



Universidad Nacional Autónoma de México

FACULTAD DE INGENIERIA

SEGMENTACION ESTADISTICA DE SERIES DE TIEMPO Y SUS
APLICACIONES EN EL PROCESAMIENTO DIGITAL DE VOZ

TESIS PROFESIONAL

Que para obtener el Título de
INGENIERO EN COMPUTACION
P r e s e n t a n

W U Z H E N Y U
Y O U J I A N Y U

DIRECTOR DE TESIS:

LUIS ANDRES BUZO DE LA PEÑA



México, D. F.

1984



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

INDICE

CAPITULO I.	INTRODUCCION	
I.1	Resumen	1
I.2	Estructura del trabajo	3
CAPITULO II.	CONCEPTOS BASICOS DE PROCESOS ESTOCASTICOS	
II.1	Algunas definiciones	5
II.2	Procesos estocásticos estacionarios	10
II.3	Procesos gaussianos	20
II.4	Procesos cuasi-estacionarios	23
CAPITULO III.	CARACTERIZACION DE PROCESOS ESTACIONARIOS	
III.1	Predicción lineal	25
III.2	Estimación espectral	34
III.3	Modelado de procesos	40
CAPITULO IV.	MEDIDAS DE DISTORSION	
IV.1	Propiedades generales de una medida	45
IV.2	Medida de distorsión Itakúra-Saito	48
CAPITULO V.	ALGORITMO DE SEGMENTACION	
V.1	Planteamiento del problema	53
V.2	Solución teórica	56
V.3	Consideraciones adicionales del algoritmo	60
V.4	Implantación computacional	67
CAPITULO VI.	DISCUSION SOBRE RESULTADOS DE SEGMENTACION	
VI.1	Efectos por variación de parámetros	72
VI.2	Comparación con otros algoritmos	79
CAPITULO VII.	APLICACIONES EN SEGMENTACION	
VII.1	Codificación LPC de tasa de transmisión variable	88

VII.2	Reconocimiento de voz basado en segmentación .	94
CAPITULO VIII.	CONCLUSIONES Y SUGERENCIAS	97
APENDICE I.	PROGRAMAS DE SEGMENTACION	99
APENDICE II.	PROGRAMAS DE SINTESIS DE TASA VARIABLE	108
REFERENCIAS	114

CAPITULO I

INTRODUCCION

I-1. Resumen

En este trabajo se presenta un algoritmo de segmentación de señal de voz por medio de la utilización de la medida de distorsión Itakura-Saito, y algunas aplicaciones en el procesamiento digital de señal de voz.

El algoritmo de segmentación trata de una técnica empleada con éxito para resolver el problema de la segmentación de señales cuasi-estacionarias, estimando el espectro de potencia de procesos estacionarios discretos en el tiempo, y para procesar la estadística de segundo orden de señales digitales. La aplicación que se describe aquí es adecuada cuando el criterio de segmentación está basado en la detección de cambios en la estadística de segundo orden de los procesos estocásticos. Las mismas técnicas pueden ser empleadas en cualquier contexto y con diversos tipos de señales cuasi-estacionarias, siempre y cuando sea la estadística de segundo orden el factor más importante para llevar a cabo la segmentación. Entre las aplicaciones, se pueden encontrar procesamiento de voz, procesamiento de imágenes, análisis automático de señales biomédicas tales como electroencefalogramas y electrocardiogramas, y geofísicas. Las señales que se usan para mostrar la aplicabilidad del método son señales de voz, debido a que son del tipo descrito y que la segmentación de estas señales es de suma importancia para el reconocimiento, el análisis y la síntesis de voz.

En el trabajo, se presentan también dos aplicaciones del algoritmo de segmentación en el procesamiento de señal de voz. Una de ellas es la codificación LPC de tasa de transmisión variable, y la otra es el reconocimiento de voz basado en segmentación.

Tanto el algoritmo de segmentación como los sistemas de aplicación son implantados en el Laboratorio de Cálculo

**Automatizado para el Diseño (CAD) de la División de Estudios de
Posgrado de la Facultad de Ingeniería (DEPFI), UNAM.**

I-2. Estructura del trabajo

El trabajo está compuesto principalmente por dos partes. En la primera parte se revisan los aspectos teóricos indispensables para la estructuración del algoritmo de segmentación. La segunda se dedica a desarrollar el algoritmo y sus aplicaciones.

La primera parte consiste en los capítulos II, III y IV. En el capítulo II, se introducen algunos conceptos básicos tales como la probabilidad, variables aleatorias, procesos estocásticos estacionarios, función de correlación, densidad espectral y los procesos gaussianos. Los lectores que tengan conocimientos sobre estos conceptos, se pueden saltar las primeras tres secciones del capítulo. En el mismo capítulo, se propone un modelo de fuente compuesta para los procesos cuasi-estacionarios.

El capítulo III consta de las técnicas que caracterizan los procesos estacionarios tales como la predicción lineal, estimación espectral y el modelado óptimo de los mismos, así como el algoritmo de Levinson para la solución del sistema de ecuaciones Yule-Walker.

En el capítulo IV se describen las propiedades básicas de la medida de distorsión Itakura-Saito.

La parte de desarrollo está formada por los capítulos V, VI y VII. El capítulo V es el más importante de este trabajo, en donde no solamente se plantea el problema de segmentación y la solución teórica, sino también se toma en cuenta las características más importantes de la señal de voz en el procedimiento de la segmentación. En la última sección de este capítulo se trata la implantación computacional del algoritmo.

En el capítulo VI, se discute el efecto sobre el funcionamiento del algoritmo de segmentación por variación de parámetros. Además una comparación del algoritmo desarrollado con algunos otros existentes.

Las aplicaciones del procedimiento de segmentación se presentan en el capítulo VII. La codificación por predicción lineal (LPC)

de tasa de transmisión variable y el reconocimiento de voz basado en la segmentación son dos ejemplos de muchas posibles aplicaciones de la segmentación.

En el capítulo VIII se encuentran las conclusiones y sugerencias al respecto del trabajo desarrollado.

Los programas que realizan las tareas de segmentación y síntesis de señal de voz se anexan en los apéndices I y II, respectivamente.

CAPITULO II

CONCEPTOS BASICOS DE PROCESOS ESTOCASTICOS

En este capítulo, introducimos los conceptos elementales de procesos estocásticos. En la sección II-1, definimos la variable aleatoria, procesos estocásticos y sus caracterizaciones ([1],[2],[3]). En II-2, limitamos exclusivamente en procesos estocásticos estacionarios, discutimos la representación en el dominio de frecuencia ([1],[4],[5],[6]). En II-3, estudiamos los procesos normales ([1],[2]). Finalmente en II-4, definimos los procesos cuasi-estacionarios y proponemos un modelo para representarlos.

II-1. Algunas Definiciones

Intuitivamente, una variable aleatoria X es una función que se toma valor real aleatoriamente que tiene la propiedad de que para cada conjunto B de los números reales, existe una probabilidad denotada por $P(X \text{ está en } B)$. En la teoría de probabilidad, una variable aleatoria se define como una función univaluada en un espacio muestral.

En orden a dar una definición formal de una variable aleatoria, debemos introducir los conceptos de:

- (1) Un espacio muestral
- (2) Un evento
- (3) Una probabilidad

El *espacio muestral* S de un fenómeno aleatorio es el conjunto de todos los posibles resultados del fenómeno.

Un *evento* E es un conjunto de muestras de S .

La *probabilidad* de un evento E se define como un número real $P(E)$ asociado a tal evento donde $E \in S$. Este número satisface tres postulados:

- (1) $P(E) \geq 0, \forall E \in S$.

- (2) $P(S) = 1$ para el evento seguro S .
- (3) Si E_1 y E_2 son 2 eventos mutuamente exclusivos, entonces
- $$P(E_1 + E_2) = P(E_1) + P(E_2)$$

Definición 1 :

Una variable aleatoria X es una función real cuyo dominio es el espacio muestral S y tal que :

- (1) El conjunto $\{ X \leq x \}$ es un evento para cualquier número real x .
- (2) La probabilidad de los eventos $\{ X = -\infty \}$ y $\{ X = +\infty \}$ es cero:

$$P(X = -\infty) = P(X = +\infty) = 0.$$

En otras palabras X es un mapeo tal : $X : S \rightarrow \mathbb{R}$.

Una variable aleatoria se caracteriza por su función de distribución de probabilidad $F_X(x)$, la cual está definida por :

$$F_X(x) = P(X \leq x).$$

Cuando la derivada de $F_X(x)$ existe, se le denomina función de densidad de probabilidad, denotada por :

$$f_X(x) = dF_X(x)/dx.$$

Generalizando el concepto de variables aleatorias, podemos definir un vector de variables aleatorias.

Definición 2 :

Un vector de variables aleatorias

$$\underline{X} = (X_1, X_2, \dots, X_n)$$

es una función que asigna a cada elemento de S un punto en \mathbb{R}^n , esto es

$$\underline{X} : S \rightarrow \mathbb{R}^n.$$

A \underline{X} también se le llama variables aleatorias conjuntamente distribuidas.

En este caso, la función de distribución de probabilidad se define como:

$$F_{\underline{X}}(\underline{x}) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

y la función de densidad de probabilidad es:

$$f_{\underline{x}}(\underline{x}) = a^n F_{\underline{x}}(\underline{x}) / \partial x_1 \dots \partial x_n$$

si ésta existe.

