Universidad Nacional de Trujillo

RECONOCIMIENTO AUTOMÁTICO DEL HABLA USANDO WAVELETS

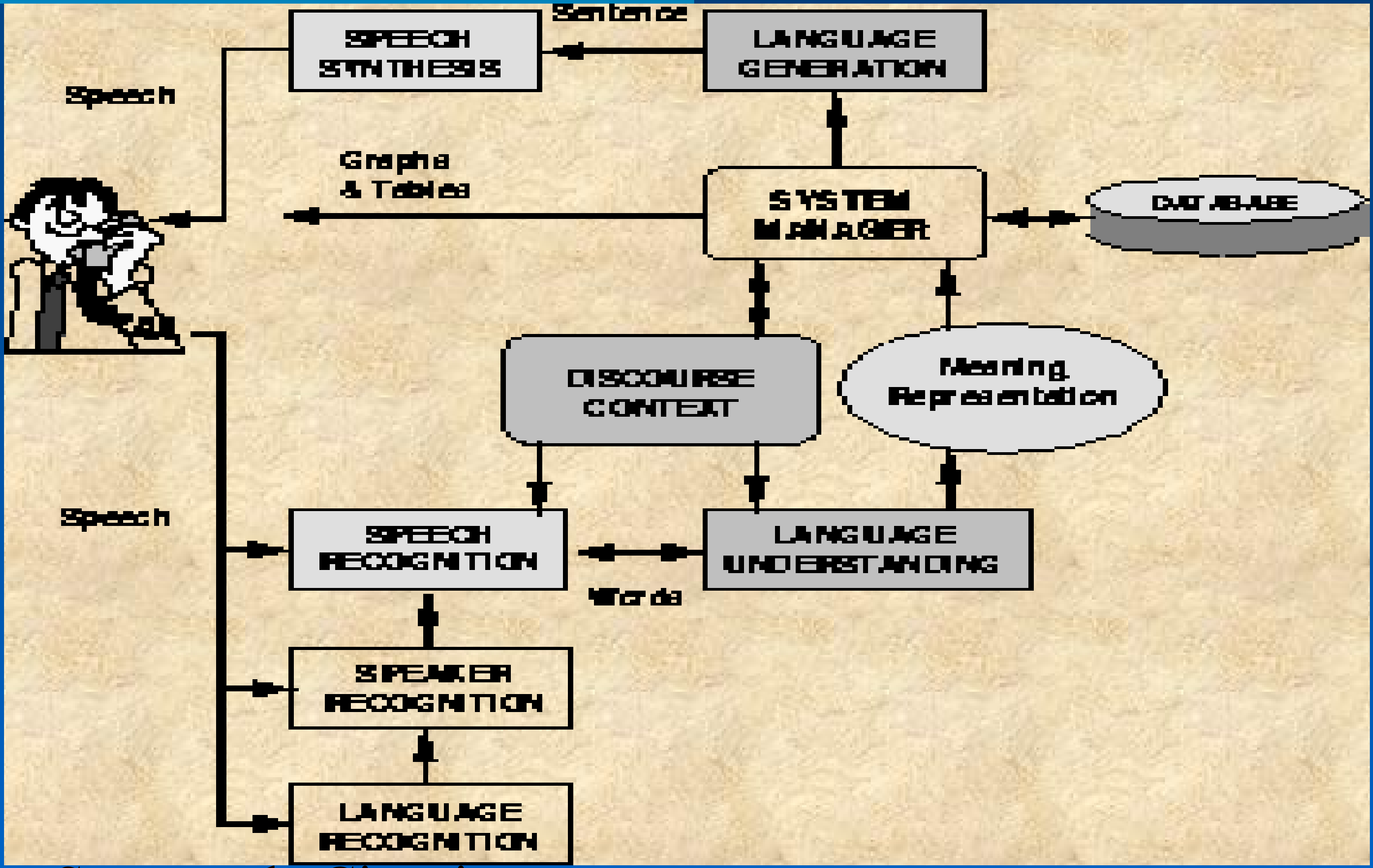
Guevara Díaz Jorge Luis (Informática)
Salazar Campos Juan Orlando (Informática)

Facultad de Ciencias Físicas y Matemáticas

Esc. de Informática

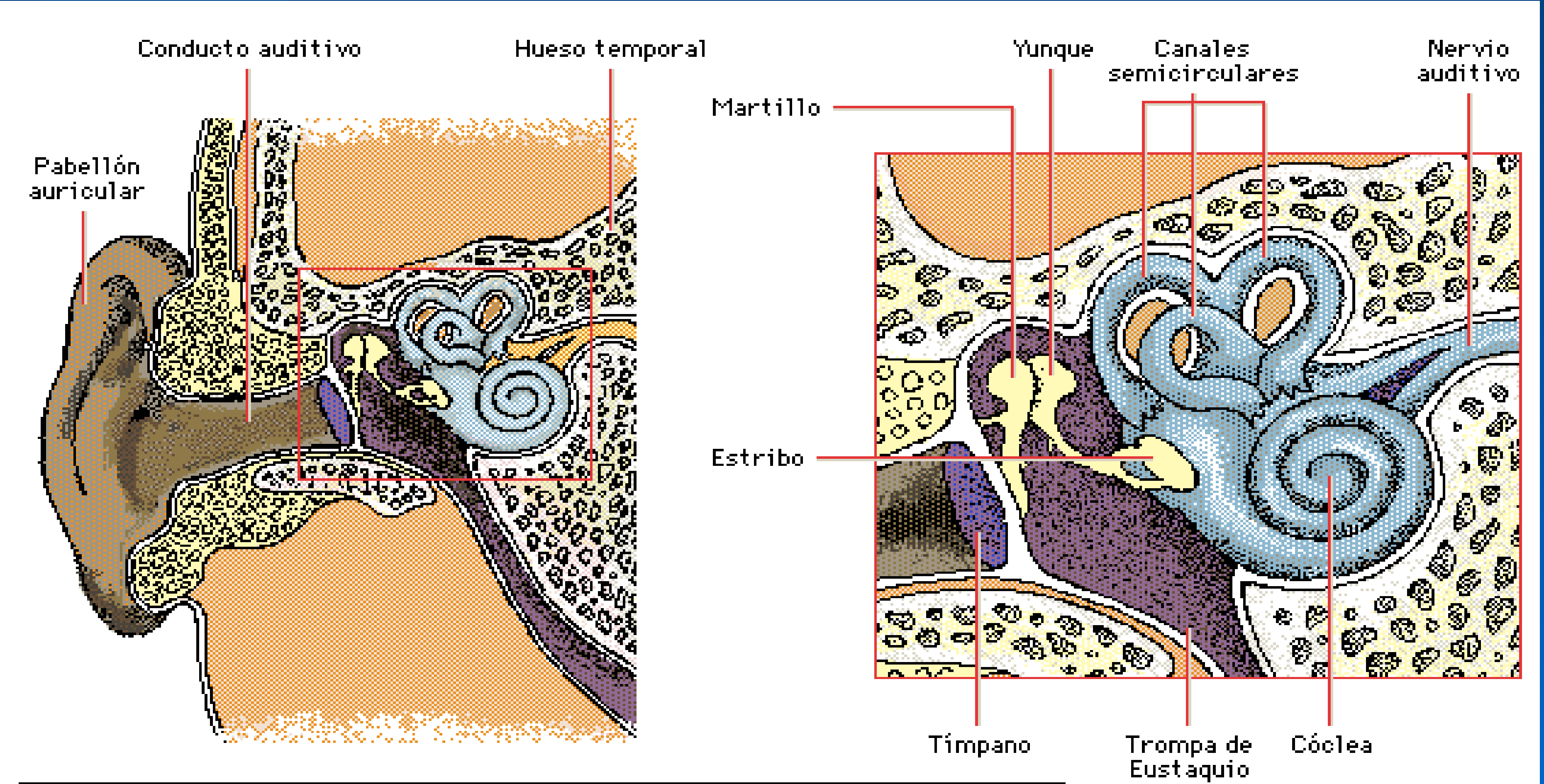
Tercera Semana de Ciencia y

Reconocimiento del Habla utilizando Wavelets.

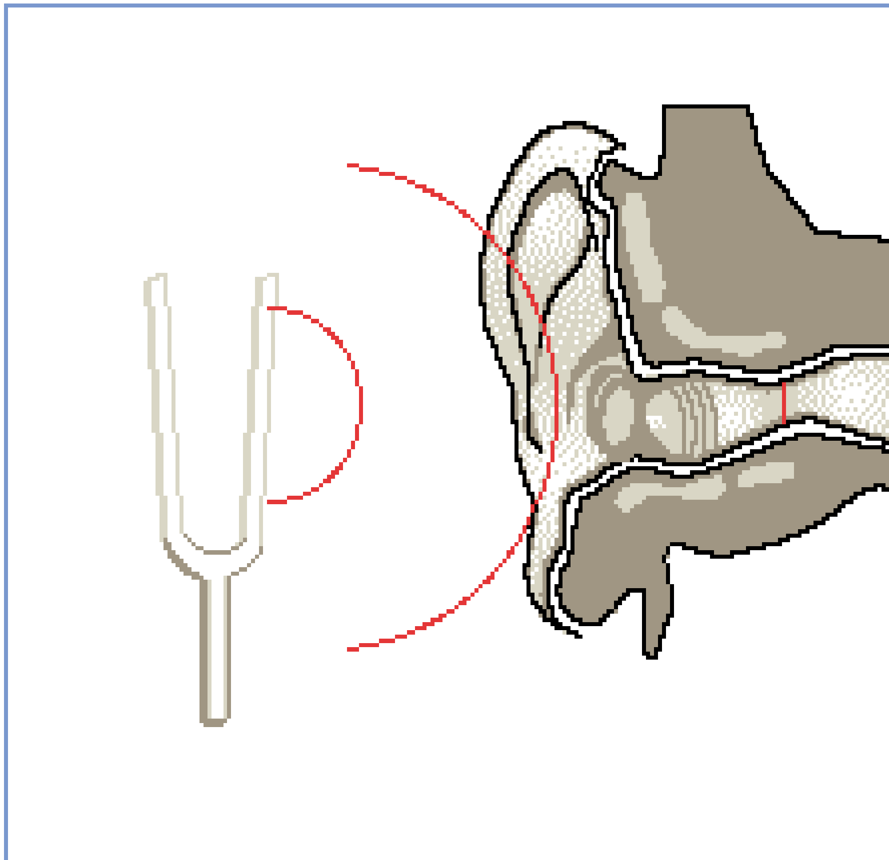


Reconocimiento del Habla utilizando Wavelets.

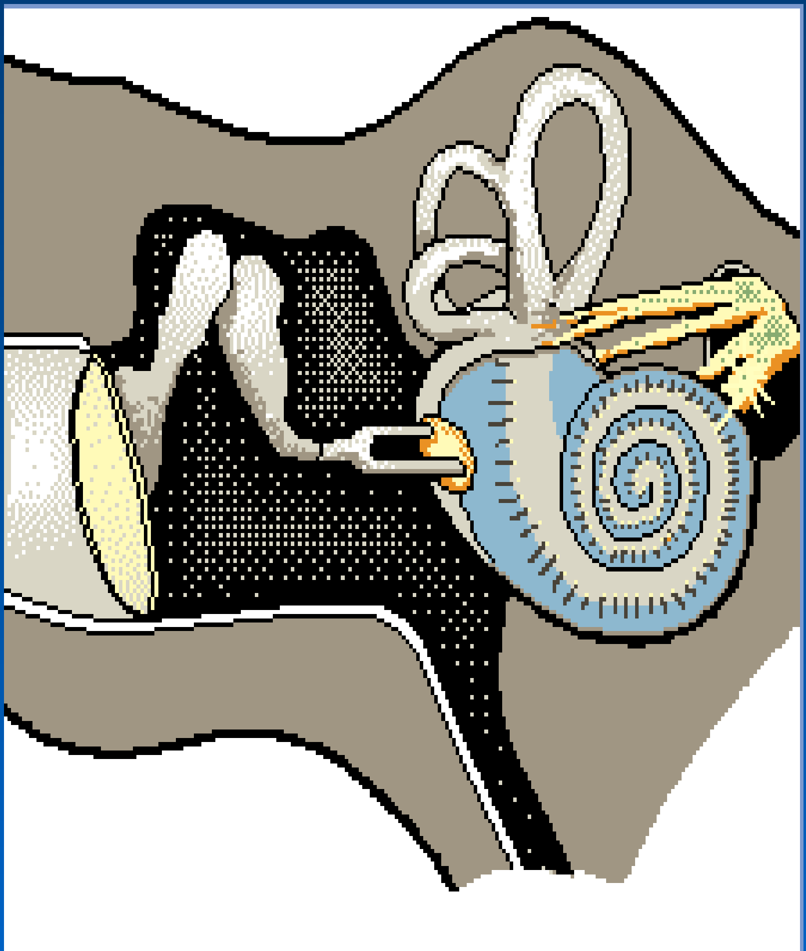
- **<u>Percepción del habla</u>:** OIDO



Reconocimiento del Habla utilizando Wavelets.



