Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería Mecánica Eléctrica

SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB

Genoveva Velásquez Ramírez

Asesorado por el Ing. MsEE. PhD. Enrique Edmundo Ruiz Carballo

Guatemala, abril de 2008

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR:

GENOVEVA VELÁSQUEZ RAMÍREZ

ASESORADO POR EL ING. MSEE. PHD. ENRIQUE EDMUNDO
RUIZ CARBALLO
AL CONFERÍRSELE EL TÍTULO DE

INGENIERA ELECTRÓNICA

GUATEMALA, ABRIL DE 2008

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Murphy Olympo Paiz Recinos
VOCAL I	Inga. Glenda Patrícia García Soria
VOCAL II	Inga. Alba Maritza Guerrero de López
VOCAL III	Ing. Miguel Ángel Dávila Calderón
VOCAL IV	Br. Kenneth Issur Estrada Ruiz
VOCAL V	
SECRETARIA	Inga. Marcia Ivonne Véliz Vargas

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Murphy Olympo Paiz Recinos
EXAMINADOR	Ing. Enrique Edmundo Ruiz Carballo
EXAMINADOR	Ing. Luis Eduardo Durán Córdoba
EXAMINADOR	Ing. Jose Alonso Rivera Carrillo
SECRETARIA	Inga. Marcia Ivonne Véliz Vargas

HONORABLE TRIBUNAL EXAMINADOR

Cumpliendo con los preceptos que establece la Ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB,

tema que me fuera asignado por la Dirección de la Escuela de Ingeniería Mecánica Eléctrica, con fecha 26 de febrero de 2007.

Genoveva Velásquez Ramírez

Guatemala, 23 de octubre de 2007

UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA

FACULTAD DE INGENIERIA

Ingeniero Julio César Solares Peñate
Coordinador Área de Electrónica
Escuela de Ingenieria Mecánica-Eléctrica
Facultad de Ingenieria
Universidad de San Carlos de Guatemala

Estimado Ingeniero Solares:

Por este medio le informo que he revisado el trabajo de graduación titulado: **SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB**, elaborado por la estudiante *Genoveva Velásquez Ramírez*.

El mencionado trabajo llena los requisitos para dar mi aprobación, e indicarle que el autor y mi persona somos responsables por el contenido y conclusiones del mismo.

Le saludo atentamente,

Ing. Enrique Edmundo Ruiz Carballo
Asesor

UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA

FACULTAD DE INGENIERÍA

Guatemala, 9 de noviembre 2007.

Señor Director
Ing. Mario Renato Escobedo Martìnez
Escuela de Ingenierìa Mecànica Elèctrica
Facultad de Ingenierìa, USAC.

Señor Director:

Me permito dar aprobación al trabajo de Graduación titulado: **SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB, de la** estudiante; Genoveva Velásquez Ramìrez, por considerar que cumple con los requisitos establecidos para tal fin.

Sin otro particular, aprovecho la oportunidad para saludarle.

Atentamente,

ID Y ENSEÑAD A TODOS

Ing. Julio César Solares Peñate
Coordinador del Area Electrònica

JCSP/sro

UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA

FACULTAD DE INGENIERÍA

El Director de la Escuela de Ingenierìa Mecànica Elèctrica, después de conocer el dictamen del Asesor, con el Visto Bueno del Coordinador de Area, al trabajo de Graduación del estudiante; Genoveva Velásquez Ramìrez titulado: **SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB,** procede a la autorización del mismo.

Ing. Mario Renato Escobedo Martìnez
DIRECTOR

GUATEMALA, 12 DE NOVIEMBRE 2,007.

Universidad de San Carlos
de Guatemala

Facultad de Ingeniería
Decanato

Ref. DTG.116.08

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería Mecánica Eléctrica, al trabajo de graduación titulado: **SISTEMA DE RECONOCIMIENTO DE VOZ EN MATLAB,** presentado por la estudiante universitaria **Genoveva Velásquez Ramírez,** autoriza la impresión del mismo.

IMPRÍMASE.

Ing. Murphy Olympo Paiz Recinos
DECANO

Guatemala, abril de 2008

/cc
c.c. archivo.

DEDICATORIA

Dedico este trabajo a todas las personas que nunca han dejado de creer, a las que han hecho historia con su lucha, a las que siempre han estado y a las que siempre estarán.

AGRADECIMIENTOS A:

Mi familia, en especial a mi mamá.

A San Judas Tadeo.

Al Ing. Enrique Ruiz Carballo

Mis amigos.

A la Universidad de San Carlos de Guatemala.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

FIGURAS

TABLAS

RESUMEN

El Sistema de Reconocimiento de Voz permite que el usuario grabe una palabra por medio de un micrófono y ésta sea reconocida en la base de datos existente en ese momento. El sistema en sí posee un entorno gráfico en la computadora, que proporciona las selecciones de grabación, donde la señal de voz es ingresada a la computadora y es procesada por los algoritmos del programa que modifican la señal, obteniendo los parámetros significativos de la señal de voz, para luego ser almacenados en la computadora. La selección de reconocimiento permite que la palabra sea comparada con la base de datos almacenada en la computadora, dicha base de datos ya fue procesada digitalmente por el programa. Esta selección reconoce la palabra.

El entorno gráfico proporciona, por otra parte, un análisis gráfico de las palabras grabadas y reconocidas. Como el Sistema de Reconocimiento de Voz es un sistema de procesamiento digital de señales de voz, el análisis gráfico en el reconocimiento es un análisis del espectro de frecuencias de la señal de voz. El presente trabajo es una recopilación de los métodos de procesamiento digital y una explicación de los algoritmos utilizados en el programa del Sistema de Reconocimiento de Voz. Son explicados varios métodos de procesamiento digital de voz y los recursos necesarios para la elaboración del sistema. Además, se incluye la explicación de las características o parámetros relevantes en el procesamiento digital de voz, y el porque de la selección de los procedimientos utilizados en los algoritmos que constituyen el programa. Terminando con un esquema gráfico general del programa, la evaluación del mismo y una presentación del funcionamiento del entorno gráfico del programa.

OBJETIVOS

- **General**

Brindar un sistema que proporcione el reconocimiento de señales de voz, por medio de la interacción entre el usuario y la computadora.

- **Específicos**

1. Explorar algoritmos de procesamiento digital de voz, que permitan un tratamiento sencillo de información relevante de las señales de voz.

2. Lograr la interacción automática humano/computadora por medio de un sistema simple de utilizar.

INTRODUCCIÓN

El habla es una de las partes más importantes de la expresión humana, es algo que nos diferencia del resto de seres vivos en planeta, ya que sin el habla el pensamiento mismo del hombre no sería posible. No se trata simplemente de un sistema para transmitir información, aunque sea claro una de sus funciones. Pero es por medio de los sonidos que se presenta la esencia espiritual del hombre.

Dada la importancia del habla, el presente trabajo de graduación pretende crear una interacción entre una de las expresiones esenciales del hombre con la computadora, creando así un Sistema de Reconocimiento de Voz.

El procesamiento digital de señales de voz tiene una gran variedad de aplicaciones, existe una base para el tratamiento digital de señales, que puede ser implementada para lograr obtener lo que nos interese según la aplicación.

El Sistema de Reconocimiento de Voz es una de las aplicaciones del procesamiento digital de señales de voz. El sistema consiste en obtener una señal de voz que permita reconocer qué palabra se esta hablando. Consta de una interfaz gráfica que permite la interacción del usuario por medio de un micrófono con la computadora, la que procesa automáticamente los datos adquiridos. Basado en los resultados de este sistema, se puede ver como se plantea la base del procesamiento digital de señales de voz y queda a la libre imaginación como puede ser utilizado para otras aplicaciones, además del de reconocimiento de voz.

1. SISTEMA DE RECONOCIMIENTO DE VOZ

1.1. Componentes del sistema

1.1.1 Micrófono

Es un transductor electroacústico, que tiene como función transformar o traducir la presión acústica ejercida sobre su capsula por las ondas sonoras en energía eléctrica. La calidad de cada micrófono viene dada por sus características, las cuales se describen a continuación:

1. Sensibilidad: es la eficiencia del micrófono, la relación entre la presión sonora que incide (expresada en Pascales) y la tensión eléctrica de salida (expresada en voltios). O sea, expresa que tan bien convierte el micrófono la presión acústica en voltaje de salida. La sensibilidad se expresa en milivoltios por Pascal.

Al utilizar el milivoltio, la sensibilidad puede ser representada en un voltímetro de la siguiente manera: a mayor voltaje, mayor sensibilidad.

2. Fidelidad: indica la variación de sensibilidad con respecto a la frecuencia. Además, la fidelidad, viene definida como la respuesta en frecuencia del micrófono, cuanto mas lineal sea la respuesta en frecuencia mayor fidelidad tendrá el micrófono. La fidelidad se expresa en dB.

En función de esta respuesta en frecuencia o fidelidad se elabora la llamada Curva de respuesta de un micrófono, que es la representación gráfica del nivel obtenido en la captación de sonidos de igual intensidad, pero de distinta frecuencia.

3. Directividad: esta característica determina en que dirección capta mejor el sonido un micrófono, es decir indica la sensibilidad del micrófono a las diferentes direcciones.

El diagrama polar es una representación gráfica que indica qué tan sensitivo es un micrófono a los sonidos que llegan a el desde diferentes ángulos alrededor de su eje central.

Dependiendo de la directividad se encuentran diferentes tipos de micrófonos:

- Omnidireccionales: captan todos los sonidos, sin importar la dirección desde donde lleguen.

- Bidireccionales: captan tanto el sonido que llega por su parte frontal, como por su parte posterior.

- Unidireccionales o direccionales: captan el sonido en una sola dirección mientras que son relativamente sordos a las otras direcciones.

4. Ruido de fondo: es la tensión que entrega el micrófono sin que exista ningún sonido incidiendo sobre él. Este ruido se produce por el movimiento térmico de los electrones en la carcasa que no tiene masa. El ruido de fondo debe estar en torno a los 60dB, pero mientras más bajo sea, mejor calidad ofrece el micrófono.

5. Rango dinámico: se puede definir de dos maneras:

- La primera definición es el margen que hay entre el nivel de referencia de salida máxima y el ruido de fondo de un determinado sistema, medido en decibelios. En este caso, el rango dinámico y relación señal/ruido son sinónimos.

- Como segunda definición es el margen que hay desde el nivel de pico y el nivel de ruido de fondo, también indicado en dB. En este caso, rango dinámico y relación señal/ruido no son igualables.

Las dos maneras son validas, por lo que generalmente los fabricantes incluyen la referencia de salida máxima y la referencia de nivel de pico en las especificaciones del micrófono. Para aclarar mejor esta característica nos referiremos a los siguientes dos términos.

- La relación señal/ruido: esta es la relación entre la señal útil dada, o sea, la señal de referencia, y el ruido de fondo del micrófono.
- Nivel máximo o nivel de pico: es la diferencia entre el nivel máximo admisible y el nivel del ruido de fondo expresada en dB. Se trata del nivel máximo admisible por el micrófono correspondiente a una distorsión armónica de la señal de 0.5% a 1000Hz.

6. Impedancia interna: es la resistencia que opone el micrófono al paso de la corriente. La impedancia según su valor viene caracterizada por baja, alta y muy alta impedancia.

- Lo-Z Baja impedancia (alrededor de 200 Ohmios)
- Hi-Z Alta impedancia (1 K Ω o 3 K Ω e incluso 600 Ω)
- VHi-Z Muy alta impedancia (más de 3 K Ω).