Definición 3 :

Las variables aleatorias conjuntamente distribuidas X_1, X_2, \dots, X_n son independientes estadísticamente si y sólo si, para todos los números reales x_1, x_2, \dots, x_n , se cumplen que

$$\begin{aligned} F_{\underline{x}}(\underline{x}) &= F_{x_1, x_2, \dots, x_n}(x_1, x_2, \dots, x_n) \\ &= F_{x_1}(x_1) F_{x_2}(x_2) \dots F_{x_n}(x_n) \end{aligned}$$

Definición 4 :

El valor esperado de una función $g(\underline{x})$ está definido por

$$\begin{aligned} E(g(\underline{x})) &= \int_{-\infty}^{\infty} g(\underline{x}) dF(\underline{x}) \\ &= \int_{-\infty}^{\infty} g(\underline{x}) f(\underline{x}) d\underline{x} \quad \text{si } f(\underline{x}) \text{ existe;} \end{aligned}$$

o $E(g(\underline{x})) = \sum_{\forall \underline{x}: P(\underline{x}) > 0} g(\underline{x}) P(\underline{x}).$

En particular,

$$E(\underline{x}) = m = \text{Media de } \underline{x}$$

$$E(\underline{x} - E(\underline{x}))^T (\underline{x} - E(\underline{x})) = V(\underline{x}) = \text{Matriz de covariancia}$$

y $E(\underline{x})^T (\underline{x}) = R(\underline{x}) = \text{Matriz de correlación.}$

Cuando $E(\underline{x}^T \underline{x}) = E(\underline{x}^T) E(\underline{x})$, las componentes de \underline{x} son no correlacionadas entre sí.

Definición 5 :

Un *proceso estocástico* es una colección de variables aleatorias $X(t)$, donde $t \in T$.

En otras palabras, un proceso estocástico es una función que mapea los elementos del espacio muestral S al conjunto de funciones reales definidas en T si restringimos que $X(t)$ sea real.

T es el conjunto de índices. En general, se refiere al tiempo.

Si $T = \{n : -\infty < n < +\infty\}$, el proceso estocástico es discreto

en tiempo.

Si $T = \mathbb{R}$, el proceso es continuo en el tiempo.

De acuerdo con la definición anterior, notamos que $\{X(t_0)\}$ es una variable aleatoria para cualquier tiempo específico t_0 ; y es una función en T para cualquier realización del proceso.

Un proceso estocástico se determina estadísticamente si se conoce la función de distribución de orden n :

$$F(x_1, \dots, x_n; t_1, \dots, t_n) \\ = P(X(t_1) \leq x_1, \dots, X(t_n) \leq x_n)$$

para todo n y t_1, \dots, t_n .

La función de densidad (si existe)

$$f(x_1, \dots, x_n; t_1, \dots, t_n)$$

del proceso $X(t)$ se obtiene derivando la función de distribución con respecto a todas las variables x_k .

La media, la función de autocorrelación y autocovariancia de un proceso están definidas como

$$m_X(t) = m(t) = E(X(t));$$

$$R(t, s) = E(X(t), X(s));$$

$$C(t, s) = E([X(t) - m(t)][X(s) - m(s)]) = R(t, s) - m(t)m(s)$$

respectivamente.

La correlación cruzada de dos procesos $X(t)$ y $Y(t)$ está dada por

$$R_{XY}(t, s) = E(X(t)Y(s))$$

y su covariancia cruzada por

$$C_{XY}(t, s) = R_{XY}(t, s) - m_X(t)m_Y(s).$$

Dos procesos $X(t)$ y $Y(t)$ son no correlacionados si para cualquier t y s , tenemos

$$R_{XY}(t, s) = m_X(t)m_Y(s)$$

i.e. $C_{XY}(t, s) = 0$.

Ellos son ortogonales si

$$R_{XY}(t, s) = 0.$$

Se dice que dos procesos son estadísticamente independientes si

$$\underline{X} = (X(t_1), \dots, X(t_n))$$

es independiente de

$Y = (Y(t_1), \dots, Y(t_n))$
para cualquier $t_1, \dots, t_n, t_1, \dots, t_n$.

11-2. Procesos Estocásticos Estacionarios

Un *proceso estocástico estacionario* es un proceso cuyas estadísticas no cambian con el tiempo. Especificando esta idea podemos dar algunas definiciones de estacionaridad.

Definición 6 :

Un proceso estocástico $X(t)$ es *estacionario en el sentido estricto*, si sus estadísticas no se alteran por un corrimiento en el tiempo. Esto significa que dos procesos

$$X(t) \quad \text{y} \quad X(t+r)$$

tienen las mismas estadísticas para cualquier r .

$$F(x_1, \dots, x_n; t_1, \dots, t_n)$$

$$= F(x_1, \dots, x_n; t_1, \dots, t_n)$$

donde $t_j = t_j + r \quad \forall r, n, t_j, j=1, \dots, n$.

En particular, esto implica que para un proceso estacionario, todas las funciones de distribución de primer orden deben ser idénticas, i.e., $F_1(x)$ no dependen de t , y todas las funciones de distribución de segundo orden sólo depende de la diferencia $t_2 - t_1$. En la práctica, salvo casos muy especiales, solamente podemos conocer las estadísticas de primer y segundo orden. Por lo tanto, es necesario introducir otra definición menos estricta.

Definición 7 :

Un proceso estocástico $X(t)$ es *estacionario en sentido amplio*, si sus estadísticas de primer orden no dependen del tiempo, y las de segundo orden son exclusivamente función de la diferencia $t_2 - t_1$.

Esta definición es equivalente a decir que el valor esperado de $X(t)$ es una constante y su función de autocorrelación depende únicamente de $t_2 - t_1$, esto es

$$E(X(t)) = m = \text{constante.}$$

$$E(X(t_2)X(t_1)) = R_x(t_2 - t_1) = R_x(s).$$

La estacionaridad dada por la definición 6 es mucho más fuerte que la de definición 7. Las dos definiciones son equivalentes

solamente en muy pocos procesos especiales tales como los procesos gaussianos, los cuales se definirán posteriormente.

En las literaturas actuales, existen otros tipos de estacionariedad. Entre ellos, encontramos procesos asintóticamente estacionarios, de incrementos estacionarios, periódicamente estacionarios y estacionarios en un intervalo dado, etc. ([1]).

En los siguientes capítulos, los procesos estacionarios referidos son de sentido amplio. La estacionariedad se caracteriza por sus estadísticas de primer y segundo orden. Aún así, es difícil de calcular la media y la correlación de un proceso en la práctica, porque en general solamente conocemos una realización del proceso, y es muy costoso o imposible de repetirlo. Entonces cuándo y cómo determinar la media y la correlación con una realización del proceso es un problema de interés. El teorema de ergodicidad responde esta pregunta.

Teorema 1 :

Sea $x(t)$ una realización de un proceso $X(t)$, definimos

$$m_T = \frac{1}{2T} \int_{-T}^T x(t) dt$$

$$R_T(s) = \frac{1}{2T} \int_{-T}^T x(t+s)x(t) dt.$$

Entonces

$$\lim_{T \rightarrow \infty} m_T = m = EX(t)$$

si y sólo si

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R(s) ds = m^2;$$

y
$$\lim_{T \rightarrow \infty} R_T(s) = R(s) = EX(t+s)X(t)$$

si y sólo si

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R_V(r) dr = R^Z(s)$$

donde $Y(t) = X(t+s)X(t)$.

La versión discreta del teorema de ergodicidad está dada a continuación.

Teorema 1' :

Sea $x(n)$ una realización de un proceso $X(n)$, definimos

$$m_N = \frac{1}{2N+1} \sum_{n=-N}^N x(n)$$

$$R_N(s) = \frac{1}{2N+1} \sum_{n=-N}^N x(n+m)x(n).$$

Entonces

$$\lim_{N \rightarrow \infty} m_N = m = EX(n)$$

$N \rightarrow \infty$

si y sólo si

$$\lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N R(k) = m^2;$$

y $\lim_{N \rightarrow \infty} R_N(m) = R(m) = EX(n+m)X(n)$

$N \rightarrow \infty$

si y sólo si

$$\lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{k=-N}^N R_V(k) = R^Z(m)$$

donde $Y(n) = X(n+m)X(n)$.

En los procesos estacionarios en sentido amplio, la función de correlación tiene un desempeño muy importante. A continuación, listamos algunas propiedades de ella.

Sea $R(s)$ la función de autocorrelación de un proceso estacionario $X(t)$

$$R(s) = R(t_2 - t_1) = E(X(t_2)X(t_1))$$

donde $s = t_2 - t_1$.

Las propiedades de $R(s)$ se resumen como sigue:

1. $R(0) = m^2 = (E(X(t)))^2$.

Entonces el valor cuadrático medio de un proceso se puede obtener simplemente haciendo $s=0$ en $R(s)$.

2. $R(s) = R(-s)$.

La función de autocorrelación es par con respecto a s .

3. $|R(s)| \leq R(0)$.

El máximo valor de autocorrelación siempre ocurre en $s=0$. Esto se demuestra fácilmente considerando

$$\begin{aligned} & E((X_1 + X_2)^2) \\ &= E(X_1^2 + X_2^2 + 2X_1X_2) \geq 0 \\ & E(X_1^2 + X_2^2) = 2R(0) \\ & \geq |E(2X_1X_2)| = |2R(s)| \end{aligned}$$

por lo tanto :

$$R(0) \geq |R(s)|.$$

4. Si la media de $X(t)$ no es cero, $R(s)$ tiene una componente constante:

$$R(s) = m^2 + C(s).$$

5. Si $X(t)$ tiene un componente periódico, $R(s)$ también tiene una componente periódica con el mismo periodo.

6. Si $X(t)$ es ergódico y con media cero, y no tiene componentes periódicos, entonces

$$\lim_{|s| \rightarrow \infty} R(s) = 0.$$

$$|s| \rightarrow \infty$$

7. Las funciones de autocorrelación no pueden ser arbitrarias, sino cumplen necesariamente

$$F(R(s)) \geq 0 \quad \forall \omega$$

donde

$$F(R(s)) = \int_{-\infty}^{\infty} R(s) e^{-j\omega s} ds$$

es la transformada de Fourier de $R(s)$.

Esta propiedad equivale a decir que $R(s)$ es positiva

semi-definida.

