)). En términos de tiempo de ejecución, la transformada de fourier puede llegar a superar al error medio cuadrado. Sin embargo, esto depende de la cantidad de datos que se utilicen. Para arreglos de datos pequeños, no existe una diferencia significativa entre cual algoritmo la diferencia entre complejidad temporal y espacial es despreciable. Sin embargo, en términos de escalabilidad, los algoritmos con complejidad logarítmica, pueden presentar problemas con arreglos muy grandes, mientras que los algoritmos con complejidad lineal no presentan problemas de complejidad espacial. De ahí que se utilicen este tipo de algoritmos en campos como deep learning donde se suelen tener datasets con millones de muestras.
Naturalmente, como los algoritmos con complejidad lineal operan mejor con largas bases de datos entonces, nuestra solución ofrecería una mejor adaptación si se quiere escalar el proyecto.

Propuestas de mejora a largo plazo:

La propuesta de solución tiene propuestas a mejorar en cuestión a la complejidad que puede presentar si se implementan nuevos métodos de análisis para tener un resultado con mejor precisión, donde en ello se pueden implementar patrones de palabras que permitan aún mejorar el sistema sobre el lenguaje natural que incluso puede desglosar el uso de otros idiomas ya que en algunos casos la entonación de voz cambia y puede ser un factor que puede alterar en la búsqueda de reconocimiento de la voz de una persona, pero tomando en cuenta que el error puede ser lo mínimo ya que los rasgos físicos, tanto fonéticos como morfológicos, dan a cada voz características peculiares que son específicas de cada persona.

Dado los scripts de Matlab estos pueden ser aún más eficaces implementando un nuevo algoritmo que permita eficientar el cálculo computacional donde el tiempo puede significar un recurso valioso en relación al tiempo de respuesta deseado, hoy en la industria se busca un proceso eficiente, pero dado que nuestra propuesta soluciona la problemática cubriendo lo necesario se considera exitoso, pero claro que puede mejorar, donde incluso el diagrama de bloques puede ser de otra manera, ya que hay muchas manera para implementar la recolección de datos, de las referencias y las muestras con las que se quieren trabajar.

La integración de más métodos también puede ser una mejora en relación a la precisión del diferenciador de las muestras, ya que entre más comparaciones se haga, más exacto puede llegar a ser el modelo, pero considerando lo necesario ya que también esto lleva un gasto computacional que puede considerarse como consumo de tiempo, por tanto solamente se recomendaría agregar unos cuantos métodos, cubriendo las necesidades que se requieren o incluso si se requiere de un consumo tan grande en la medida lograr el menos costo posible dependiendo lo que un usuario busca.

Otro punto destacable puede ser el concepto del dominio del tiempo a términos de frecuencia donde a partir de comparaciones todos tienen sus ventajas, pero en este proyecto con los métodos de estadística descriptivas, herramientas como el de cálculo del error medio presenta una precisión similar en los diferentes dominios ya mencionados, viendo más allá de la escalabilidad que puede presentar el proyecto, esto puede implementarse para el guardado de diferentes voces, donde la base de datos puede ser enorme y hacer un gran número de diferencias respecto a una sola muestra, y a partir de ello implementar el dominio que más convenga y ver la complejidad de los algoritmos planteados.

Este sistema puede ir de la mano con otros sistemas de identificación, tal puede ser el reconocimiento de huellas dactilares o incluso fáciles, esto para prevenir imitaciones, tal puede ser una grabación de voz, donde ello se verifica si es una persona la que está tratando de validar su identidad, donde en ello a partir de otros sistemas, estos modelos de seguridad pueden brindar una mayor precisión y protección de datos personales y prevenir violaciones ante la privacidad de alguien, donde en colaboración con otro sistema de seguridad ambos pueden validar si es correcto la identificación, ya dependiendo de la necesidad del usuario.

Conclusiones:

Algunas observaciones finales fue la manera de organizar el diagrama de bloques y llegar a una propuesta final ya que se tenía pensado de una manera más sofisticada a la que se está planteando, además de la búsqueda exhaustiva de comandos de software matlab para hacer el análisis y también de la información sobre las técnicas de estadística descriptiva que se iban a ocupar para la implementación de la solución. Cabe destacar que para implementar comandos se tuvo que investigar sobre cómo operaban, sobre el significado del valor que se arrojaba y otros conceptos relacionados para dar un mejor entendimiento del análisis.