Los sonidos entran en el oído



Generalidades de Wavelets

- Una wavelet es una función, cuyas dilataciones y traslaciones dan origen a una familia de funciones, la que constituye una base de un espacio de funciones.
- Son usadas para la representación de datos y otras funciones.
- Técnicas wavelet nos permiten descomponer una función complicada en otras mas simples y estudiarlas separadamente

Sistemas Wavelets

- Una expansión wavelet de una señal o función es la combinación lineal:

$$f(t)=\sum_k c_k \psi_k(t)$$

- Si para cada función f(t) los coeficientes c_k son únicos, la familia $\{\psi_k(t)\}$ es una base.
- Esta base es ortogonal si: $<\psi_k(t),\psi_l(t)>=0 \quad k\neq l$, y los coeficientes se pueden calcular como: $c_k =< f(t),\psi_k(t)>$.
- Los coeficientes de la expansión de f(t)

$$f(t)=\sum_{j,k} c_{j,k}\psi_{j,k}(t)$$

son llamados LA TRANSFORMADA WAVELET DISCRETA (TWD) de f(t).

- El conjunto de expansión wavelet no es único. Existen muchos que pueden ser usados efectivamente.
- Características generales que debe satisfacer un Sistema Wavelet:

  - Si una familia de wavelets $\{\psi_{j,k}(t)\}_{j,k=1,2,...}$ es dada, entonces existen coeficientes $c_{j,k}$ tales que $f(t)=\sum_{j,k} c_{j,k}\psi_{j,k}(t)$.

  - La expansión wavelet da una localización tiempo-frecuencia de la señal.

  - El cálculo de los coeficientes de la señal puede ser hecho eficientemente.

- Características mas específicas:

  - Todos los sistemas wavelets son generados una sola función escala o wavelet, por escalamiento y traslación. La parametrización bidimensional es alcanzada desde la wavelet madre $\psi(t)$ por medio de

$$\psi_{j,k}(t)=2^{j/2}\psi(2^{j}t-k) \qquad j,k\in Z$$

- Los sistemas wavelet satisfacen las condiciones de multiresolución. Es decir, si un conjunto de señales puede ser representado por una suma pesada de $\varphi(t-k)$, entonces un conjunto mas grande puede ser representado por la suma pesada

$\varphi(2t-k)$

- Los coeficientes de baja resolución pueden ser calculados desde coeficientes de resoluciones mas altas por medio de un algoritmo de estructura de árbol (piramidal).

- La wavelet madre ψ(t), trae siempre asociada consigo una función escala φ(t). Con estas dos funciones podremos aproximar cualquier función o señal *f(t)* , mediante una de las funciones o mediante ambas, de la forma:

$$f(t) = \sum_k \sum_j c_{j,k} \varphi_{j,k}(t) + \sum_k \sum_j d_{j,k} \psi_{j,k}(t)$$

Función Escala

- **Dada una función $\varphi(t) \in L_2(R)$, llamada función escala básica se genera la familia de funciones escala definidas por:**

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k)$$

- **Las funciones escala generan los subespacios V_j de $L_2(R)$, de la siguiente manera:**

$$V_j = \overline{span_{k \in Z}\{\varphi_{j,k}(t)\}}$$

- **Una función $\varphi(t) \in L_2(R)$ es considerada como una "buena" función escala si cumple con las siguientes condiciones:**
  - **La familia $\{\varphi_{j,k}(t)\}$ es una base ortonormal de V_j.**
  - $V_j \subset V_{j+1}, \quad \forall j \in Z$
  - **La función $\varphi(t) \in L_2(R)$ tiene soporte compacto.**

Función Wavelet

- Definimos a

$$W_j = \overline{span_{k\in Z}\{\psi_{j,k}(t)\}}$$

como el complemento ortogonal de Vj en Vj+1, esto significa que todos los miembros de Vj son ortogonales a todos los miembros de Wj . Estos son subespacios de $L^2(R)$ y son llamados Espacios Detalle.

- Las relaciones entre los diferentes espacios puede verse de la siguiente manera:

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset L^2(R)$$

según la definición

$$V_1 = V_0 \oplus W_0, \quad V_2 = V_0 \oplus W_0 \oplus W_1, \quad \cdots, \quad L^2(R) = V_0 \oplus W_0 \oplus W \oplus \cdots$$

- De la última relación, toda $f(t) \in L_2(R)$ puede ser representada como

$$f(t) = \sum_k c_k \varphi_k(t) + \sum_k \sum_j d_{j,k} \psi_{j,k}(t)$$

Wavelets de Haar

- **La función escala básica de Haar se define:**

$$\varphi(t) = \begin{cases} 1, & si\ x \in [0,1) \\ 0, & otro\ valor \end{cases}$$

- **La función Madre de Haar se define:**

$$\psi(t) = \begin{cases} 1, & t \in [0, \frac{1}{2}) \\ -1, & t \in [\frac{1}{2}, 1) \\ 0, & otro\ valor \end{cases}$$

- **Las wavelets de Haar tienen las siguientes propiedades: base ortogonal, de soporte compacto, función escala simétrica y las wavelets antisimétricas, momento de decaimiento 1, y es una base interpolante.**

Reconocimiento del Habla utilizando Wavelets.

Localización tiempo frecuencia.

Trasformada de fourier de una función f(t)

$$(Ff)(\omega) = \frac{1}{\sqrt{2\pi}} \int dt\, e^{-i\omega t} f(t)$$

Localizando en el tiempo.

$$\left(T^{win} f\right)(\omega,t) = \int ds \quad f(s) \quad g(x-t) \quad e^{-i\omega t}$$

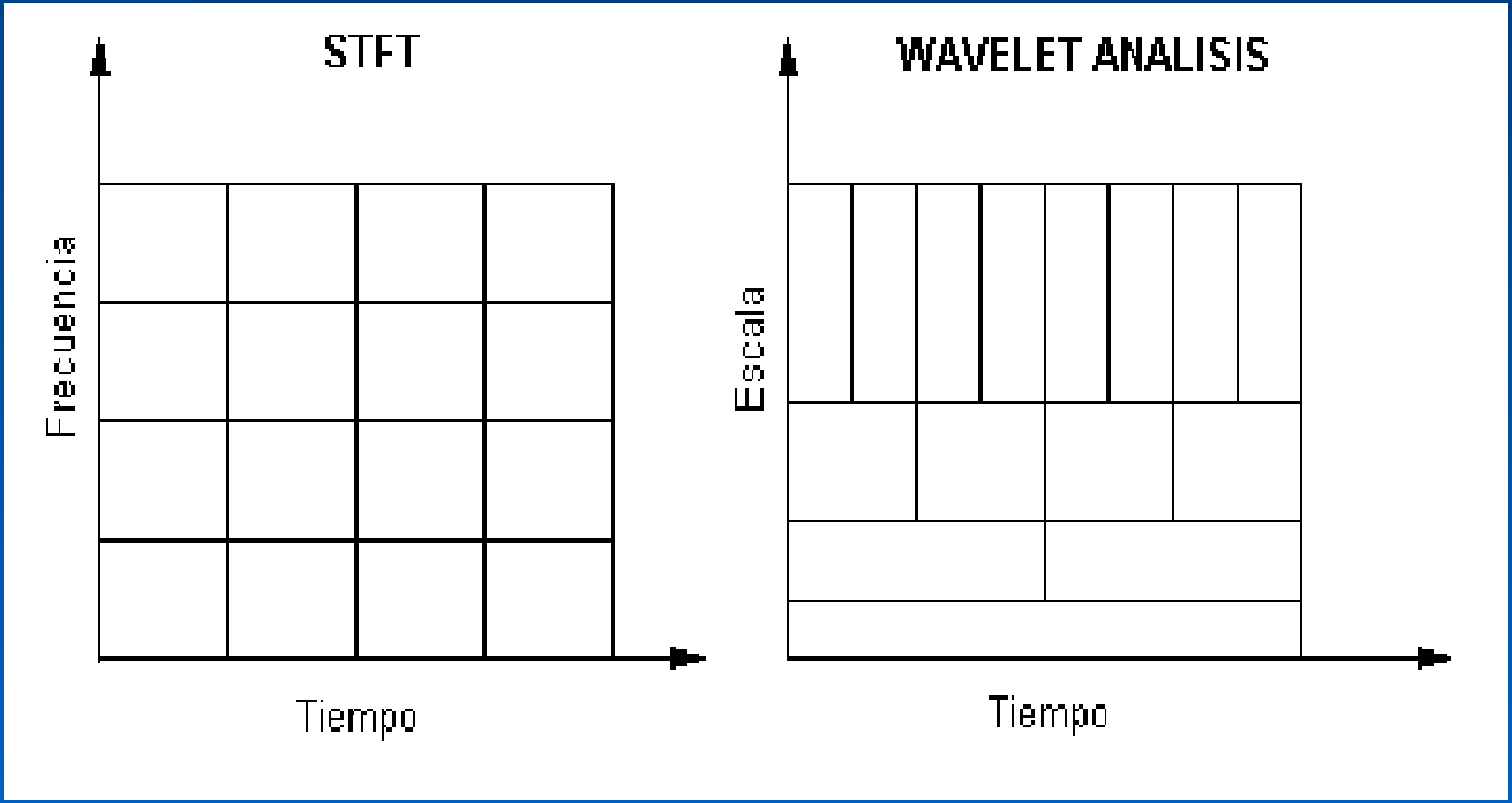
Reconocimiento del Habla utilizando Wavelets.

- Transformada wavelet.
- Utiliza una ventana igual que STFT, pero focaliza el análisis solo en ese intervalo.

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi(\frac{t-b}{a})dt$$

$$CWT(a,b) = \frac{1}{2\pi\sqrt{a}} \int_{-\infty}^{\infty} X(\omega)\Psi(a\omega)e^{-i\omega b}dt$$

Reconocimiento del Habla utilizando Wavelets.



Reconocimiento del Habla utilizando Wavelets.

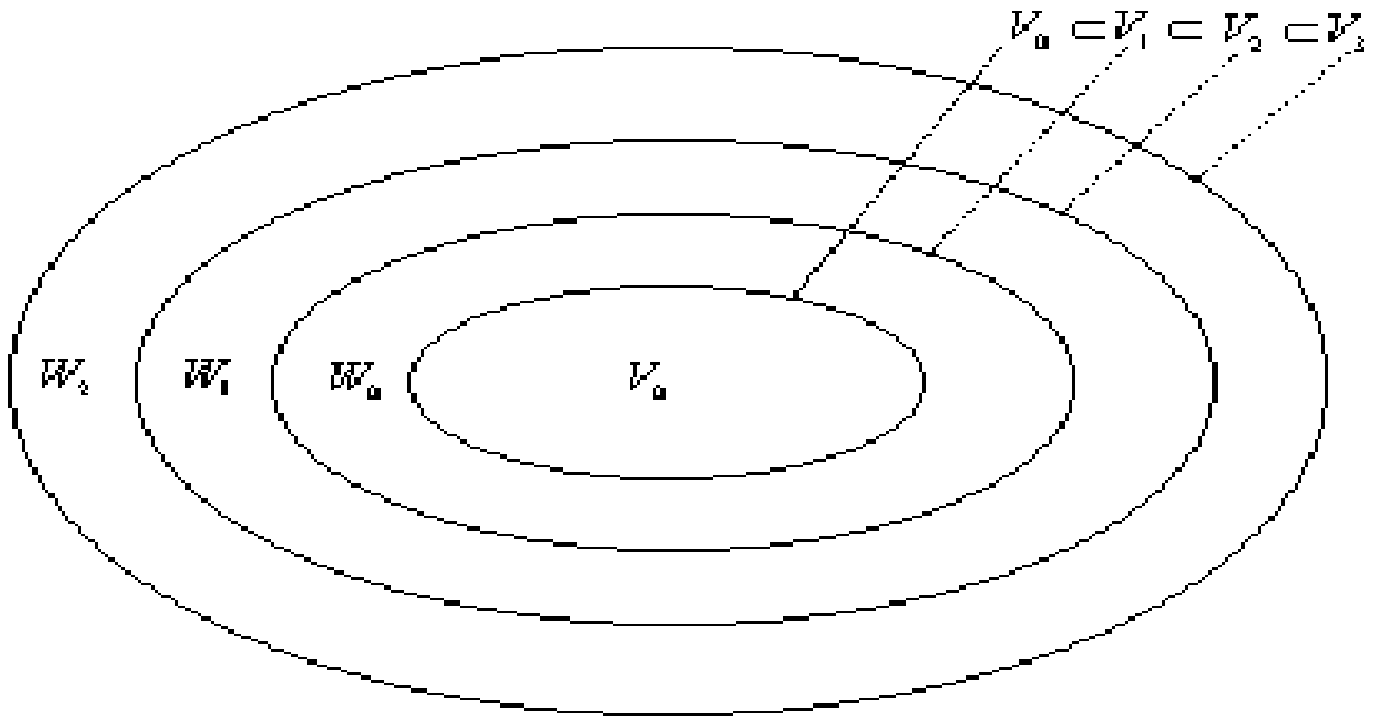
➢ Análisis multiresolución:

El análisis multi-resolución consiste básicamente en aproximar una función $f(t)$ en distintos niveles de resolución $\{f_1(t), f_2(t), f_3(t), \ldots\}$, lo que nos entrega una descomposición multi-escala de la forma:

$$f(t) = f_0(t) + \sum_{j \geq 0} g_j(t)$$

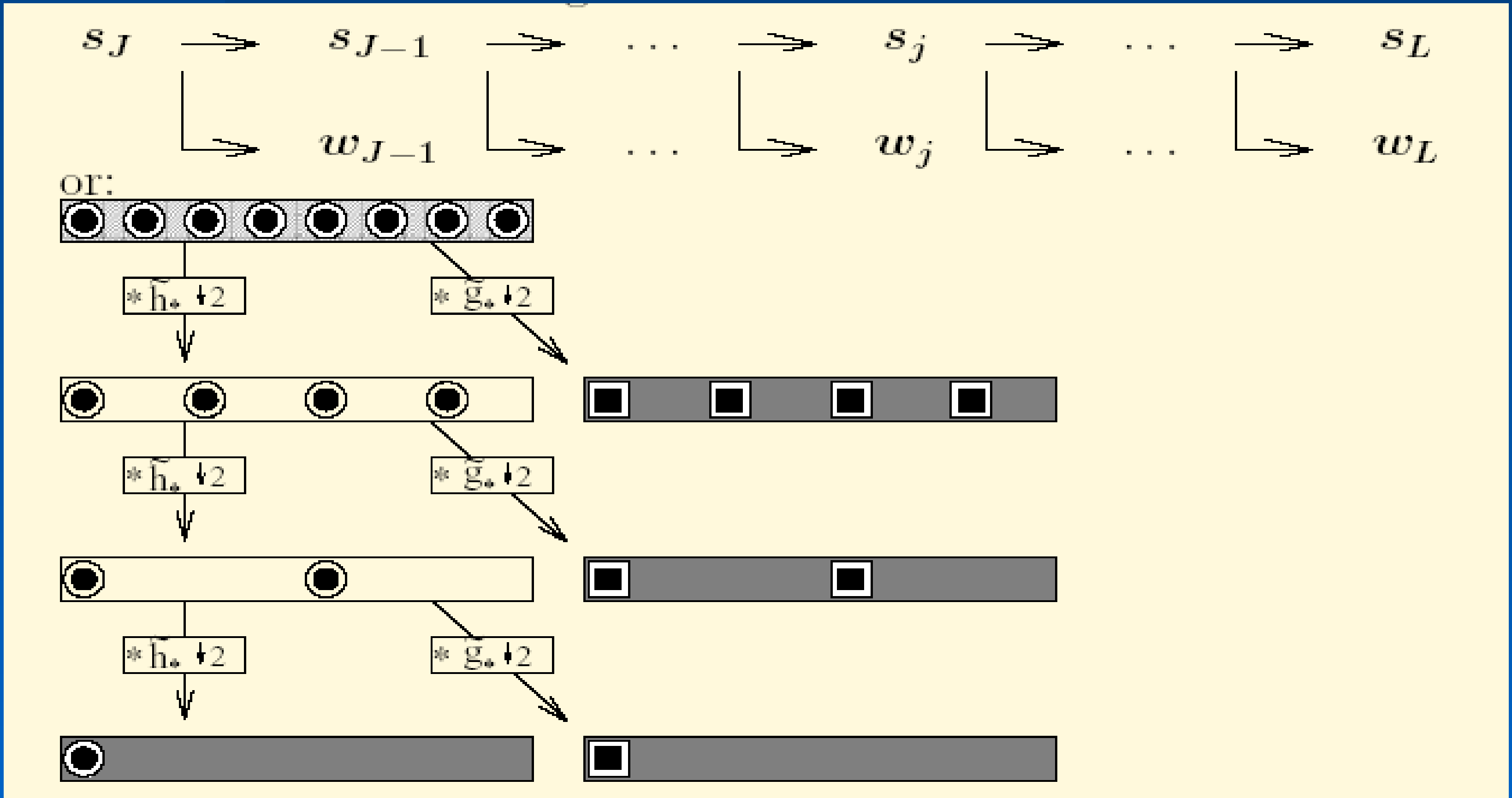
donde cada $g_j(t) = f_{j+1}(t) - f_j(t)$ representa el error en que se incurre al aproximar $f_{j+1}(t)$ mediante $f_j(t)$, o lo que es lo mismo, la fluctuación entre dos niveles sucesivos de resolución.

Reconocimiento del Habla utilizando Wavelets.



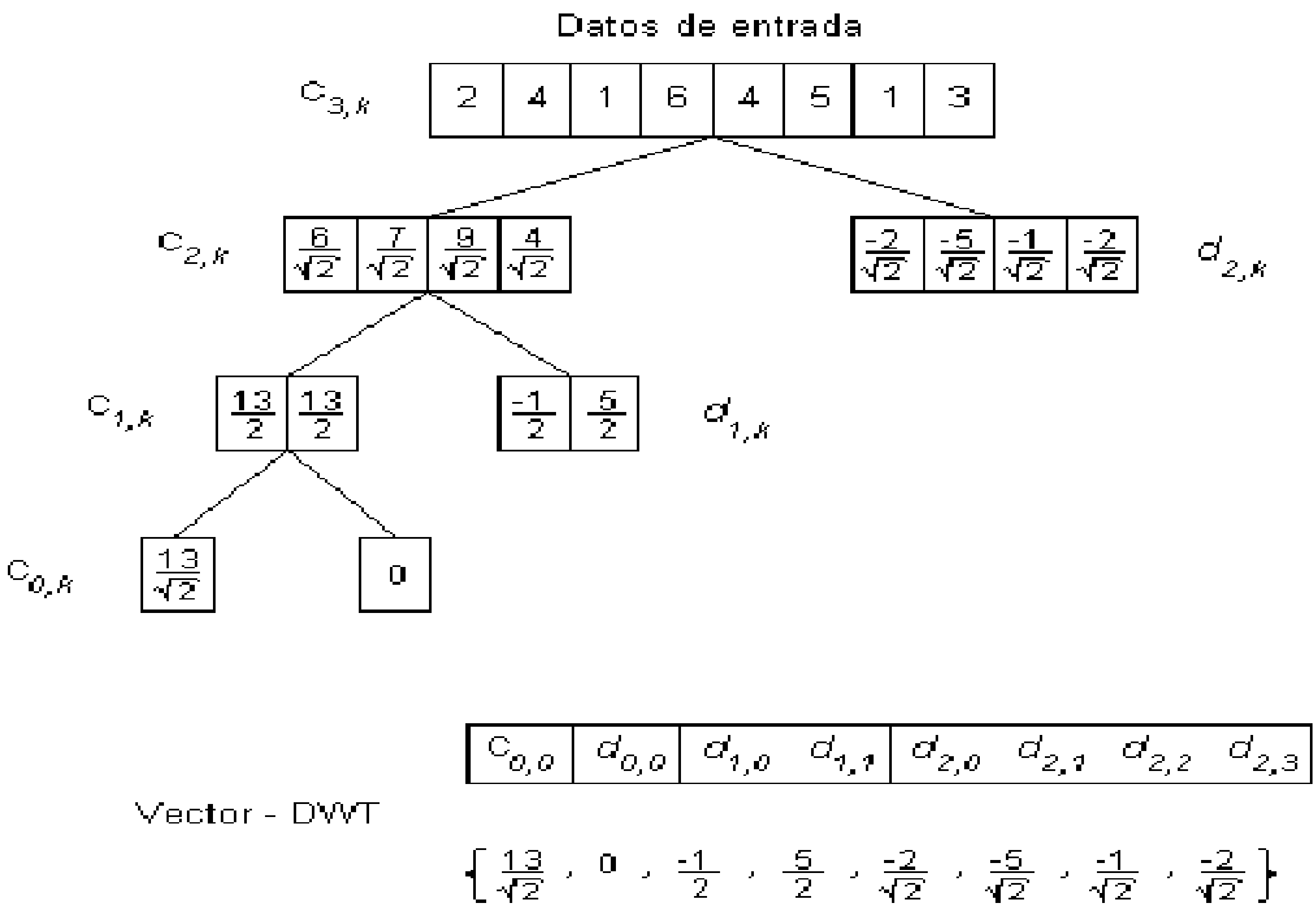
Reconocimiento del Habla utilizando Wavelets.

Bancos de filtros recursivos:



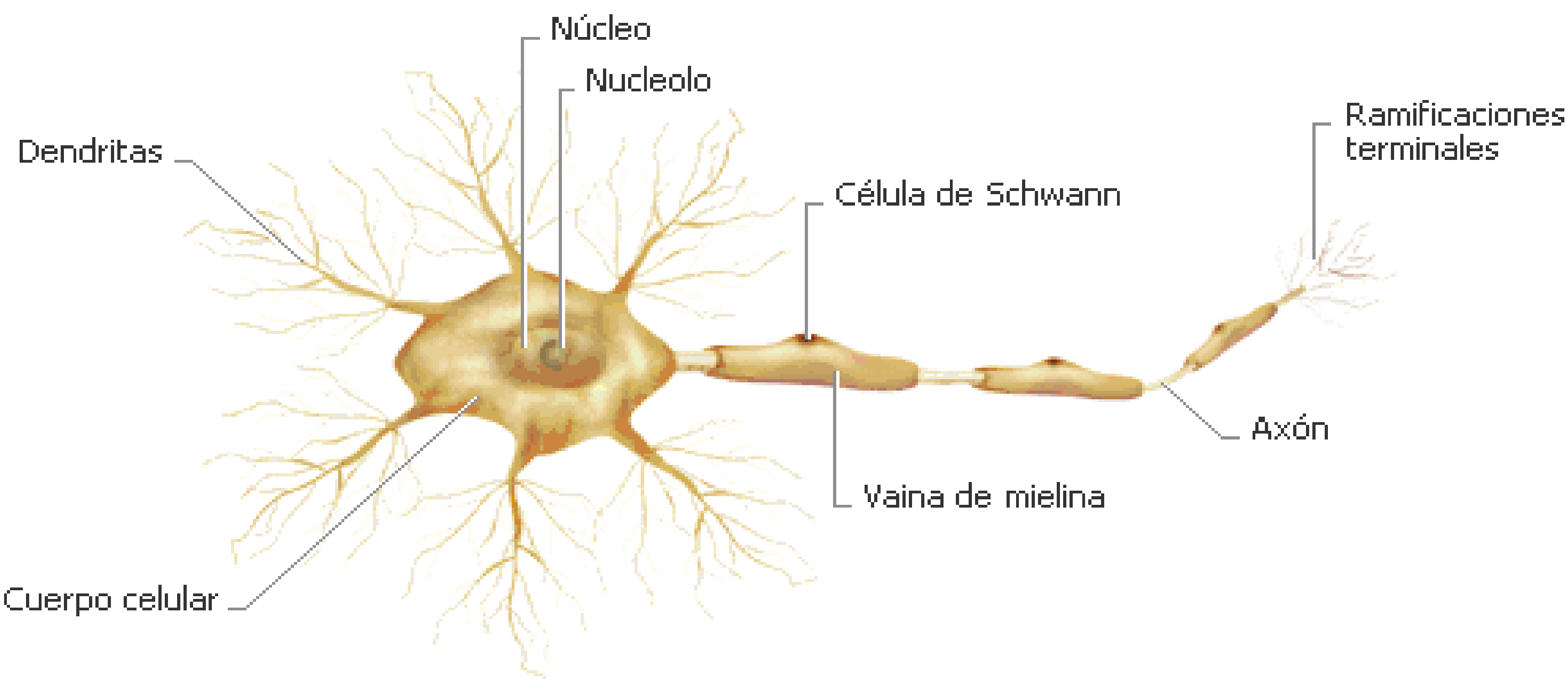
Jorge L. Guevara D. - Juan O. Salazar C.

