Evaluación comparativa de sistemas de verificación de locutor basados en LPC y MFCC

Oscar Alberto Rocha Arcos y Héctor Daniel Juárez Leonel
Co. Dra. Yesenia Eleonor González Navarro

RIIAA 2.0

Introducción.

Este estudio es una evaluación comparativa de los sistemas de verificación de locutor basados en los algoritmos de extracción de características de voz LPC (Linear Predictive Coding) y MFCC (Mel-Frequency Cepstral Coeficientes).

La evaluación se llevó a cabo con una metodología cuantitativa experimental: usando muestras de voz (señales de entrada del sistema) y variando su relación señal a ruido (SNR) con fuentes de ruido gaussiano y de bullicio.

El resultado de la evaluación comparativa servirá para la elección del algoritmo de extracción de características de voz para la construcción de sistemas de verificación, tomando en cuenta el entorno en donde se ocupará el sistema (diferentes tipos de ruido).

Algoritmos de extracción de características de la voz.

La idea básica del algoritmo LPC es dada una muestra de voz en un tiempo *n*, de puede aproximar mediante una combinación lineal de las últimas *p* muestras de voz, de manera que:

$$S(n) = -a_1 s(n-1) - a_2 s(n-2) - a_3 s(n-3) - \ldots - a_p s(n-p)$$

Donde los coeficientes a_1, a_2, a_3 ,...., a_p representan los coeficientes LPC [1].

El algoritmo MFCC son los coeficientes que representan al espectro de potencia a corto plazo de un sonido, estos coeficientes están en escala Mel, la cual es una escala perceptual propuesta para el sonido, que busca analogar la manera en que percibe el oído humano.

Ruido de Bullicio y Gaussiano.

El ruido gaussiano fue generado de manera artificial mientras, que el ruido de bullicio se obtuvo realizando grabaciones en la UPIITA-IPN en donde se conglomeraban multitudes (cafetería, explanada, salones, etc).

Datos y método de clasificación.

Los datos están compuestos por archivos de audio de los integrantes del proyecto, que son pronunciaciones de las palabras :

"gama, mano, llave, barra, niño, lobo, mayo, daga, lava, murciélago, aurelion"

Las cuales proporcionan más información para caracterizar al hablante [2].

A su vez, los datos se van a corromper con los diferentes tipos de ruido (gaussiano y bullicio) para compararlos con los datos previos y ver el rendimiento de cada algoritmo de extracción.

El método de clasificación es una Red Neuronal Multicapa Perceptron.

Resultados.

Se observan los porcentajes de reconocimiento de todas las palabras propuestas.