Según los datos presentados podemos concluir que el proyecto si entregó los resultados deseados, donde a partir de distintas pruebas exitosamente cumple con los requerimientos a solucionar de la situación problema, además de involucrar el uso de distintas herramientas tanto analiticas como de software para implementación de la propuesta, en ello se involucran conceptos en la unidad de formación que fueron de utilidad para comprender mejor los conceptos y presentar una solución que puede ser escalable para el campo laboral.

En relación a la vida cotidiana nos damos cuenta que el tema del reconocimiento de voz hoy en día es un tema que un futuro próximo pueda estar incorporado como herramienta del día a día, donde pueda ser utilizado en actividades tan simples para reconocer a una persona, dado que hoy en día está pensado para temas de seguridad e identificación personal, el contexto histórico se basa en conceptos tan sencillos que para lograr algo tan complejo se necesita investigar y entender conceptos tan fundamentales como los que se vieron en la unidad de formación, además de involucrar conocimientos de pasadas unidades de formación que se van acomulando para implementar cada día un proyecto más sofisticado. Este proyecto permitió explorar conocimientos para implementar una metodología de investigación y así dar una propuesta clara y precisa, además de tomar las consideraciones con las que se quería implementar el proyecto tomando en cuenta los requerimientos solicitados

Reflexiones Finales:

- Luis Felipe:

Dada la unidad de formación se pudieron apreciar diferentes conocimientos, donde fueron de gran ayuda para resolver la situación problema, en ello fue el tema del sistema de reconocimiento de voz, donde para poner una propuesta en la mesa se tuvo que ver algunos conceptos de sistemas de señales para tener un panorama general de lo que estábamos enfrentando, en ello se emplearon herramientas de estadística descriptiva y la analítica de datos, donde nosotros teníamos que identificar cuál podría ser la herramientas más conveniente para presentar un diferenciador de voz entre dos personas, también se tuvo que hacer diversas pruebas, además del feedback de profesor que fue de gran ayuda para clarificar la idea, en ello se forma nuestra propuesta de solución donde en combinación con nuestro conocimiento empirico y cientifico se formuló la idea para atacar esta situación problema, además de aprender a darle un plus a nuestra idea para que este pueda ser un proyecto escalable en el futuro, agregando cosas que pueden mejor para tener una mejor

precisión del modelo e identificando errores que puedan ser duros para la implementación del sistema, donde por mi parte la unidad de formación dió mucho de qué hablar, además de que son conceptos que se usan el día con día y que estos recursos pueden servir para proyectos a futuro.

- Luis Dario

En esta unidad de formación se presentó la oportunidad de poner a prueba tanto conceptos nuevos como conceptos que ya habían sido adquiridos con anterioridad. En particular, use técnicas de analítica de datos para procesar una señal de audio, y use técnicas fundamentales de machine learning para la clasificación de datos. La libertad operativa del proyecto fue abierta, por lo que pudimos experimentar el desarrollo de un proyecto a un nivel más cercano a la industria con requerimientos específicos, ciñéndonos a un time to market definido, y adaptándonos a dar la mejor solución posible con los recursos que tenemos a nuestra disposición. Gracias a la retroalimentación del maestro, me di cuenta que para que no solo basta entregar un proyecto en tiempo y forma, sino que también hay que probarlo en diferentes escenarios para poder definir las condiciones de operación que garanticen el mejor desempeño posible. Esto no solo es importante desde el punto de vista ingenieril. Desde la perspectiva de inteligencia de negocios, los aprendizajes de la experimentación pueden proveer insights, de la mano con los detalles técnicos de la implementación, los materiales, y el costo, sobre cuál podría ser el posible segmento de mercado al cual va dirigido el producto, y donde se posiciona con respecto a la competencia. Me di cuenta de que la mejor forma de defender un proyecto en el mundo profesional es conociendo tanto las ventajas como las desventajas del proyecto para poder argumentar por qué nuestra solución es la mejor dada una determinada situación. También, me di cuenta de que con la experimentación se descubren errores y áreas de oportunidad. En particular, en sistemas biométricos es importante entender los errores que podrían relevar vulnerabilidades a los atacantes informáticos.

