

Match de voz usando un algoritmo genético

Rubén A. Morales Hdez.
Inteligencia Artificial 401
Universidad de Xalapa
Veracruz, México
16 de enero del 2026

Abstract—El reconocimiento de voz es un problema relevante dentro del procesamiento digital de señales, debido a la variabilidad introducida por el ruido ambiental, las condiciones de grabación y las diferencias naturales en la pronunciación humana. En este trabajo se presenta un sistema de comparación y reconocimiento de voz basado en Algoritmos Genéticos, cuyo objetivo es generar una señal genética resultante que se aproxime de forma óptima a una señal de voz de entrada.

El método propuesto utiliza operadores genéticos clásicos como selección mediante ruleta, cruce de un solo punto y mutación por inversión, junto con una función de aptitud definida como el cuadrado del grado de similitud entre señales. La comparación final entre la señal genética resultante y la señal de entrada se realiza utilizando dos enfoques: un método basado en umbral y la distancia euclidiana.

Los resultados experimentales muestran que el uso de Algoritmos Genéticos mejora la capacidad de coincidencia en escenarios donde la comparación directa entre señales no es suficiente. Asimismo, se observa que la distancia euclidiana proporciona una medida más robusta y consistente en comparación con el método basado en umbral, especialmente en presencia de ruido.

Index Terms—Algoritmo genético, Reconocimiento de voz, Distancia euclidiana

I. INTRODUCCIÓN

El reconocimiento de voz es un área fundamental dentro del procesamiento digital de señales y la inteligencia artificial, con aplicaciones en sistemas de seguridad, autenticación biométrica y control por voz. Sin embargo, la variabilidad inherente de las señales de voz, causada por ruido ambiental, diferencias en pronunciación, velocidad y condiciones de grabación, representa un reto significativo para los métodos tradicionales de comparación directa.

En este trabajo se presenta la implementación de un sistema de comparación de voz basado en Algoritmos Genéticos (GA, por sus siglas en inglés), los cuales son especialmente adecuados para manejar datos ruidosos y funciones de optimización complejas. El enfoque propuesto utiliza operadores genéticos clásicos como selección, cruce y mutación para generar una señal genética resultante que se aproxima de forma óptima a una señal de voz de entrada, permitiendo posteriormente su comparación mediante métodos basados en umbral y distancia euclidiana.

II. MARCO TEÓRICO

A. Señales de Voz y Representación Digital

La voz humana es una señal analógica de naturaleza compleja, generada por la vibración de las cuerdas vocales y modulada por el tracto vocal. Para su procesamiento mediante sistemas computacionales, esta señal debe ser digitalizada a través de un proceso de muestreo y cuantificación, dando lugar a una secuencia discreta de valores de amplitud.

Un archivo de audio en formato WAV almacena estas muestras de amplitud junto con información adicional como la frecuencia de muestreo y el número de canales. En grabaciones estéreo, la señal se compone de dos canales independientes (izquierdo y derecho), los cuales suelen contener información similar. Por esta razón, para muchos sistemas de análisis de voz es suficiente seleccionar uno de los canales sin afectar significativamente los resultados.

La comparación directa de señales de voz puede realizarse evaluando las diferencias entre los valores de amplitud correspondientes. No obstante, este enfoque es altamente sensible al ruido, a pequeñas variaciones temporales y a cambios en la intensidad de la señal, lo que dificulta una coincidencia exacta incluso entre grabaciones de la misma persona.

B. Problemas del Reconocimiento de Voz en Presencia de Ruido

En entornos reales, las señales de voz suelen estar contaminadas por ruido proveniente de diversas fuentes, como maquinaria, tráfico o interferencias electrónicas. Estas perturbaciones provocan que dos grabaciones de la misma palabra, pronunciadas por la misma persona, presenten diferencias apreciables en sus valores de amplitud.

Debido a esta variabilidad, los métodos deterministas de comparación directa pueden fallar, ya que no consideran la naturaleza aproximada del problema. Esto motiva el uso de técnicas de optimización y búsqueda heurística, como los Algoritmos Genéticos, que permiten encontrar soluciones cercanas al óptimo aun cuando los datos son ruidosos o incompletos.