Si el micrófono es de alta impedancia y se tiene un cable largo se produce una pérdida muy grande. Si se tiene una impedancia baja se puede utilizar un cable muy largo y no se pierde tanto la señal. Por último, se puede bajar la resistencia para evitar pérdidas en altas frecuencias.

7. Factor de directividad: es la relación entre la intensidad sonora del sonido directo con respecto a la del ruido ambiente recogido en todas las direcciones.

1.1.1.1. Clasificación de los micrófonos

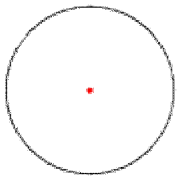
Se pueden dividir según:

1. La Directividad

Como ya se mencionó en las características, según la directividad hay tres tipos de micrófonos:

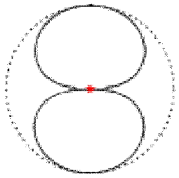
- Micrófono omnidireccional: este es aquel cuyo diagrama polar es considerado como un círculo perfecto (véase figura 1). Aunque esto es en el caso ideal.

Figura 1. Patrón omnidireccional.



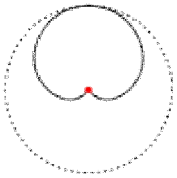
- Micrófono bidireccional: en este caso el diagrama polar muestra como captan por la parte frontal y la lateral (veáse figura 2), como ejemplo tenemos el micrófono de gradiente de presión.

Figura 2. **Patrón bidireccional**



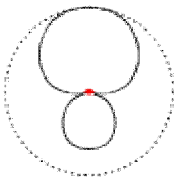
- Micrófono unidireccional: existe gran número de patrones polares para este tipo, los más comunes:
  - Micrófono cardioide: llamado así porque su patrón de sensitividad tiene la forma de un corazón o un cardioIde (véase figura 3).

Figura 3. Patrón cardioide



  - Micrófono hipercardioide: es similar al cardioide, pero con un área más apretada de sensitividad frontal y un pequeño lóbulo de sensitividad posterior. Este patrón es mostrado en la figura 4.

Figura 4. Patrón hipercardioide



2. El Transductor

Se encuentran tres grandes grupos según el tipo de transductor.

o Micrófono de Condensador o Capacitor: lo que sucede con este micrófono es que las ondas sonoras provocan el movimiento oscilatorio del diafragma, el cual actúa como una de las placas de un capacitor y la vibración provoca una variación en la energía almacenada en el condensador que forma el núcleo de la capsula microfónica. Esta variación genera una tensión eléctrica que es la señal de salida del sistema. La señal de salida de este sistema es análoga. Según el principio de operación:

- Micrófono de condensador DC.
- Micrófono de condensador electret.
- Micrófono de condensador de radiofrecuencia (RF).

Para nuestra aplicación en MATLAB utilizamos un micrófono simple de tipo electret, ya que son los menos costos.

o Micrófono Dinámico: trabajan por medio de inducción electromagnética, la vibración del diafragma provoca el movimiento de una bobina móvil o cinta corrugada ancladas a un imán permanente que genera un campo magnético que a su vez genera una tensión eléctrica, que es la señal de salida. Esta señal eléctrica es análoga. Hay dos tipos básicos:

- Micrófono de bobina móvil o dinámico
- Micrófono de cinta

o Micrófono piezoeléctrico: utilizan el fenómeno de piezoelectricidad, cuando las ondas sonoras hacen vibrar el diafragma el movimiento de este hace que se mueva el material contenido en su interior (cuarzo, sales de Rochélle, carbón, etc.). La fricción entre estas partículas generan sobre la superficie del material una tensión eléctrica.

La respuesta en frecuencia de estos micrófonos es muy irregular, ya que son micrófonos piezoeléctricos:

- Micrófono de carbón
- Micrófono de cristal
- Micrófono de cerámica

3. La Utilidad

Existen cinco tipos principales de micrófono según la utilidad:

1. Micrófono de mano o de bastón: como su nombre lo dice esta hecho para estar sujeto con la mano, además esta diseñado de forma que amortigua los golpes y ruidos de manipulación.

2. Micrófonos de estudio: no tienen protección contra la manipulación, pero están situados en una posición fija y protegido con gomas de las vibraciones.

3. Micrófono de contacto: toman el sonido estando en contacto físico con la onda sonora.

4. Micrófono de corbata, de solapa o Lavalier: micrófono en miniatura que poseen filtros para evitar las altas frecuencias que produce el rozo del micro con la ropa.

5. Micrófono inalámbrico: la particularidad de este micrófono es que se puede utilizar sin cable. No necesitan el cable porque poseen un transmisor de FM o de AM, pero el más habitual es el de FM.

1.1.2. MATLAB

MATLAB es el nombre abreviado de “MATrix LABoratory”. Es un lenguaje de alto nivel y de ambiente interactivo que permite realizar tareas intensas y con una mayor velocidad que los lenguajes de programación comúnmente usados.

MATLAB se especializa en cálculos numéricos con vectores y matrices, como casos particulares puede trabajar también con otras estructuras de información. Aunque cada objeto es considerado como un arreglo.

El lenguaje esta construido por código llamado M-code que puede ser fácilmente ejecutado en la ventana de comandos. Con lo cual se pueden crear funciones, etc. Pero la razón principal para la elección de este lenguaje de programación son las herramientas que proporciona para el procesamiento de señales, y el conjunto de funciones para el procesamiento digital.

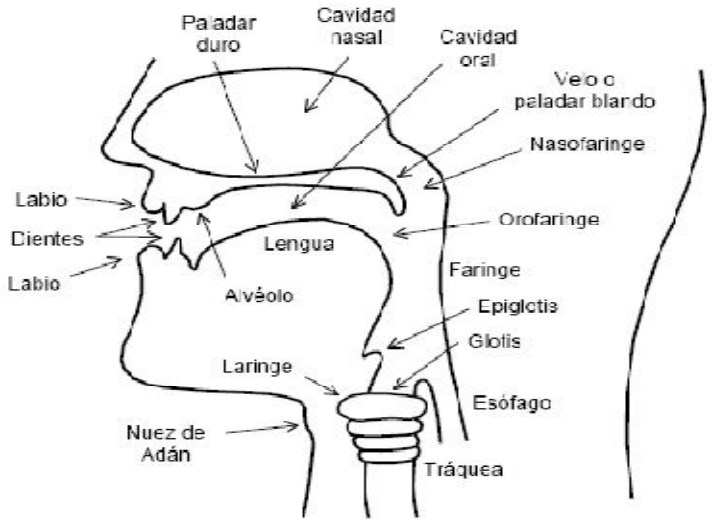
Además, para crear entornos gráficos se puede utilizar el GUIDE de MATLAB, que provee herramientas para crear GUIs, ‘Graphical User Interface’, con lo cual se puede crear la forma del entorno grafico así como asociar funciones a los elementos del GUI. MATLAB también incluye funciones para manipular archivos.

1.1.3. Señal de voz

1.1.3.1 Breve anatomía del aparato fonatorio

La voz humana se produce por medio del aparato fonatorio. Este esta formado por los pulmones como fuente de energía, en forma de flujo de aire, la laringe que contiene las cuerdas vocales, la faringe, las cavidades oral y nasal y una serie de elementos articulatorios: los labios, los dientes, el alveolo, el paladar, el velo del paladar y la lengua(véase figura 5).

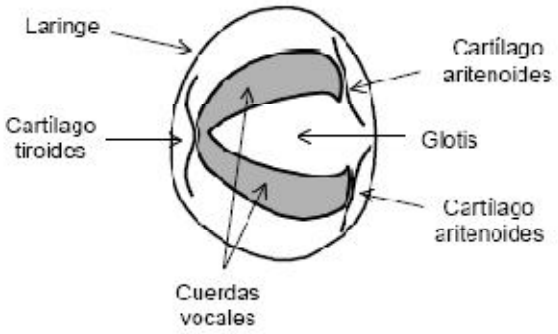
Figura 5. Aparato Fonatorio Humano



Fuente: http://www.eie.fceia.unr.edu.ar/~acustica/biblio/biblio.htm#4a

Las cuerdas vocales son dos membranas dentro de la laringe orientadas de adelante hacia atrás. Por delante se unen en el cartílago tiroides, por detrás, cada una esta sujeta a uno de los dos cartílagos aritenoides, los cuales pueden separarse voluntariamente por medio de músculos. La abertura entre ambas cuerdas se denomina glotis (véase figura 6).

Figura 6. Corte esquemático de la laringe según un plano horizontal.



Cuando las cuerdas vocales se encuentran separadas la glotis adopta una forma triangular. El aire pasa libremente y casi no se produce sonido; este es el caso de la respiración. Cuando la glotis comienza a cerrarse, el aire que la atraviesa proveniente de los pulmones experimenta una turbulencia, produciendo un ruido conocido como aspiración. Ahora, al cerrarse más, las cuerdas vocales comienzan a vibrar de modo audible, produciéndose un sonido tonal, es decir periódico. La frecuencia de este sonido depende de varios factores, entre otros del tamaño y la masa de las cuerdas vocales, de la tensión de las cuerdas vocales, de la tensión que se les aplique y de la velocidad del flujo del aire proveniente de los pulmones. A mayor tamaño, menor frecuencia de vibración, A mayor tensión la frecuencia aumenta, siendo los sonidos más agudos. También aumenta la frecuencia al crecer la velocidad del flujo de aire. Finalmente, es posible obturar la glotis completamente, en cual caso no se produce sonido. Sobre la glotis se encuentra la epiglotis, un cartílago de la faringe que permite tapar la glotis durante la deglución para evitar que el alimento ingerido se introduzca en el tracto respiratorio. La porción que incluye las cavidades faringea, oral y nasal junto con los elementos articulatorios se denomina cavidad supraglotica en tanto que los espacios por debajo de la laringe, es decir la traquea, los bronquios y los pulmones, se denominan cavidades infragloticas.

Varios de los elementos de la cavidad supraglotica se controlan a voluntad, permitiendo modificar dentro de márgenes muy amplios los sonidos producidos por las cuerdas vocales o agregar partes distintivas a estos, y hasta producir sonidos propios. Esto se efectúa con dos mecanismos principales: el filtrado y la articulación.

El filtrado actúa modificando el espectro del sonido. Tiene lugar en las cuatro cavidades supragloticas principales: la faringe, la cavidad nasal, la cavidad oral y la cavidad labial. Las mismas constituyen resonadores acústicos que enfatizan determinadas bandas frecuenciales del espectro generado por las cuerdas vocales, conduciendo al concepto de formantes, es decir que se refuerza la amplitud de grupos de armónicos situados alrededor de una determinada frecuencia.

En resumen, en el habla los formantes se determinan por el proceso de filtrado que se produce en el tracto vocal por la configuración de los articuladores.

Dada la anterior explicación es necesaria una clasificación acústica, como la que se resume en la tabla I.

Tabla I. Resumen de la clasificación acústica de los sonidos.

Sonidos periódicos compuestos o complejos.	Vibración de las cuerdas vocales (frecuencia del fundamental, F0) y resonancia (armónicos).	Vocales, nasales, laterales.
Sonidos aperiódicos impulsionales.	Cierre y explosión en el tracto vocal.	Oclusivas.
Sonidos aperiódicos continuos.	Fricción en el tracto vocal.	Fricativas.