Como funciones determinísticas, un proceso estacionario puede expandirse en series. Existen muchas formas de expandirlo, como expansión en serie de Taylor, en serie de Fourier, en sus muestras ([1]) y expansión Karhunen-Loeve ([7]). Pero tenemos interés particular en la siguiente expansión, la cual nos lleva a una representación en el dominio de frecuencia de un proceso estacionario ([4],[6]).

Teorema 2 :

Sea $X(t)$ un proceso estocástico estacionario, y para $\forall E > 0, T > 0$

existen variables aleatorias $x_0, x_1, \dots, x_n, \dots$ y constantes (distintas) $w_0, w_1, \dots, w_n, \dots$

$$E x_n = 0 \quad \forall n$$

$$E x_n x_m = b_n \delta_{n-m} \quad \forall n, m$$

donde δ_{n-m} es la delta Kronecker :

$$\delta_n = 1 \quad \text{si } n=0$$

$$\delta_n = 0 \quad \text{si } n \neq 0$$

y $N = N(E, T)$, cuando $n > N$

$$E |X(t) - \sum_{k=1}^n x_k e^{j\omega_k t}|^2 < E \quad -T < t < T.$$

La demostración del teorema será más fácil después de la siguiente discusión.

Definición 8 :

La función $Z(\omega)$ está definida como

$$Z(\omega) = \sum_{k: \omega_k \in \omega} w_k$$

donde ω es un intervalo conjunto de ω ;

Cuando $\omega = (-\infty, \omega)$, se denota

$$Z(\omega) = Z((-\infty, \omega)).$$

La función Z tiene las siguientes propiedades:

1. $E(Z(\omega)) = 0, \quad \forall \omega;$

$$E(Z(w)) = 0, \quad \forall w.$$

$$2. Z(\Delta w_1 + \Delta w_2) = Z(\Delta w_1) + Z(\Delta w_2)$$

si Δw_1 y Δw_2 son disjuntos.

$$3. E(Z(\Delta w_1) Z^*(\Delta w_2)) = 0$$

si Δw_1 y Δw_2 son disjuntos; ó

$$E\{[Z(w_1 + \Delta w_1) - Z(w_1)] \cdot$$

$$[Z^*(w_2 + \Delta w_2) - Z^*(w_2)]\} = 0$$

si $(w_1, w_1 + \Delta w_1)$ y $(w_2, w_2 + \Delta w_2)$ son disjuntos,

donde f^* significa el conjugado de f .

$$4. E|Z(\Delta w)|^2 = \sum_{k: w_k \in \Delta w} E|x_k|^2; \text{ ó}$$

$$E|Z(w)|^2 = \sum_{k: w_k < w} E|x_k|^2.$$

$$5. dZ(w) = Z(dw).$$

Definición 9 :

La integral Fourier-Stieltjes está definida como

$$\int_{-\infty}^{\infty} e^{j\omega t} Z(dw) = \lim_{T \rightarrow \infty} \int_{-T}^T e^{j\omega t} Z(dw)$$

donde

$$\int_{-T}^T e^{j\omega t} Z(dw) = \lim_{\max(w_k - w_{k-1}) \rightarrow 0} \sum_{k=1}^N e^{j\omega t} Z(\Delta w_k).$$

Del teorema 2 y la definición 9, inmediatamente deducimos que cualquier proceso estacionario $X(t)$ se puede expresar como

$$X(t) = \int_{-\infty}^{\infty} e^{j\omega t} Z(dw) = \int_{-\infty}^{\infty} e^{j\omega t} dZ(w).$$

Entonces

$$R(s) = E\{x(t+s)x^*(t)\}$$

$$\begin{aligned}
&= E \left\{ \int_{-\infty}^{\infty} e^{j\omega_1(t+\tau)} dZ(\omega_1) \cdot \int_{-\infty}^{\infty} e^{-j\omega_2\tau} dZ^*(\omega_2) \right\} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{j(\omega_1(t+\tau) - \omega_2\tau)} E dZ(\omega_1) dZ^*(\omega_2) \\
&= \int_{-\infty}^{\infty} e^{j\omega\tau} E |dZ(\omega)|^2 \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega\tau} dF(\omega)
\end{aligned}$$

donde $dF(\omega) = \frac{1}{2\pi} E |dZ(\omega)|^2$.

Cuando $dF(\omega)$ se puede expresar como

$$dF(\omega) = S(\omega) d\omega \quad \text{ó}$$

$$F(\omega) = \int_{-\infty}^{\omega} S(y) dy$$

llegamos a una expresión muy interesante

$$\begin{aligned}
R(s) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega\tau} dF(\omega) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega\tau} S(\omega) d\omega
\end{aligned}$$

A la función $S(\omega)$ se le denomina la densidad espectral de potencia del proceso estacionario $X(t)$.

La última ecuación es exactamente la transformada inversa de Fourier, por lo tanto, $S(\omega)$ es la transformada de Fourier de $R(s)$.

$$S(\omega) = \int_{-\infty}^{\infty} R(s) e^{-j\omega s} ds$$

Hemos mencionado antes que $R(s)$ es positiva semi-definida, es decir que su transformada de Fourier $S(\omega)$ es no negativa. Este resultado se justifica fácilmente de la siguiente manera

$$dF(\omega) = S(\omega) d\omega \quad \text{pero}$$

$$dF(\omega) = \frac{1}{2\pi} E\{dZ(\omega)\}^2 \geq 0 \quad \text{y}$$

$$d\omega > 0$$

por lo tanto

$$S(\omega) \geq 0 \quad \forall \omega.$$

Ahora damos un bosquejo de la demostración del teorema 2, el cual es equivalente a decir

$$X(t) = \int_{-\infty}^{\infty} e^{j\omega t} dZ(\omega).$$

Sea $X(t)$ un proceso estacionario, definimos

$$Z_T(\omega) = \frac{1}{2\pi} \int_{-T}^T \frac{e^{-j\omega t} - 1}{-jt} X(t) dt$$

Primero tenemos que demostrar la existencia de la integral cuando $T \rightarrow \infty$. Esto no es más que demostrar que para todo T, S

$$E\{Z_T(\omega) - Z_S(\omega)\}^2 = 0 \quad T, S \rightarrow \infty$$

Después probamos que

$$Z(\omega) = Z_T(\omega) \quad T \rightarrow \infty$$

cumple las propiedades listadas anteriormente. Finalmente demostramos que

$$E\{X(t) - \int_{-\infty}^{\infty} e^{j\omega t} dZ(\omega)\}^2 = 0.$$

La demostración detallada se encuentra en [4]. Otra forma interesante de demostrarlo ([6]) se basa en mapeos en espacio de

Hilbert.

Hasta ahora no hemos analizado cómo afecta un sistema lineal a un proceso estacionario. Sea L un sistema lineal invariante con el tiempo con respuesta al impulso $h(t)$, $X(t)$ un proceso estacionario como entrada al sistema y $Y(t)$ la salida correspondiente. Entonces

$$\begin{aligned}
 Y(t) &= \int_{-\infty}^{\infty} h(s)X(t-s)ds \\
 &= \int_{-\infty}^{\infty} h(s) \left[\int_{-\infty}^{\infty} e^{j\omega(t-s)} dZ_X(\omega) \right] ds \\
 &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(s)e^{-j\omega s} ds \right] e^{j\omega t} dZ_X(\omega) \\
 &= \int_{-\infty}^{\infty} e^{j\omega t} H(\omega) dZ_X(\omega) \\
 &= \int_{-\infty}^{\infty} e^{j\omega t} dZ_Y(\omega)
 \end{aligned}$$

donde $H(\omega)$ es la transformada de Fourier de $h(t)$ y $dZ_Y(\omega) = H(\omega)dZ_X(\omega)$.

Se observa que $Y(t)$ también es un proceso estacionario, con un espectro

$$\begin{aligned}
 dF_Y(\omega) &= |H(\omega)|^2 dF_X(\omega) \\
 f_Y(\omega) &= |h(\omega)|^2 f_X(\omega).
 \end{aligned}$$

Cuando el proceso es discreto, podemos tener de forma análoga los resultados similares:

$$\begin{aligned}
 \text{si } X(n) &= \int_{-n}^n e^{j\omega n} dZ(\omega) \\
 R(m) &= EX(n+m)X^*(n) \\
 &= EX(n+m)X(n) \text{ si se trata de procesos reales,}
 \end{aligned}$$

entonces

$$S(w) = \sum_m e^{-jw m} R(m)$$

$$R(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jw m} S(w) dw$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jw m} dF(w)$$

donde $dF(w) = \frac{1}{2\pi} E |dZ(w)|^2$.

Cuando excitamos $X(n)$ a un sistema lineal invariante con el tiempo, con respuesta al impulso $h(n)$, entonces la función de densidad espectral de potencia de la salida $Y(n)$ se obtiene por

$$S_Y(w) = |H(w)|^2 S_X(w)$$

donde

$$H(w) = \sum_{n=-\infty}^{\infty} h(n) e^{-jw n}$$

es la transformada Z de $h(n)$ evaluada en el círculo unitario.

Los métodos convencionales de análisis de sistemas lineales como la convolución, transformada de Fourier y transformada Z, se puede ver en [8].

Finalmente, damos una definición de ruido blanco.

Definición 10 :

Se dice que un proceso estacionario es un ruido blanco si su densidad espectral de potencia es una constante.

Esta definición es equivalente a decir que $X(t_1)$ y $X(t_2)$ son no correlacionados entre sí cuando $t_1 \neq t_2$.

Un ruido blanco puede ser continuo o discreto.

II-3. Procesos Gaussianos

Definición 11 :

X_1, \dots, X_n son variables aleatorias conjuntamente gaussianas, si la función de densidad de probabilidad está dada por:

$$f_{\underline{x}}(\underline{x}) = \frac{1}{(2\pi)^{n/2} |C|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{m})^T C^{-1}(\underline{x}-\underline{m})\right\}$$

donde $\underline{m} = [m_i] = [E(X_i)]$

$C = [C_{ij}] = [\text{Covariancia}(X_i, X_j)]$

$|C|$ es la determinante de la matriz C.