Bibliografía:

- Electronics Projects Focus. (2021). Understanding Voice Recognition. Recovered on August, 20, 2021. Retrieved from: https://www.elprocus.com/understanding-voice-recognition/
- BBVA.(2021). Biometría de voz: la huella vocal será el gran aliado de la banca ‘online’. Recovered on August 20, 2021. Retrieved from: https://www.bbva.com/es/biometria-de-voz-la-huella-vocal-sera-el-gran-aliado-de-la-banca-online/

Anexos:

1. Anexo 1: Código fuente de muestra (proporcionado en canvas).

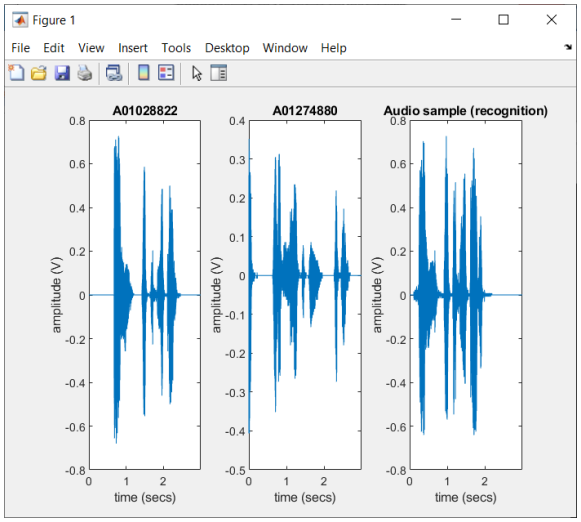
```
clear
clc

q=5;
f1=2500;
%% Record your voice for q seconds.
recObj = audiorecorder;
disp('Start speaking.')
recordblocking(recObj, q);
disp('End of Recording.');

% Play back the recording.
play(recObj);

% Store data in double-precision array.
myRecording = getaudiodata(recObj);
% Time axis
qa=recObj.TotalSamples;
t=(0:q/qa:q-q/qa)';

%% for the frrequency axis
Ts=q/qa; %sampling time
fs=1/Ts; %sampling frequency
[na,nb]=size(t(:)); % na=number of points of signal
ff=fs*[0:na-1]/na-fs/2;

% Plot the waveform.
plot(t,myRecording);
xlabel('time (secs)')
ylabel('amplitude (V)')
figure

cs=cos(2*pi*f1*t);
mods=cs.*myRecording;
plot(ff,fftshift(abs(fft(myRecording))),'r')
% hold on
% plot(ff,fftshift(abs(fft(mods))),'k')
% hold off
xlabel('frequency (Hz)')
ylabel('Magnitude (V)')
```

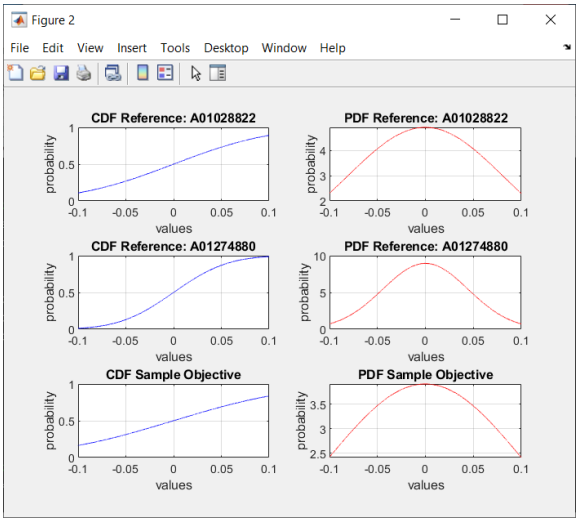
2. Anexo 2: Ejemplos de resultados (gráficas comparativas).