Fuente: http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/fon_acust.html

1.1.3.2. Formantes

Los formantes son elementos que sirven para distinguir componentes del habla humana, principalmente, las vocales y sonidos sonantes. El formante con la frecuencia más baja se llama F1, el segundo F2, el tercero F3, etc. Normalmente sólo los dos primeros son necesarios para caracterizar una vocal, aunque la pueden caracterizar más formantes. Generalmente, los formantes posteriores determinan propiedades acústicas como el timbre.

Los dos primeros formantes se determinan principalmente por la posición de la lengua. Sucediendo que F1 tiene una frecuencia más alta cuanto mas baja esta la lengua, es decir una mayor abertura. Para el F2 tiene mayor frecuencia cuanto mas hacia delante esta posicionada la lengua.

No todos los sonidos se componen de formantes definidos. Solamente aparecen en sonantes, que incluyen los sonidos pulmonares: vocales, aproximantes y nasales. Las nasales tienen un formante adicional F3, en torno a los 1500 Hz.

Si la frecuencia fundamental es mayor que la frecuencia de los formantes, entonces el carácter del sonido se pierde y se vuelven difíciles de distinguir, por lo cual son difíciles de reconocer.

Aquí están algunos anchos de banda entre los cuales se localizan las vocales:

Tabla II. Formantes Vocálicos

Formantes Vocálicos	
Vocal	Región principal formantica
/u/	200 a 400 Hz
/o/	400 a 600 Hz
/a/	800 a 1200 Hz
/e/	400 a 600 y 2200 a 2600 Hz
/i/	200 a 400 y 3000 a 3500 Hz

Fuente: http://es.wikipedia.org/wiki/Formante

1.2 Procesamiento digital de señales

1.2.1. Transformada discreta de Fourier (*DFT*)

La transformada discreta de Fourier (*DFT* por sus siglas en inglés) permite evaluar la transformada de Fourier de secuencias de duración finita. La *DTF* es una secuencia compleja que es obtenida por medio de muestrear un período de la transformada de Fourier de la señal a un numero finito de puntos de frecuencia, es decir, que corresponde a muestras igualmente espaciadas en frecuencia de la transformada de Fourier de la señal discreta. La *DFT* es importante por dos razones. Primero, permite determinar el contenido frecuencial de la señal de voz, o sea, realizar análisis espectral. La segunda razón de importancia es realizar operaciones de filtrado en el dominio de la frecuencia. La eficiencia es la razón principal por la cual se procesan las señales en el dominio de la frecuencia.

1.2.1.1. Muestreando la transformada de Fourier

Consideremos una secuencia periódica x[n] con su transformada de Fourier $X(e^{j\omega})$ y asumamos que una secuencia X[k] es obtenida al muestrear $X(e^{j\omega})$ a frecuencias $\omega_k = \frac{2\pi k}{N}$, como sigue:

Ecuación 1

$$X[k] = X\left(e^{j\omega}\right)\Big|_{\omega=\frac{2\pi k}{N}} = X\left(e^{j\left(\frac{2\pi k}{N}\right)}\right)$$

Como la transformada de Fourier es periódica en ω con período 2π, la resultante secuencia es periódica en k con periodo N. La secuencia de muestras es periódica dado a que los N puntos están igualmente espaciados iniciando desde cero. Por lo que la misma secuencia se repite mientras k varia en el rango de $0 \le k \le N-1$.

Se puede notar que la secuencia de muestras X[k], siendo periódica con período N, podría ser la secuencia de los coeficientes discretos de la serie de Fourier de la secuencia x^[n], la cual tiene la relación con x[n] del siguiente modo:

Ecuación 2

$$x\hat{}[n] = \sum_{r=-\infty}^{\infty} x[n-rN] = x[n] * \sum_{r=-\infty}^{\infty} \delta[n-rN]$$

Esto es x^[n] es la secuencia periódica que resulta de la convolución aperiódica de x[n] con un tren periódico de impulsos unitarios. Por tanto, la secuencia periódica x^[n], correspondiente a X[k], que es obtenida de muestrear $X(e^{j\omega})$, esta formada de x[n] por medio de sumar juntos un numero infinito de replicas cambiadas de x[n]. Estos cambios son todos los enteros positivos y negativos múltiplos de N. El período de la secuencia X[k].

Equivalentemente, x[n] se puede recuperar a partir de la correspondiente secuencia periódica x^[n] a través de la siguiente ecuación:

Ecuación 3

$$x[n] = \begin{cases} x\hat{}[n], 0 \le n \le N-1 \\ 0, otherwise. \end{cases}$$

Alternativamente, dada la secuencia de los coeficientes de Fourier X[k], se puede encontrar x^[n] y utilizar la ecuación anterior para obtener x[n]. Cuando las series de Fourier son utilizadas de este modo para representar secuencias de duración finita, es llamada la transformada discreta de Fourier.

1.2.1.2. Definición de la transformada discreta de Fourier

Esta definida como la secuencia de frecuencia-discreta de duración-finita que es obtenida de muestrear un período de la transformada de Fourier. Este muestreo como ya se mencionó es convencionalmente hecho a N puntos igualmente espaciados sobre un período que se extiende desde $0 \le \omega \le 2\pi$, o con lo siguiente

Ecuación 4

$$\omega_k = \frac{2\pi k}{N} \qquad \text{para} \quad 0 \le k \le N-1$$

Si x[n] es una secuencia discreta en el tiempo con transformada de Fourier $X(e^{j\omega})$, entonces la transformada discreta de Fourier, denotada por X[k], se define como:

Ecuación 5

$$X[k] = X(e^{j\omega})|_{\omega = \omega k = \frac{2\pi k}{N}} = \sum_{n=0}^{N-1} x[n] e^{-2\pi n k / N} \quad \text{para} \quad 0 \le k \le N-1$$

Es importante notar una cosa, siendo M la duración de la señal x[n] y N el período de x^[n], si M≤ N entonces la señal x[n] puede ser recuperada a partir de x^[n] y el exceso de número de puntos en un período de x^[n] son iguales a cero. Por otro lado, si M ≥ N sucede un traslape al formarse la extensión periódica, entonces se dice que ocurre *time-aliasing*.

A partir de x[n] o x^[n] se puede obtener una o la otra según las ecuaciones anteriores, pero la distinción entre una y la otra se hacen más evidentes cuando se analizan las propiedades de la *DTF*. Como la *DTF* consiste en muestras de la transformada de Fourier, las propiedades de linealidad, periodicidad y simetría de la transformada de Fourier son ciertas también para la *DTF*. Además para todas las expresiones de las propiedades dadas se especifican los rangos, para x[n] de $0 \le n \le N-1$ y para X[k] de $0 \le k \le N-1$, y ambas son cero fuera de estos rangos. A continuación se resumen estas propiedades.

1.2.1.3. Propiedades de la transformada discreta de Fourier

1. Definición: Para una secuencia de duración finita x[n]

Ecuación 6

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{j 2\pi kn / N} \quad \textit{para} \quad 0 \le k \le N-1 \quad (\textit{DTF})$$

y

Ecuación 7

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j 2\pi kn / N} \ \textit{para} \quad 0 \le n \le N-1 \quad (\textit{IDTF})$$

2. Extensión Periódica:

Ecuación 8

$$x\hat{}[n] = \sum_{m=0}^{\infty} x[n - mN]$$

3. Linealidad:

Ecuación 9

$$Si\ \{x[n]\} = \{x_1[n]\} + \{x_2[n]\}, entonces, X[k] = X_1[k] + X_2[k]$$

4. Periodicidad:

Ecuación 10

$$X[k] = X[k+N]$$

5. Funciones de magnitud y fase: Si

Ecuación 11

$$X[k]=X_R[k] + j\ X_I[k]$$

Entonces,

Ecuación 12

$$\left|X[k]\right|^2 = X_R^2[k] + X_I^2[k]$$

Ecuación 13

$$Arg\left|X[k]\right| = \arctan\left|X_I[k] / X_R[k]\right|$$

6. Transformada de Fourier de una secuencia con retraso:

Ecuación 14

$$Si\ \{y[n]\} = \{x[n-n_0]\}\ \ entonces\ \ Y[k] = X[k]e^{\frac{j2\pi kn_0}{N}}$$

7. Transformada de Fourier de la convolución de dos secuencias:

Ecuación 15

$$Si\ \{y[n]\} = \{h[n]\} * \{x[n]\} \qquad entonces\ \ Y[k] = H[k]\ X[k]$$

Una convolución lineal de h[n] y x[n] resulta cuando las secuencias *DTF* de N_y-puntos son computadas para h[n] y x[n], donde N_y es la duración de y[n]. De otro modo, una convolución circular resulta.

8. Transformada de Fourier del producto de dos secuencias:

Ecuación 16

$$Si\ \{y[n]\} = \{h[n]x[n]\}\ \ entonces\ \ Y[k] = H[k] \otimes X[k]\ \ (convolución\ circular)$$

1.2.2. Convolución circular

Dado un h[n], para $0 \le n \le N-1$, y {x[n]}, su convolución es igual a:

Ecuación 17

$$y[n] = \sum_{k=0}^{N-1} h[k]x[n-k] \qquad para\ todo\ n.$$

Denotamos la ecuación anterior como la convolución lineal. Se sabe que la transformada de Fourier es útil para convertir una operación de convolución en una multiplicación.

En este caso, el interés se centra en obtener {y[n]} a partir de la inversa DFT de {Y[k]}, donde Y[k] = H[k] X[k], para $0 \le k \le N-1$. Se debe de tener especial cuidado al aplicar las relaciones de la DFT porque las secuencias son periódicas con periodo N, ya que son una extensión periódica de las secuencias originales. La convolución que resulta es llamada entonces convolución circular, y se define como:

Ecuación 18

$$y\hat{}[n] = \sum_{k=0}^{N-1} h\hat{}[k]x\hat{}[n-k] \quad \textit{para} \quad 0 \le n \le N-1$$

O mejor dicho,

Ecuación 19

$$\{y\hat{}[n]\} = \{h\hat{}[n]\} \otimes \{x\hat{}[n]\}$$

Donde $\otimes$ denota la operación de convolución circular. Los pasos utilizados para computar esta convolución son idénticos a aquellos utilizados en la convolución lineal, excepto que la suma es tomada sobre un solo periodo. La {y^[n]} resultante también es periódica, con período N.

La mayor consecuencia producida por la naturaleza periódica de las secuencias en la convolución es que el cambio producido por el índice (n – k) actualmente representa una rotación.

1.2.3. Transformada Rápida de Fourier

La transformada rápida de Fourier tiene gran importancia en una gran variedad de aplicaciones, como ejemplo el procesamiento digital de señales. FFT es la abreviatura usual (de sus siglas en inglés Fast Fourier Transform), y es un eficiente algoritmo que permite calcular la transformada discreta de Fourier y su inversa dados vectores de longitud N por las siguientes ecuaciones:

Ecuación 20

$$X(k) = \sum_{j=1}^{N} x(j)\omega_N^{(j-1)(k-1)}$$

Ecuación 21

$$x(j) = \left(\frac{1}{N}\right)\sum_{k=1}^{N} X(k)\omega_N^{-(j-1)(k-1)} \qquad \textit{donde} \qquad \omega_N = e^{(-2\pi i)/N}$$

es una N-ésima raíz de unidad.