Definición 12 :

Un proceso $\{X(t), t \in T\}$ es gaussiano si para cualquier entero n y cualquier subconjunto $\{t_1, \dots, t_n\}$ de T, las n variables $X(t_1), \dots, X(t_n)$ son conjuntamente gaussianas. Es decir

$$f_{\underline{x}}(\underline{x}) = \frac{1}{(2\pi)^{n/2} |C|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}(t))^T C^{-1} \underline{x}(t)\right\}$$

para cualquier $\underline{x}(t) = (X(t_1), \dots, X(t_n))$

donde $C = [C_{ij}] = [\text{Cov}(X(t_i), X(t_j))] \forall i, j = 1, \dots, n$.

De la definición anterior, observamos inmediatamente que conociendo la media y la función de autocorrelación (o variancia) de \underline{X} , podemos obtener la densidad de probabilidad, y luego conocer las estadísticas de cualquier orden del proceso gaussiano.

Una propiedad fundamental de procesos gaussianos es:

Si la entrada de un sistema lineal es un proceso gaussiano, la salida sigue siendo un proceso gaussiano. Esta propiedad es una consecuencia del siguiente teorema.

Sean X_1, \dots, X_n variables aleatorias conjuntamente gaussianas. Sean $Y_1, \dots, Y_m = \underline{Y}$ funciones lineales de \underline{X} :

$$\underline{Y} = A\underline{X} + \underline{b} \quad \text{donde } A \text{ es una matriz de } n \times m, \\ \underline{b} \text{ es un vector columna.}$$

Entonces Y_1, \dots, Y_m también son gaussianas y

conjuntamente distribuidas. La demostración se encuentra en (p.90 [2]).

Otra propiedad interesante es que las variables gaussianas no correlacionadas son independientes. La demostración es sencilla por lo siguiente:

Si las variables son no correlacionadas,

$$C_{ij} = 0 \quad \forall i \neq j, \text{ y } C_{ii} \neq 0$$

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{X}) &= \frac{1}{(2\pi)^{n/2} (C_{11} \dots C_{nn})^{1/2}} \exp\left\{-\sum_{i=1}^n \frac{(x_i - m_i)^2}{2C_{ii}}\right\} \\ &= \prod_{k=1}^n \frac{1}{(2\pi C_{kk})^{1/2}} \exp\left\{-\frac{(x_k - m_k)^2}{2C_{kk}}\right\} \\ &= f_{X_1}(X_1) \dots f_{X_n}(X_n) \end{aligned}$$

Concepto de Entropía:

Este concepto es una herramienta útil en los temas posteriores. Aquí sólo describimos algunas definiciones y el principio, se puede consultar los detalles en las referencias ([9], [10]).

Consideremos un espacio de probabilidad S y una partición A de S . Esto es, la unión de los eventos mutuamente exclusivos A_k es igual a S .

Definición 13 :

La entropía $H(A)$ de A es:

$$H(A) = -\sum_{k=1}^n P_k \ln P_k \quad \text{donde } P_k = P(A_k)$$

No es difícil de observar que los eventos seguros y los con probabilidad cero contribuyen entropía cero. Además $H(A)$ llega su valor máximo cuando $P_k = 1/n$, $\forall k = 1, \dots, n$.

Para una variable x continua, la entropía es:

$$H(x) = - \int f(x) \ln f(x) dx = -E(\ln(f(x))).$$

Generalizando el concepto, se define la entropía de un vector aleatorio en lo siguiente:

$$H(X_1, \dots, X_r) = -E(\ln f(X_1, \dots, X_r)).$$

Definición 14 :

Sea $X(n)$ un proceso estacionario discreto, la razón de entropía H_n de tal proceso es el límite

$$H_n = \frac{1}{r} H(X_1, \dots, X_r) \quad \text{cuando } r \rightarrow \infty$$

Cuando deseamos estimar P_i en la ausencia de cualquier información a priori, un criterio de optimalidad es maximizar la entropía para obtener P_i .

Dadas la media y la variancia de una variable X , se puede determinar la función de densidad de probabilidad más justa usando el principio de máxima entropía:

$$\begin{aligned} &\text{Maximizar } - \int f(x) \ln f(x) dx \\ &\text{Sujeto a } \begin{cases} \int x f(x) dx = m \\ \int (x-m)^2 f(x) dx = \text{Var}(x) = \text{cte.} \end{cases} \end{aligned}$$

La solución del problema es

$$f(x) = \frac{1}{(2\pi \text{Var}(x))^{1/2}} \exp\left(-\frac{(x-m)^2}{2\text{Var}(x)}\right).$$

De la misma manera, podemos deducir que conociendo la media y la función de autocorrelación de un proceso estocástico $X(n)$, la razón de entropía llega al valor máximo cuando $X(n)$ es gaussiano. En particular, el ruido blanco gaussiano es el proceso que maximiza la entropía conociendo la media y la potencia promedia.

11-4. Procesos Cuasi-estacionarios

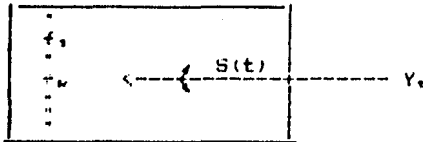
En la realidad, hay muchos procesos que no son estacionarios durante todo el tiempo. Pero algunos de ellos se comportan parecidos a los procesos estacionarios. Por tanto, introducimos otro tipo de estacionariedad (111).

Definición 15 :

Un proceso $X(t)$ es *cuasi-estacionario*, si la ecuación (11.1) se cumple para todo el tiempo de un intervalo dado.

En una gran diversidad de aplicaciones prácticas, se presenta la necesidad de extraer información de señales cuasi-estacionarias o de señales que pueden ser modeladas adecuadamente como procesos estocásticos cuasi-estacionarios.

Para describir una señal cuasi-estacionaria desde el punto de vista de teoría de información se requiere del concepto de fuente compuesta ([11], [12]), la cual está formada por una familia (posiblemente no contable) de subfuentes $f_k, k \in I$, donde I es un conjunto de índices. Aquí cada f_k representa la descripción probabilística completa de una fuente que genera un proceso estacionario (x_k) , así como de un proceso compuesto o proceso conmutador $(S(t))$, donde $S(t)$ toma valores en el conjunto I . La fuente compuesta tiene como salida la señal generada por la subfuente f_k cuando $S(t)=k$, es decir, el conmutador cambia en forma aleatoria de posición entre las posibles subfuentes f_k , produciendo como salida el proceso generado por la fuente a la cual está apuntando. Esto se ilustra en la siguiente figura:



Escribiendo en forma algebraica, la función de esta fuente compuesta es:

$$Y_t = X_{f_i}, \quad t \in [t_1, t_2]$$

donde X_{f_i} es la salida de la subfuente f_i en el tiempo t y t_1, t_2 son dos instantes en que el switch cambia de su posición. Se puede ver que, el proceso Y_t es la concatenación de los procesos X_{f_i} .

Si el conmutador permanece en una misma subfuente durante intervalos de tiempo grandes, el proceso generado puede servir como modelo de proceso cuasi-estacionario.

CAPITULO III

CARACTERIZACION DE PROCESOS ESTACIONARIOS

De este capítulo en adelante, los procesos son discretos excepto aclaraciones específicas. Además, sin perder la generalidad, suponemos que los procesos tienen media cero. Esta suposición es válida porque cuando la media no es cero podemos siempre restarle una constante y trabajar con el nuevo proceso sin alterar los resultados.

III-1. Predicción Lineal

Sea $X(n)$ un proceso estacionario discreto con media cero, el problema de predicción consiste en estimar $x(n)$ con un conjunto de las muestras pasadas, de acuerdo con cierto criterio de optimalidad. Esto es

$$\text{Min } I(x(n), x(n))$$

$$x(n)$$

donde $x(n) \in M \subset \{x(k) : k < n\}$.

Cuando f es lineal, es decir que M es generado por combinaciones lineales de los elementos de un subconjunto de $\{x(k) : k < n\}$, entonces la predicción es lineal.

Existen muchos criterios de optimalidad, uno de los más comunes y usados es el de los mínimos cuadrados. A continuación, este va a ser nuestro criterio de optimalidad para predicción lineal. En este caso, el problema de predicción lineal se puede plantear como

$$\text{Min } E|x(n) - \hat{x}(n)|^2$$

$$x(n)$$

donde $x(n) \in M \subset V\{x(k) : k < n\}$.

$V(A)$ es el espacio lineal generado por los elementos del conjunto A .

Leorema 1 (Principio de ortogonalidad) ([6]):

Sean $X(n)$ un proceso estacionario, $M \subset \{x(k) : k < n\}$,

entonces $\hat{x}(n|M)$ es la mejor predicción lineal de $x(n)$ dado M en el sentido de mínimos cuadrados, si y sólo si

$$e(n) = x(n) - \hat{x}(n|M)$$

es ortogonal a M , esto es

$$E(e(n)e(m)) = 0 \quad \forall m \in M.$$

El teorema se demuestra simplemente aplicando la definición de mínimos cuadrados.

Para un proceso $X(n)$ dado, el error cuadrático medio no puede ser arbitrariamente pequeño, sino tiene una cota inferior. La cota se establece cuando la predicción lineal se basa en todas las muestras anteriores, es decir

$$\text{Min } E|x(n) - \hat{x}(n)|^2$$

$$\hat{x}(n)$$

donde $\hat{x}(n) \in M = V(x(k) : k \leq n)$.

Para encontrar esta cota inferior, necesitamos introducir unos conceptos antes.