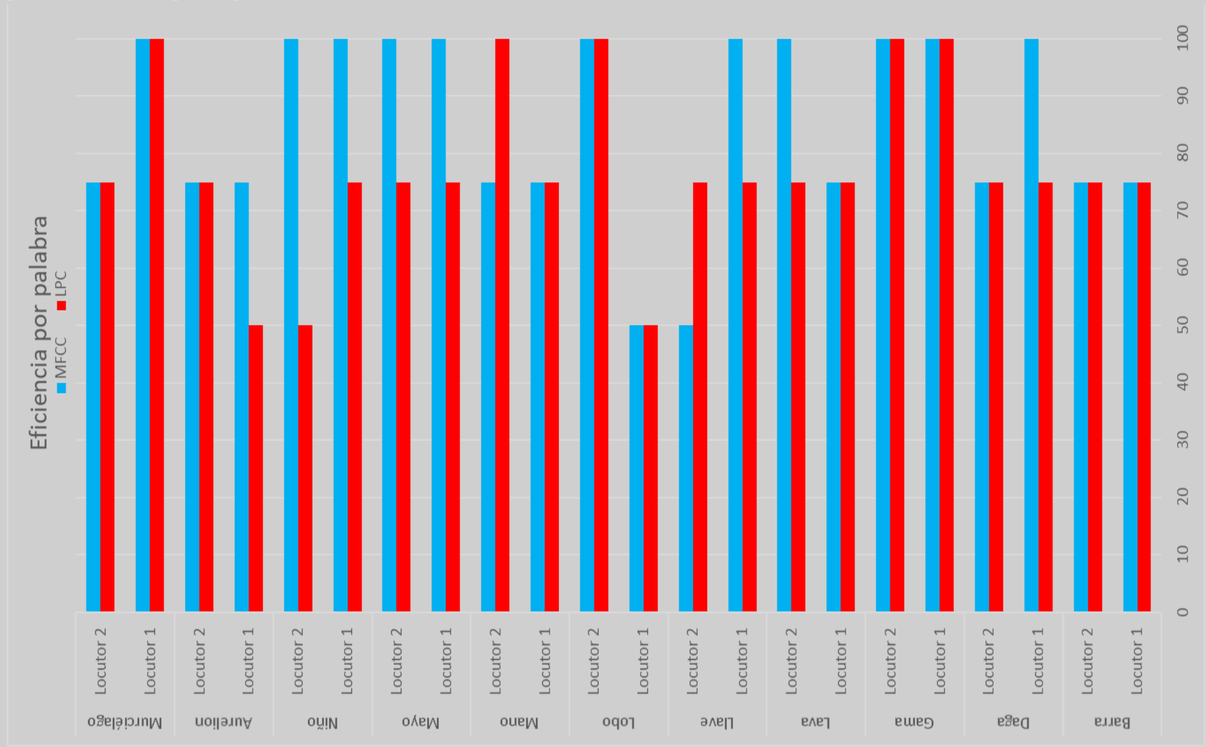

Figura 1. Porcentaje de reconocimiento por palabra y locutor.

Se realiza la comparación de los algoritmos LPC y MFCC para las muestras de voz de cada integrante del proyecto, en un rango de SNR de 50 a 0 dB para el ruido gaussiano y de bullicio.

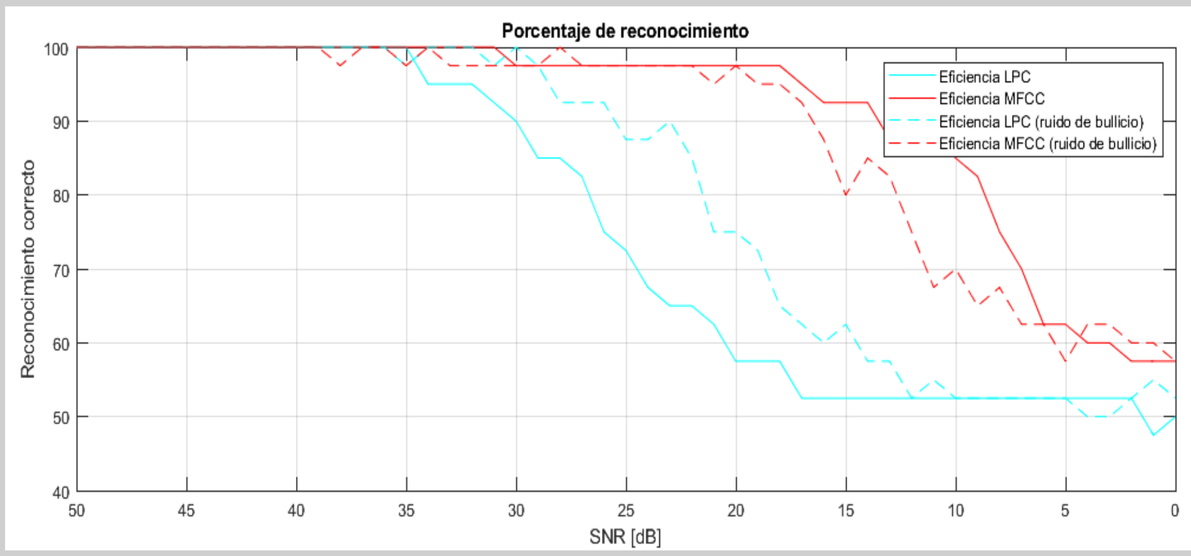

Figura 2. Porcentaje de reconocimiento en rango de SNR para ruido gaussiano y de bullicio.

Conclusiones.

- LPC requiere menos épocas y coeficientes que MFCC para obtener mejores resultados de reconocimiento.
- Emplear más coeficientes en los algoritmos, no implica más resistencia al ruido.
- El comportamiento de los algoritmos con ruido de bullicio es más caótico que con ruido gaussiano.
- Una palabra con varias repeticiones mejor que varias palabras con pocas repeticiones para la diferenciación de los hablantes.
- El número de capas ocultas de la red, es un factor importante para el incrementar el reconocimiento del hablante.
- No es rentable el incremento en el número de neuronas de la red, para obtener mayor porcentaje de reconocimiento.

Referencias.

[1] L. Rabier y B. H. Juang, Fundamental of Speech Recognition, Englewood Cliffs: Prentice Hall Internatonal, 1993.

[2] C. S. Aguilar Orozco y M. W. Marín Benítez, Sistema certificador de locutor por voz, Instituto Politécnico Nacional, 2003.