

**VII REUNION DE TRABAJO  
EN PROCESAMIENTO  
DE LA INFORMACION  
Y CONTROL**

**RPIC 97**

**Volumen I**

**San Juan, 17 al 19 de Setiembre de 1997**

**Instituto de Automática. Facultad de Ingeniería.  
Universidad Nacional de San Juan.**

## INDICE GENERAL - VOLUMEN I

### MT1 - Inteligencia artificial 1

Comparación entre dos aproximaciones diferentes de diagnosis basada en conocimiento <i>Acosta, G.</i> .....	1
Sistema para análisis y síntesis del habla <i>Estienne, C., Campiglio, G.</i> .....	7
Vinculando el razonamiento revisable con los modelos de razonamiento científico <i>Delrieux, C.</i> .....	13
Detección de diagramas causales en sistemas de funciones no monotónicas <i>Delrieux, C.</i> .....	19
Tecnologías de inteligencia artificial aplicadas a problemas industriales complejos <i>Rueda, L., Arias, I., Ibáñez, F., Forradellas, R.</i> .....	25

### MT2 - Procesamiento de señales 1

Low noise constant-input limit cycle-free second order digital filters <i>Sarcinelli, M., de Oliveira, M.</i> .....	31
Análisis de Métodos para el Reconocimiento Ruidoso del Habla <i>Graciarena, M., Arini, P.</i> .....	37
Parsing prosodico simplificado para conversão texto-fala <i>da Silva, C., Violaro, F., Runstein, F.</i> .....	43
Estudio de las transiciones entre fonemas rioplatenses para la síntesis de voz utilizando algoritmos temporales <i>Barbuzza, R., Doorn, J.</i> .....	48
Análisis del ruido 1/f en señales de emisión acústica <i>Piotrkowski, R., Gómez, M., Ruzzante, J.</i> .....	53

### MT3 - Modelado y simulación 1

Regularidad de sistemas diferenciales algebraicos semiexplícitos <i>Etchechoury, M., Muravchik, C.</i> .....	58
Analysis of contagion models and their asymptotic distributions <i>Barraza, N., Cernuschi-Frías, B., Cernuschi, F.</i> .....	64

## ANÁLISIS DE MÉTODOS PARA EL RECONOCIMIENTO RUIDOSO DE HABLA

Martín Graciarena, Pedro David Arini

Instituto de Ingeniería Biomédica (IIBM).  
Facultad de Ingeniería - Universidad de Buenos Aires  
Paseo Colón 850 piso 5to - c.p: 1063  
Buenos Aires - Argentina  
e-mail: mgracia@iibm.uba.ar  
arini@iibm.uba.ar

**Abstract:** El propósito de este trabajo es el análisis de métodos propuestos para el reconocimiento ruidoso de voz. Estos métodos se basan en modificar la medida de distancia euclídea utilizada en los modelos de ocultos Markov (HMMs) con objeto de compensar la disminución del módulo del vector de características Mel Cepstral. Demostramos que esta modificación implica, si se cumplen cierta condición en la matriz de covarianza, que la frontera de discriminación de clases pasa por el origen. También se muestra experimentalmente que existe una variación de fase no uniforme de estos vectores para diferentes fonemas de una palabra, debido a una diferente relación señal a ruido local.

**Keywords:** Reconocimiento de Voz, HMM,

### 1. INTRODUCCION

El propósito de este trabajo, es analizar algunas técnicas propuestas que integran el estado del arte en reconocimiento ruidoso, cuyo objetivo es mantener constante su índice de reconocimiento en un determinado rango de condiciones adversas. Consideramos que las conclusiones obtenidas de este análisis, son relevantes para mejorar el índice de reconocimiento del habla en condiciones adversas.

En el reconocimiento del habla en condiciones ideales se logran altos índices de reconocimiento. Sin embargo, estos índices se ven fuertemente degradados cuando las señales de voz a reconocer están contaminadas con algún tipo de ruido. El origen de esta degradación, se debe a las diferentes condiciones que tienen las señales de voz de la base de datos de entrenamiento (*condiciones controladas*) con respecto a las que tienen en el reconocimiento (*condiciones desconocidas*). Algunos autores le

asignan a este fenómeno el nombre de "*condition mismatch*".