1.2.4. Ventaneo

En el proceso de procesamiento de voz, se asume que la señal es estacionaria en intervalos de tiempo lo suficientemente cortos. Durante la pronunciación de un fonema la señal es cuasiestacionaria, nosotros asumiremos que en trozos de 20-40ms es estacionaria. Lo cual es útil para el análisis de estos "trozos de voz estacionarios", ya que se realizan transformadas de Fourier a cada intervalo. La solución para obtener los intervalos o trozos es por medio del ventaneo, que consiste en multiplicar la señal por una función ventana cuyo valor fuera de un determinado rango es cero.

Es importante analizar el efecto de cada una de las ventanas, ya que permite disminuir los efectos de las discontinuidades, los tipos de ventanas mas conocidas son:

1. Ventana Rectangular: se define como

Ecuación 22

$$w_n = \begin{cases} 1, 0 \le n \le N \\ 0, otherwise \end{cases}$$

Pero hay que considerar las discontinuidades que esto produce en los extremos de la señal, las distorsiones, por otro lado el parámetro de la amplitud se ve siempre un poco afectado.

2. Ventana Hanning: se define como

Ecuación 23

$$vh(k) = 0.5(1 - \cos(2\pi \frac{k}{n+1})), k = 1,...n$$

Esta ventana tiene un efecto en el dominio del tiempo y de la frecuencia. En el dominio del tiempo podemos decir que la ventana hace disminuir la amplitud de la señal cerca de los bordes de la ventana lo cual ayuda a eliminar las discontinuidades.

3. Ventana Hamming: en el análisis de señales de voz la ventana más común es ésta, se define como

Ecuación 24

$$w_n = \begin{cases} 0.54 - 0.46\cos(2\pi n/(N-1)), 0 \le n \le N \\ 0, otherwise \end{cases}$$

La distorsión producida por las discontinuidades se ve atenuada en este tipo de ventana.

4. Ventana Blackman: se define mediante la siguiente ecuación

Ecuación 25

$$w[k] = 0.42 - 0.5\cos(2\pi(\frac{k-1}{n-1})) + 0.08\cos(4\pi(\frac{k-1}{n-1})), k = 1,...,n$$

Los efectos en el dominio del tiempo para esta ventana no son muy diferentes de los anteriores, son los efectos en dominio de la frecuencia con los que hay que tener especial cuidado.

Para notar la diferencia en la eficiencia de cada una de las ventanas se hace la siguiente comparación:

Con respecto al parámetro de amplitud todas las ventanas presentan un pequeño error, aunque este error no es muy notorio. En cuanto a otros parámetros como el ancho de banda del lóbulo principal, se puede decir que las ventanas rectangular y hamming tienen un lóbulo principal muy definido, pero la atenuación de frecuencias parasitas no es tan eficiente como el de otras ventanas. Estas frecuencias parasitas son los lóbulos secundarios. Al comparar, las otras dos ventanas restantes; Hanning y Blackman vamos a encontrar similitud en cuanto a atenuación de frecuencias parasitas, amplitud y ancho de banda del lóbulo principal. En el caso de la ventana Blackman se observa que la aparición de frecuencias parasitas es menor pero tenemos un ancho de banda mayor para cada pulso de frecuencia.

2. RECONOCIMIENTO DE VOZ

2.1 Planteamiento del Problema

El Procesado de voz es el estudio de las señales de voz y las técnicas de procesado de estas señales. Las señales se digitalizan con el propósito de manipular su información, lo cual es llamado procesamiento digital de voz.

El procesamiento digital de voz se puede dividir en varias categorías, la de nuestro interés es el reconocimiento de voz.

En el reconocimiento de voz el problema es identificar las palabras habladas, sin importar el hablante. Bajo este esquema, se pre-procesan las señales de voz, se obtienen las características, y lo que se trata al final es capturar las similitudes entre las palabras habladas.

2.2. Modelado de la voz

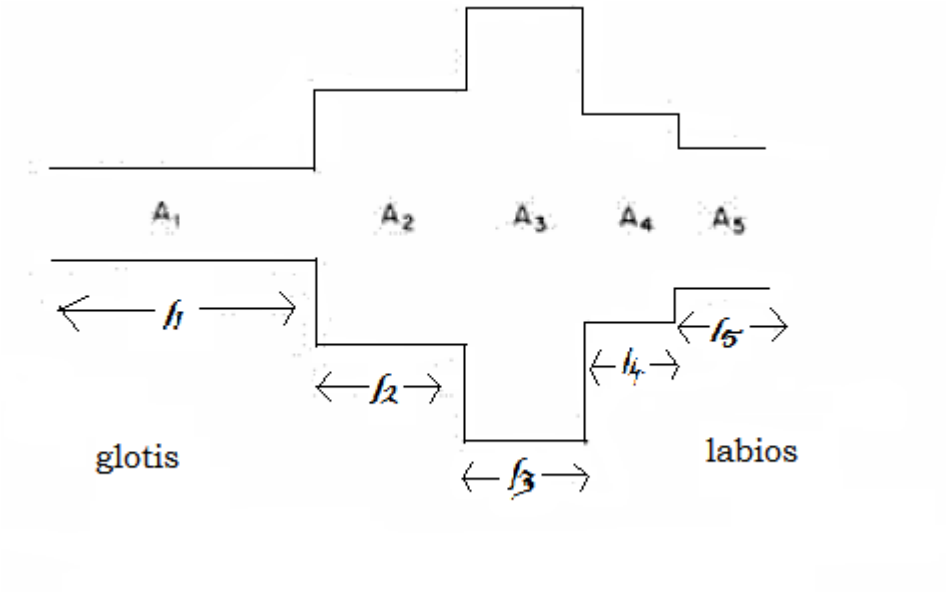
Las ecuaciones fundamentales que se aplican en acústica son lineales, por lo que se pueden utilizar sistemas lineales en el modelado de la voz para conseguir una precisión aceptable. Estos modelos lineales son aproximaciones de gran utilidad ya que utilizar modelos no lineales resulta demasiado complejo.

En resumen, el habla es producida por la modulación del flujo de aire a través del tracto vocal. Por un lado, la tensión de las cuerdas vocales se gobierna por la musculatura, que funciona como un control de entrada.

En este caso, la tensión de las cuerdas afecta la frecuencia de la señal de voz por lo que la señal de control será parecida a la portadora en una modulación. Por otro lado, el tono de voz no es necesario para saber la información que se esta transmitiendo. Generalmente los modelos suelen formarse utilizando un filtro, para separar las partes trascendentales de la señal de voz.

El tracto vocal es modelado como la concatenación de tubos acústicos de distinto diámetro, con o sin pérdidas. Lo cual resulta en un modelo lineal inestacionario, ya que las secciones de los tubos van cambiando de acuerdo al fonema que se esta emitiendo. Se puede decir entonces que, el tracto vocal actúa como una cavidad resonante formando regiones donde el sonido producido es filtrado. La figura muestra lo anterior:

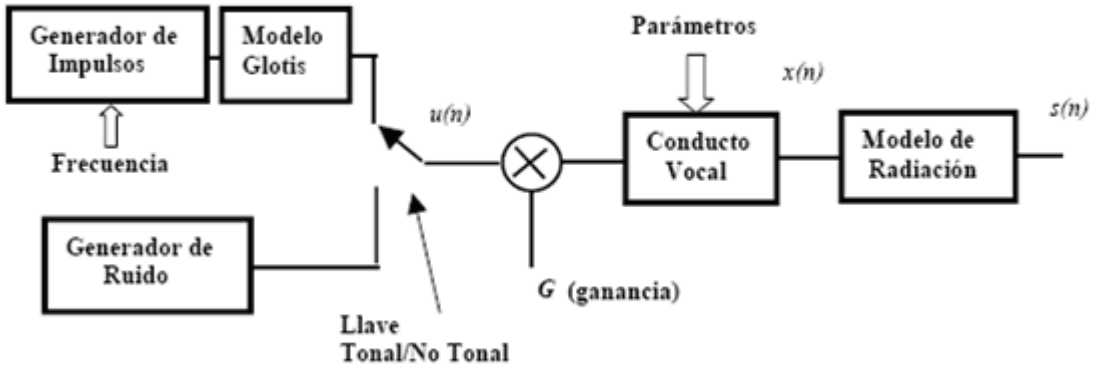
Figura 7. Modelado acústico del tracto vocal



Fuente: www.pas.deusto.es/recursos/

A continuación se muestra el diagrama del modelo en tiempo discreto del sistema de producción de voz:

Figura 8. Modelo de producción de voz



Fuente: http://www.ee.columbia.edu/~dpwe/e6820/lectures/E6820-L05-speechmodels.pdf

El conducto vocal se representa por un sistema lineal (en general inestacionario) que es excitado a través de una llave que selecciona entre una fuente de impulsos cuasi periódicos para el caso de sonidos *tonales*, o una fuente de ruido aleatorio para el caso de sonidos *no tonales*. La ganancia apropiada de la fuente, *G*, es estimada a partir de la señal de voz, y la señal escalada es usada como entrada del modelo del conducto vocal.

Modelo de radiación**:** describe la impedancia de radiación vista por la presión de aire cuando abandona los labios, que puede ser razonablemente aproximada por una ecuación en diferencias de primer orden, o equivalentemente por una función de transferencia de la forma

Ecuación 26

$$R(z) = (1 - z^{-1})$$

Modelo de glotis**:** existen diferentes modelos de la glotis, para el caso en que es excitada por pulsos. Un modelo simple es el denominado *modelo exponencial* representado por una función transferencia *Z* de la forma

Ecuación 27

$$G(z) = \frac{-ae\ln(a)z^{-1}}{(1 - az^{-1})^2}$$

Donde *e* es la base de los logaritmos neperianos. El numerador se selecciona de manera que $g(n) = Z^{-1}\{G(z)\}$ tenga un valor máximo aproximadamente igual a 1. El modelo está inspirado en mediciones de la respuesta de la glotis a impulsos, que se asemejan a la respuesta de un sistema de segundo orden.

Clasificación de los sonidos**:**

Sonidos Sonoros o Tonales: en ellos las cuerdas vocales vibran y el aire pasa a través del tracto vocal sin impedimentos importantes. Además, son de alta energía, la información se encuentra entre los primeros 300Hz a 4kHz y poseen cierta periodicidad. Puede ser modelado matemáticamente como un tren de impulsos.

Sonidos Sordos o no Tonales: en ellos las cuerdas vocales no vibran y existen restricciones importantes al paso del aire, por lo que son de amplitud menor y de naturaleza más ruidosa. Por lo que son de baja energía, componente frecuencial uniforme y cierta aleatoriedad. Matemáticamente son modelables como ruido blanco.

2.3. Obtención de información mediante micrófono

Micrófono: El micrófono es un transductor electroacústico. Su función es la de transformar (traducir) la presión acústica ejercida sobre su capsula por las ondas sonoras en energía eléctrica.

El audio es un fenómeno analógico. Para grabar una señal de voz se hace la conversión de la señal analógica del micrófono en una señal digital por medio del conversor A/D en la tarjeta de sonido. Cuando un micrófono esta operando las ondas de sonido hacen que vibre el elemento magnético del micrófono causando una corriente eléctrica hacia la tarjeta de sonido, donde el conversor A/D básicamente graba los voltajes eléctricos en intervalos específicos.

