

Programacion III

Reconocimiento de voz

**Gutierrez Julian
Karraz Facundo**

09/10/2018

Trabajo de investigación

Tema: Reconocimiento de voz

Resumen:

El trabajo consiste en un sistema destinado al reconocimiento de palabras aisladas mediante el uso de un dsPIC30F4013, dado que tiene una EEPROM mayor a la del dsPIC33FJ32MC202.

Marco teórico:

Diagrama en bloques del proceso en cuestión:



Entrada de voz

En la etapa de entrada de voz la señal analógica a la salida del micrófono es muestreada para poder ser procesada por el DSP.

El rango de frecuencias conversacionales de la voz humana está entre 250 y 3000 [Hz], por lo tanto, para cumplir con el teorema de Nyquist, utilizando una frecuencia de muestreo F_s igual a 8000 [Hz] se puede realizar una buena representación discreta de la señal de entrada {8}.

Pre - procesamiento

Diagrama en bloques



Ver {1}, {2} y {3}.

En esta etapa, la señal ya discretizada, pasa por un filtro cuya finalidad es la de eliminar los momentos de silencios por medio del cálculo de energía en corto tiempo descrito en {4}.

Dado que la voz presenta por naturaleza una atenuación en las frecuencias altas, se aplica un filtro pre - énfasis cuya finalidad es la de resaltar estas para poder hacer un análisis más exacto de la voz. Cuya salida y función de transferencia estarán designadas por:

$$y[n] = x[n] - a \cdot x[n-1]$$

$$H(z) = 1 - a \cdot z^{-1}$$

La voz, en periodos de entre 20 y 30 [ms] es cuasi - estacionaria, de manera que para poder analizarla, debemos hacerlo en intervalos de esos tiempos. Por lo tanto se hace una segmentación de 30 [ms] con un solapamiento de 10 [ms] para no perder información. El ventaneo se realiza con una ventana Hamming igual a:

$$s[n] = 0.54 - 0.46 \cdot \cos(2\pi n/N) \text{ siendo } N = \text{número de muestras del segmento.}$$

Dado que $F_s = 8000$ [Hz] y tenemos segmentos de 300 [ms] entonces $N = 8000 \cdot 0.03 = 240$ muestras.

Extracción de parámetros

Hay varios métodos para extraer parámetros de una señal, algunos más eficientes que otros pero con la implicación de un mayor tiempo computacional, un método que cumple con eficiencia y rapidez es el llamado LPC explicado en {5} y {6}.

Mediante estudios se ha llegado a la conclusión de que la cantidad de elementos de un vector LPC para una correcta caracterización de la señal ronda entre los 15 y 25, es por eso que se utilizara un total de 20 elementos.

Reconocimiento y decisión

Dado que la duración de las palabras no siempre es la misma, a la hora de comparar la similitud entre una y otra se vuelve una tarea complicada de desarrollar de una manera "lineal" o continua, es por esto, que para esta etapa se utiliza DTW {7}.

- {1} <https://scielo.conicyt.cl/pdf/rfacing/v12n1/art02.pdf> (paper a seguir)
- {2} Dialnet-ReconocimientoDeVozHumanaAplicadoALaDomotica
<https://dialnet.unirioja.es/descarga/articulo/5038438.pdf> (paper a seguir)
- {3} Reconocimiento de Palabras aisladas utilizando LPC Y DTW, para control de navegación de un mini-robot. http://www.alvarestech.com/temp/4/robot_voz_unillanos.pdf (paper a seguir)
- {4} <https://arxiv.org/pdf/1804.06159.pdf> (endpoint speech)
- {5} http://profesores.fi-b.unam.mx/procvoz/Capitulo3/3_3.doc (Explicacion de prediccion lineal)
- {6} https://www.youtube.com/watch?v=RTaNej-gz7k&list=PLH2VulvVYXkZcNYdjs5GheCXSg_vHhN55 (Explicacion de prediccion lineal)
- {7} https://en.wikipedia.org/wiki/Dynamic_time_warping (explicacion de DTW)
- {8} <http://elastixtech.com/fundamentos-de-telefonía/transmision-de-la-voz/>

Micrófono:

Se propone utilizar un micrófono con conexión TRRS de tipo condensador electret (micrófono de auriculares) dicho micrófono se alimentara con 2 V.

Los micrófonos electret tienen una respuesta en frecuencia de 50 a 15.000 Hz y una sensibilidad entre -50 dB y -70 dB

<http://electronicayciencia.blogspot.com/2010/06/utilizar-un-microfono-electret.html>

<https://www.youtube.com/watch?v=jAZGgzxb5Ss> (Proyecto de Speech Recognition en DSP silla de ruedas)

<http://amateurselectronics.blogspot.com/> (podria llegar a servir, LPC)

http://biblioteca.usac.edu.gt/tesis/08/08_0223_EO.pdf (Un poquito de todo)

http://www.iitg.ac.in/samudravijaya/tutorials/fundamentalOfASR_picone96.pdf (Base teorica de todo)

<http://ww1.microchip.com/downloads/en/devicedoc/51456b.pdf> (libreria de DSP MPLAB)