Una posible solución sería entrenar el sistema con muestras de voz ruidosa. Esta técnica, denominada "*inmunización al ruido*", presenta un alto costo computacional ya que el tamaño de la base de datos de entrenamiento se multiplica por la cantidad de niveles de relación señal a ruido consideradas. Además, los modelos necesitan mayor cantidad de parámetros debido a la variación ruidosa de las características y supone disponible las muestras de ruido que afectan la señal de voz.

Las técnicas que analizamos se basan en la incorporación de invarianzas en el clasificador, frente a la distorsión de los parámetros. Esta distorsión surge durante el reconocimiento, debido a que el clasificador se entrena únicamente con muestras de voz limpia. Además suponemos conocida la distribución estadística del ruido (*no sus muestras actuales*).

Un antecedente importante para alcanzar dicho objetivo, es la demostración matemática que el módulo de los coeficientes Mel Cepstrum disminuye, cuando se contamina a la señal de voz con ruido blanco, (Mansour and Huang, 1989). La compensación a esta disminución se logra implementando el método de las proyecciones WPM (*Weighted Projection Measure*) (Carlson and Clements, 1994). En este, se calcula la distancia euclídea entre el vector a clasificar (*afectado de ruido blanco*) y el valor medio corregido de la gaussiana. Esta corrección surge de un coeficiente que se obtiene de la proyección del vector a clasificar sobre la media del sistema entrenado.

Nos proponemos realizar un estudio teórico y experimental sobre esta propuesta. Primero desarrollaremos el método de WPM, luego realizaremos el análisis teórico citado y luego presentaremos dos pequeños problemas de este método, como son el problema de las proyecciones negativas y la variación no uniforme de fase para diferentes fonemas de un palabra. Finalmente presentaremos una aplicación del método citado y lo compararemos con el caso del entrenamiento ruidoso.

## 2. ANALISIS DISCRIMINATIVO

### 2.1 Desarrollo de WPM.

El objetivo de modificar la medida de distancia convencional (WED), es la de lograr una medida de distancia que sea robusta a los cambios de amplitud de uno de los vectores. El mecanismo para alcanzar este objetivo es el de multiplicar a uno de los vectores por un coeficiente que modifique la norma del vector.

Si definimos a  $\mathbf{x} \in \mathfrak{R}^N$ , como el vector de coeficientes cepstrales y a  $\mathbf{u} \in \mathfrak{R}^N$  como el vector correspondiente al valor medio de la distribución gaussiana, y  $\Sigma^{-1}$  como la inversa de la matriz de autocovarianza, entonces la distancia euclídea ponderada (WED) será:

$$\text{Dist. Euclídea Pond} = (\mathbf{x} - \mathbf{u})^T \Sigma^{-1} (\mathbf{x} - \mathbf{u}) \quad (1)$$

Si afectamos al vector de medias  $\mathbf{u}$  de un coeficiente  $\lambda$ , de forma de minimizar la distancia al vector  $\lambda \mathbf{u}$  tenemos la distancia ponderada por proyección (WPM):

$$\text{Dist. Proy.} = (\mathbf{x} - \lambda \mathbf{u})^T \Sigma^{-1} (\mathbf{x} - \lambda \mathbf{u}) \quad (2)$$

El coeficiente  $\lambda$  podemos calcularlo derivando la distancia e igualando a cero. El coeficiente  $\lambda^*$  óptimo resulta:

$$\lambda^* = \mathbf{x}^T \Sigma^{-1} \mathbf{u} / \mathbf{u}^T \Sigma^{-1} \mathbf{u} \quad (3)$$

La interpretación geométrica de esta solución es para minimizar la distancia el coeficiente  $\lambda$  debe ser la proyección ortogonal del vector  $\mathbf{x}$  en el vector  $\mathbf{u}$ , de ahí su nombre de medida de distancia por proyección.