Definición 1 :

Sea $X(n)$ un proceso estacionario con media cero, definimos

$$M_m = V(x(k) : k \leq m) \quad \forall m.$$

Además

$$M_{\infty} = \bigcup_m M_m = \text{Unión de todos los } M_m;$$

$$M_{-\infty} = \bigcap_m M_m = \text{Intersección de los } M_m.$$

Definición 2 :

Sea $X(n)$ un proceso estacionario, entonces

$X(n)$ es regular si $M_{-\infty} = \emptyset = \text{conjunto vacío};$

$X(n)$ es singular si $M_{\infty} = M_{-\infty}$.

Cuando $X(n)$ es singular, se puede lograr una predicción perfecta; cuando $X(n)$ es regular, $x(m) \rightarrow 0$, cuando $m \rightarrow \infty$. En general, un proceso no es singular ni regular, pero siempre es posible descomponer el proceso en dos ortogonales en M_{∞} , uno regular y el otro singular. Como la parte singular se puede predecir exactamente, solo nos interesa la parte regular. Por lo tanto

podemos suponer que los procesos son regulares sin perder la generalidad.

Para un proceso estacionario no-singular

$$\begin{aligned} M_k &\neq M_{k+1} && \forall \\ M_k &\subset M_{k+1} && \forall k. \end{aligned}$$

Este resultado se debe a la estacionaridad del proceso.

Teorema 2 (Descomposición de Wold) ([6]):

Sea $X(n)$ un proceso estacionario con media cero, existen constantes c_0, c_1, \dots y variables aleatorias ortonormales con media cero $u(n), u(n-1), \dots$ tales que

$$x(n) = \sum_{k=0}^{\infty} c_k u(n-k) \quad \forall n. \quad (\text{III.1})$$

Demostración :

Sea $x(n|M_{n-1})$ la mejor predicción lineal de $x(n)$ dado M_{n-1} , entonces por el principio de ortogonalidad

$$\begin{aligned} x(n) &= x(n|M_{n-1}) + e(n) \\ &= x(n|M_{n-1}) + c_0 u(n) \end{aligned}$$

donde $E|u(n)|^2 = 1$ y $u(n)$ es ortogonal a M_{n-1} ,

$$M_n = V(M_{n-1}, u(n));$$

De la misma forma

$$M_{n-1} = V(M_{n-2}, u(n-1))$$

tomando en cuenta que $Eu(n)u(n-1) = 0$;

Siguiendo el mismo procedimiento, podemos llegar a

$$M_n = V(M_{-\infty}, u(k) : k \leq n).$$

Pero $M_{-\infty} = \emptyset$, por lo tanto

$$M_n = V(u(k) : k \leq n). \quad (\text{III.2})$$

Por la estacionaridad del proceso, las constantes no dependen de n

$$x(n) = \sum_{k=0}^{\infty} c_k u(n-k) \quad \forall n.$$

Una vez que se tenga la descomposición de Wold, la mejor predicción se obtiene en forma muy sencilla :

$$x(n|M_m) = \sum_{k=0}^{\infty} c_k u(n-k) \quad n > m \quad (III.3)$$

$$y \quad E|x(n) - x(n|M_m)|^2 = \sum_{k=0}^{n-m-1} c_k. \quad (III.4)$$

Cuando $m=n-1$, tenemos la cota inferior de error cuadrático medio,

$$E\bar{\epsilon}_0 = E|x(n) - x(n|M_{n-1})|^2 \\ = c_{\bar{\epsilon}}.$$

Ahora el problema es encontrar $c_{\bar{\epsilon}}$.

Como $u(n)$ es un elemento de M_n , $u(n)$ puede ser expresado en combinaciones lineales de $x(m)$, $m \leq n$.

$$u(n) = \sum_{k=0}^{\infty} d_k x(n-k) \\ = \int_{-\infty}^{\infty} e^{j(n-k)\omega} W(\omega) dZ(\omega) \quad (III.5)$$

$$\text{donde } W(\omega) = \sum_{k=0}^{\infty} d_k e^{-jk\omega}. \quad (III.6)$$

Ahora demostramos que

$$S(\omega)W(\omega)W^*(\omega) = 1 \quad \text{si } S(\omega) \text{ existe.}$$

Por la definición de $u(n)$, sabemos que $\{u(n)\}$ es un conjunto de variables aleatorias ortonormales, esto es

$$Eu(n)u^*(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(n-m)\omega} W(\omega)W^*(\omega)S(\omega) d\omega \\ = \delta_{n-m}$$

Por lo tanto

$$S(\omega)W(\omega)W^*(\omega) = 1. \quad (III.7)$$

Sea $S_1(w) = \frac{1}{W(w)}$, entonces

$$S(w) = S_1(w)S_1^*(w). \quad (III.8)$$

Como $\text{Ex}(n-k)u(n) = 0 \quad \forall k > 0$, i.e.

$$\begin{aligned} \text{Ex}(n-k)u(n) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-jkw} S(w) W^*(w) dw \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-jkw} S_1(w) dw \\ &= 0 \end{aligned}$$

Esto equivale a decir que $S_1(w) \notin V(e^{-jkw}, k \geq 0)$

$$S_1(w) = \sum_{k=0}^{\infty} a_k e^{-jkw}. \quad (III.9)$$

Pero

$$\begin{aligned} x(n) &= \int_{-\pi}^{\pi} e^{jn\omega} dZ(w) \\ &= \int_{-\pi}^{\pi} e^{jn\omega} S_1(w) W(w) dZ(w) \\ &= \sum_{k=0}^{\infty} a_k \int_{-\pi}^{\pi} e^{j(n-k)\omega} W(w) dZ(w) \\ &= \sum_{k=0}^{\infty} a_k u(n-k) \end{aligned}$$

Por lo tanto :

$$a_k = c_k \quad \forall k \geq 0. \quad (III.10)$$

Suponemos que $S(w) > 0$, entonces existe $\ln S(w)$ con una expansión en serie de Fourier

$$\begin{aligned}
 \ln S(w) &= \sum_{k=-\infty}^{\infty} b_k e^{jkw} \\
 &= \left(\frac{1}{2} b_0 + \sum_{k=-\infty}^{-1} b_k e^{jkw} \right) + \\
 &\quad \left(\frac{1}{2} b_0 + \sum_{k=1}^{\infty} b_k e^{jkw} \right) \\
 &= B(w) + B^*(w) \tag{III.11}
 \end{aligned}$$

i.e.
$$\begin{aligned}
 S(w) &= e^{B(w)} e^{B^*(w)} \\
 &= S_1(w) S_1^*(w)
 \end{aligned}$$

entonces

$$\begin{aligned}
 a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{B(w)} dw \\
 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp \left(\frac{1}{2} b_0 + \sum_{k=-\infty}^{-1} b_k e^{jkw} \right) dw \\
 &= \exp \left(\frac{b_0}{2} \right)
 \end{aligned}$$

pero
$$b_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S(w) dw$$

por lo tanto, cuando $S(w) > 0$

$$\begin{aligned}
 E B_0 &= E \left[\ln(n) - \ln(n! M_{n-1}) \right]^2 \\
 &= |a_0|^2 \\
 &= \exp(b_0) \\
 &= \exp \left(\int_{-\pi}^{\pi} \ln S(w) \frac{dw}{2\pi} \right) \tag{III.12}
 \end{aligned}$$

Cuando $S(w)$ se puede descomponer en $S_1(w) S_1^*(w)$,
 pues

$$W(w) = \frac{1}{S_x(w)} = \sum_{k=0}^{\infty} d_k e^{-jwk}$$

Entonces, el mejor predictor está dado por

$$\hat{x}(n|M_{n-1}) = - \sum_{k=1}^{\infty} d_k x(n-k)$$

donde $d_k = d_k/d_0$.

Sin embargo, el predictor lineal con todo el pasado (también se llama el predictor de un paso hacia adelante) tiene poco interés práctico por varias razones:

- 1) difícil de conocer todo el pasado;
- 2) problemas de almacenaje de las muestras;
- 3) no-correlación entre las muestras lejanas, si el proceso es ergódico.

Por lo tanto, es necesario introducir un predictor lineal más práctico ([13], [14]).

Definición 3 :

El predictor lineal de orden p está definido como sigue:

Sean $M = V(x(n-1), x(n-2), \dots, x(n-p))$,

$$\hat{x}(n) = \hat{x}(n|M)$$

$$= - \sum_{k=1}^p a_k x(n-k). \quad (III.13)$$

Los coeficientes a_k se determinan por

$$\min_{a_k} E |x(n) - \hat{x}(n)|^2$$

El mínimo error cuadrático medio se denota por $E\hat{\epsilon}$.

El problema se puede resolver siguiendo el procedimiento similar que usamos para encontrar $E\hat{\epsilon}_0$, pero el procedimiento es muy complicado y difícil. Afortunadamente no es necesario resolver de esta forma sino una manera mucho más sencilla.

Aplicando el principio de ortogonalidad, el error

$$e(n) = x(n) - \hat{x}(n)$$

es ortogonal a M cuando el predictor es óptimo, esto es

$$\begin{aligned}
 E(x(n-k)e(n)) &= E(x(n-k) \sum_{m=0}^p a_m x(n-m)) \\
 &= \sum_{m=0}^p a_m R(k-m) \\
 &= 0 \quad k=1,2,\dots,p.
 \end{aligned}
 \tag{III.14}$$

Además

$$\begin{aligned}
 E(e(n))^2 &= E(x(n)e(n)) \\
 &= E(x(n) \sum_{k=0}^p a_k x(n-k)) \\
 &= \sum_{k=0}^p a_k R(k) \\
 &= E\hat{\epsilon}
 \end{aligned}
 \tag{III.15}$$

Escribimos en forma matricial, tomando en cuenta que $R(m)=R(-m)$, tenemos lo siguiente:

$$\begin{bmatrix}
 R(0), R(1), \dots, R(p) \\
 R(1), R(0), \dots, R(p-1) \\
 \vdots \\
 R(p), R(p-1), \dots, R(0)
 \end{bmatrix}
 \begin{bmatrix}
 a_0 \\
 a_1 \\
 \vdots \\
 a_p
 \end{bmatrix}
 =
 \begin{bmatrix}
 E\hat{\epsilon} \\
 0 \\
 \vdots \\
 0
 \end{bmatrix}
 \tag{III.16}$$

En las literaturas estadísticas, el sistema de ecuaciones (III.16) se conoce como ecuaciones de Yule-Walker o ecuaciones normales. Ahora el mejor predictor lineal se obtiene resolviendo las ecuaciones normales. Por la forma especial que tiene la matriz de

correlacion, el sistema (III.16) siempre tiene solución única, además se puede resolverlo de una manera muy eficiente la cual se deduce en III-3. Antes de obtener la solución, tratamos de llegar a las mismas ecuaciones por diferentes enfoques.