C. Algoritmos Genéticos

Los Algoritmos Genéticos (GA) son métodos de optimización inspirados en los principios de la evolución

biológica y la selección natural. En un GA, cada posible solución al problema se representa como un individuo dentro de una población, y cada individuo posee un valor de aptitud que indica qué tan buena es la solución que representa.

A lo largo de múltiples generaciones, los individuos con mayor aptitud tienen mayor probabilidad de ser seleccionados para reproducirse, generando nuevos individuos mediante operadores genéticos. Este proceso iterativo permite que la población evolucione progresivamente hacia soluciones de mayor calidad.

En el contexto del reconocimiento de voz, cada individuo puede interpretarse como una señal de audio candidata, y el objetivo del algoritmo es generar una señal que se aproxime lo más posible a la señal de entrada.

D. Operadores Genéticos

Los principales operadores genéticos utilizados en este trabajo son los siguientes:

1) *Selección*: La selección es el proceso mediante el cual se eligen individuos de la población para participar en la generación de nuevos descendientes. En este trabajo se utiliza el mecanismo de selección por ruleta (*Roulette Wheel Selection*), en el cual la probabilidad de selección de cada individuo es proporcional a su valor de aptitud. Este método permite que individuos con mejor desempeño tengan mayor probabilidad de ser seleccionados, sin excluir completamente a los de menor aptitud.

2) *Crossover*: El operador de cruce o *crossover* combina la información genética de dos individuos padres para generar nuevos individuos descendientes. En este trabajo se emplea el cruce de un solo punto, donde se selecciona un punto específico de la señal y se intercambian los segmentos posteriores a dicho punto entre ambos padres. Este operador permite explorar nuevas combinaciones de características presentes en las señales originales.

3) *Mutación*: La mutación introduce cambios aleatorios en un individuo con el fin de preservar la diversidad genética de la población y evitar la convergencia prematura a soluciones subóptimas. En este trabajo se utiliza la mutación por inversión (*flip*), donde se altera el valor de una muestra específica de la señal seleccionada como punto de mutación.

E. Función de Aptitud

La función de aptitud es el criterio mediante el cual se evalúa la calidad de cada individuo. En el enfoque propuesto, la aptitud se define como el cuadrado del grado de similitud entre la señal candidata y la señal de entrada, es decir:

$$f(x) = x^2$$

donde x representa la proporción de coincidencias entre las señales comparadas. Esta función enfatiza las soluciones con

mayor similitud, favoreciendo la selección de individuos que se aproximan más a la señal de voz de entrada.

F. Medidas de Similitud

Para evaluar la similitud entre señales de voz se utilizan dos métricas principales:

1) *Comparación Basada en Umbral*: Este método consiste en calcular la diferencia absoluta entre los valores de amplitud correspondientes de dos señales y compararla con un umbral predefinido. Si la diferencia es menor que el umbral, se considera una coincidencia. Cuando más de la mitad de las muestras cumplen esta condición, las señales se clasifican como correspondientes a la misma voz.

2) *Distancia Euclidiana*: La distancia euclidiana es una métrica ampliamente utilizada para medir la similitud entre vectores en espacios multidimensionales. En el caso de señales de voz, cada señal se interpreta como un vector de amplitudes, y la distancia euclidiana proporciona una medida global de diferencia entre ambas señales. Una distancia menor indica una mayor similitud entre las voces comparadas.

III. METODOLOGÍA

La metodología seguida en este trabajo se divide en varias etapas claramente definidas:

A. Adquisición y Preprocesamiento

Las señales de voz utilizadas como referencias fueron grabadas en formato WAV utilizando un teléfono móvil, garantizando una frecuencia de muestreo constante y una calidad de audio suficiente para el análisis posterior. Las grabaciones se realizaron en un entorno controlado para minimizar, en la medida de lo posible, la presencia de ruido ambiental.

Posteriormente, las señales de audio fueron transferidas a una computadora portátil, donde se llevó a cabo el proceso de normalización y ajuste de duración utilizando el software *FL Studio 2025*. Durante esta etapa se eliminaron silencios innecesarios, se normalizó la amplitud de las señales y se ajustó la longitud de todos los audios para garantizar una alineación temporal uniforme entre las señales de entrada y las señales de referencia.