Las distancias anteriores se pueden relacionar con el logaritmo de una función de densidad de probabilidad gaussiana, de la forma:

$$b(\mathbf{x}, \mathbf{u}) = (2\pi)^{-N/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{u})^T \Sigma^{-1} (\mathbf{x} - \mathbf{u})\right) \quad (4)$$

si extraemos el logaritmo de la ecuación anterior, tenemos:

$$\log(b(\mathbf{x}, \mathbf{u})) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma|) + (\mathbf{x} - \mathbf{u})^T \Sigma^{-1} (\mathbf{x} - \mathbf{u}) \quad (5)$$

Podemos simplificar el primer término ya que es igual para todas las clases. Finalmente nos queda la log distancia WED:

$$b_{\text{WED}}(\mathbf{x}, \mathbf{u}) = -\frac{1}{2} \log(|\Sigma|) + (\mathbf{x} - \mathbf{u})^T \Sigma^{-1} (\mathbf{x} - \mathbf{u}) \quad (6)$$

Si incorporamos en (6) la medida de distancia por proyección (2) y (3), resulta la log distancia WPM

$$b_{\text{WPM}}(\mathbf{x}, \mathbf{u}) = -\frac{1}{2} \log(|\Sigma|) + (\mathbf{x} - \left(\frac{\mathbf{x}^T \Sigma^{-1} \mathbf{u}}{\mathbf{u}^T \Sigma^{-1} \mathbf{u}}\right) \mathbf{u})^T \Sigma^{-1} (\mathbf{x} - \left(\frac{\mathbf{x}^T \Sigma^{-1} \mathbf{u}}{\mathbf{u}^T \Sigma^{-1} \mathbf{u}}\right) \mathbf{u}) \quad (7)$$

### 2.2 Fronteras Discriminativas de WPM.

El objeto de WPM es independizar el cálculo de las distancias cuando las normas de los vectores cepstrales se contraen por efecto del ruido. Para lograr esta independencia, las fronteras de clasificación, siguiendo la notación de Duda y Hart, (1973), deben ser rectas que pasan por el origen. Para dejar esto claro daremos la siguiente definición:

**DEFINICIÓN:** Para que un clasificador sea *Invariante a Dilataciones o Contracciones* de los vectores a ser clasificados (es decir, que no cambie la clase de los mismos si su norma aumenta o se reduce), sus fronteras de discriminación entre clases deben ser rectas que pasan por el origen.

La definición anterior es sencilla de demostrar. Si los hiperplanos de separación no pasan por el origen entonces existe una contracción donde el vector cruza el hiperplano de separación y por lo tanto cambia de clase.

El propósito del siguiente teorema es demostrar que un clasificador de dos clases con medida de distancia WPM cumple la condición de invariancia, sólo si las dos matrices de covarianza son iguales.

Antes de realizar esta demostración, introduciremos la matriz de proyección de un vector sobre otro. Sean

$u, z, \Sigma^{-1/2}x \in \mathcal{R}^2$  y  $\alpha \in \mathcal{R}$ , se cumple que  $u^T z = 0$ . Si  $z = \Sigma^{-1/2}x + \alpha u$ , se deduce, despejando  $\alpha$  y por la condición de ortogonalidad, (Noble y Daniel, 1989) que  $z = Hx$ , donde  $H = [\Sigma^{-1/2} - (\Sigma^{-1/2}uu^T\Sigma^{-1}) / (u^T\Sigma^{-1}u)]$  es la matriz de proyección de  $x$  en la recta ortogonal a  $u$ .

**TEOREMA 1:** Los puntos de la frontera de clasificación, solución de la igualdad de distancias WPM a ambas medias  $u_1$  y  $u_2$ , son rectas en  $\mathcal{R}^2$ , si las matrices de covarianza de ambas clases  $\Sigma_1$  y  $\Sigma_2$  son iguales.