Hay dos factores importantes durante este proceso. Primero esta la taza de muestreo o que tan seguido los valores de voltaje son grabados. Segundo, son los bits por segundo, o que tan exactamente los valores son grabados. Un tercero podría ser el número de canales (mono o estereo), pero para las aplicaciones de reconocimiento de voz un canal mono es suficiente. La mayoría de aplicaciones vienen con valores pre-determinados, para desarrollo del código se debería de cambiar los parámetros para ver lo que mejor funciona en el algoritmo.

Dado a que el habla es relativamente de bajas frecuencias (entre 100Hz-8kHz), una frecuencia de muestreo de 16000 muestras/seg provee una mayor exactitud en la adquisición de la información, la frecuencia de nyquist.

La obtención de la información mediante micrófono en MATLAB se realiza con la función wavrecord(n,Fs) graba n muestras de la señal de audio, muestreadas a una frecuencia de Fs, utilizamos la frecuencia de 11025 Hz ya que se adapta bien a nuestro algoritmo y no se pierde información. La señal obtenida es de canal mono, que es el valor predeterminado en la función, 1 para mono y 2 para stereo.

2.4. Preprocesado

Convertir la entrada de voz a una forma que el reconocedor pueda procesar o que la señal sea más accesible para procesar luego.

2.4.1. Muestreo y cuantificación

Muestreo consiste en el proceso de conversión de señales continuas a señales discretas en el tiempo, es un paso para digitalizar una señal analógica. Este proceso se realiza midiendo la señal en momentos periódicos del tiempo.

Teorema de nyquist:

Si x[n] es una secuencia de muestras obtenida a partir de una señal continua en el tiempo x(t), por medio de la relación

Ecuación 28

$$x[n] = x(nT), \quad para \ -\infty \le n \le \infty$$

donde T es el período de muestreo, y su reciproco, $f_s = \frac{1}{T}$ es la frecuencia de muestreo, en muestras por segundo. También podemos expresar la frecuencia de muestreo como $\Omega_s = \frac{2\pi}{T}$ en radianes por segundo.

Entonces el teorema de muestreo de nyquist esta definido como: sea x(t) una señal limitada en banda por:

Ecuación 29

$$X(j\Omega) = 0 \qquad \text{para } |\Omega| \ge \Omega_N$$

Entonces x(t) esta únicamente determinada por sus muestras x[n] = x(nT), n = $0, \pm 1, \pm 2, \ldots$ si $\Omega_s = \frac{2\pi}{T} \ge 2\Omega_N$.

La frecuencia Ω_N es comúnmente referida como la frecuencia de Nyquist, y la frecuencia $2\Omega_N$ que tiene que ser excedida por la frecuencia de muestreo es llamada la razón de Nyquist.

Cuantificación:

En la cuantificación el valor de cada muestra de la señal se representa como un valor elegido de entre un conjunto finito de posibles valores.

Se conoce como error de cuantificación (o ruido), a la diferencia entre la señal de entrada (sin cuantificar) y la señal de salida (ya cuantificada), interesa que el ruido sea lo más bajo posible. Para conseguir esto y según sea la aplicación a desarrollar, se pueden usar distintas técnicas de cuantificación:

- Cuantificación uniforme
- Cuantificación logarítmica
- Cuantificación no uniforme
- Cuantificación vectorial

Cuantificación uniforme:

En los cuantificadores uniformes o lineales la distancia entre los niveles de reconstrucción es siempre la misma, la mayoría usan un número de niveles que es una potencia de 2. No hacen ninguna suposición acerca de la señal a cuantificar, de allí que no proporcionen los mejores resultados. Pero son los mas fáciles y menos costosos a implementar.

Cuantificación logarítmica:

Para evitar desperdicio de niveles de reconstrucción y de ancho de banda se utiliza un método sencillo para mejorar el incremento de la distancia entre los niveles de reconstrucción conforme aumenta la amplitud de la señal. Para conseguir esto se hace pasar la señal por un compresor logarítmico antes de la cuantificación. Esta señal comprimida puede ser cuantificada uniformemente. A la salida del sistema la señal pasa por un expansor. A esta técnica se le llama compresión.

Cuantificación no uniforme:

Este cuantificador utiliza la función de la distribución de probabilidad, conociendo esto se puede ajustar los niveles de reconstrucción a la distribución de forma que se minimice el error cuadrático medio.

Cuantificación vectorial:

Este método cuantifica los datos en bloques de N muestras. En este tipo de cuantificación, el bloque de N muestras se trata como un vector N-dimensional.

2.4.2. Eliminación del ruido

La señal digitalizada es escaneada y las zonas de silencio son removidas por medio del cálculo de energía en corto tiempo. Segmentos de 10ms se escogieron para este propósito. En un segmento la energía promedio es menor que un valor umbral proporcional a la energía promedio de la señal entera es descartado. Las siguientes fórmulas se utilizaron:

Ecuación 30

$$En = \sum_{k=1}^{Wn} |x[k]|^2 w[n-k]$$

Ecuación 31

$$E_{avg} = \frac{1}{N} \sum_{k=1}^{N} |x[k]|^2$$

Donde *En* es la energía promedio de cada segmento y E_{avg} es la energía promedio de la señal entera. El valor umbral escogido THRES=0.2.

2.4.3. Filtro de Pre-Énfasis

Se aplica un filtro digital pasa altas de primer orden a la señal, para enfatizar las frecuencias altas de los formantes por dos razones, primero para que no se pierda información durante la segmentación, ya que la mayoría de la información esta contenida en las frecuencias bajas, en segundo remueve la componente DC de la señal, aplanando espectralmente la señal. Uno de los filtros de pre-énfasis más utilizados tiene la ecuación:

Ecuación 32

$$H(z) = (1 - az^{-1})$$

a = 0.95 en nuestro caso.

2.4.4. Segmentación

La segmentación consiste en cortar la señal en segmentos de análisis. La señal de voz es asumida como estacionaria en estos segmentos.

Durante la segmentación los segmentos son guardados cada uno como la columna de una matriz, para el posterior procesamiento de la señal de voz.

Para el proceso una ventana de Hamming de 30ms es aplicada a la señal de voz, enfatizada previamente con el filtro de pre-énfasis. Con un desplazamiento típico 10ms entre cada ventaneo.

Se realiza el algoritmo en base a las siguientes fórmulas:

Ecuación 33

$$Q_n = \sum_{k=-\infty}^{\infty} x[k]w[n-k]$$

Q_n es cada nth cuadro de segmentación.

Ecuación 34

$$w[n] = 0.54 - 0.46\cos\left\{\frac{2\pi(n)}{N}\right\}$$

En la ecuación de la ventana de Hamming, N es el largo de cada cuadro o segmento de análisis.

Para los valores de los parámetros que hacen falta para la implementación del algoritmo, se utiliza la siguiente tabla:

Tabla III. Valores de Parámetros para el reconocimiento de voz

Parámetro	Valor
N número de muestras en el segmento de análisis.	240 (30*ms*)
M número de muestras entre cada segmento.	80 (10*ms*)
p LPC orden de análisis.	10
Q dimensión del vector cepstral derivado del LPC.	15

2.5. Extracción de características

En el reconocimiento del habla, la señal de voz pre-procesada se ingresa a un nuevo procesamiento para producir una representación de la voz en forma de secuencia de vectores o agrupaciones de valores que se denominan parámetros, que deben representar la información contenida en la envolvente del espectro.

Hay que tener en cuenta que el número de parámetros debe ser reducido, para no saturar la base de datos, ya que mientras más parámetros tenga la representación menos fiables son los resultados y mas costosa la implementación.

Existen distintos métodos de análisis para la extracción de características, y se concentran en diferentes aspectos representativos. En este caso analizaremos los dos de mayor importancia para el análisis de la voz:

- Análisis de predicción lineal (LPC)
- Análisis cepstral

2.5.1. Predicción lineal

Se trata de una de las técnicas más potentes de análisis de voz, y uno de los métodos más útiles para codificar voz con buena calidad.

Su función es representar la envolvente espectral de una señal digital de voz en una forma comprimida, utilizando la información de un modelo lineal, con lo cual se proporcionan unas aproximaciones a los parámetros de la voz muy precisas.

Se fundamenta en establecer un modelo de filtro de tipo todo polo, para la fuente de sonido. La principal motivación del modelo todo polo viene dada porque permite describir la función de transferencia de un tubo, que sin perdidas esta formado por diferentes secciones.

El modelo recibe este nombre porque pretende extrapolar el valor de la siguiente muestra de voz s(n) como la suma ponderada de muestras pasadas s(n-1), s(n-2),…, s(n-K):

Ecuación 35

$$s(n) \approx -\sum_{k=1}^{p} \alpha_k s(n-k)$$

Incluyendo un término de excitación *Gu(n),* la ecuación puede escribirse como una igualdad:

Ecuación 36

$$s(n) = -\sum_{k=1}^{p} \alpha_k s(n-k) + Gu(n)$$

Siendo α_k los denominados coeficientes de predicción lineal (LPC), y G, la ganancia de excitación. Por otro lado en el dominio Z la ecuación puede escribirse como:

Ecuación 37

$$S(z) = -\sum_{k=1}^{p} \alpha_k z^{-k} S(z) + GU(z)$$

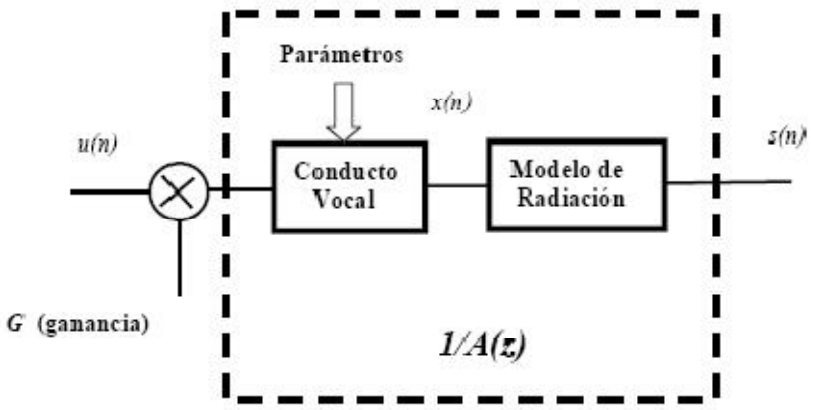
Lo que conduce a una función de transferencia

Ecuación 38

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 + \sum_{k=1}^{p} \alpha_k z^{-k}} = \frac{1}{A(z)},$$

La del tipo todo polo. Una interpretación de esta ecuación, que es una versión simplificada, esta dada en la figura 9.

Figura 9. Modelo de producción de voz basado en LPC



Fuente: www.pas.deusto.es/recursos/

H(z) representa la función transferencia de un modelo lineal del conducto vocal + radiación. Los parámetros del filtro digital *H(z)* son controlados por la señal de voz que está siendo producida y los coeficientes de este filtro son los LPC.

Estimación de los LPC

Una estima (o predicción) de *s*(*n*) basada en *p* muestras anteriores, puede calcularse como $s\hat{}(n) = -\sum_{k=1}^{p} \alpha_k s(n-k)$ y el error de estimación (predicción) puede entonces definirse como $\varepsilon(n) = s(n) - s\hat{}(n)$, resultando el error de predicción:

Ecuación 39

$$\varepsilon(n) = s(n) + \sum_{k=1}^{p} \alpha_k s(n-k)$$

Los LPC se obtienen minimizando un criterio cuadrático en los errores de predicción, para cada cuadro en que es dividido el segmento de voz.