A continuación, se muestran algunos ejemplos de los resultados arrojados por los experimentos en diferentes escenarios:

a. Condiciones controladas:

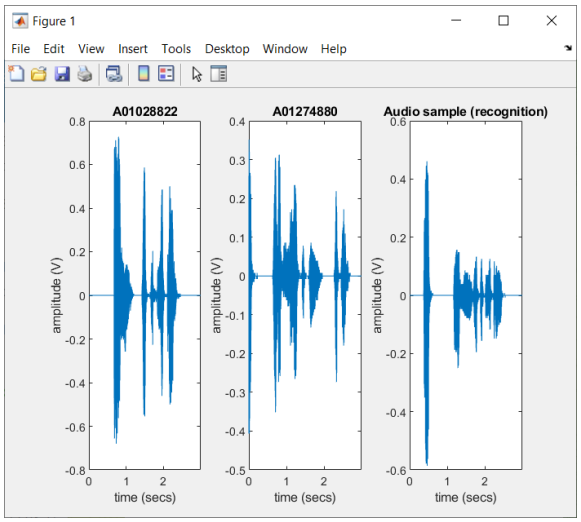
Primer reconocimiento de voz: **Luis Dario**

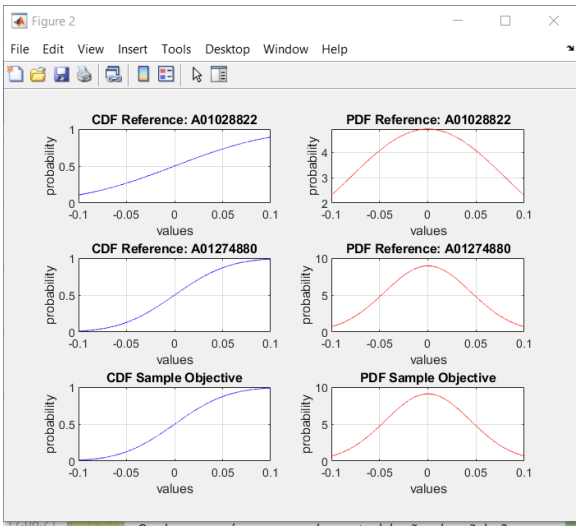





```
The audio sample is closer to reference: A01028822
It is more probable that the voice belongs to A01028822
>> ReconocimientoVoz
```

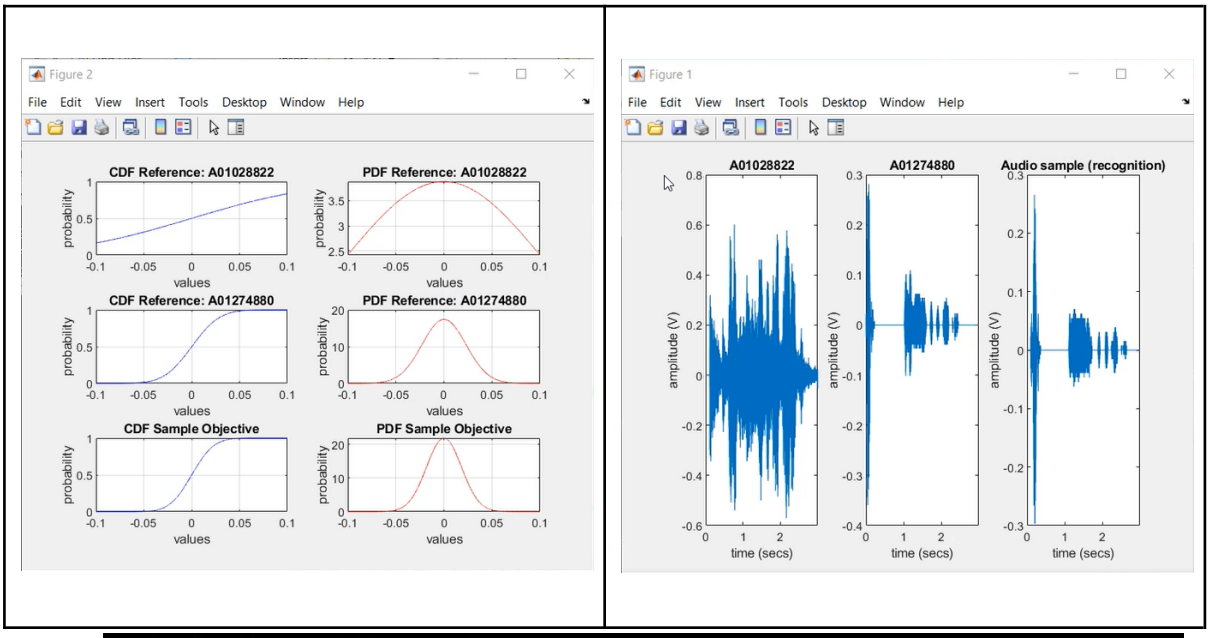
Segundo reconocimiento de voz: **Juan Pablo**





```
End of Recording.
The audio sample is closer to reference: A01274880
It is more probable that the voice belongs to A01274880
```