*Demostración:* La frontera de clasificación son los puntos donde se igualan las distancias a las medias de cada clase. Desarrollaremos primero la ecuación de la distancia para una clase genérica de una forma útil. Operando:

$$\begin{aligned} (x - \lambda u)^T \Sigma^{-1} (x - \lambda u) &= \\ &= (\Sigma^{-1/2}x)^T (\Sigma^{-1/2} - \frac{\Sigma^{-1/2}uu^T\Sigma^{-1}}{u^T\Sigma^{-1}u})x \end{aligned} \quad (8)$$

Entonces si igualamos las distancias WPM a dos medias  $u_1$  y  $u_2$ . Tenemos:

$$\begin{aligned} -\log(|\Sigma_1|) + (\Sigma_1^{-1/2}x)^T (\Sigma_1^{-1/2} - \frac{\Sigma_1^{-1/2}u_1u_1^T\Sigma_1^{-1}}{u_1^T\Sigma_1^{-1}u_1})x &= \\ = -\log(|\Sigma_2|) + (\Sigma_2^{-1/2}x)^T (\Sigma_2^{-1/2} - \frac{\Sigma_2^{-1/2}u_2u_2^T\Sigma_2^{-1}}{u_2^T\Sigma_2^{-1}u_2})x \end{aligned} \quad (9)$$

pasando todos los términos a la izquierda de la igualdad y agrupamos los términos de  $x$  y utilizando la hipótesis que las matrices de covarianza son iguales:  $\Sigma_1 = \Sigma_2 = \Sigma$ .

$$\begin{aligned} (\Sigma^{-1/2}x)^T (\Sigma^{-1/2} - \frac{\Sigma^{-1/2}u_1u_1^T\Sigma^{-1}}{u_1^T\Sigma^{-1}u_1})x - \\ - (\Sigma^{-1/2}x)^T (\Sigma^{-1/2} - \frac{\Sigma^{-1/2}u_2u_2^T\Sigma^{-1}}{u_2^T\Sigma^{-1}u_2})x = 0 \end{aligned} \quad (10)$$

Comparando la ecuación anterior con la expresión de la matriz de proyección de un vector sobre una recta presentada anteriormente, tenemos:

$$\begin{aligned} (\Sigma^{-1/2}x)^T [H_1x - H_2x] &= \\ = x^T [\Sigma^{-1/2}(H_1x - H_2x)] &= 0 \end{aligned} \quad (11)$$

Podemos interpretar la ecuación anterior de la siguiente forma. El producto escalar de  $x$  con la resta de sus proyecciones sobre las rectas ortogonales a  $u_1$  y  $u_2$ , afectado de  $\Sigma^{-1/2}$ , es nulo.

Las posibles soluciones a esta ecuación son dos y ambas son rectas. Gráficamente presentamos las dos soluciones a esta ecuación.

En el primer gráfico, si llamamos  $x$  al vector que pasa por el medio angular de  $u_1$  y  $u_2$ , su proyección sobre la recta ortogonal a  $u_1$  la llamaremos  $x_1$ , ídem con la proyección sobre la recta ortogonal a  $u_2$ . Podemos observar que el vector  $x$  es ortogonal a la resta de las proyecciones  $x_1$  y  $x_2$ , (cuando  $\Sigma = I$ ). En el segundo gráfico el análisis es semejante, solo que la recta pasa por el medio angular de las rectas ortogonales. Por lo tanto las rectas definidas por los puntos  $x$  son solución de la ecuación 11. Demostraremos a continuación analíticamente que estas rectas son solución de la ecuación 11. Para eso, reordenaremos esta ecuación de forma de facilitar la demostración.

$$\begin{aligned} (\Sigma^{-1/2}x)^T [(\Sigma^{-1/2} - \frac{\Sigma^{-1/2}u_1u_1^T\Sigma^{-1}}{u_1^T\Sigma^{-1}u_1}) - \\ - (\Sigma^{-1/2} - \frac{\Sigma^{-1/2}u_2u_2^T\Sigma^{-1}}{u_2^T\Sigma^{-1}u_2})] (\Sigma^{-1/2}x) &= \\ (x)^T [-\Sigma^{-1/2}(\frac{u_1u_1^T}{u_1^T\Sigma^{-1}u_1} + \frac{u_2u_2^T}{u_2^T\Sigma^{-1}u_2})\Sigma^{-1/2}(x)] &= 0 \end{aligned} \quad (12)$$

En la ecuación anterior hemos agrupado los términos de forma de dejar al vector incógnita  $x$ , como un único vector.