Reconocimiento del Habla utilizando Wavelets.



➢ Redes Neuronales:

Modelo Biológico

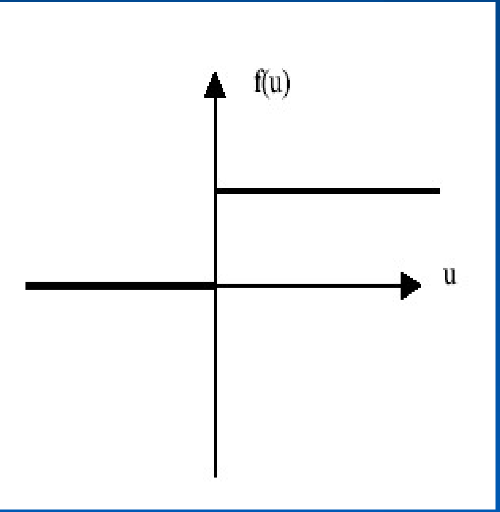


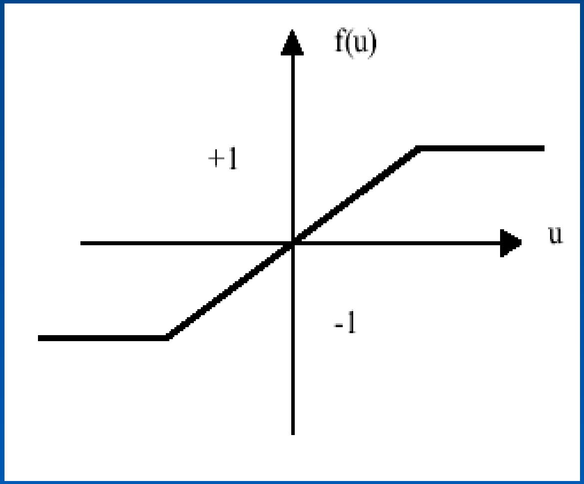
➢ Funciones de activación:

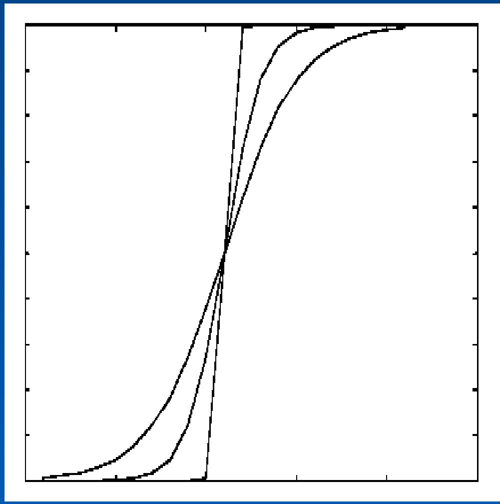
$$y = f(u) = \begin{cases} 0 & if \quad u < 0 \\ 1 & if \quad u \geq 0 \end{cases}$$

$$y = f(u) = \begin{cases} -1 & if \quad u < -1 \\ u & if \quad -1 \geq u \geq 1 \\ 1 & if \quad u \geq 1 \end{cases}$$

$$f(x) = \frac{1}{1+e^{-\alpha x}}, \; 0 \leq f(x) \leq 1$$

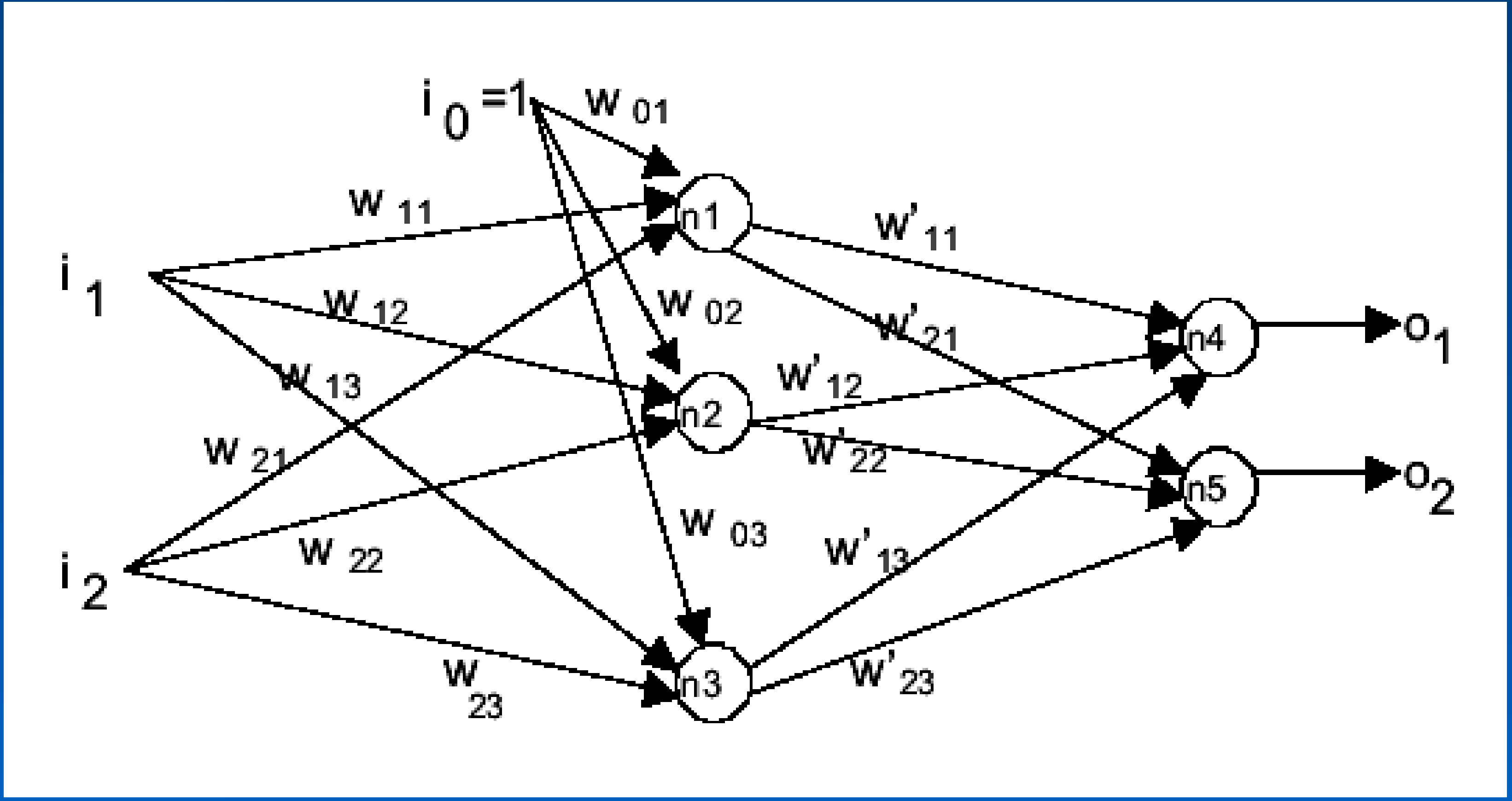






$$f(x) = \frac{e^{\alpha x} - e^{-\alpha x}}{e^{\alpha x} + e^{-\alpha x}}$$

Reconocimiento del Habla utilizando Wavelets.



Modelo Informático:

- Captura de la señal
- Procesamiento de la Señal
- Caracterización de la señal
- Clasificación de la Señal

Reconocimiento del Habla utilizando Wavelets.

- Para evaluar el modelo hemos tomado muestras de personas comprendidas entre los 18 y 60 años.
- Las muestras tomadas corresponden a las vocales.

- **Captura de la señal:**

Teorema del muestreo

$$Fs = 2 * Fa$$

Fs=frecuencia de sampleo

Fa=más alta frecuencia de la señal

$$T = 1/Fs$$

$$x(t) \longrightarrow x(nT), \quad n \text{ es entero}$$

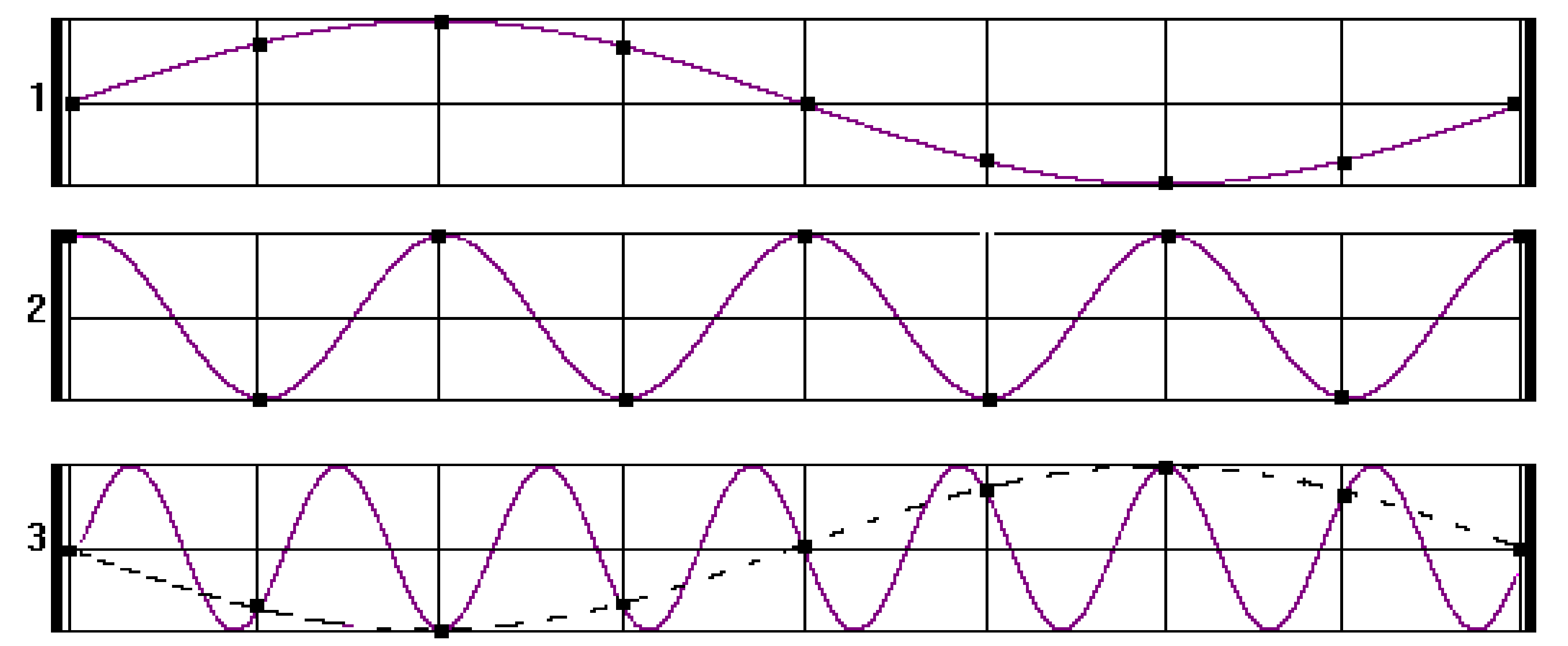
Reconocimiento del Habla utilizando Wavelets.

- **Etapas:**

  Filtro pasa baja ⟶ eliminar frecuencias > mitad de Fs

  Medir amplitud instantánea

  Convertir cada medicion en un valor numerico



T

Reconocimiento del Habla utilizando Wavelets.

En la señal de voz la mayor parte de información se tiene en frecuencias bajas usualmente se usa Fs entre 8Kz y 16Kz con lo que podemos capturar señales en un rango de frecuencia máximas aproximadamente entre 4Kz y 8Kz.

Muestreo	**:**	Frecuencia: 8000 Hz.
Cuantificación	**:**	8 bits.