III-2. Estimación Espectral

La densidad espectral de potencia de un proceso estacionario $X(n)$ se define como

$$S(\omega) = \sum_{m=-\infty}^{\infty} R(m) e^{-j\omega m} \quad (III.17)$$

donde $R(m) = EX(n+m)X(n)$. (III.18)

En casos prácticos, no se conoce la función de autocorrelación. Entonces una suposición adicional se debe hacer que el proceso es ergódico en primer y segundo orden. La propiedad de ergodicidad nos permite la sustitución del promedio estadístico por el promedio en el tiempo. Bajo la suposición de ergodicidad del proceso, la función de autocorrelación se puede evaluar como

$$R(m) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n+m)x(n). \quad (III.19)$$

Es posible demostrar ([1],[15],[16]) que con la expresión anterior, la densidad espectral se puede expresar equivalentemente como

$$S(\omega) = \lim_{N \rightarrow \infty} E \left(\frac{1}{2N+1} \left| \sum_{n=-N}^N x(n) e^{-j\omega n} \right|^2 \right) \quad (III.20)$$

El valor esperado en esta expresión se requiere, porque la propiedad ergódica de $R(m)$ no acopla con la transformada de Fourier; es decir, el límite no converge en ningún sentido estadístico sin tomar el valor esperado.

En la práctica, la señal $X(n)$ es conocida solamente en un intervalo finito, o solamente nos interesa en un intervalo finito. El problema de estimación espectral consiste en obtener una densidad espectral aproximada teniendo disponible un intervalo finito $\{x(n), 0 \leq n \leq N-1\}$, donde N es el número total de muestras. De la definición de la densidad espectral, dos técnicas de estimación ([17]) son posibles basadas en (III.20) y (III.19), (III.17).

El método de periodograma aplica directamente (III.20) sin tomar el límite ni el valor esperado:

$$S(\omega) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\omega n} \quad (III.21)$$

El uso de la transformada rápida de Fourier (FFT) ([18], [19], [20]) permite la evaluación de $S(\omega)$ en un conjunto discreto de N frecuencias igualmente espaciadas $\omega_m = m\omega$, para $m=0, 1, \dots, N-1$ y $\omega = 2\pi/N$. Entonces

$$S(\omega_m) = \frac{1}{N} |X(\omega_m)|^2 \quad (III.22)$$

donde $X(\omega_m)$ es la transformada discreta de Fourier de $x(n)$, $0 \leq n \leq N-1$.

El otro método es la técnica convencional Blackman-Tukey, la cual estima la densidad espectral mediante una estimación de los coeficientes de correlación. Con una secuencia finita de señal, sólo se puede estimar un número finito de coeficientes de correlación. La densidad espectral obtiene por

$$S(\omega) = \sum_{m=-M}^M R(m) e^{-j\omega m} \quad (III.23)$$

Una expansión obvia de (III.19) para la estimación de los coeficientes de correlación es

$$R(s) = \frac{1}{N} \sum_{n=0}^{N-m-1} x(n+m)x(n) \quad (III.24)$$

y $R(-m) = R(m)$, $m=0, 1, \dots, M$; $M \leq N-1$.

Una generalización de (III.24) es predistorsionar la señal por una ventana de la misma longitud $w(n)$, $0 \leq n \leq N-1$, es decir

$$R(s) = \frac{1}{N} \sum_{n=0}^{N-m-1} x'(n+m)x'(n) \quad (III.25)$$

donde $x'(n) = x(n)w(n)$ $0 \leq n \leq N-1$ (III.26)

Existen muchos tipos de ventanas, un buen resumen de las ventanas discretas se encuentra en [21]. La ecuación (III.24) es un caso especial de (III.25) usando una ventana rectangular. También se puede aplicar ventanas a (III.23) para mejorar algunas características de la densidad espectral estimada, en [15] y [22] se discute el efecto de algunas ventanas.

La evaluación de (III.23) en un conjunto de N frecuencias igualmente espaciadas en $[0, 2\pi)$ se puede realizar mediante el uso de FFT.

Las dos técnicas mencionadas anteriormente tienen la ventaja de ser computacionalmente eficientes, pero el uso implícito de ventana limita la resolución del espectro. La resolución del espectro estimado puede ser mejorada por el uso de extrapolación de la función de correlación. Un criterio de optimalidad para la extrapolación es el principio de máxima entropía. La idea de máxima entropía es extrapolar los coeficientes desconocidos de la función de correlación de tal forma que la entropía de la función de densidad de probabilidad correspondiente se maximiza en cada paso de la extrapolación ([9],[10],[22],[23]). El planteamiento concreto consiste en lo siguiente:

Dados los primeros $p+1$ valores $R(0), R(1), \dots, R(p)$ de la función de correlación de un proceso $X(n)$, se desea estimar su densidad espectral $S(\omega)$. Las estadísticas de $X(n)$ son determinadas en términos de la densidad conjunta de $x(n), x(n-1), \dots, x(n-r)$. Entonces, aplicando el principio de máxima entropía, debemos determinar los valores desconocidos de $R(m)$ tales que maximizan la entropía $H(s(0), s(1), \dots, s(r))$ de ese proceso y encontrarlo cuando $r \rightarrow \infty$. Esto es equivalente a maximizar la razón de entropía H_x de $X(n)$, sujeto a las restricciones dadas. El problema no se resuelve fácilmente para un proceso no gaussiano. Sin embargo, cuando el proceso es normal, podemos demostrar que H_x es máxima si $X(n)$ es generado por un sistema de todos polos de orden p excitando por un ruido blanco gaussiano, o bien un proceso gaussiano AR de orden p .

Hay varias formas de demostrar la proposición anterior ([10]), vamos a dar una de estas demostraciones.

Sea $\hat{x}_p(n)$ el predictor lineal de orden p de $x(n)$, y el error de predicción está dado por

$$\begin{aligned} e(n) &= x(n) - \hat{x}_p(n) \\ &= x(n) + \sum_{k=1}^p a_k x(n-k) \end{aligned}$$

Es claro que $e(n)$ es la salida del filtro

$$A(Z) = \sum_{k=0}^p a_k e^{j\omega k} \quad \text{donde } a_0 = 1$$

con entrada $x(n)$. Entonces la razón de entropía de la salida está dada por:

$$H_e = H_x + \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln |H(e^{j\omega})|^2 d\omega$$

Al maximizar H_x , es suficiente maximizar H_e debido a que la integral se especifica en términos de los valores dados de $R(m)$. Como sabemos, la potencia promedio de $e(n)$ es igual a

$$E_{\tilde{e}} = E\{e(n)^2\} = \sum_{k=0}^p a_k R(k).$$

Como $e(n)$ es gaussiano cuando $x(n)$ lo es, H_e se maximiza si el proceso $e(n)$ es ruido blanco. Por lo tanto, el proceso gaussiano $X(n)$ que maximiza la entropía conociendo los primeros $p+1$ valores de $R(m)$ es un proceso AR de orden p . En este caso la densidad espectral se estima como

$$S(\omega) = \frac{E_{\tilde{e}}}{\left| \sum_{k=0}^p a_k e^{j\omega k} \right|^2}$$

Para un proceso AR, es decir que $X(n)$ se genera por

$$x(n) = - \sum_{k=1}^p a_k x(n-k) + e(n)$$

donde $e(n)$ es un ruido blanco gaussiano con potencia promedio $E\{e^2}\sigma^2$, podemos tener las siguientes expresiones:

$$\begin{aligned} E(x(n-k)e(n)) &= E(x(n-k) \sum_{m=0}^p a_m x(n-m)) \\ &= \sum_{m=0}^p a_m R(k-m) \\ &= 0 \quad k \geq 0. \end{aligned}$$

Una cosa interesante aquí es que la ecuación anterior se cumple para toda k no negativa; una forma equivalente de decir es

$$E\{e^2\} = E\{e_0^2\}$$

para procesos AR de orden p . Además

$$\begin{aligned} E\{e^2\} &= E(e(n)e(n)) \\ &= E(x(n)e(n)) \\ &= E(x(n) \sum_{k=0}^p a_k x(n-k)) \\ &= \sum_{k=0}^p a_k R(k) \end{aligned}$$

Escribiendo las primeras p ecuaciones y la ecuación de $E\{e^2\}$ en forma matricial, tenemos

$$\begin{bmatrix} R(0), R(1), \dots, R(p) \\ R(1), R(0), \dots, R(p-1) \\ \vdots \\ R(p), R(p-1), \dots, R(0) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} E z^p \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Notamos que estas son las mismas ecuaciones normales (III.16) que deducimos en la sección anterior. La técnica de máxima entropía tiene aparentemente una restricción de que $X(n)$ debe ser gaussiano, pero cuando se desconoce la densidad de probabilidad del proceso $X(n)$, la suposición de $X(n)$ normal, maximiza la entropía, y por otra parte los procesos tratados en la práctica pueden ser modelados adecuadamente como gaussianos en general. Entonces, la estimación espectral de máxima entropía nos lleva a un modelo AR de orden p para representar un proceso gaussiano conociendo los primeros $p+1$ valores de $R(m)$. Cuando se desconocen estos valores, se los estiman con (III.24) o (III.25).