*Solución 1:* Demostraremos que la recta representada por  $k^*x$ , donde  $k \in \mathcal{R}$  y  $x = \Sigma^{-1/2}u_1 / (\sqrt{u_1^T\Sigma^{-1}u_1}) + \Sigma^{-1/2}u_2 / (\sqrt{u_2^T\Sigma^{-1}u_2})$ , siendo  $x$  la suma de los dos vectores de medias  $u_1$  y  $u_2$  normalizados, es solución de la ecuación 11. Si llamamos  $m_1^2 = u_1^T\Sigma^{-1}u_1$ ;  $m_2^2 = u_2^T\Sigma^{-1}u_2$ , entonces  $x = \Sigma^{-1/2}u_1 / m_1 + \Sigma^{-1/2}u_2 / m_2$ , luego:

$$\begin{aligned} (\frac{\Sigma^{-1/2}u_1}{m_1} + \frac{\Sigma^{-1/2}u_2}{m_2})^T \cdot \\ [(\Sigma^{-1/2}(\frac{u_1u_1^T}{m_1^2} - \frac{u_2u_2^T}{m_2^2})\Sigma^{-1/2})(\frac{\Sigma^{-1/2}u_1}{m_1} + \frac{\Sigma^{-1/2}u_2}{m_2})] &= \\ = 1 + (\frac{u_1^T\Sigma^{-1}u_2}{m_1m_2} - \frac{u_1^T\Sigma^{-1}u_2}{m_1m_2}) + \\ + (-\frac{u_1^T\Sigma^{-1}u_2u_2^T\Sigma^{-1}u_1}{m_2^2m_1^2} + \frac{u_1^T\Sigma^{-1}u_2u_2^T\Sigma^{-1}u_1}{m_2^2m_1^2}) + \\ + (-\frac{u_2^T\Sigma^{-1}u_1}{m_1m_2} + \frac{u_2^T\Sigma^{-1}u_1}{m_1m_2}) - 1 = 1 + 0 + 0 + 0 - 1 = 0 \end{aligned} \quad (13)$$

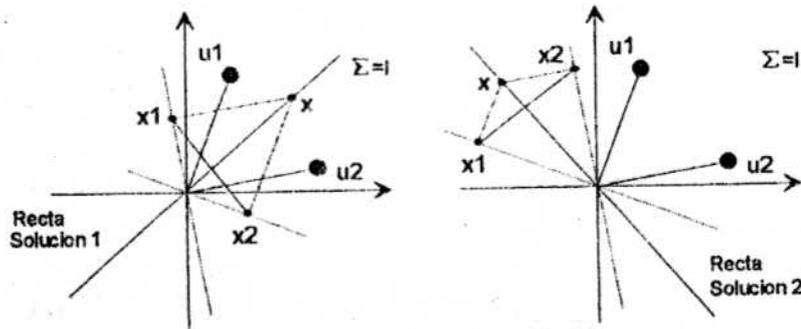


Fig. 1: Soluciones a la ecuación 11

Esto demuestra que la solución propuesta es una recta del tipo  $k^*x$  es solución de la ecuación anterior.

**Solución 2:** Demostraremos que la recta representada por  $k^*x$ , donde  $k \in \mathbb{R}$  y  $x = \Sigma^{-1/2}x_1 / (\sqrt{x_1^T \Sigma^{-1}x_1}) + \Sigma^{-1/2}x_2 / (\sqrt{x_2^T \Sigma^{-1}x_2})$ , y se verifica además que  $z_1^T \Sigma^{-1}u_1 = 0$ ;  $z_2^T \Sigma^{-1}u_2 = 0$  es solución de la ecuación 11. Si llamamos  $n_1^2 = x_1^T \Sigma^{-1}x_1$ ;  $n_2^2 = x_2^T \Sigma^{-1}x_2$ , entonces  $x = \Sigma^{-1/2}x_1 / n_1 + \Sigma^{-1/2}x_2 / n_2$ , luego:

$$\begin{aligned} & \left( \frac{\Sigma^{-1/2}x_1}{n_1} + \frac{\Sigma^{-1/2}x_2}{n_2} \right)^T \\ & \left[ (\Sigma^{-1/2} \left( \frac{u_1 u_1^T}{m_1^2} - \frac{u_2 u_2^T}{m_2^2} \right) \Sigma^{-1/2}) \left( \frac{\Sigma^{-1/2}x_1}{n_1} + \frac{\Sigma^{-1/2}x_2}{n_2} \right) \right] = \\ & = \left( \frac{x_1^T \Sigma^{-1} u_1 u_1^T \Sigma^{-1} x_1}{m_1^2 n_1^2} + \frac{x_1^T \Sigma^{-1} u_1 u_1^T \Sigma^{-1} x_2}{m_1^2 n_1 n_2} + \right. \\ & \quad \left. \frac{x_2^T \Sigma^{-1} u_1 u_1^T \Sigma^{-1} x_1}{m_1^2 n_2 n_1} + \frac{x_2^T \Sigma^{-1} u_1 u_1^T \Sigma^{-1} x_2}{m_1^2 n_2^2} - \right. \\ & \quad \left. \frac{x_1^T \Sigma^{-1} u_2 u_2^T \Sigma^{-1} x_1}{m_2^2 n_1^2} - \frac{x_1^T \Sigma^{-1} u_2 u_2^T \Sigma^{-1} x_2}{m_2^2 n_1 n_2} - \right. \\ & \quad \left. \frac{x_2^T \Sigma^{-1} u_2 u_2^T \Sigma^{-1} x_1}{m_2^2 n_2 n_1} - \frac{x_2^T \Sigma^{-1} u_2 u_2^T \Sigma^{-1} x_2}{m_2^2 n_2^2} \right) \\ & = 0 + \frac{x_1^T \Sigma^{-1} u_1 u_1^T \Sigma^{-1} x_2}{m_1^2 n_2^2} - \frac{x_1^T \Sigma^{-1} u_2 u_2^T \Sigma^{-1} x_1}{m_2^2 n_1^2} = 0 \end{aligned} \quad (14)$$

Un aspecto importante a verificar es que aunque ambas rectas son solución de la ecuación 11, solo una (la primera) discrimina las dos clases. Esta es entonces la frontera de clasificación como la recta separadora de las clases.

Observamos que los valores medios de las clases sólo afectan la dirección de la frontera de discriminación, pero la relación es sólo con la dirección de estos vectores y no con sus amplitudes.

**Corolario 1:** Si un clasificador tiene varianzas iguales en todas las clases, entonces todas las

fronteras de clasificación son rectas que pasan por el origen.

Esto es una extensión directa del teorema anterior.

**Corolario 2:** Si en un clasificador la medida de distancia utilizada ignora el determinante de la matriz de covarianza, es decir utiliza la distancia euclídea ponderada presentada en la ecuación 2, entonces las fronteras de clasificación son rectas que pasan por el origen, *cualquiera sea las matrices de covarianza de las clases.*

Esto se puede visualizar a partir de la ecuación 9, donde si se ignora el determinante de la matriz de covarianza, aunque estos sean diferentes, el desarrollo es totalmente análogo al realizado anteriormente sin necesidad de imponer la condición de igualdad de matrices de covarianza.

### 3 EL PROBLEMA DE LAS PROYECCIONES NEGATIVAS

Otro aspecto importante de estudiar sobre la medida por proyección es el problema de las proyecciones negativas. De acuerdo a cómo fue definido el coeficiente de proyección en la ecuación 3, este puede ser tanto positivo como negativo ya que el numerador puede tener ambos signos. El caso donde es positivo es trivial y es el caso deseable. El caso donde es negativo puede tener un efecto indeseado en la clasificación.

Un ejemplo de este último caso se presenta en el gráfico siguiente. En el mismo se desca la clasificación del patrón  $x$ , en una de dos clases, representada cada una por un vector  $u_1$  y  $u_2$ .

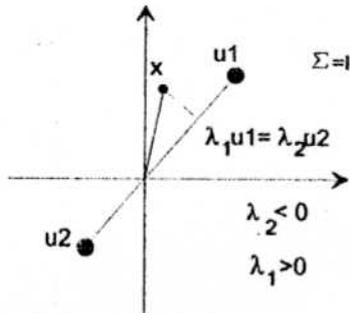


Fig. 2: Problema del Coeficiente de Proyección Negativo

Observamos que este problema no puede ser resuelto debido a que si  $\lambda_1$  es positivo y  $\lambda_2$  es negativo, las dos distancias son iguales.

