

RECONOCIMIENTO AUTOMÁTICO DEL HABLA UTILIZANDO LA TRANSFORMADA DE FOURIER Y REDES NEURONALES

Jose D. Pardo¹, Jose. A. Castro¹, Gilberto. P. Iseda¹, Cesar. Torres M², Lorenzo Mattos².

¹ *Estudiantes de Ingeniería Electronica, Universidad Popular de Cesar*

² *Laboratorio de óptica e Informática, Universidad Popular de Cesar.*

(Recibido 09 de Sep.2005; Aceptado 20 de Jun. 2006; Publicado 20 de Nov. 2006)

RESUMEN

En el presente artículo se estudia el reconocimiento automático de las señales de voz utilizando para esto la transformada de Fourier y las redes neuronales. Se discriminarán las características del habla en el dominio de la frecuencia y se procederá a entrenar una red neuronal que sea capaz de distinguir los diferentes patrones. Con la red entrenada se clasifican los diversos patrones para utilizarlos como comandos para realizar funciones de control.

Palabras claves: Reconocimiento automático de la voz, Transformada de Fourier, Redes neuronales.

ABSTRACT

In the present article it is studied the automatic recognition of the signal of speech using for this the Fourier's transform and the Neural. Networks. The characteristics of the speech will be discriminated in the domain of the frequency and we will proceed to train a Neural. Networks that is capable of distinguishing the different patterns. With the trained network it is classify the diverse patterns to use them as commands to realize functions of control.

Key words: Automatic Speech Recognition, Fourier Transform, Neural Network.

1. Introducción

Los sistemas de reconocimiento automático de voz tienen gran auge en la actualidad dada la amplia comodidad y versatilidad que brindan. Entre sus virtudes está el ofrecer la posibilidad que el hombre interactúe en su lenguaje natural con equipos electrónicos y eléctricos, además es un proceso casi intuitivo el operar estos equipos por la sencillez de los comandos que se utilizan, que por lo general son palabras cortas como “encender”, “apagar”, entre otras instrucciones que describen explícitamente su función. El proyecto que se propone es vanguardista a nivel regional, porque en la costa atlántica el trabajo en esta área es poco.

El reconocimiento automático de la voz dota a las máquinas de la capacidad de recibir mensajes orales. Tomando como entrada la señal acústica recogida por un micrófono, este proceso de reconocimiento del habla tiene como objetivo final decodificar el mensaje contenido en la onda acústica para realizar las acciones pertinentes.

2. Procedimiento

2.1. Componentes de un sistema de Reconocimiento Automático de Voz (RAV)

Estos sistemas se pueden describir como la concatenación de cuatro subsistemas Ver Figura No. 1 La adquisición de la entrada, la representación de la misma, una ordenación local y un decodificador general

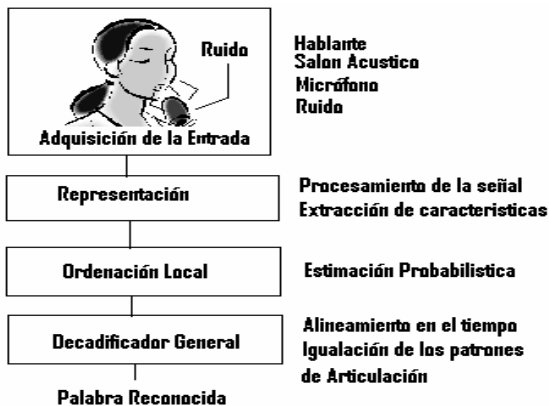


Fig.1. Diagrama de bloques de un sistema de reconocimiento de voz.