Gracias a este preprocesamiento previo, todas las señales utilizadas en el sistema presentan la misma duración y están correctamente alineadas, lo que permite realizar comparaciones directas de amplitud sin requerir ajustes adicionales durante la ejecución del algoritmo genético.

B. Selección de Individuos

Inicialmente, la señal de voz de entrada se compara directamente con las señales de referencia almacenadas en la base de datos mediante una comparación basada en umbral.

Para este propósito, se calcula la diferencia absoluta entre los valores de amplitud correspondientes de ambas señales y se compara cada diferencia con un umbral predefinido.

El valor del umbral se establece en 0.09, el cual fue seleccionado empíricamente por los autores del trabajo original tras la realización de múltiples pruebas experimentales. Si la diferencia absoluta es menor que dicho umbral, se considera una coincidencia. En caso de que más del 50% de las muestras cumplan esta condición, se concluye que ambas señales de voz corresponden a la misma persona.

Si la señal de entrada no coincide con ninguna de las referencias bajo este criterio, se procede a la selección de dos señales de referencia utilizando el mecanismo de selección por Ruleta (Roulette Wheel Selection), donde la probabilidad de selección de cada individuo es proporcional a su valor de aptitud.

C. Crossover

Se aplica un cruce de un solo punto entre las dos señales seleccionadas. El punto de cruce se calcula como:

$$\text{punto de cruce} = \log_2(n + 1)$$

donde n es el número de muestras de la señal. A partir de este punto, se intercambian los segmentos de las señales para generar nuevos descendientes.

D. Mutación

La mutación se realiza mediante un operador de inversión (flip), donde se selecciona un punto de mutación calculado de forma similar al punto de cruce. Este proceso permite introducir variaciones adicionales y evitar convergencia prematura.

E. Criterio de Selección del Mejor Individuo

Tras el cruce y la mutación, se compara la aptitud de los nuevos individuos con la de sus padres. El individuo con mayor valor de aptitud se selecciona como resultado genético.

IV. RESULTADOS EXPERIMENTALES

Se realizaron diversos experimentos para evaluar el desempeño del sistema propuesto.

A. Experimento 1: Voces Iguales

En este experimento, se compararon dos señales correspondientes a la misma palabra pronunciada por la misma persona. Los resultados mostraron que la mayoría de las diferencias de amplitud se mantuvieron por debajo del umbral establecido, lo que permitió clasificar correctamente ambas voces como iguales.

B. Experimento 2: Voces Diferentes

Cuando la señal de entrada no coincidía con ninguna referencia, el sistema aplicó el algoritmo genético para generar una señal genética resultante. Posteriormente, esta señal se comparó con la señal de entrada utilizando ambos métodos de comparación.

1) *Comparación por Umbral*: Los resultados indicaron que, tras aplicar el algoritmo genético, la mayoría de las diferencias de amplitud no superaron el umbral, permitiendo una correcta identificación de similitud.

2) *Comparación por Distancia Euclidiana*: La distancia euclidiana mostró una mayor capacidad para discriminar entre señales, proporcionando resultados más estables y consistentes, especialmente en condiciones de ruido moderado.

V. CONCLUSIONES

En este trabajo se implementó un sistema de reconocimiento de voz basado en algoritmos genéticos, demostrando su efectividad para manejar señales ruidosas y variaciones naturales en la voz humana. Los resultados experimentales indican que la comparación mediante distancia euclidiana ofrece mayor robustez y precisión en comparación con el método basado en umbral.

El uso de algoritmos genéticos permitió generar señales genéticas resultantes que se aproximan de forma óptima a la señal de entrada, mejorando la tasa de coincidencia en casos donde la comparación directa no era suficiente. Como trabajo futuro, se propone la incorporación de técnicas de extracción de características y filtrado avanzado de ruido para mejorar aún más el desempeño del sistema.

REFERENCES

- [1] Randy L. Haupt (2004) *Practical genetic Algorithm*. John Wiley and Sons Inc.
- [2] David J. White, Andrew P. King, And Shan D. Duncan (2002) *Voice Recognition Technology, As A Tool For Behavioral Research*. Indiana University
- [3] Abhishek Bal, Nilima Paul, Suvasree Chakraborty, Sonali Sen (2014) *Voice Matching Using Genetic Algorithm*. International Journal of Advanced Computer Research