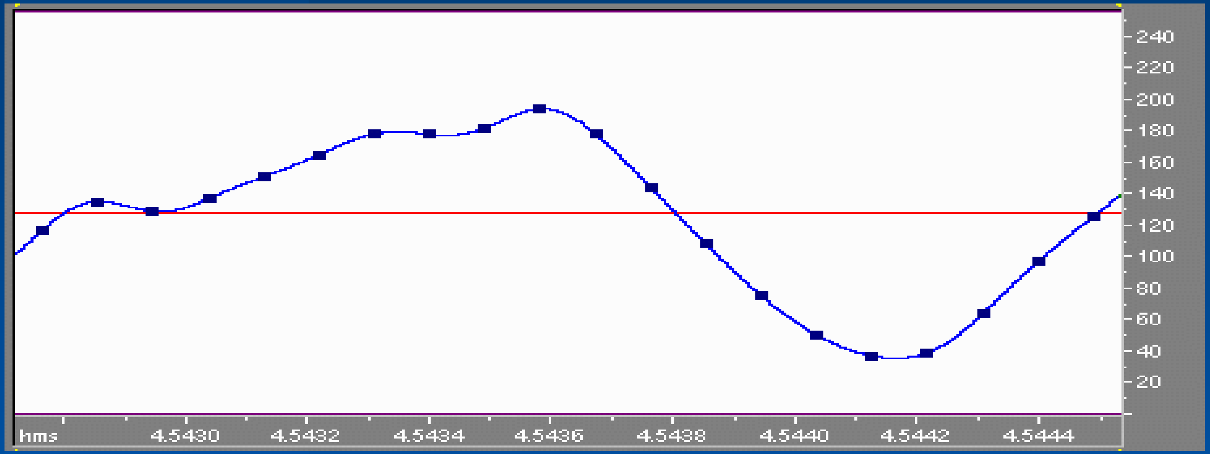
Reconocimiento del Habla utilizando Wavelets.

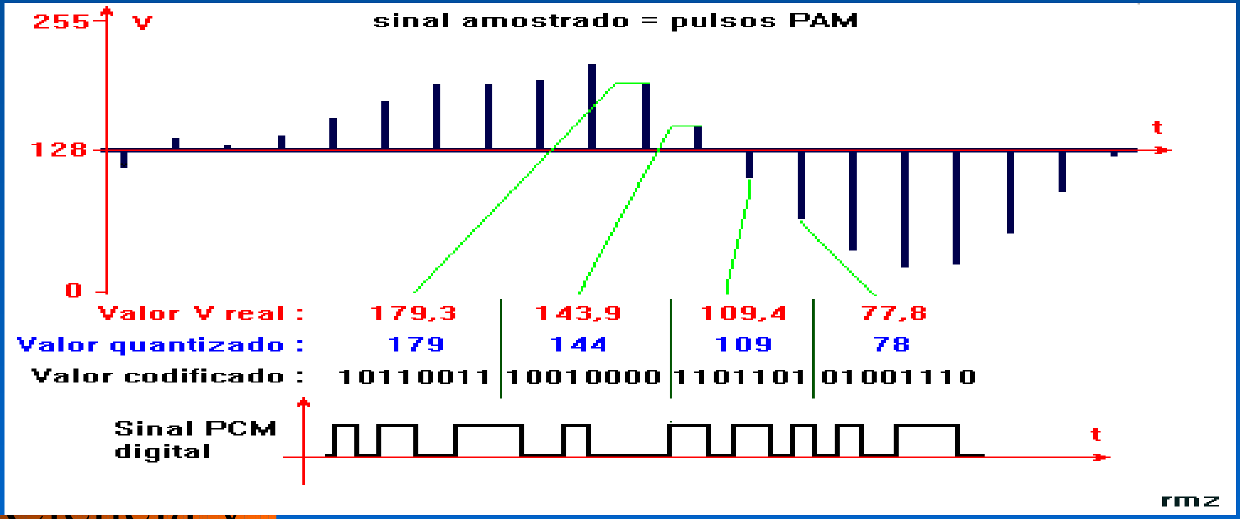
V= vector de la señal de voz en su forma digital

Para obtener el vector se tendrá que construir un programa que trabaje capturando las intensidades de las señales.

Jorge L. Guevara D. - Juan O. Salazar C.

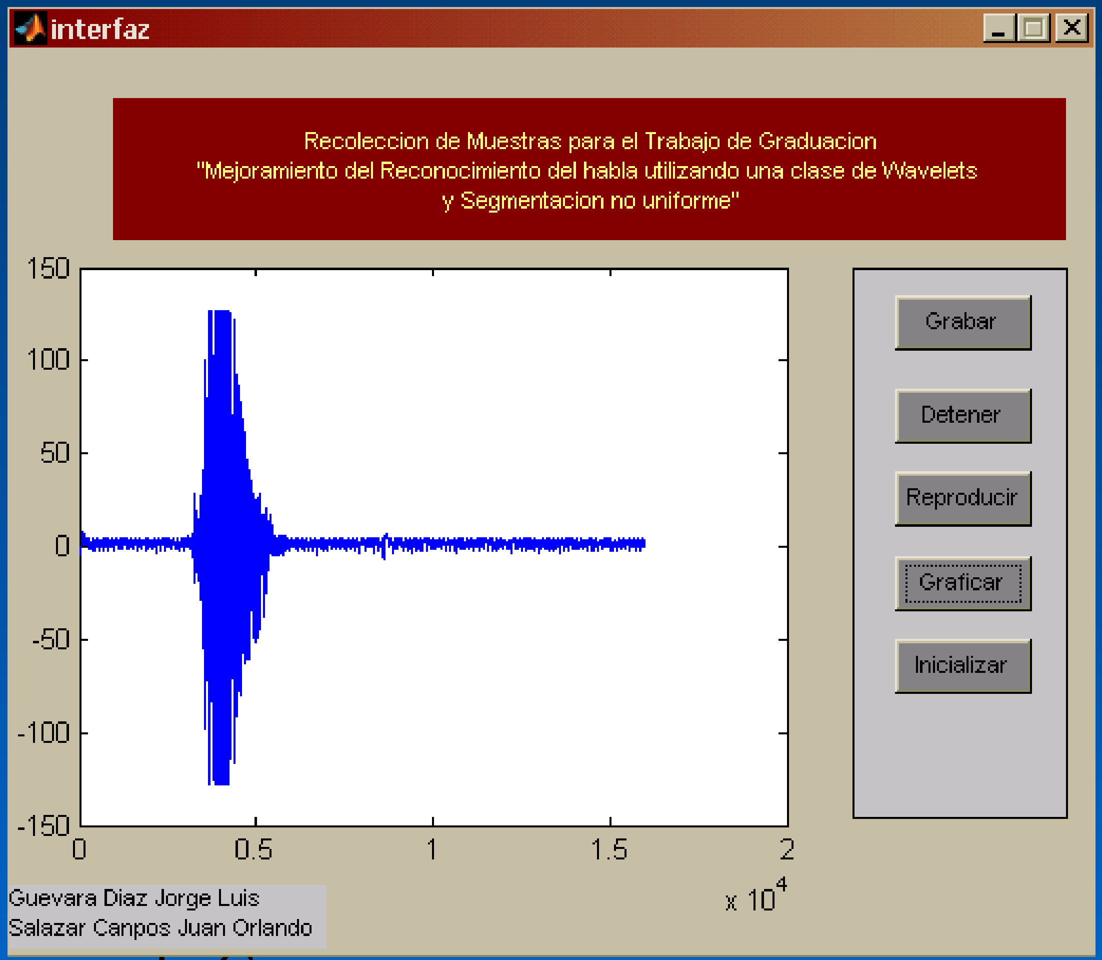
Conversor analógico digital:





Reconocimiento del Habla utilizando Wavelets.

Vocal “a”



Reconocimiento del Habla utilizando Wavelets.

- **<u>Procesamiento digital de la señal:</u>**

  - Eliminación de segmentos inservibles.
  - Normalización de la señal.
  - Análisis de componentes de frecuencia.
  - Patrón de características.

Reconocimiento del Habla utilizando Wavelets.

- **Eliminación de segmentos inservibles:**
  - **Obtener umbral:**

valor = $$1/radio * \sum_{i}^{radio+i} |x|$$

umbral = (valorInic + valorFinal)/2 + n

n = valor entre 5 y 10 aproximadamente

Reconocimiento del Habla utilizando Wavelets.

- **<u>Algoritmo</u>:**

Recorrer toda la señal desde el inicio (i)

para una ventana de tamaño = radio

valor = $1/radio * \sum_{i}^{radio+i} |x|$

if valor<=umbral

eliminar segmento ;

señal(i)=señal(i+radio)

etiquetar inicio

Reconocimiento del Habla utilizando Wavelets.

Recorrer toda la señal desde el final (i)

para una ventana de tamaño = radio

valor = $1/radio * \sum_{i}^{radio+i} |x|$

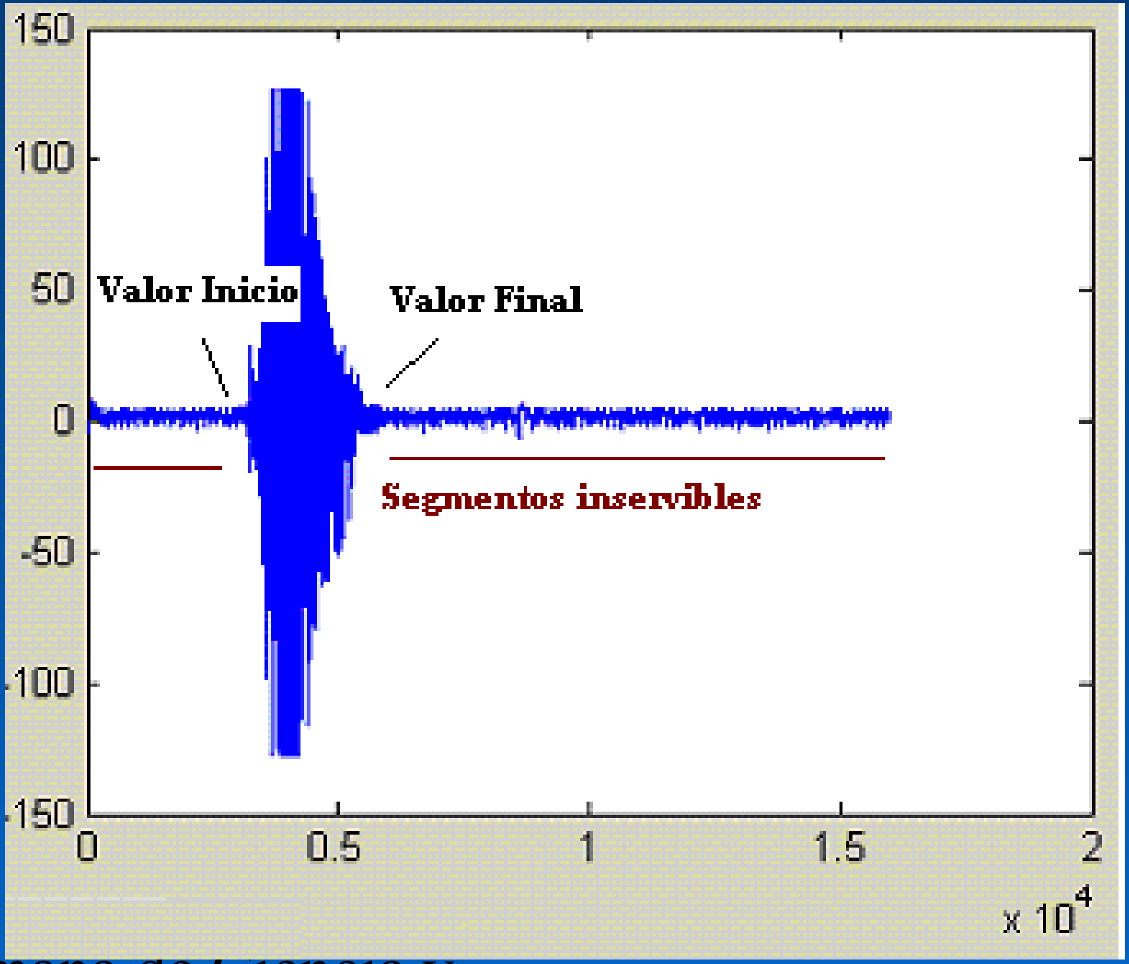
if valor<=umbral

eliminar segmento ;

señal(i)=señal(i-radio)

etiquetar fin

Reconocimiento del Habla utilizando Wavelets.