2.2. Adquisición de señales de voz

La señal se capturó con la ayuda de un micrófono conectado a la tarjeta de sonido del computador; muestreando con un intervalo T , a una frecuencia que es igual a dos veces la componente frecuencial de la señal de mayor orden de acuerdo con el teorema de Nyquist, para el caso de las señales de voz la frecuencia de muestreo es igual a 8000 muestras por segundo, la salida $y(nT)$ es dada por:

$$y(nT) = \int_{-\infty}^{\infty} h(t - nT) x(t) dt \tag{1}$$

Donde, $\tilde{h}(t)$ es la respuesta impulsional del micrófono; $y(t)$ es la salida del micrófono; T es el intervalo de muestreo; $y_s(t)$ es la versión muestreada de $y(t)$; $y(nT)$ es la salida, $n \in \mathbb{Z}$ y son los valores de las muestras. El bloque del C/D es un conversor análogo a digital. La adquisición de las señales se realiza utilizando el DAQ (data acquisition) de Matlab.

2.3. Tratamiento de las señales

Después de la adquisición la señal es filtrada y normalizada.

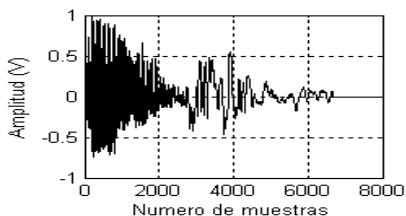


Fig.2. Señal de voz filtrada y normalizada.

La FFT es una versión eficiente de la DTF (transformada discreta de Fourier). La DTF de una secuencia de duración finita [1].

Utilizando el algoritmo de la FFT se visualiza la señal mostrada en la Figura No. 3.

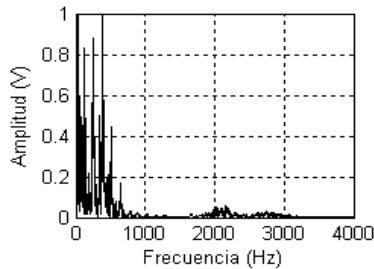


Fig.3. Espectro de la señal.

Para disminuir el costo computacional y aumentar la velocidad de procesamiento la señal es promediada mediante un filtro de corrimiento.

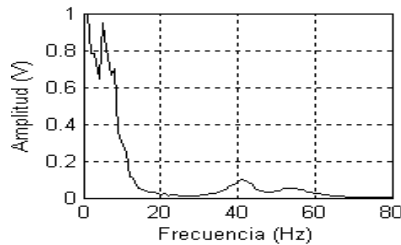


Fig.4. Espectro normalizado y promediado.

2.4. Implementación de redes neuronales

Las redes neuronales son los algoritmos encargados de realizar la ordenación local o decir si la señal de entrada es similar a las señales que hacen parte de la base de datos con que se entrena la red neuronal. [2]

Para entrenar la red neuronal se tomaron 2400 muestras de 30 palabras de un mismo locutor, 10 palabras ‘tele’, 10 ‘canal’ y 10 ‘volumen’. La red que se entrenó fue una perceptrón multicapa con retropropagación de error. Se utilizó el toolbox de redes neuronales de matlab

3. Resultados

Tabla No.1. Porcentajes de reconocimiento por palabra

	Tele	Canal	Volumen
Tele	91.3	3.5	4.1
Canal	3.4	92.5	5.3
Volumen	6.4	2.8	90.2

Con la red neuronal entrenada satisfactoriamente, se concatenó cada una de las etapas; adquisición de la señal, procesamiento de la misma y extracción de características, resultando un reconocedor de tres palabras (tele, canal y volumen).

Conclusiones

- La transformada discreta unidimensional de Fourier brinda una herramienta útil para la extracción de características de las señales de voz. Para que esa extracción mejore se debe realizar un pre-procesamiento.
- Las redes neuronales proveen una excelente ayuda en la diferenciación de las señales, su implementación en Matlab se hace de manera sencilla, sin embargo se debe estudiar otras topologías para optimizar los resultados obtenidos.

REFERENCIAS

- [1] Cooley, J. W., y Tukey, J., "An algorithm for machine computation of complex Fourier series", Math comput. 19: 297-301, 1965
- [2] DELGADO, Alberto. Inteligencia Artificial y Mini robots. ECOE Ediciones. 1998.