Suponiendo que en cada cuadro hay $m+1 >> p$ muestras, y definiendo lo siguiente:

Ecuación 40

$$\alpha = [\ \alpha_1 \quad \alpha_2 \quad \ldots \quad \alpha_p\ ]^T$$

$$\Phi^T(n) = [\text{-}s(n\text{-}1) \quad \text{-}s(n\text{-}2) \ldots \quad \text{-}s(n\text{-}p)]$$

la ecuación $s\hat{}(n) = -\sum_{k=1}^{p} \alpha_k s(n-k)$ puede escribirse matricialmente como,

Ecuación 41

$$\begin{bmatrix} s(n) \\ s(n+1) \\ \vdots \\ s(n+m) \end{bmatrix} = \begin{bmatrix} -s(n-1) & \cdots & -s(n-p) \\ -s(n) & \cdots & -s(n-p+1) \\ \vdots & \vdots & \vdots \\ -s(n+m-1) & \cdots & -s(n+m-1-p) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}$$

Donde el vector α son los coeficientes; también se puede escribir lo anterior como,

Ecuación 42

$$S_m(n) = \Phi^T(n)\alpha.$$

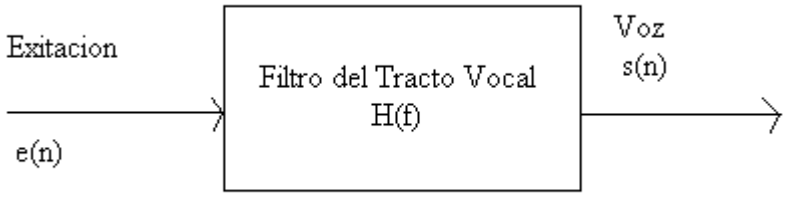
2.5.2. Cepstrum

Como se sabe los sonidos de la voz se pueden representar mediante un espectrograma, que indica las componentes frecuenciales de la señal de voz. Es así entonces como el espectro nos proporciona información acerca de los parámetros del modelo de producción de voz, tanto de la excitación como del filtro que representa el tracto vocal.

Desde el principio de la década de los 70 los sistemas homo mórficos han tenido una gran importancia en los sistemas de reconocimiento de voz. Estos sistemas homo mórficos son una clase de sistemas no lineales que obedecen a un principio de superposición. De estos los sistemas lineales son un caso especial.

La razón para realizar un procesado homomórfico del habla se resume en la figura 10.

Figura 10. Modelo de la técnica Homomórfica



La señal de voz s(n) se descompone en una parte de excitación e(n) y en un filtro lineal H($e^{j\theta}$), como se menciono anteriormente. Así, en el dominio de la frecuencia tenemos $S(e^{j\theta}) = H(e^{j\theta})E(e^{j\theta})$.

En el dominio logarítmico, por su parte, las dos componentes anteriores pueden separarse empleando técnicas convencionales del procesamiento de señal.

Eso se logra del siguiente modo:

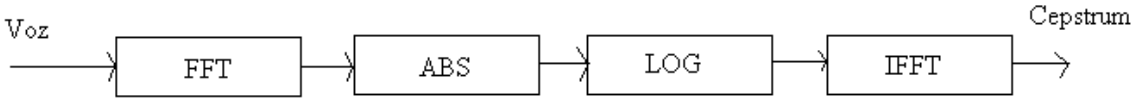
Ecuación 43

$$\log(|S(e^{j\theta})|) = \log(|H(e^{j\theta})|) + \log(|E(e^{j\theta})|)$$

Para la mayoría de aplicaciones de voz solamente necesitamos la amplitud espectral.

El proceso anterior se puede describir con un diagrama (véase figura 11).

Figura 11. Modelo Coeficientes Cesptrales



En la salida de este sistema tenemos entonces:

Ecuación 44

$$c(n) = \frac{1}{N_s} \sum_{k=0}^{N_s-1} \log\left|S_{med}(k)\right| e^{j\frac{2\pi}{N_s}kn} \qquad para \;\; 0 \le n \le N_s - 1$$

En cual caso, el valor c(n) se conoce como coeficientes cepstrales derivados de la transformada de Fourier. N_s es el número de puntos con que se calcula la transformada. Esta ecuación puede ser convenientemente simplificada teniendo en cuenta que el espectro logarítmico es una función real simétrica.

Ecuación 45

$$c(n) = \frac{2}{N_s} \sum_{k=1}^{N_s} S_{med}(I(k)) \cos(\frac{2\pi}{N_s} kn)$$

En los cálculos lo habitual es usar solamente los primeros términos ($n \le 20$). Por otro lado, *I(k)* representa una función que traduce la posición de un valor en frecuencia al intervalo donde este contenido.

Es posible, a la hora de calcular un coeficiente cepstral, transformar el espectro utilizando bandas definidas según escalas de Mel. En cual caso este tipo de parámetro se conoce como coeficientes cepstrales con frecuencia en escala de Mel (*MFCC*).

Partiendo del análisis de predicción lineal también es posible obtener la expresión de los coeficientes cepstrales asociados:

Ecuación 46

$$c(0) = \log(1) = 0$$

Ecuación 47

$$c(i) = -\alpha(i) - \sum_{j=1}^{i-1}(1 - \frac{j}{i})\alpha(j)c(i-j) \qquad 1 \le i \le N_c$$

En el sistema de reconocimiento de voz en MATLAB existe una función para obtener los coeficientes cepstrales utilizando *la FFT* . La función utilizada es la *rceps*, que nos proporciona el cepstrum real de la función ingresada, por medio del algoritmo mostrado en la figura 10. O sea que es la implementación del algoritmo mostrado anteriormente. La razón principal para utilizar los coeficientes cepstrales es que tienen la ventaja adicional que uno puede derivar de ellos una serie de parámetros que son invariantes sin importar las distorsiones que puedan ser introducidas por el micrófono o por cualquier sistema de transmisión.

Características:

Por último, los coeficientes son normalizados para reducir variabilidades espectrales durante largos periodos de tiempo. Los coeficientes son expandidos por medio de una representación polinomial ortogonal durante intervalos de 90ms cada 10ms. Este intervalo es adecuado para preservar información de transición entre fonemas. Solamente los dos primeros coeficientes ortogonales polinomiales son utilizados. Las siguientes ecuaciones se utilizan en el algoritmo:

Ecuación 48

$$P_{0j} = 1$$
$$P_{1j} = j - 5$$

Los primeros dos coeficientes de la representación ortogonal polinomial son:

Ecuación 49

$$a = \frac{(\sum_{j}^{9} xj)}{9} \qquad b = \frac{(\sum_{j}^{9} x_j P_{1j})}{(\sum_{j}^{9} P_{1j}^2)}$$

Los coeficientes a y b representan el promedio, de la función de tiempo de cada coeficiente cepstral en cada segmento respectivamente. Dicha representación es una función del tiempo de los coeficientes cepstrales $x_t(i)$ y los coeficientes polinomiales de primer orden que están representados por $b_t(i)$, donde *t* es el número de segmento e *i* es el índice de los coeficientes cepstrales. Como el valor de p es escogido como 10, la representación resultante es una función del tiempo de 20 elementos de características.

2.6 Medida de distancia

Una característica fundamental de los sistemas de reconocimiento es la forma en que los vectores característicos son combinados y comparados con los patrones de referencia.

Para poder realizar estas operaciones es necesario definir una medida de distancia entre los vectores característicos. Algunas de las medidas de distancia mas utilizadas son las distancias o métricas inducidas por las normas en espacios Lp.

En el algoritmo de reconocimiento en MATLAB se utiliza una distancia Euclidea, definida del siguiente modo: por ejemplo si f_i y f_i', con $i=0, 1, 2,\ldots, D$ son las componentes de dos vectores característicos f y f', puede definirse la siguiente métrica inducida por la norma Lp:

Ecuación 50

$$d = \sqrt{\sum_{i=1}^{D} \left| f_i - f_i' \right|^2}$$

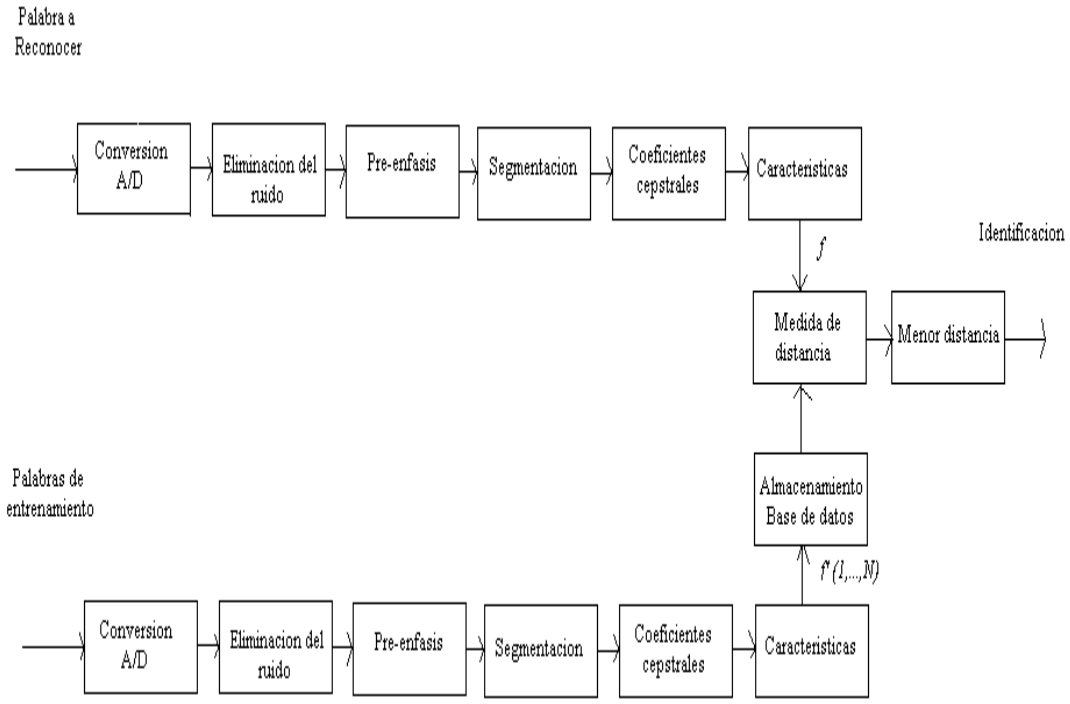
En el algoritmo primero se define el tamaño del mayor vector, y se calcula con la formula anterior la distancia entre el vector de la palabra a reconocer y cada uno de los vectores de referencia en la base de datos, luego se hacen las condiciones para obtener la menor distancia, con lo cual se encuentra la palabra identificada en la base de datos.

3. PROGRAMA DE RECONOCIMIENTO DE VOZ

3.1. Diagrama de bloques del sistema

El sistema de reconocimiento de voz se puede resumir en el siguiente diagrama esquemático (véase figura 12).

Figura 12. Diagrama esquemático del Sistema de Reconocimiento de Voz



3.2. Resultados de evaluación

La base de datos de entrenamiento consiste en muestras almacenadas en un archivo .mat de cuatro nombres distintos, todas estas muestreas fueron tomadas con un mismo hablante. La base de datos se almacena una vez y puede ser modificada o ampliada según sea la necesidad ya que posee su propio M-file. Las señales fueron tomadas utilizando un micrófono de bajo costo.

El programa implementado en MATLAB utiliza las herramientas de procesamiento de señales así como sus funciones.