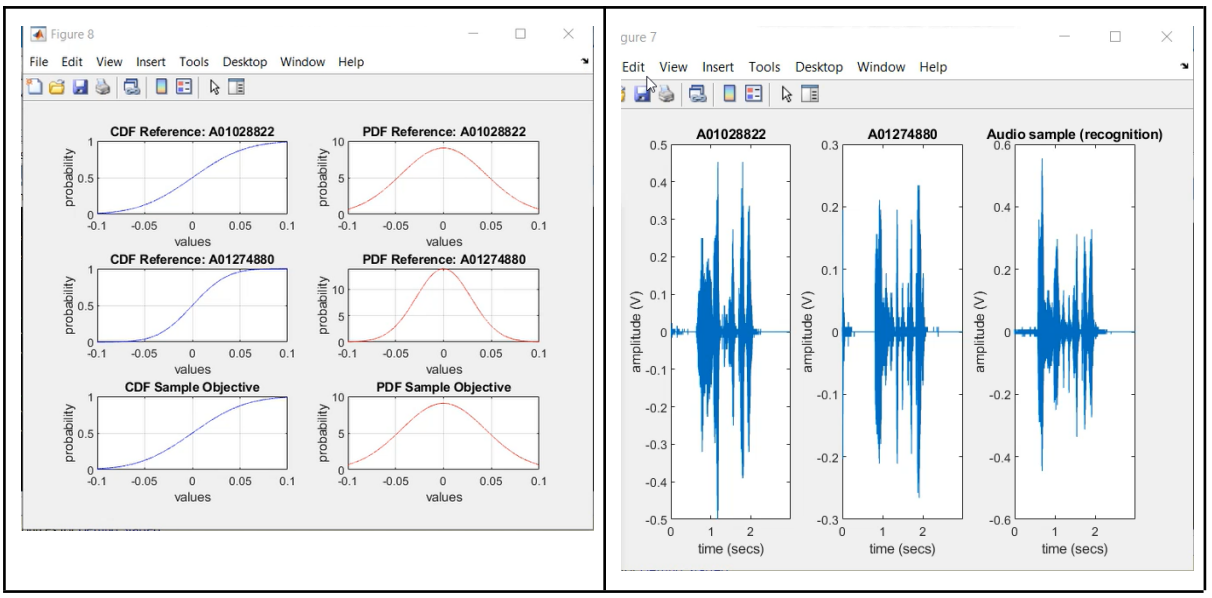
b. Voz Femenina:




```
Retrieving Data...
Data Retreived
Start speaking.
End of Recording.
The audio sample is closer to reference: A01274880
It is more probable that the voice belongs to A01274880
```