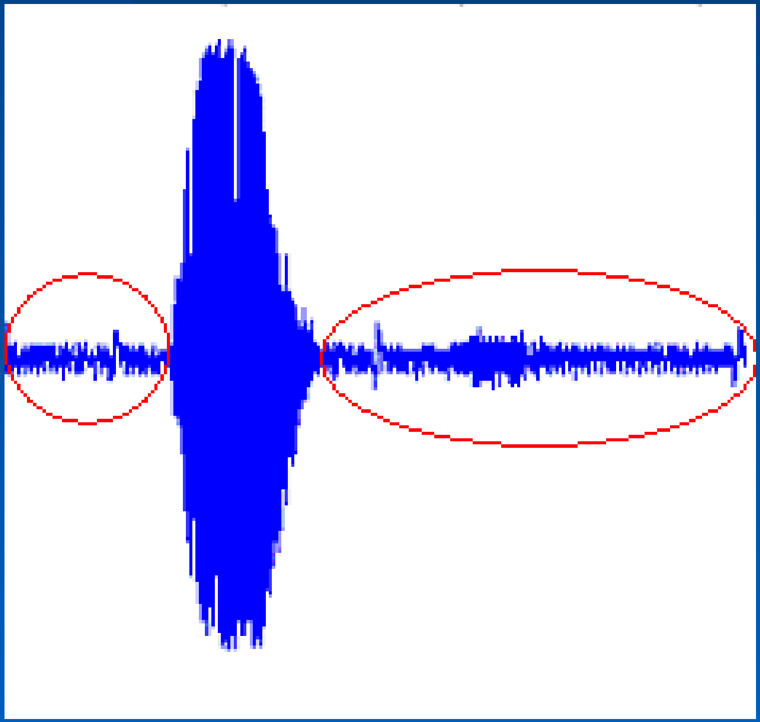


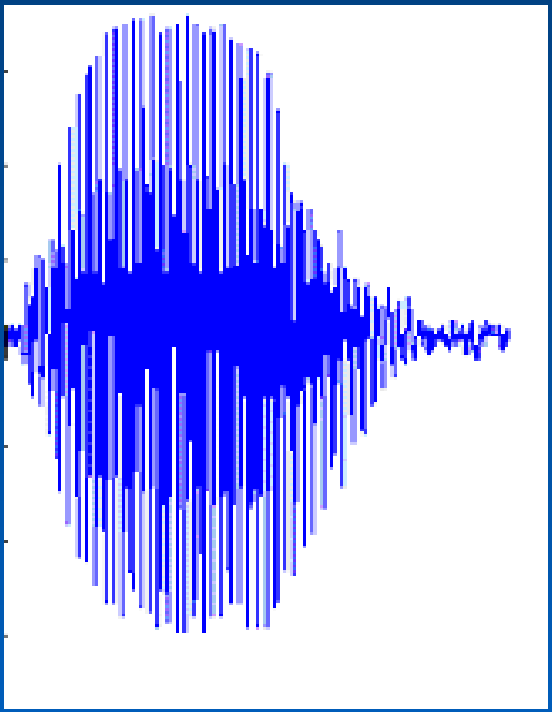
Reconocimiento del Habla utilizando Wavelets.

$$\underbrace{\boxed{10}\boxed{8}\boxed{4}\boxed{5}\boxed{7}}_{10}\boxed{9}\boxed{6}\boxed{\quad \ldots \quad}\boxed{10}\boxed{11}\underbrace{\boxed{8}\boxed{9}\boxed{15}\boxed{13}\boxed{12}}_{16}$$

$$\text{Umbral} = \frac{10+16}{2} + 5 = 18$$

Reconocimiento del Habla utilizando Wavelets.





➢**<u>Normalización de la señal</u>:**

1° Fijamos valores estándares tanto para el valor el mínimo como para el valor máximo donde trabajaremos.

2° Ingresamos a una función el vector obtenido para llevarlo a estos rangos de trabajo.

Reconocimiento del Habla utilizando Wavelets.

- **<u>Normalización de la señal</u>:**

Método 1:

$$[a,b] \rightarrow f \rightarrow [c,d] \rightarrow T \rightarrow [m,n]$$

T(y)=Ay+B

T(c)=m

T(d)=n

A=(n-m)(d-c) B=m-(n-m)/(d-c)

T(y)=(n-m)y+md-nc/(d-c)

- **<u>Normalización de la señal</u>:**

Método 2:

$$\text{Longitud} = \sqrt{\sum_{i=1}^{n} f^{2}(x)}$$

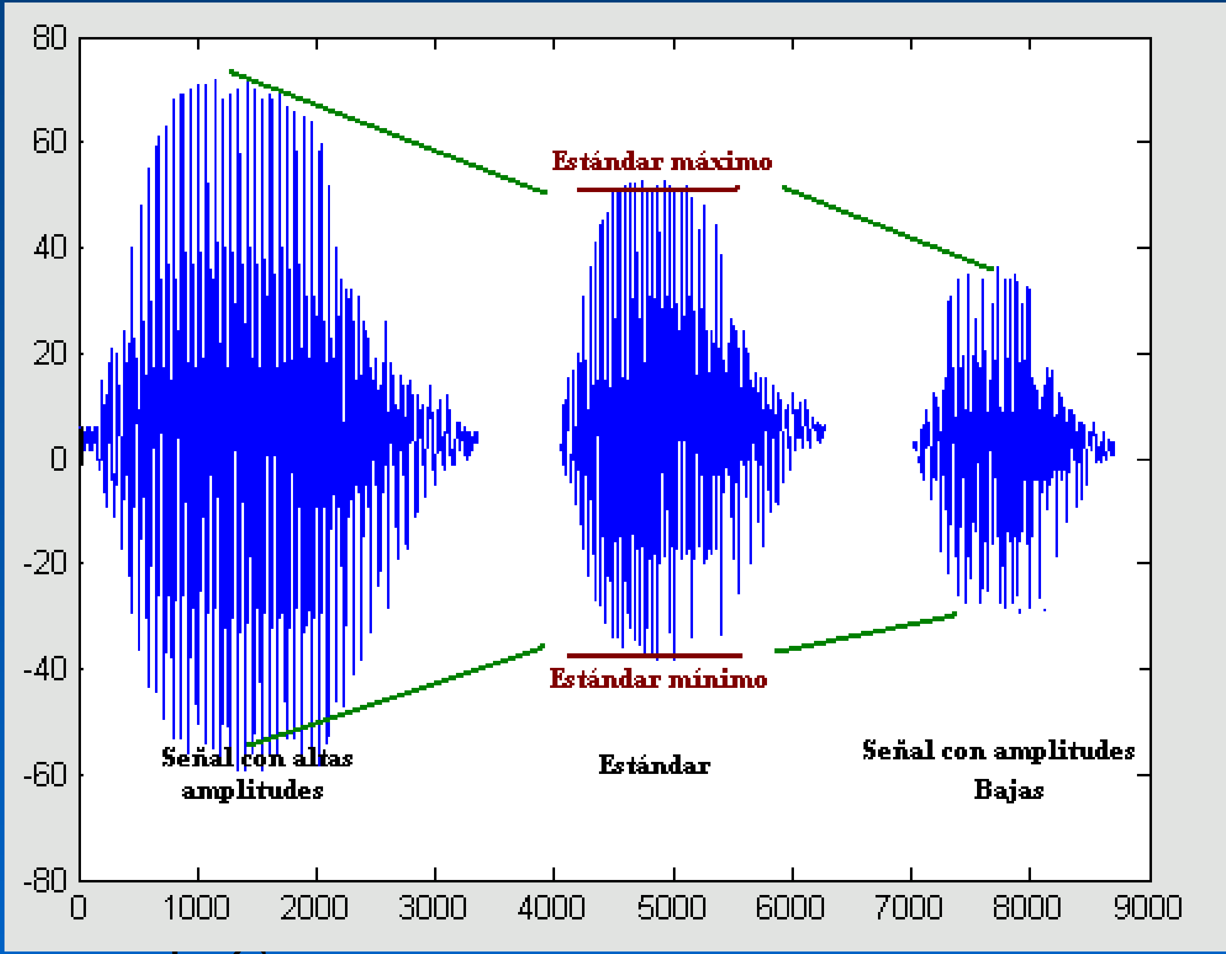
$$f(x)=f(x)/\text{longitud}$$

Reconocimiento del Habla utilizando Wavelets.

- **Normalización:**

Se esta trabajando con el primer método y el rango de valores es de 10 a -10

Reconocimiento del Habla utilizando Wavelets.



Reconocimiento del Habla utilizando Wavelets.

- **Segmentación de la señal:**

  Nos servirá para capturar características locales de la señal de habla.

  Segmentación uniforme

  Dividir en espacios iguales

  Segmentación no uniforme

  Varios criterios

- **<u>Segmentación Uniforme</u>:**

  Número de segmentos.
  Manejo del Tamaño del vector.
  Potencia de 2.

  Tamaño del vector es: 2^11 * NumSeg.

Reconocimiento del Habla utilizando Wavelets.