En las literaturas actuales, existen muchas otras formas de estimar la densidad espectral, por ejemplo, [15] y [16] proporcionan varias de ellas. Además, en [17] se encuentra un resumen de las técnicas modernas de estimación espectral.

111-3. Modelado de Procesos

En las secciones anteriores, las ecuaciones Yule-Walker fueron encontradas mediante búsquedas del predictor lineal que minimiza el error cuadrático medio, y la estimación espectral que maximiza la entropía. En esta sección, deducimos las mismas ecuaciones con un enfoque de aproximación en el dominio de la frecuencia a un proceso AR, usando la medida de distorsión de Itakura-Saito ([24], [25], [26]) como el criterio de optimalidad.

Sean $X(n)$ un proceso estacionario con densidad espectral $S_X(\omega)$, y $Y(n)$ un proceso AR de orden p con densidad espectral

$$S_Y(\omega) = \frac{P_Y}{\left| \sum_{k=0}^p a_k e^{-j\omega k} \right|^2}$$

$$= \frac{P_Y}{|A(e^{j\omega})|^2}$$

donde $a_0 = 1$.

La distorsión Itakura-Saito entre dos procesos $X(n)$ y $Y(n)$ está definida como

$$D(X, Y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{S_X(\omega)}{S_Y(\omega)} - \ln \frac{S_X(\omega)}{S_Y(\omega)} - 1 \right) d\omega$$

Definimos

$$P_X = E\sigma^2$$

$$= \exp \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S_X d\omega \right)$$

entonces

$$D(X, Y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{S_X(\omega)}{S_Y(\omega)} - \ln \frac{P_X}{P_Y} - 1 \right) d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} |A(e^{j\omega})|^2 \frac{S_X(\omega)}{P_V} d\omega - \ln \frac{P_X}{P_V} - 1$$

$D(X,Y)$ se minimiza sobre $a_k, k=1, \dots, p$ y P_V . Tomamos la derivada parcial de D con respecto a $a_k, k=1, \dots, p$ e igualándolas a cero.

$$\begin{aligned} D_{a_k} &= \frac{1}{2\pi P_V} \int_{-\pi}^{\pi} S_X(\omega) \sum_{m=1}^p 2a_m \cos((k-m)\omega) d\omega \\ &= \frac{2}{P_V} \sum_{m=0}^p a_m R(k-m) = 0 \quad k=1, 2, \dots, p. \end{aligned}$$

Tomando en cuenta las ecuaciones anteriores,

$$D(X,Y) = \frac{\sum_{m=0}^p a_m R(k-m)}{P_V} - \ln \frac{P_X}{P_V} - 1.$$

Minimizando D sobre P_V , tenemos

$$P_V = \sum_{m=0}^p a_m R(m) = E\hat{\epsilon}^2$$

$$y \quad D(X,Y) = \ln \frac{E\hat{\epsilon}^2}{E\sigma^2}$$

Otra vez tenemos las mismas ecuaciones normales que encontramos antes. Ahora tratamos de resolver este sistema de ecuaciones. Es posible resolverlo por métodos convencionales, pero por la forma especial del sistema existe una manera muy eficiente de resolverlo llamada el algoritmo Levinson-Durbin ([27], [28], [14]).

Suponemos que tenemos el mejor modelo AR de orden m que está descrito por:

$$\begin{bmatrix} R(0), R(1), \dots, R(m) \\ R(1), R(0), \dots, R(m-1) \\ \vdots \\ R(m), R(m-1), \dots, R(0) \end{bmatrix} \begin{bmatrix} a_{0m} \\ a_{1m} \\ \vdots \\ a_{mm} \end{bmatrix} = \begin{bmatrix} E_m \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

y queremos encontrar el mejor modelo AR de orden $m+1$. Notamos

$$\underline{a}_{m+1} = \begin{bmatrix} a_{0m} \\ a_{1m} \\ \vdots \\ a_{mm} \\ 0 \end{bmatrix} + k_{m+1} \begin{bmatrix} 0 \\ a_{mm} \\ \vdots \\ a_{1m} \\ a_{0m} \end{bmatrix}$$

donde k_{m+1} es una constante indeterminada. Entonces

$$\begin{bmatrix} R(0), R(1), \dots, R(m+1) \\ R(1), R(0), \dots, R(m) \\ \vdots \\ R(m), R(m-1), \dots, R(1) \\ R(m+1), R(m), \dots, R(0) \end{bmatrix} \underline{a}_{m+1} = \begin{bmatrix} E_{m+1}^z \\ 0 \\ \vdots \\ 0 \\ z \end{bmatrix} + K_{m+1} \begin{bmatrix} z \\ 0 \\ \vdots \\ 0 \\ E_{m+1}^z \end{bmatrix}$$

donde $z = \sum_{k=0}^m R(m+1-k) a_{k,m}$.

Si escogemos K_{m+1} tal que $z + K_{m+1} E_{m+1}^z = 0$

i.e. $K_{m+1} = \frac{\sum_{k=0}^m R(m+1-k) a_{k,m}}{E_{m+1}^z}$

entonces, el mejor modelo AR de orden $m+1$ está determinado por \underline{a}_{m+1} definido anteriormente, y la ganancia E_{m+1}^z , los cuales se calculan por:

$$\begin{aligned} \underline{a}_{m+1, m+1} &= K_{m+1} \\ \underline{a}_{j, m+1} &= a_{j,m} + K_{m+1} \underline{a}_{m+1-j, m} \\ E_{m+1}^z &= E_{m+1}^z + K_{m+1} z \\ &= E_{m+1}^z (1 - K_{m+1} \bar{r}) \end{aligned}$$

donde $a_{0,m} = 1$, $j=2, \dots, m$.

El algoritmo se inicializa con $m=0$

$$\begin{aligned} \underline{a}_{00} &= 1 \\ E_0^z &= R(0) \end{aligned}$$

con la recursión para $m=1, \dots, p-1$.

El algoritmo es muy eficiente computacionalmente, sólo requiere operaciones de orden p^2 , en contraste las de eliminación gaussiana que son de p^3 . Además, éste tiene una propiedad muy útil cuando no se conoce de antemano el orden correcto del modelo, se puede parar el algoritmo hasta un ϵ^2 aceptable.

Comúnmente a K_1, K_2, \dots, K_p se les conocen como los coeficientes de reflexión ([13],[14]). Ellos son no mayores a 1 en valor absoluto si y sólo si la secuencia asociada de correlaciones $R(0), R(1), \dots, R(p)$ es válida, i.e. la matriz de correlación es positiva semi-definida. En este caso se garantiza la estabilidad del filtro $A(z)$ ([9]). Los coeficientes de reflexión tienen una interpretación en el contexto de la formulación "lattice" de error de predicción lineal ([17], [10]). La relación entre los coeficientes $\{a_m\}$ y $\{K_m\}$ son uno a uno, ya sabemos cómo obtener $\{a_m\}$ en función de los $\{K_m\}$, la relación inversa se puede ver en [14].

CAPITULO IV

MEDIDAS DE DISTORSION

IV-1. Propiedades Generales de una Medida de Distorsión

Con el fin de segmentar las señales cuasi-estacionarias, en nuestro caso la señal de voz, introducimos el concepto de fuentes compuestas en II.4 para modelar adecuadamente este tipo de señales. Bajo este modelo, la segmentación consiste en identificar los cambios de sub-fuentes y la caracterización de las mismas. Para lograr la solución de estos problemas, es indispensable el establecimiento de medidas cuantitativas las cuales describen la distorsión o la similitud entre dos señales estacionarias.

Una medida de distorsión es una asignación de un número no-negativo a un par de entrada/salida de un sistema. La distorsión entre una entrada o señal original y una salida o reproducción, representa un costo o distorsión resultante cuando la entrada está reproducida por la salida. Tales medidas tienen una gran variedad de aplicaciones en el diseño y comparación de sistemas.

Para ser útil, una medida de distorsión debe poseer por lo menos a cierto grado las siguientes propiedades:

- 1) A cada pareja de señales o segmentos X y Y , la función de distorsión $d(X,Y)$ le debe asignar un número no-negativo; éste es cero si y sólo si $X=Y$.
- 2) $d(.,.)$ debe ser subjetivamente sensible, o lo que es lo mismo, a valores pequeños de $d(.,.)$ le debe corresponder una mejor calidad subjetiva que a valores grandes de $d(.,.)$. Este aspecto es de mucha importancia debido a que no hay una medida de distorsión universalmente aplicable que siempre sea subjetivamente representativa.
- 3) $d(.,.)$ debe ser analíticamente tratable en el sentido de que permite diseñar y analizar distintos criterios de segmentación.
- 4) $d(.,.)$ debe ser computable, lo cual indica que el valor numérico de esta función debe ser relativamente fácil de

calcular para que dicha medida sea de alguna utilidad en el terreno práctico.

También es deseable pero no necesario que una medida de distorsión cumple otras propiedades tales como la simetría y la desigualdad de triángulo para que sea una medida de distancia.

En este trabajo, evitamos el uso del término "medida de distancia", y usamos "medida de distorsión". La razón es precisamente no ser restringidas las medidas utilizadas por las propiedades de simetría y desigualdad de triángulo.

La medida de distorsión más común es el tradicional error cuadrático. El logro de gran popularidad de ésta se debe a su tratabilidad y computabilidad. Sin embargo, para muchos sistemas, entre ellos los sistemas de voz con razón de transmisión baja, ésta no es subjetivamente significativa.

Para no tener este tipo de problemas, se han introducido muchas medidas de distorsión las cuales cuentan con mayor significado subjetivo. Como los procesos tratados en este trabajo son estacionarios en el sentido amplio, sólo discutimos las medidas de distorsión que dependen exclusivamente de sus propiedades de segundo orden : la función de correlación o la densidad espectral. Este tipo de distorsión es generalmente más fácil ser definidas en el dominio de la frecuencia.