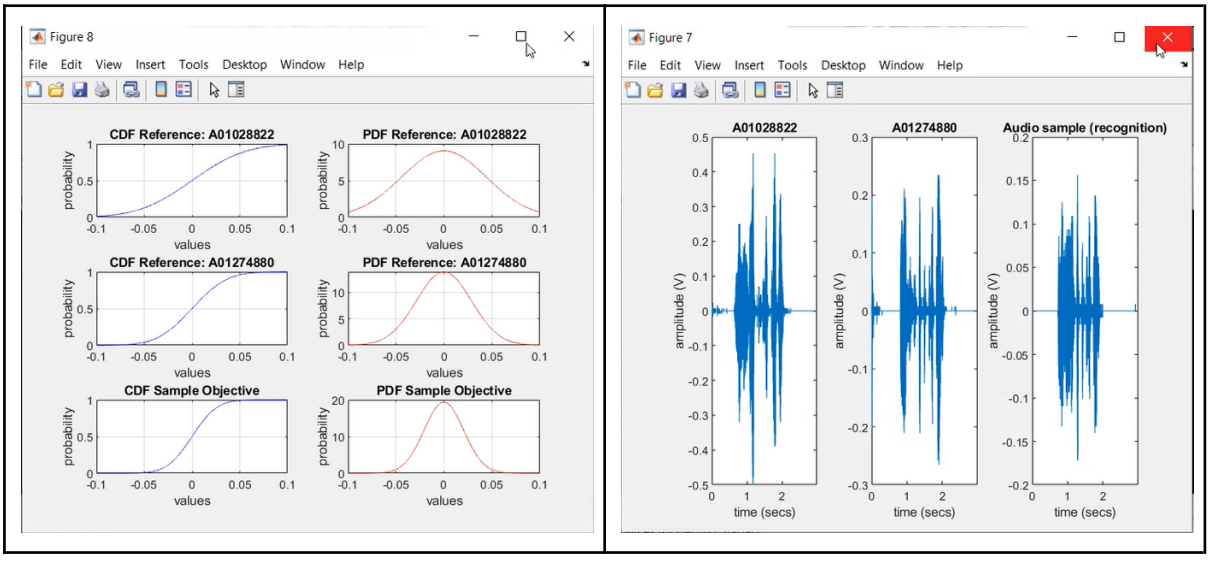
c. Entorno urbano:

Primer reconocimiento de voz: **Luis Dario**




```
Retrieving Data...
Data Retreived
Start speaking.
End of Recording.
The audio sample is closer to reference: A01028822
It is more probable that the voice belongs to A01028822
>>
```

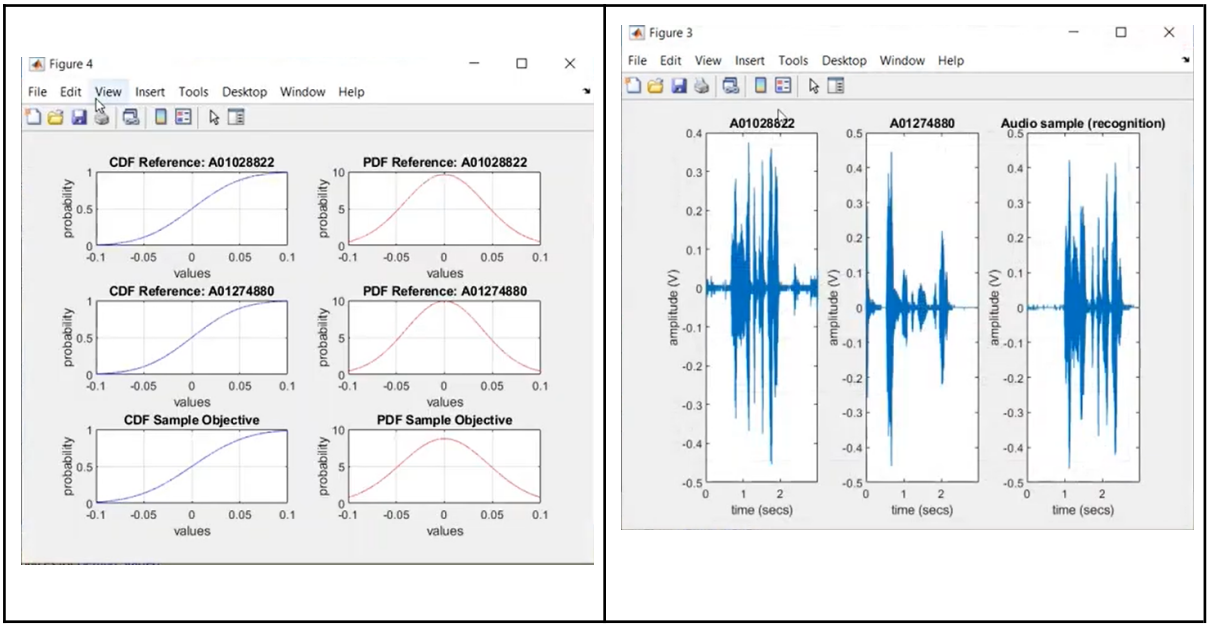
Segundo reconocimiento de voz: **Juan Pablo**



```
Retrieving Data...
Data Retreived
Start speaking.
End of Recording.
The audio sample is closer to reference: A01274880
It is more probable that the voice belongs to A01274880
```