8000 Hz

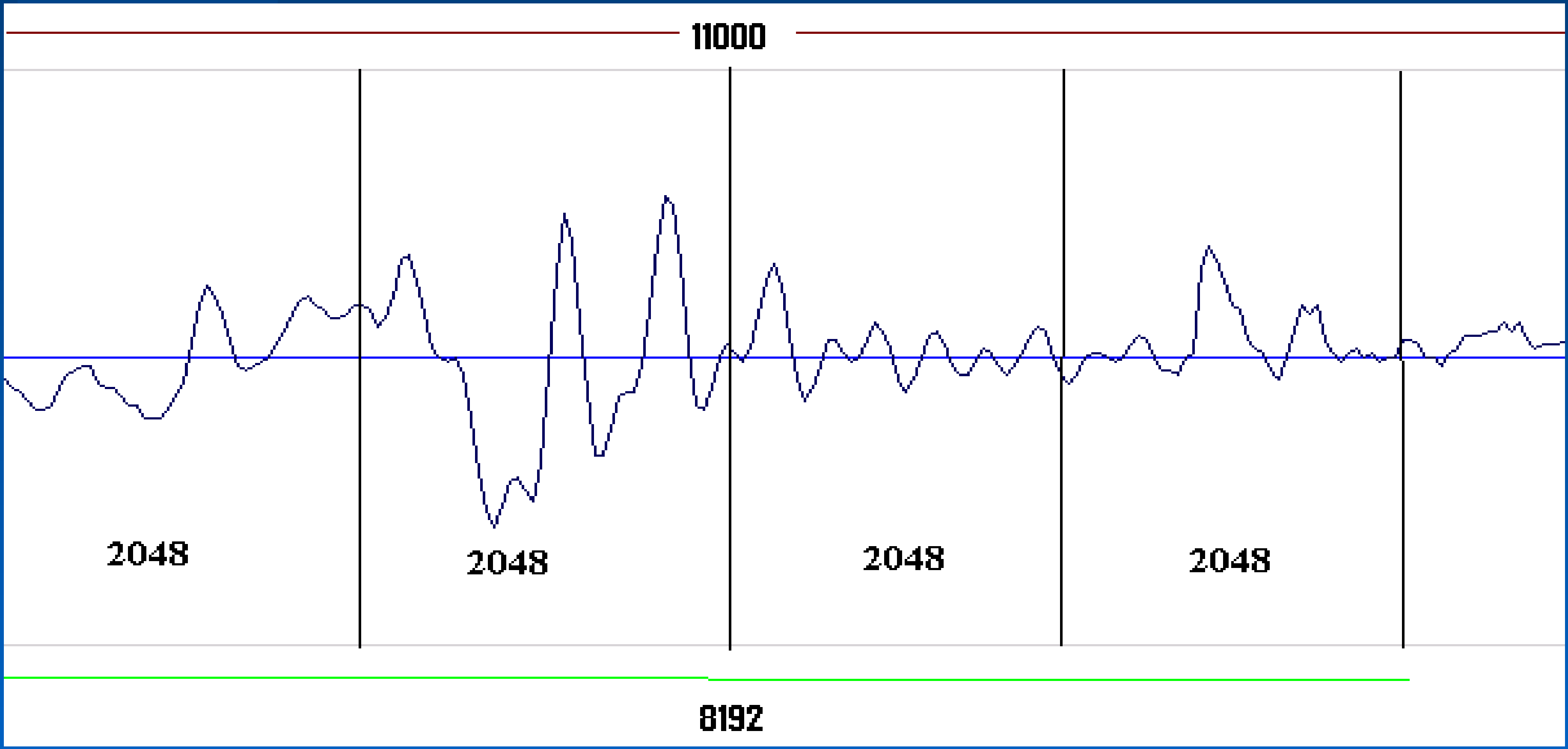
1 segundo

2400

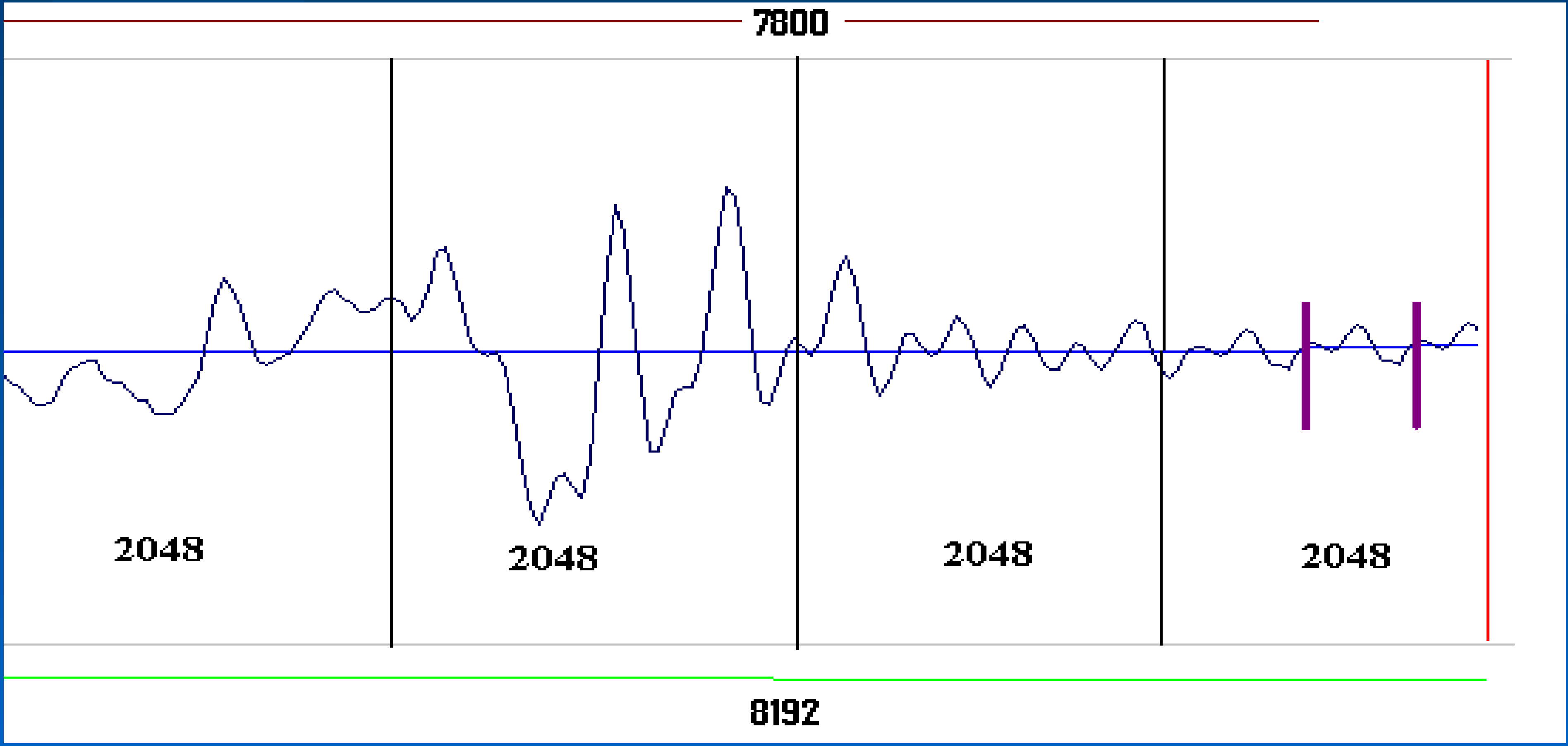
0.3 segundos

Jorge L. Guevara D. - Juan O. Salazar C.

Reconocimiento del Habla utilizando Wavelets.



Reconocimiento del Habla utilizando Wavelets.



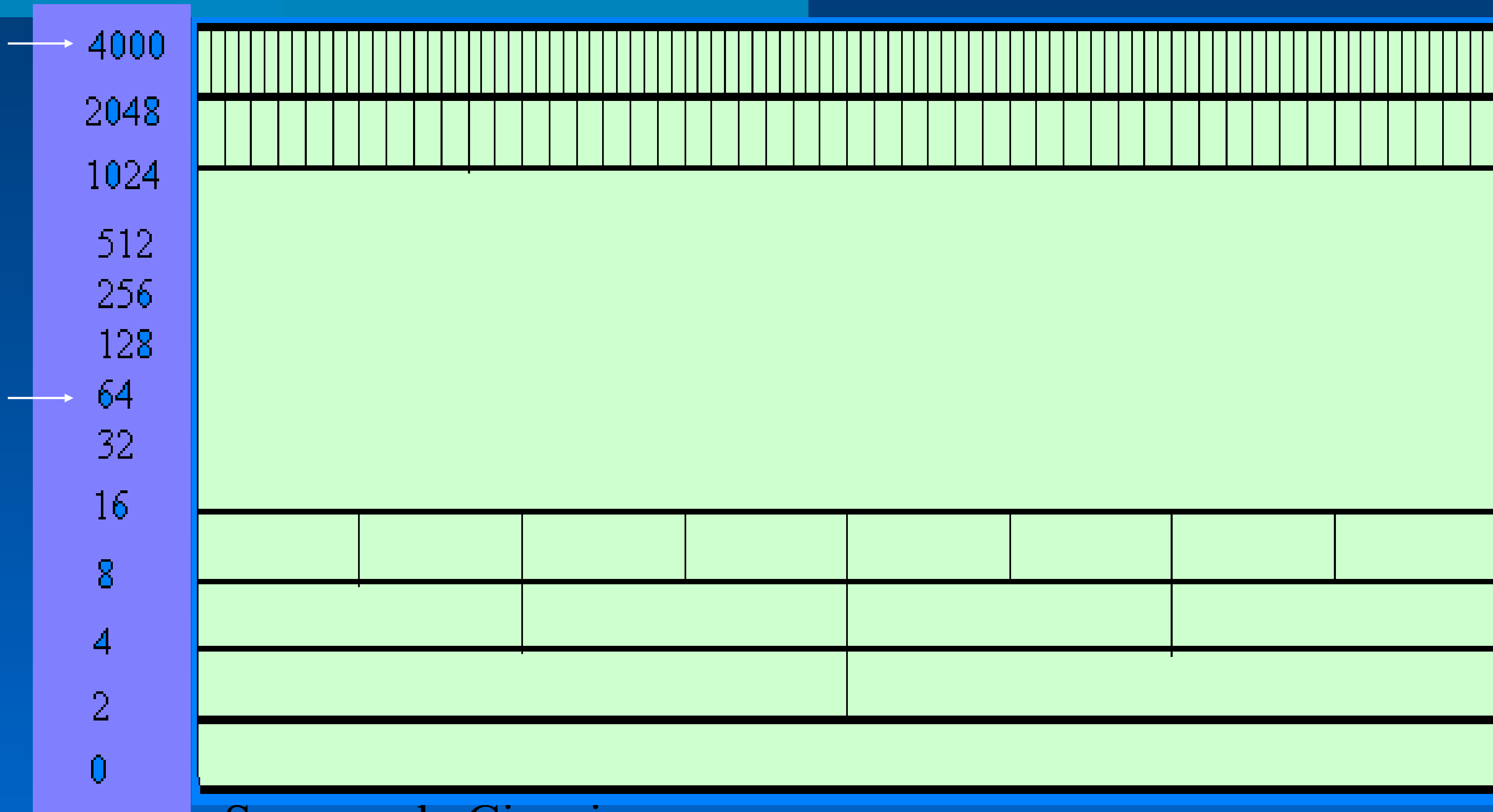
Reconocimiento del Habla utilizando Wavelets.

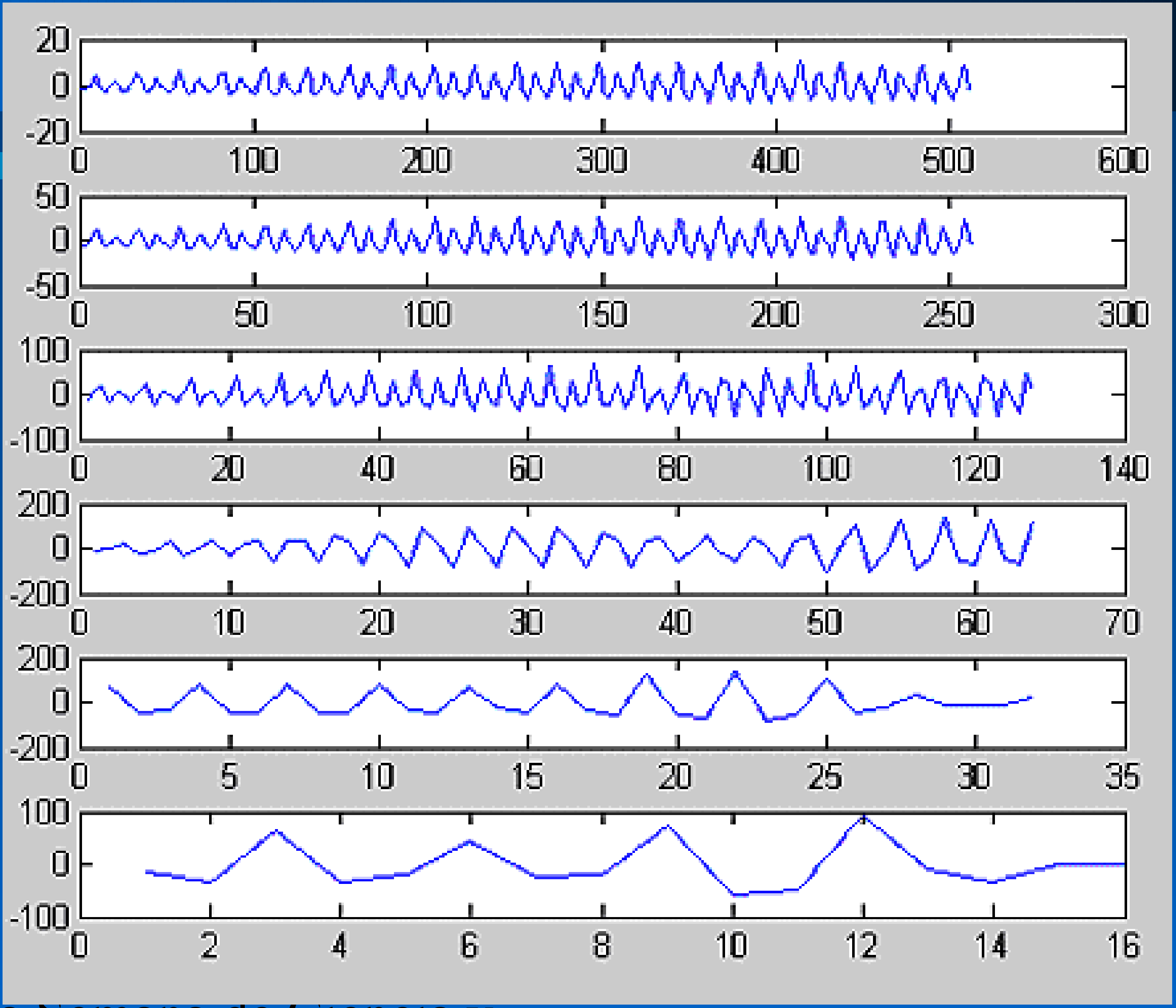
- **<u>Aplicación de la Transformada Rápida de Wavelets Haar</u>:**

Por cada segmento.

6 Niveles de descomposición.

Reconocimiento del Habla utilizando Wavelets.




20
0
-20
0 100 200 300 400 500 600
50
0
-50
0 50 100 150 200 250 300
100
0
-100
0 20 40 60 80 100 120 140
200
0
-200
0 10 20 30 40 50 60 70
200
0
-200
0 5 10 15 20 25 30 35
100
0
-100
0 2 4 6 8 10 12 14 16

- **Caracterización de la señal:**
  cálculo de las energías.

$$E(n)=1/N\sum_{n=0}^{N-1}x\left(n\right)^{2}$$

Reconocimiento del Habla utilizando Wavelets.

- Patrón formado por las energías de los distintos niveles de descomposición
- Para n niveles habrá un patrón con n elementos

Reconocimiento del Habla utilizando Wavelets.

