

Verificación de Hablantes a través de la Voz

Trabajo Final de Procesamiento Digital de Señales

Iván F. Schweikofski, Camila Saucedo y Darién J. Ramírez
Tutor: Matias F. Gerard

Introducción

Identificación automática del hablante

Identificación del hablante

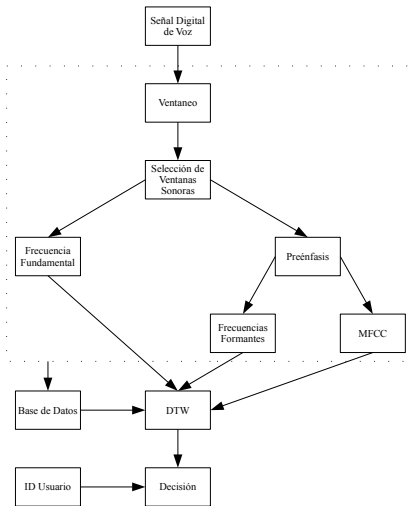
Decidir si la persona está o no dentro de un conjunto de personas.

Verificación del hablante

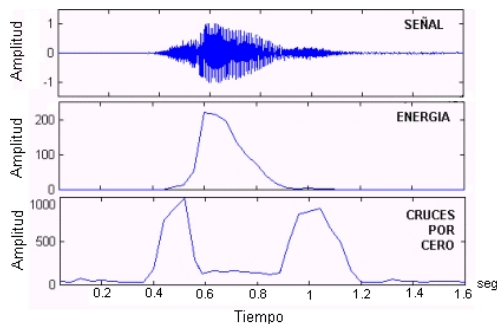
Decidir si el hablante es quien dice ser.

Implementación

- 1 Base de datos.
- 2 Señal de entrada.
- 3 Ventaneo de la señal de entrada.
- 4 Selección de ventanas sonoras.
- 5 Extracción de características (F_0 , Formantes y MFCC).
- 6 Comparación mediante DTW.
- 7 Decisión.



Sonidos sonoros y sonidos sordos



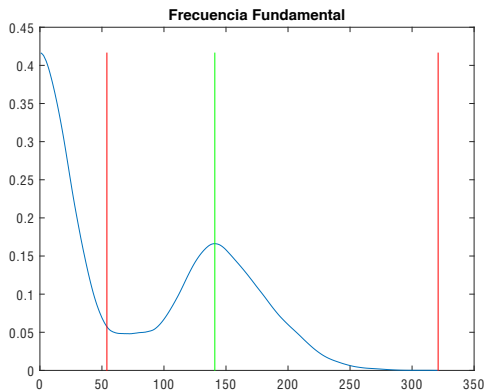
Ventaneo con ventanas sonoras.

- 1 *Sonidos sonoros*: baja cantidad de cruces por cero y alta energía. Periodicidad.
- 2 *Sonidos sordos*: mayor densidad de cruces por cero y menor energía.

Frecuencia fundamental F_0

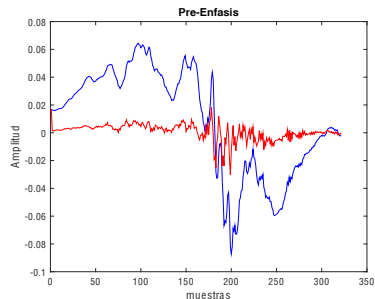
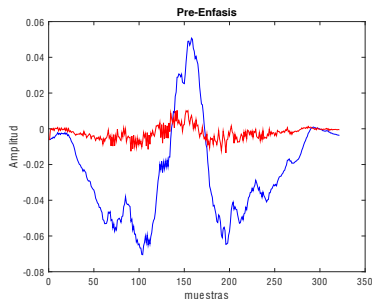
① Autocorrelación.

②
$$\frac{1}{f_{max}} = \frac{1}{300} \leq T_0 \leq \frac{1}{50} = \frac{1}{f_{min}}$$

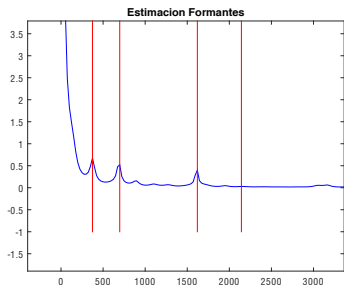
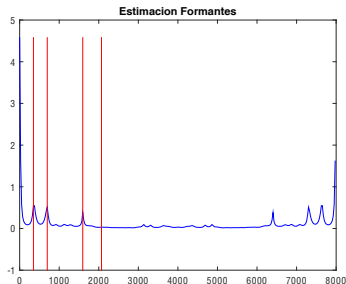