Para evaluar el sistema, se obtiene una señal de muestra y se sacan las características de esta señal para ser comparada con cada una de las características almacenadas en la base de datos. Para esta comparación se utiliza una medida de distancia Euclidiana. La menor medida de distancia representa la de mayor similitud.

En conclusión se presento una introducción al problema de reconocimiento de voz. Se discutieron los problemas del modelado del habla y de la comparación de patrones. Se presenta una implementación en MATLAB de un sistema identificador de palabras utilizando algoritmos para los coeficientes Cepstrales y las medidas de distancia; además que se explica tal implementación para futuras modificaciones.

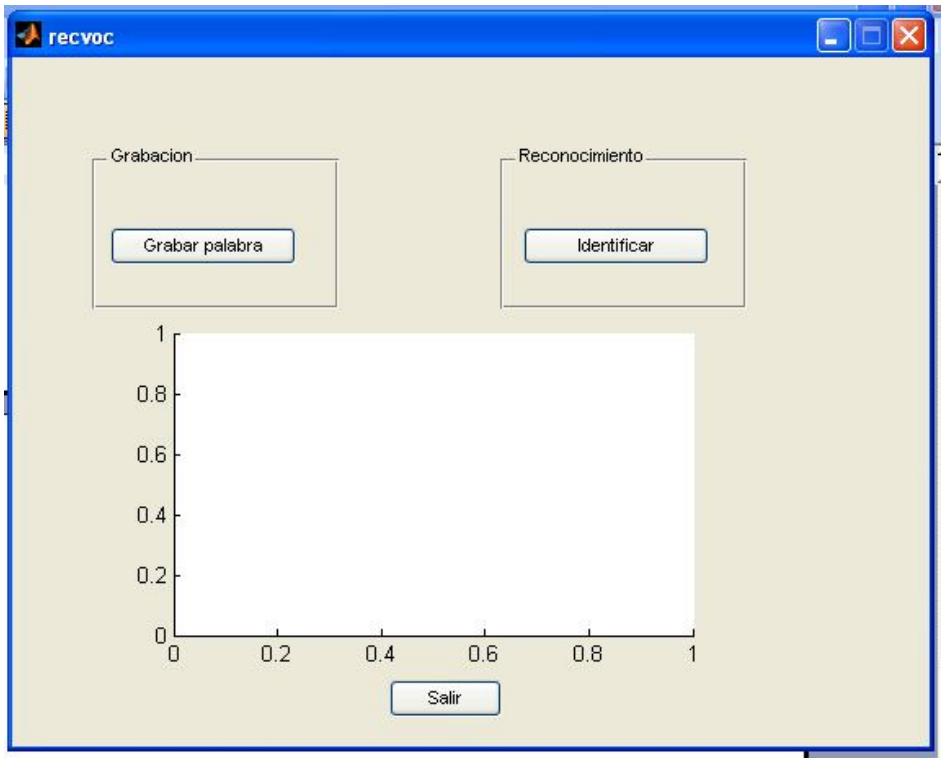
3.3. Programa

El programa tiene una interfaz grafica, muy fácil de utilizar. Con una base de datos de prueba de 4 nombres, esta base de datos puede ser aumentada en cualquier cantidad. Los nombres en la base de datos son:

- Geno
- Flavio
- Rosa
- Andres

El entorno gráfico es como se muestra (véase figura 13).

Figura 13. Interfaz gráfica

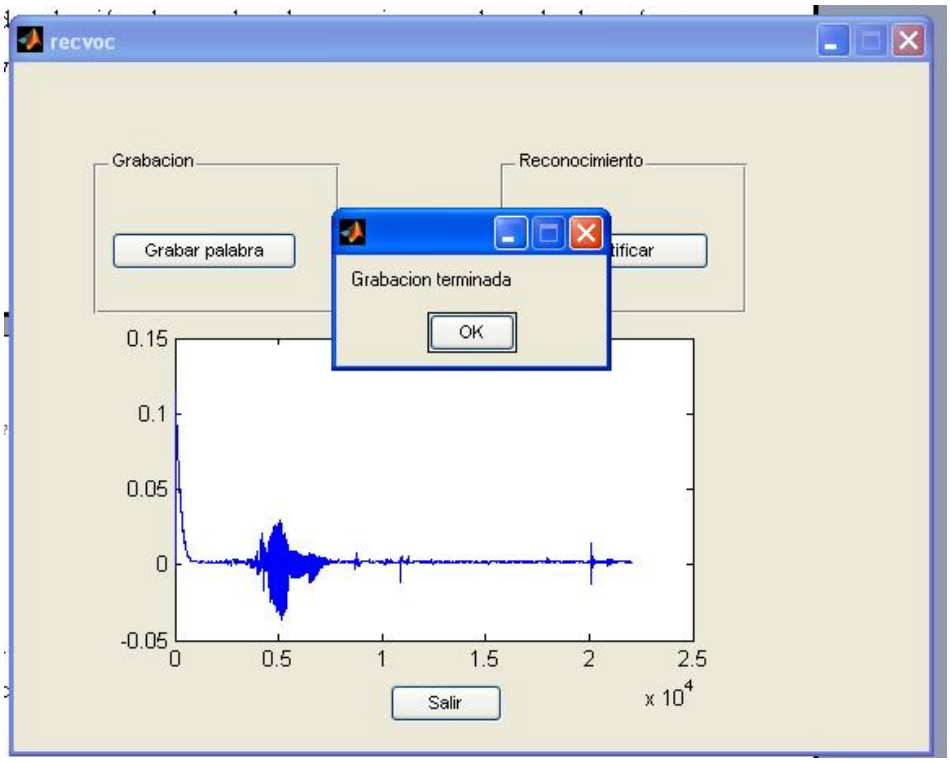


Consta de 3 cuadros de selección, dos cuadros de mensaje y un plano donde serán graficadas las señales de voz en tiempo continuo y el espectro de frecuencias de la señal discreta ya procesada.

Grabación

En este cuadro nos aparece la opción de *grabar la palabra* para ingresarla al reconocedor de voz, cuando la palabra es grabada aparece un cuadro de mensaje indicando que la grabación ha sido terminada. Por otra parte en el plano aparece una gráfica en el tiempo (eje X) y en amplitud (eje Y), de la señal de voz grabada por el usuario (véase figura 14).

Figura 14. Entorno gráfico modo de grabación



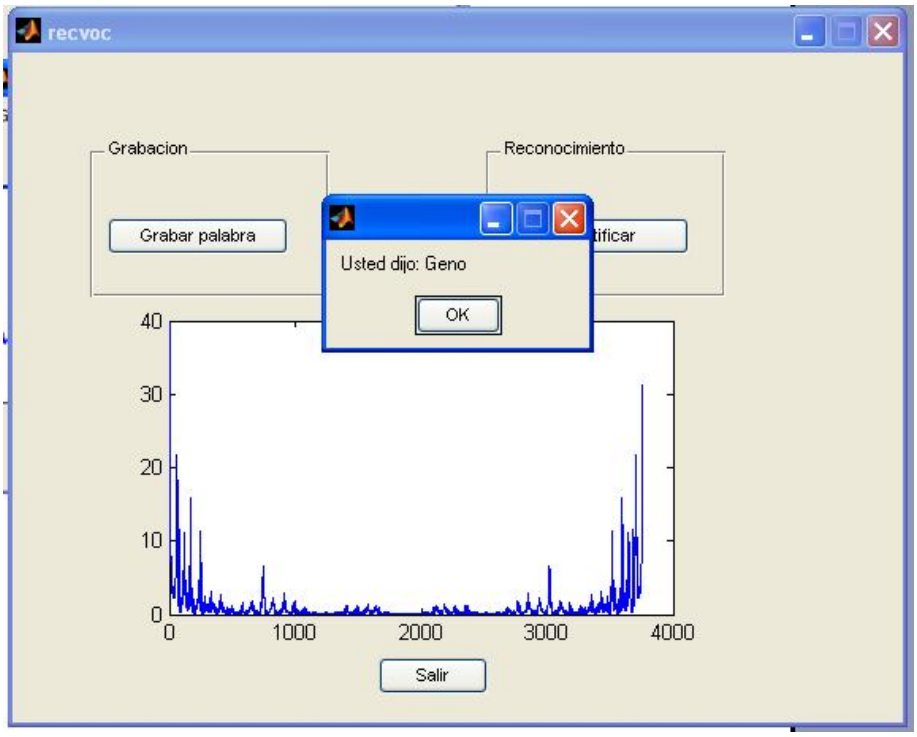
Salir

Este botón nos permite salir completamente del Sistema de Reconocimiento de Voz.

Reconocimiento

El cuadro de reconocimiento brinda los resultados del sistema, con un cuadro de mensaje indicando cual palabra ha sido reconocida y la palabra grabada es reproducida. Además, nos da una representación gráfica en el plano con la información del espectro de frecuencias de la señal del usuario reconocido (véase figura 14).

Figura 15. Entorno gráfico modo de reconocimiento



CONCLUSIONES

1. El reconocimiento de voz es una de las aplicaciones del procesamiento digital de señales que permite interacción entre seres humanos y computadoras.

2. Con la herramienta MATLAB se reduce la complejidad del procesamiento digital de señales.

3. El espectro de la señal brinda la información relevante de las señales de voz.

4. Es necesario filtrar las señales de voz para enfatizar las características acústicas, llamadas formantes.

5. Las medidas de distancia euclidianas son más sencillas y eficaces de implementar en el sistema.

6. El sistema para reconocer palabras es muy simple de modificar, para otras aplicaciones.

RECOMENDACIONES

1. Investigar en Guatemala que tan factible es la utilización del Sistema de Reconocimiento de Voz en la industria de sistemas de seguridad.

2. Realizar pruebas del Sistema de Reconocimiento de Voz con un mayor número de personas en la base de datos, para ver si se mejora el rendimiento del programa.

3. Estudiar el potencial del Sistema de Reconocimiento de Voz como la base para distintas aplicaciones, o sea la posibilidad de ser modificado e innovado para futuras investigaciones.

4. Investigar con respecto al rendimiento del programa, distintos métodos de extracción de características y medidas de distancia en el procesamiento digital de señales de voz.

BIBLIOGRAFÍA

1. Oppenheim, Alan V., Schafer, R.W., Buck, J.R. **Discrete-time signal processing.** 2ª edición. Prentice-Hall, Inc. New Jersey, 1989

2. Kuc, Roman. **Introduction to digital signal processing.** Prentice-Hall
   Singapore, 1982

3. Taub, H., Schilling, D.J. **Principles Of Communication Systems.** 2ª
   edición. McGraw-Hill, Singapore, 1986

4. Miyara, Federico. **La Voz Humana.** UNR Editora, Rosario. Argentina.

BIBLIOGRAFÍA ELECTRÓNICA

1. Lopez, E. y Hernandez L. 2006. **Laboratorio de Tratamiento Digital de Voz.** Disponible en: http://www.gaps.ssr.upm.es/TDV/prelaiinfr.html

2. Gomez, J.C. **Procesamiento digital de Señales de Voz.** Apuntes: Modelos de producción de Voz. http://www.eie.fceia.unr.edu.ar/%7Eprodivoz/apuntes_index.html

3. Gomez, J.C. **Apuntes: Medidas de distancia y Reconocimiento de Vocales.** Disponible en: http://www.eie.fceia.unr.edu.ar/%7Eprodivoz/apuntes_index/pdf/

4. MATLAB 7.0. Help

5. **Micrófono** http://es.wikipedia.org/wiki/Micr%C3%B3fono

6. **Formante** http://es.wikipedia.org/wiki/Formante

7. **Modelado de la señal de Voz.** Grupo PAS. Universidad de Deusto, España http://www.pas.deusto.es/recursos/Modelado%20de%20la%20Se%C3%B1al%20de_%20la%20Voz%20-%20Grupo%20PAS.ppt.