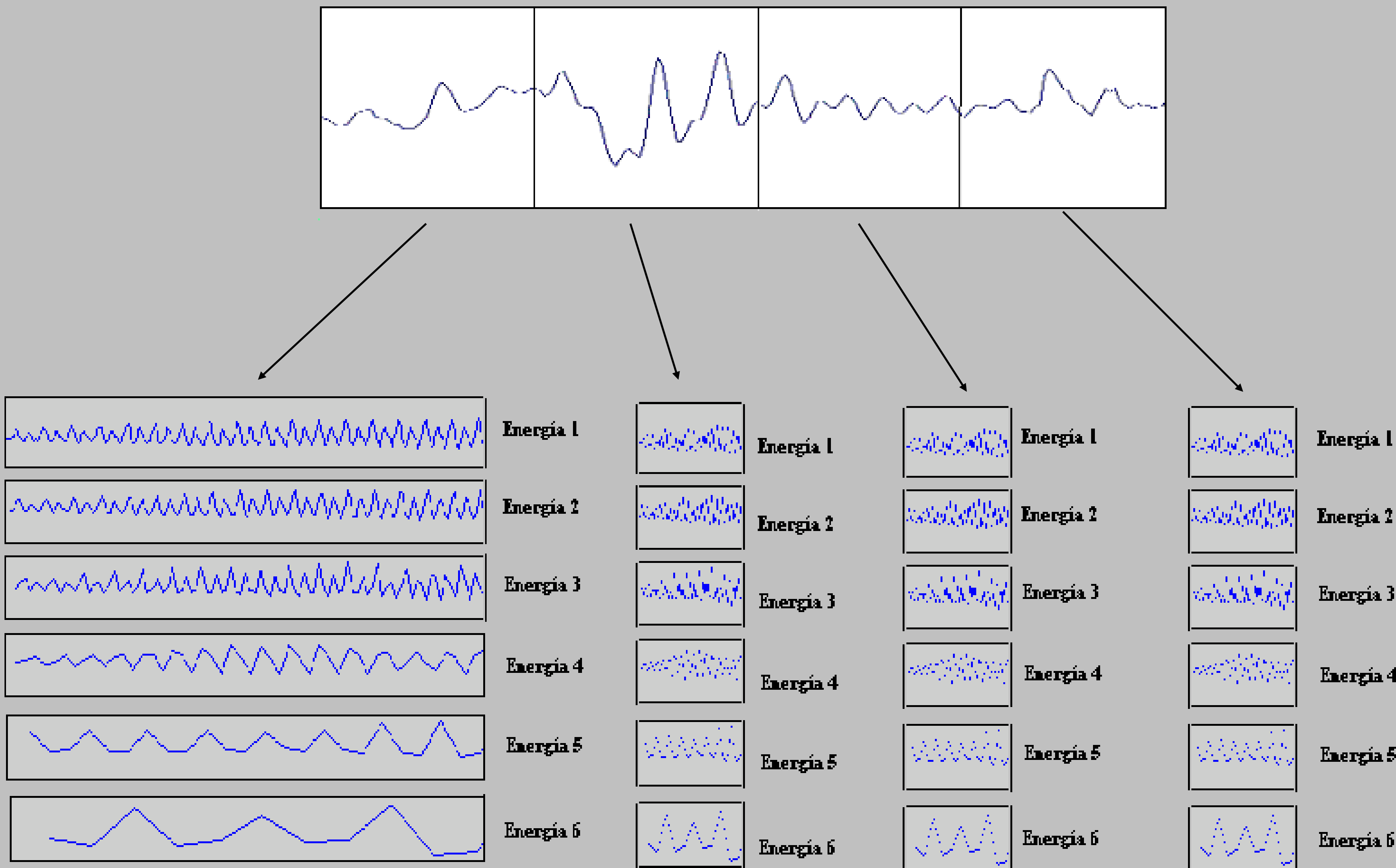
- **Formación del patrón:**

Calculamos la energía de cada nivel:

$$\frac{1}{\|x\|}\sum_{i=1}^{k}(x_i^2)$$

Vector con 6 energías por segmento.

Juntando los segmentos obtendremos un vector.



Energia 1
Energia 2
Energia 3
Energia 4
Energia 5
Energia 6
Energia 1
Energia 2
Energia 3
Energia 4
Energia 5
Energia 6
Energia 1
Energia 2
Energia 3
Energia 4
Energia 5
Energia 6
Energia 1
Energia 2
Energia 3
Energia 4
Energia 5
Energia 6

Reconocimiento del Habla utilizando Wavelets.

E1	E2	E3	E4	E5	E6	...						

Patrón

Reconocimiento del Habla utilizando Wavelets.

➢ Reconocimiento:

Usar una Red Neuronal BackProgation (solo para analizar resultados).

Capas: Una de entrada, Una oculta y Una de salida.

Neuronas de entrada: 6.
Neuronas ocultas: 15.
Neuronas de salida: 6.

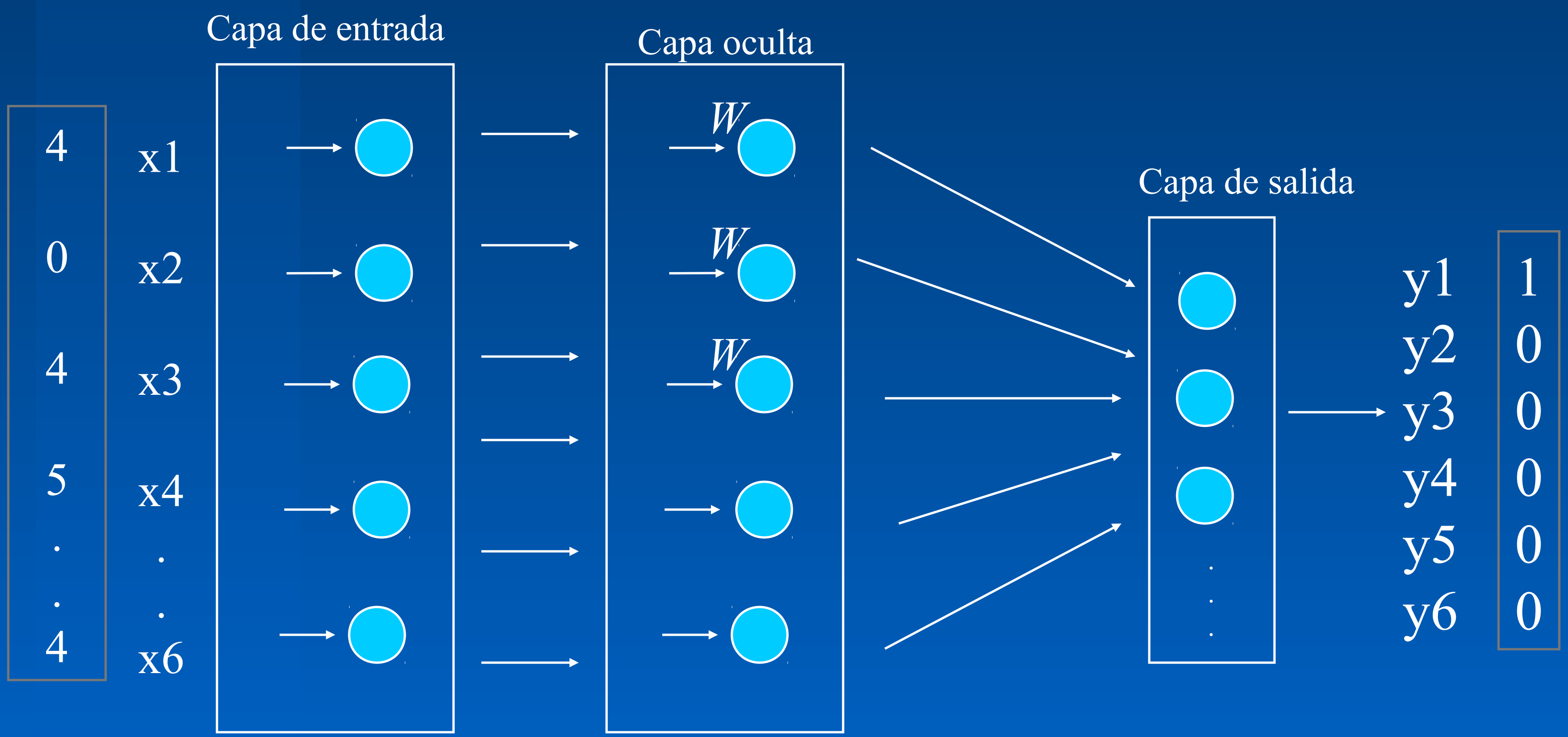
Reconocimiento del Habla utilizando Wavelets.

➢ <u>**Fase de entrenamiento**</u>**:**

Entrenamos a la red con:
5 muestras de la misma vocal pronunciada por la misma persona, es decir 25 muestras por personas, en un total de 17 personas

Todas las muestras pasan por la fase del procesamiento de señal.

Reconocimiento del Habla utilizando Wavelets.

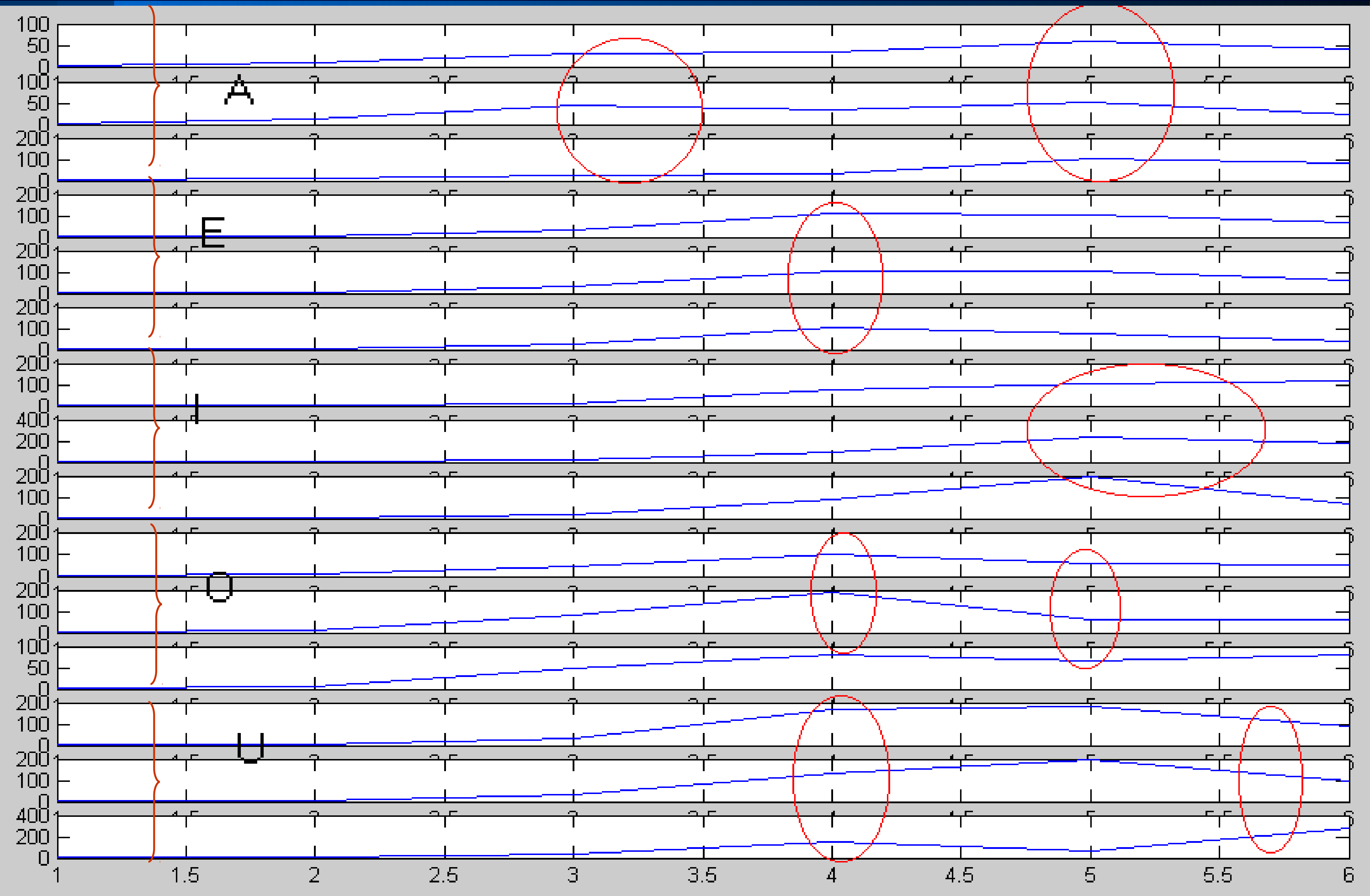


Reconocimiento del Habla utilizando Wavelets.

➢ **Fase de Emparejamiento (Aplicación):**

La Persona pronuncia una vocal, se la procesa, y obtenemos un resultado el cual nos indicará si es o no una vocal y en caso de que lo fuese nos mostrará que vocal es.

- **Algunos resultados:**



A
E
I
O
U
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
6

Conclusiones:

- Los valores capturamos dependen del muestreo y la cuantificación.
- La normalización y segmentación nos ayuda en la formación de las características de una señal.
- Wavelets nos ayudan enormemente en la representación de información de una señal.
- Costo de entrenamiento de la red neuronal tipo backpropagation es demasiado alto

Reconocimiento del Habla utilizando Wavelets.

<u>Proyecciones:</u>

- Probar con otros tipos de wavelets.
- Usar una segmentación no uniforme.
- Otro modelo de red neuronal.