Preénfasis

$$y[n] = x[n] - ax[n-1] \quad 0.9 \leq a \leq 0.97 \quad y[1] = x[1]$$



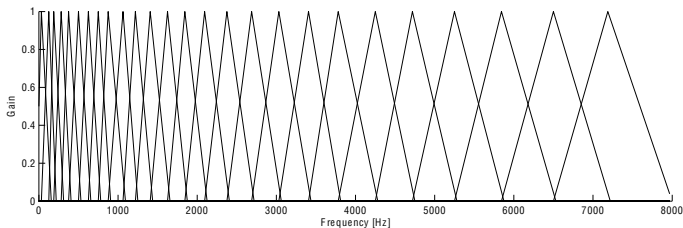
Frecuencias formantes



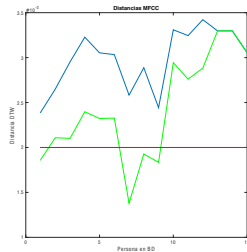
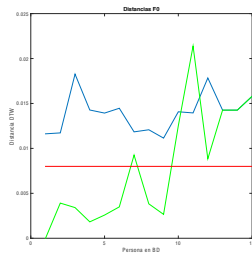
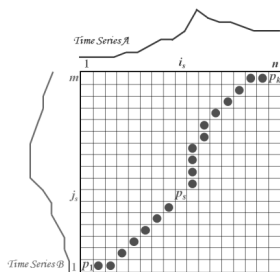
- 1 Respuesta en frecuencia.
Predicción lineal.
Wiener-Hopf.
- 2 Parámetros del sistema y
factor de ganancia.
Levinson-Durbin.
- 3 $H(z)$. Máximos locales.

MFCC

- 1 Ventana. DFT.
- 2 Filtros triangulares.
- 3 $F_{mel} = 1000 \log_2 \left(1 + \frac{F_{Hz}}{1000} \right)$
- 4 IDFT.



DTW

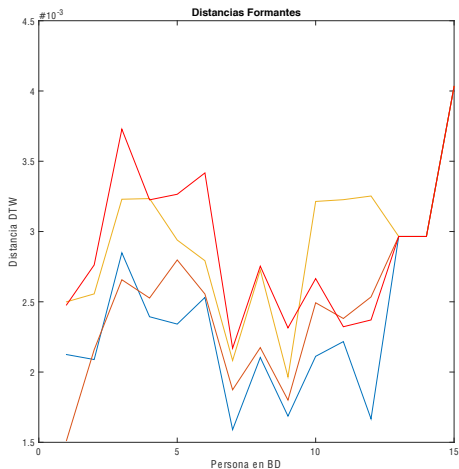


Resultados

Intento / Persona	F_0	Formantes	MFCC
1	OK	OK	OK
2	OK	ERROR	OK
3	OK	OK	OK
4	OK	ERROR	OK
5	ERROR	OK	OK
6	OK	OK	OK
7	OK	OK	OK
8	OK	ERROR	OK
9	ERROR	ERROR	ERROR
10	ERROR	OK	OK

Resultados

DTW - Formantes



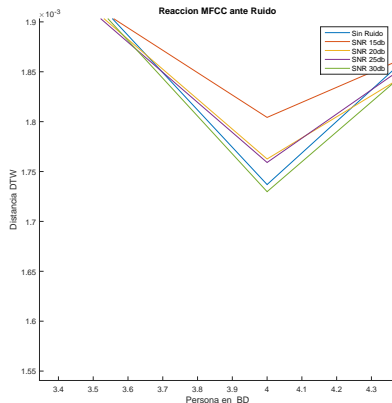
Resultados

		Valor real	
		V	F
Valor predicho	V	8	2
	F	1	9

$$\text{Sensibilidad: } \frac{VP}{(VP + FN)} = \frac{8}{8 + 2} = 0.8$$

$$\text{Especificidad: } \frac{VN}{(VN + FP)} = \frac{9}{9 + 1} = 0.9$$

Resultados



Ruido blanco:

$$\mu = 0, \sigma = 0.5$$

$$SNR = 45[dB]$$

9/10 aciertos.

Ruido ambiente:

$$SNR = 30[dB]$$

10/10 aciertos

$$SNR = 25[dB]$$

7/10 aciertos

Conclusiones

- 1 F_0 por si sola no es un buen método de verificación pero complementa.
- 2 Las *formantes* son poco precisas. Suelen arrojar falsos negativos.
- 3 Los MFCC son los que arrojan los mejores resultados para la verificación.
- 4 Todos los métodos son poco robustos al ruido, para que la verificación sea correcta, se necesita una relación señal/ruido de al menos 30 dB.
- 5 Ante la distorsión de la voz de una persona que se encontraba pregrabada en la base de datos, la verificación se muestra inestable.

Preguntas

¿Preguntas?