Si bien este es un caso que es difícil encontrar en la práctica, debido a que los espacios de características son en general multidimensionales, revela una limitación del método propuesto.

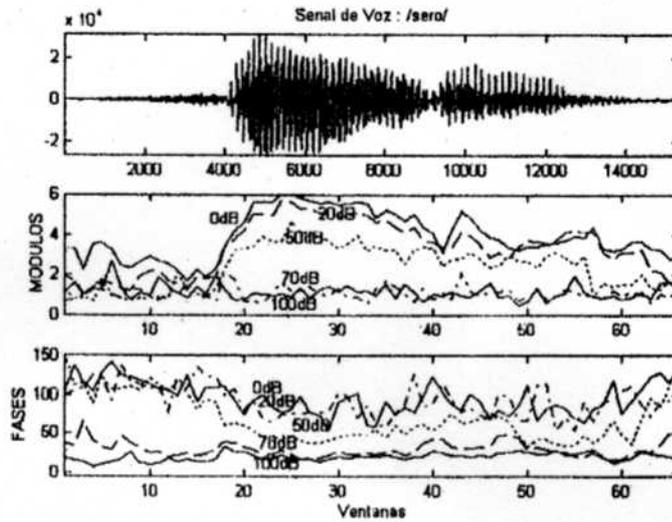


Fig. 3: Evolución no uniforme de la diferencia de fase.

Con objeto de estudiar el impacto de este problema en un caso real, se propone una modificación al método de WPM, donde el coeficiente de proyección es el módulo del coeficiente calculado en la ecuación 3. Llamamos a este método **WPM módulo**.

#### 4 DIFERENCIA DE FASE NO UNIFORME

Uno de estos problemas es que este método compensa variaciones de amplitud pero no compensa las variaciones de fase, siendo estas también significativas.

Al contaminar con ruido una señal de voz, la SNR global es función de la amplitud promedio de la palabra. Cada fonema tiene una SNR particular en función de su amplitud. Lo que implica que las SNR particulares son usualmente menores que la global para fonemas de baja amplitud (*fricativas*) y mayores en fonemas de gran amplitud (*vocales*). Por lo tanto, cuando una palabra es contaminada por ruido blanco las variaciones de los coeficientes Mel Cepstral para cada fonema serán diferentes. Presentamos el siguiente gráfico con objeto de fundamentar esta afirmación.

La primer conclusión, de varias posibles sobre el gráfico 3, es que tanto el módulo como la fase de estos vectores son fuertemente dependientes de la presencia de ruido en la señal. En este sentido se confirma el análisis de Mansour y Huang, acerca de que el módulo del vector cepstral se reduce a medida que la amplitud de ruido aumenta.

La conclusión más importante es la dependencia de la diferencia de fase del vector con la relación señal a ruido. Si además la relación señal a ruido depende de la amplitud de la señal entonces la diferencia de fase depende de la amplitud de la señal. Esto puede observarse comparando la evolución de la fase para el fonema /s/ con la evolución de la fase para el fonema /e/. Se observa que para altas SNR las diferencias de fase de estos fonemas son semejantes, pero cuando la SNR comienza a disminuir, la diferencia de fase del fonema /s/ (de menor amplitud) es mayor que la del fonema /e/ (de mayor amplitud).

El origen de este comportamiento es que la SNR es un parámetro global, calculado con el promedio de amplitudes de la palabra. Como las amplitudes de

cada fonema son diferentes las SNR de cada fonema son diferentes.

## 5 SIMULACIÓN

### 5.1 Base de Datos

Se construyó una base de datos de habla de los dígitos en castellano rioplatense con treinta hablantes masculinos, que emitieron los diez dígitos en tres oportunidades cada uno (un total de 900 emisiones). La base de datos se dividió en dos partes iguales, la primera mitad fue utilizada para el entrenamiento y el resto en la etapa de reconocimiento.