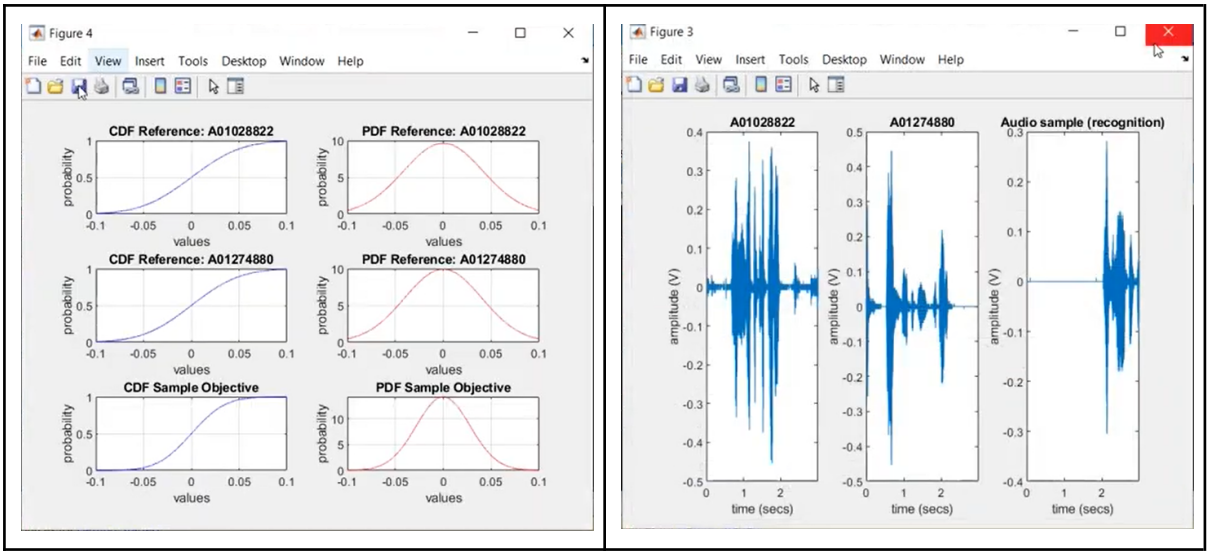
d. Tormenta:

Primer reconocimiento de voz: **Luis Dario**




```
Retrieving Data...
Data Retreived
Start speaking.
End of Recording.
The audio sample is closer to reference: A01274880
It is more probable that the voice belongs to A01274880
```