8. **The Speaker Recognition Homepage, Algorithms** http://www.speaker-recognition.org/

9. **Speech Modeling** http://www.ee.columbia.edu/~dpwe/e6820/lectures/E6820-L05-speechmodels.pdf

10. **Linear Predictive Coding** http://en.wikipedia.org/wiki/Linear_predictive_coding

11. **Cepstrum**
http://en.wikipedia.org/wiki/Cepstrum

12. **Los elementos Suprasegmentales.** Joaquim Llisterri
http://liceu.uab.es/~joaquim/phonetics/fon_prosod/suprasegmentales.html

13. **Efecto de enventanado en la obtención del espectro discreto de una señal .** Diego Alvarado
http://www.monografias.com/trabajos20/enventanado/enventanado.shtml

14. **Analisis de Fourier**
http://www.euskalnet.net/iosus/speech/fourier.html

15. **Fast Fourier Transform**
http://www.euskalnet.net/iosus/speech/fourier.html

16. **Inside Speech Recognition**
http://www.tldp.org/HOWTO/Speech-Recognition-HOWTO/inside.html

17. **Mathworks**
www.mathworks.com

APÉNDICE

CÓDIGO UTILIZADO

```
%                BASE DE DATOS

clc
Fs=11025;
y0=wavrecord(1*Fs,Fs,1)';
soundsc(y0);
word0=['geno'];
y0=chop_silencio(y0);
z0=enfasis(y0);
seg0=segmentos(z0);
c0=cepstrum(seg0);
f0=features(c0);
save geno.mat word0 f0 y0;

y1=wavrecord(1*Fs,Fs,1)';
soundsc(y1);
word1=['rosa'];
y1=chop_silencio(y1);
z1=enfasis(y1);
seg1=segmentos(z1);
c1=cepstrum(seg1);
f1=features(c1);
save rosa.mat word1 f1 y1;

y2=wavrecord(1*Fs,Fs,1)';
soundsc(y2);
word2=['flavio'];
y2=chop_silencio(y2);
z2=enfasis(y2);
seg2=segmentos(z2);
c2=cepstrum(seg2);
```

```
f2=features(c2);
save flavio.mat word2 f2 y2;

y3=wavrecord(1*Fs,Fs,1)';
soundsc(y3);
word3=['andres'];
y3=chop_silencio(y3);
z3=enfasis(y3);
seg3=segmentos(z3);
c3=cepstrum(seg3);
f3=features(c3);
save andres.mat word3 f3 y3;

%                   FUNCIÓN ELIMINACIÓN DEL RUIDO
%chop_silencio
%Corta el silencio en la señal completa
%y=chop_silencio(s)


function y = chop_silencio(s)
   len   = length(s);    % length del vector
   avg_e = sum(s.*s)/len;  %promedio señal entera
   THRES = 0.2;

   y     = [0];
   for i = 1:80:len-80         % cada 10ms
      seg = s(i:i+79);        % segmentos
      e   = sum(seg.*seg)/80; % promedio de cada segmento
      if( e> THRES*avg_e)     % si el promedio energetico es mayor que la señal
%completa por el valor umbral
         y=[y,seg(1:end)];    % almacena en y sino es eliminado como espacio en
blanco
      end;
   end;

%                   FUNCIÓN FILTRO DE PRE-ÉNFASIS
%enfasis
%filtro de enfasis para la señal
%enfasis(s)


function [y]=enfasis(x)
   b=[1 -0.95];
   y=filter(b,1,x);
```

```
%                FUNCIÓN SEGMENTACIÓN
%segmentos
%segmenta la señal en tramas de 30ms
%[segs]=segmentos(y)


function [segs]=segmentos(y)
   len          =  length(y);  %longitud del vector
   num_segments      =     floor(len/80) -2;     % redondeamos el numero de
segmentos
   segs          =       zeros(num_segments,240);        % matriz de segmentos
num_segments x 240
   win  =  hamming(240)';  % ventana hamming de 240 puntos(30ms)
win=coeficients (numero de muestras en el analisis )
   for i  =    0:num_segments-1
      inicio                =  i*80+1;
      segs(i+1,1:240)     =      (y(inicio:inicio+239).*win);  %multiplicamos  por  la
ventana %hamming
   end;

%                FUNCIÓN CEPSTRUM
%
%                (EXTRACCIÓN DE LOS COEFICIENTES CEPSTRALES)
%cepstrum
%Encuentra los coeficientes cepstrales de cada segmento
%[cc]=cepstrum(seg)


function [c] = cepstrum(segs)

%programa para encontral los coeficientes cepstrales de cada segmento
[M,N]   =   size(segs);      % M filas, N columas, para definir Mo=output frame
size(# de %frecuencias a las que se aplica FFT)
[c]     =    zeros(M,10);    % inicialiando el vector para guardar los coeficientes
cepstrales. P=10

%calculando  los coeficientes cepstrales reales
for i=1:M
%    f     = fft(segs(i,:));
%    m     = abs(f);
%    l     = log(m+1e-5);
%    in    = real(ifft(l));
%    c(i,:)= in(1:10); o resumidamente
```

```
      r     = rceps(segs(i,:));
     c(i,:)= r(1:10);
end
%
% normalizacion de los coeficientes cepstrales
sum_c     = sum(c);
avg_c     = sum_c/M;
%
%sustrayendole el minimo, solo las que estan arriba del promedio los
%fonemas
for i=1:M
   c(i,:)= c(i,:)-avg_c;   % smoothened spectrum
end

%                    FUNCIÓN FEATURES ( DE CARACTERISTICAS)
%features
%Encuentra los coeficientes polinomiales de la expansion correspondientes al
los
%coeficientes cepstrum
%[ftr]=features(c)


function ftr=features(c)
   [M,N]   =   size(c);
   ftr     =   zeros(M-8,20);

   j       =   (1:9); %col vector
   p1      =   repmat((j-5)',1,10);    %crea una matriz llena de P1j=j-5
   sum_p1  =   sum((j-5).*(j-5));      % suma total

   for i = 1:M-8
      %caracteristicas de orden cero (coeficientes cepstrales)
      ftr(i,1:10)=c(i,1:10);
      %caracteristicas de primer orden
      b = zeros(1,10);
      for j = 0:8       % aqui se sacan los coeficientes polinomiales para cada
segmento
         b = b+ p1(j+1,:).*c(i+j,:);   % numerador de la ecuacion
      end;
      ftr(i,11:20)=b/sum_p1;             % dividiendolo en el denominador
   end;                                  %se obtienen los valores de b
```

```
%                    FUNCIÓN DE DISTANCIA
%distancia
%
%[dist]=distancia(ftr_a,ftr_b)

function tot_dist = spv_dis(ftr_a,ftr_b)
[m_a,n_a]=size(ftr_a);
[m_b,n_b]=size(ftr_b);

%establece the guide and slave
if(m_a < m_b )
   guide = ftr_a;
   slave = ftr_b;
   m     = m_b;
   n     = m_a;
else
   guide = ftr_b;
   slave = ftr_a;
   m     = m_a;
   n     = m_b;
end;

%computa la matriz de distancia
dist    = zeros(m,n);
for i = 1:m
   for j = 1:n
      dist(i,j) = sqrt(sum(((guide(j,:)-slave(i,:)).*(guide(j,:)-slave(i,:))).^2));
   end;
end;

%             INICIALIZACION

function varargout = recvoc(varargin)
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                   'gui_Singleton',  gui_Singleton, ...
                   'gui_OpeningFcn', @recvoc_OpeningFcn, ...
                   'gui_OutputFcn',  @recvoc_OutputFcn, ...
                   'gui_LayoutFcn',  [] , ...
                   'gui_Callback',   []);
if nargin && ischar(varargin{1})
   gui_State.gui_Callback = str2func(varargin{1});
end
```

```
if nargout
   [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
   gui_mainfcn(gui_State, varargin{:});
end

%++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
++
%                   PROGRAMA PRINCIPAL

function recvoc_OpeningFcn(hObject, eventdata, handles, varargin)
load geno.mat;
load rosa.mat;
load flavio.mat;
load andres.mat;

load 'handel';
wavwrite(y,'handel');

rec=reconociendovoz(y0);

handles.voz_grabada=y;
handles.feature_datos=rec
handles.feature_f0=f0;
handles.feature_f1=f1;
handles.feature_f2=f2;
handles.feature_f3=f3;

handles.feature_y0=y0;
handles.feature_y1=y1;
handles.feature_y2=y2;
handles.feature_y3=y3;
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);
% --- Outputs from this function are returned to the command line.
function varargout = recvoc_OutputFcn(hObject, eventdata, handles)
% varargout  cell array for returning output args (see VARARGOUT);
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Get default command line output from handles structure
```

```
varargout{1} = handles.output;


% --- Executes on button press in Grabar.
function Grabar_Callback(hObject, eventdata, handles)
% hObject    handle to Grabar (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% obtener la palabra del microfono

Fs = 11025;   % Sampling Frequency (Hz)
n = 20;
Nseconds = 2; % largo de la señal de voz
y  = wavrecord(Nseconds*Fs, Fs, 1)';
rec1=reconociendovoz(y);

handles.voz_grabada = y;
handles.feature_datos=rec1
plot(y);
msgbox('Grabacion terminada');

guidata(hObject, handles);
%load rec.mat;
%handles.feature_voz=f;
% --- Executes on button press in identificar.
function identificar_Callback(hObject, eventdata, handles)
% hObject    handle to identificar (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

rec1=handles.feature_datos;

y0=handles.feature_y0;
y1=handles.feature_y1;
y2=handles.feature_y2;
y3=handles.feature_y3;


f0=handles.feature_f0;
f1=handles.feature_f1;
f2=handles.feature_f2;
f3=handles.feature_f3;
```

```
d1=distancia(f0,rec1)
d2=distancia(f1,rec1)
d3=distancia(f2,rec1)
d4=distancia(f3,rec1)

if (d1 < d2) & (d1 < d3) & (d1 < d4)
   plot(abs(fft(y0)));
   soundsc(y0);
   msgbox('Usted dijo: Geno');
       else if (d2 < d1) & (d2 < d3)& (d2 < d4)
          plot(abs(fft(y1)));
          soundsc(y1);
          msgbox('Identificado: Rosa')
         else if (d3 < d1) & (d3 < d2) & (d3 < d4)
          plot(abs(fft(y2)));
          soundsc(y2);
          msgbox('Identificado: Flavio');
           else if (d4 < d1) & (d4 < d2) & (d4 < d3)
             plot(abs(fft(y3)));
             soundsc(y3);
             msgbox('Identificado: Andres');
             end
          end
   end
end

% --- Executes on button press in pushbutton3.
function pushbutton3_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton3 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
exit;

%                   FUNCIÓN RECONOCIENDOVOZ
function rec=reconociendovoz(y)
% load data
y=chop_silencio(y);
z=enfasis(y);
seg=segmentos(z);
c=cepstrum(seg);
f=features(c);
rec=f;

save('rec.mat','f','y');
```