El nivel de cuatización fue de 16 bits a una frecuencia de 22.050 Hz. Los registros se hicieron con un micrófono marca *Shure SM10A* (utilizado en bases de datos de habla internacionales) y en condiciones de laboratorio (ambiente sin ruido).

### 5.2 Modelos de Markov ocultos

Para modelizar los 10 dígitos, se utilizaron modelos ocultos de Markov (HMMs). La cantidad de estados correspondiente a cada palabra o dígito se determinó utilizando criterios fonéticos. Cada estado de los HMMs, fue representado con 3 multigaussianas con el objetivo de caracterizar mejor la estadística de los mismos.

Se utilizaron ventanas de 25 mseg solapadas cada 10 mseg, obteniendo 12 coeficientes Mel Cepstral, 12 Delta y energía de ambos.

El entrenamiento de los HMMs se realizó con muestras de voz sin ruido implementando el algoritmo de Baum-Welch. Cabe agregar, que para el reconocimiento se utilizó el algoritmo de Viterbi.

Las señales de voz a reconocer, se contaminaron con ruido blanco de distribución Gaussiana con las siguientes relaciones señal a ruido (SNR) globales:  $\infty$ , 30, 20, 15, 10, 5 y 0 dB.

La tabla 1 presenta los resultados de reconocimiento ruidoso, empleando los siguientes métodos: distancia Euclídea (WED), distancia por proyecciones (WPM), distancia por proyecciones en módulo (WPM módulo) y finalmente el entrenamiento de los HMMs con ruido (Entrenamiento con ruido). Todos estos resultados se presentan en porcentaje de reconocimiento para diferentes relaciones señal a ruido.

Como comparación del caso más favorable se realizó un entrenamiento de los modelos con muestras de voz afectadas de un determinado nivel de ruido y se reconoció con el mismo nivel de ruido. Este experimento se lo llamó Entrenamiento con Ruido.

Tabla 1 Resultados de Reconocimiento Ruidoso

	SNR $\infty$ dB	30dB	20dB	15dB	10dB	5dB	0dB
Tipo							
WED	96.0	93.4	66.5	44.2	31.4	17.1	10.7
WPM	96.4	95.7	84.6	63.0	40.0	21.9	13.4
WPM módulo	96.4	95.7	85.3	61.8	29.8	20.6	15.9
Entren. con Ruido	96.0	97.0	96.4	90.4	89.5	83.1	65.5

Se observa que el índice de reconocimiento para el caso de la distancia euclídea WED, es muy dependiente de la relación señal a ruido. Apartir de 20 dB su caída a es muy pronunciada. Para el caso de la distancia WPM mejora significativamente el índice de reconocimiento de WED, sobre todo a altas relaciones señal a ruido. El caso propuesto WPM módulo tiene un comportamiento errático. En 20 dB pareciera ser mejor pero a partir de ahí, su desempeño cae respecto del de WPM, para volver a superarlo en 0 dB. Finalmente se verifica que conociendo el tipo y nivel de ruido y pudiendo aparear las condiciones de entrenamiento y reconocimiento se logran índices muy superiores a las técnicas propuestas.

## 6 CONCLUSIÓN

La contribución original del trabajo, es la demostración matemática que la frontera de discriminación de clases con la medida WPM (propuesta por Carlson and Clements, 1994) son rectas que pasan por el origen, si se cumple que las matrices de covarianza de las clases son iguales. También se presentó el problema de las proyecciones negativas, irresoluble con esta medida.

## REFERENCIAS

- Carlson B. A. y Clements B., A Projection-Based Likelihood Measure for Speech Recognition in Noise, *IEEE Transactions on Speech and Audio Processing*, 2, No 1, 97-102, Enero 1994.
- Duda R. O., Hart P. E., *Pattern Clasification and Scene Analysis*, John Wiley & Sons, New York, 1973.
- Mansour, D. and B. H. Juang, "A family of distortion measures based upon projection operation for robust speech recognition", *IEEE Trans. Acoustics, Speech and Signal Processing*, 37, 11, 1659-1671 (1989).
- Noble B. y Daniel J. W., *Algebra Lineal Aplicada tercera edición*, Prentice Hall Hispano Americana, Méjico, 1989.