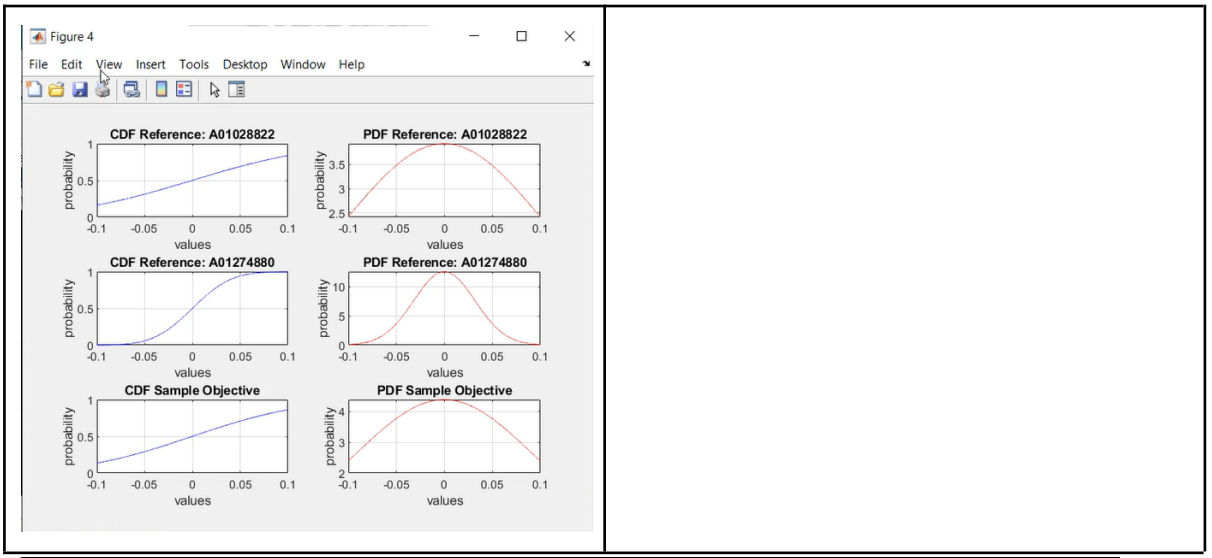
Segundo reconocimiento de voz: **Juan Pablo**




```
Data Retreived
Start speaking.
End of Recording.
The audio sample is closer to reference: A01274880
It is more probable that the voice belongs to A01274880
```

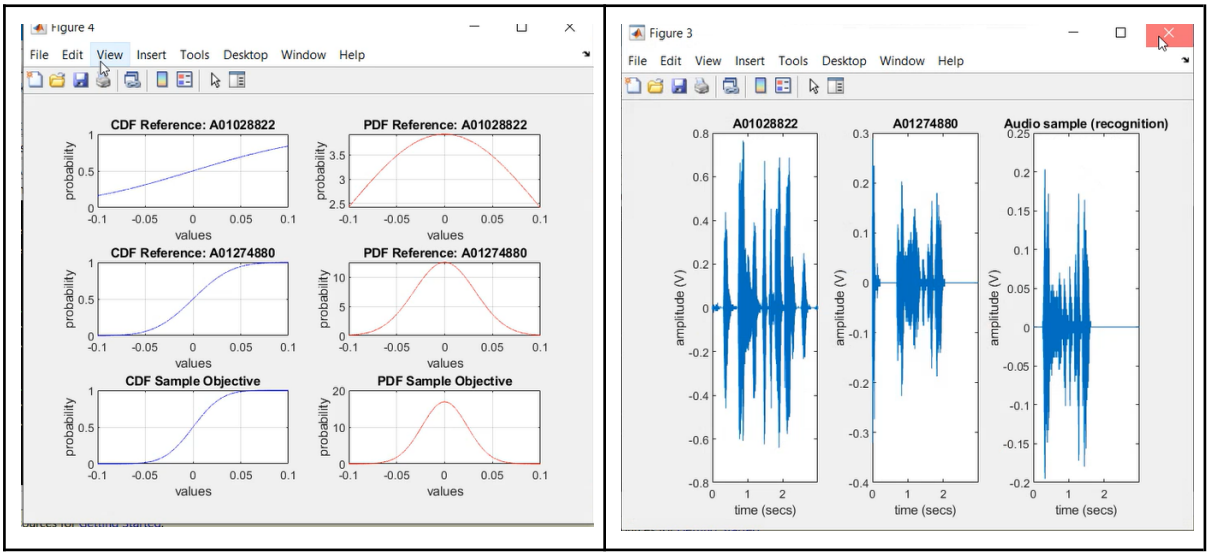
e. Ladridos de perro:

Primer reconocimiento de voz: **Luis Dario**




```
Retrieving Data...
Data Retreived
Start speaking.
End of Recording.
The audio sample is closer to reference: A01028822
It is more probable that the voice belongs to A01028822
```

Segundo reconocimiento de voz: **Juan Pablo**




```
Retrieving Data...
Data Retreived
Start speaking.
End of Recording.
The audio sample is closer to reference: A01274880
It is more probable that the voice belongs to A01028822
>>
```