Una medida de distorsión espectral es una función de dos densidades f y g , la cual asigna un número no negativo $d(f,g)$ a representar la distorsión en usar g para representar f . Entre las más comunes de tales medidas, se conoce la distorsión diferencial donde se usa una norma L_p ([29]) en la diferencia $f-g$. Bajo esta definición, las medidas de distorsión forman distancias en el sentido de que ellas satisfacen el requerimiento de la simetría y la desigualdad de triángulo.

Sin embargo, las medidas de distorsión espectral tratadas en este trabajo dependen únicamente de la diferencia del logaritmo del espectro, o bien razón de espectros. Para este tipo de medidas observamos que

$$d(f,g) = d(f,g/f) = d(f/g,1)$$

Los tipos de escalamiento son usados comunmente : la normalización de ganancia y la optimización de ganancia. Una medida de distorsión de ganancia normalizada está definida por

$$d^*(f,g) = d(f/E_f, g/E_g)$$

donde E_f y E_g son ganancias o errores de predicción lineal de un paso hacia adelante para f y g . Las medidas de distorsión de ganancia normalizada pueden ser útiles para considerar separadamente los efectos de modelos normalizados y las ganancias.

Una medida de distorsión de ganancia optimizada se define como

$$d'(f,g) = \min_{k \geq 0} d(f, kg)$$

Por la definición, tenemos

$$d(f,g) \geq d'(f,g)$$

y también

$$d^*(f,g) = d(f, g/E_g * E_f) \\ \geq d'(f,g)$$

Las últimas dos ecuaciones indican que tanto $d(f,g)$ como $d^*(f,g)$ son medidas más estrictas que $d'(f,g)$.

En la sección posterior, discutiremos detalladamente una medida de distorsión llamada Itakura-Saito que es usada en la segmentación.

IV-2. Medida de distorsión Itakura-Saito

Una medida de distorsión propuesta por Itakura y Saito ([24], [25]) es

$$d_{IS}(f,g) = \frac{1}{2n} \int_{-n}^n \left(\frac{f(w)}{g(w)} \ln \frac{f(w)}{g(w)} + 1 \right) dw$$

Esta medida apareció en la formulación de predicción lineal como una estimación aproximada de máxima similitud, y se ha usado para ilustrar las propiedades pareadas de espectros de predicción lineal y la razón de que los polos pesan más que los ceros en los espectros muestrales ([14]).

Como

$$E\tilde{f} = \exp\left(\frac{1}{2n} \int_{-n}^n \ln f(w) dw\right)$$

$d_{IS}(f,g)$ se puede expresar como

$$d_{IS}(f,g) = \frac{1}{2n} \int_{-n}^n \frac{f(w)}{g(w)} dw - \ln \frac{E\tilde{f}}{E\tilde{g}} - 1$$

donde $E\tilde{f}$ y $E\tilde{g}$ son las ganancias o errores de predicción lineal de un paso de f y g respectivamente.

Usando el hecho de que la media geométrica es menor o igual que la media aritmética (Desigualdad de Jensen), uno se puede escribir

$$\frac{1}{2n} \int_{-n}^n \frac{f(w)}{g(w)} dw \geq \exp\left(\frac{1}{2n} \int_{-n}^n \ln \frac{f(w)}{g(w)} dw\right)$$

De esto, inmediatamente se deduce que

$$d_{IS}(f,g) \geq d_{IS}(E\tilde{f}, E\tilde{g}).$$

Diciéndolo en palabras, la distorsión llega a su mínima para ganancias dadas cuando los espectros son constantes.

Un desarrollo alternativo de la distorsión Itakura-Saito aparece en el contexto de estadísticas matemáticas ([26], [30]).

Supongamos que tenemos dos procesos gaussianos de media cero con

densidades espectrales f y g respectivamente, y sean $p_N^f(X)$ y $p_N^g(X)$, $X = (x_0, \dots, x_{N-1})$, las densidades correspondientes de probabilidad de orden N . La información discriminatoria de orden N (o entropía relativa, número de Kullback-Leibler, entropía cruzada o divergencia direccionada ([31], [32])) está definida por

$$I(f, g) = \int p_N^f(X) \ln \frac{p_N^f(X)}{p_N^g(X)} dx$$

Esta función tiene aplicaciones extensivas en el problema de discriminación entre procesos gaussianos ([31]), y se usa tanto en detección como aplicaciones de teoría de información. Para procesos gaussianos la información discriminatoria es bien conocida ([32])

$$2I_N(f, g) = \ln(\det R_N(f) / \det R_N(g)) + \text{tr}(R_N(f) R_N^{-1}(g)) - N$$

donde "tr" denota la traza de una matriz, y R_N es la matriz $N \times N$ de autocorrelación asociada con la densidad espectral en su argumento. De [32], la información discriminatoria normalizada tiene el límite

$$\begin{aligned} 2I(f, g) &= \lim_{N \rightarrow \infty} \frac{2}{N} I_N(f, g) \\ &= d_{\text{It}}(f, g) \end{aligned}$$

Esto es, la distorsión de Itakura-Saito es exactamente dos veces la información discriminatoria asintótica bajo la suposición de que los procesos son gaussianos. Por esta razón esta medida es potencialmente útil como una función de costo o distorsión para discriminación de vecino más cercano de series de tiempo localmente estacionarias, donde N es grande tal que $N^{-1} I_N(f, g)$ es aproximadamente $I(f, g)$ y los espectros muestrales observados son aproximadamente iguales a los reales.

La aplicación de la distorsión a predicción lineal va a ser

más aparente si f es un espectro muestral y g es un espectro de modelo AR de orden p :

$$g(w) = \frac{E_0^2}{|A(e^{jw})|^2}$$

donde

$$A(z) = \sum_{n=0}^p a_n z^{-n} \quad \text{con } a_0 = 1,$$

suponiendo que todas las raíces de $A(z)$ están dentro del círculo unitario.

En este caso, la medida de distorsión Itakura-Saito tiene un atributo importante como desarrollamos en la sección III.3 : minimizando $d_{IS}(f,g)$ es equivalente al análisis LPC. Debido a que un sistema de LPC con una tasa alta logra una buena calidad subjetiva, pues $d_{IS}(f,g)$ es un candidato lógico para una medida de distorsión subjetivamente significativa en sistemas de procesamiento de voz.

Además, la medida de distorsión Itakura-Saito es bien tratable teórica y computacionalmente. En el capítulo III hemos desarrollado un método eficiente para encontrar el mejor filtro $f_p = E_p/A_p$ de orden p de una densidad espectral f , minimizando $d_{IS}(f, |E_p/A_p|^2)$, con la distorsión mínima :

$$d_{IS}(f, f_p) = \ln \frac{E_p^2}{E_0^2}$$

A continuación, obtenemos la fórmula para calcular la distorsión entre una densidad f y cualquier filtro E/A de orden p .

Sean $\underline{a} = (a_0, a_1, \dots, a_p)'$ donde $a_0 = 1$,

$$I_p(\underline{a}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |A(e^{jw})|^2 f(w) dw$$

Entonces, con un poco de álgebra, es fácil de llegar a:

$$T_p(\underline{a}) = \underline{a}^T R_p \underline{a}.$$

Definiendo

$$K_a(m) = \sum_{n=0}^{p-m} a_{m+n} a_n \quad m=0, \dots, p:$$

entonces

$$T_p(\underline{a}) = K_a(0)R(0) + 2 \sum_{m=1}^p K_a(m)R(m)$$

Por lo tanto, el cálculo de $d_{10}(f, |E/A|^2)$ es muy fácil:

$$d_{10}(f, |E/A|^2) = \frac{T_p(\underline{a})}{E^2} - \ln \frac{E \delta_0}{E^2} - 1.$$

Una observación importante es que el espectro f_p genera las primeras $p+1$ correlaciones idénticas que las de f , y $T_p(\underline{a})$ es función exclusivamente de \underline{a} y los primeros $p+1$ coeficientes de correlación de f . Entonces podemos obtener un resultado muy interesante:

$$\begin{aligned} d_{10}(f, |E/A|^2) &= \ln \frac{E_p^2}{E \delta_0} + \frac{T_p(\underline{a})}{E^2} - \ln \frac{E_p^2}{E^2} - 1. \\ &= d(f, f_p) + d(f_p, |E/A|^2) \end{aligned}$$

Esta propiedad de igualdad de triángulo de la distorsión Itakura-Saito implica que un sistema de análisis bajo este enfoque se puede diseñar separadamente la parte de identificación y la de cuantización, y la distorsión total es la suma de las distorsiones producidas en cada etapa. Además, en la práctica es difícil o

imposible conocer el valor exacto de E_{∞} , muchas veces se sustituye en el cálculo E_{∞} por E_p debido a que la distorsión de identificación es solamente función de f .

En muchas ocasiones, a uno sólo le interesa la forma de onda y no la energía que posee la señal, entonces la versión de la distorsión Itakura-Saito con ganancia normalizada se aplica para lograr este objetivo. Sea g de forma $|E/A|^2$, entonces

$$d\gamma_g(f, |E/A|^2) = \frac{\gamma_p(a)}{E\bar{z}_g} - 1.$$

Esta distorsión de ganancia normalizada será utilizada en la segmentación para reducir la sensibilidad sobre el cambio de potencia de la señal.

CAPITULO V ALGORITMO DE SEGMENTACION

V-1. Planteamiento del Problema

En la sección II.4, definimos los procesos cuasi-estacionarios o localmente estacionarios, y propusimos un modelo de fuentes compuestas para representarlos. En el mundo real hay muchas señales que se pueden modelar adecuadamente como procesos cuasi-estacionarios, entre ellas la señal de voz es la de nuestro interés. Una señal de esta naturaleza, se puede considerar aproximadamente como una concatenación de un conjunto de procesos estacionarios, llamados sub-fuentes, con un posible intervalo de transición. Es decir, la señal se divide en intervalos dentro de los cuales la estadística o la forma de onda no sufren cambios notorios.

Observamos una señal típica de voz (ver fig. V.1), en ocasiones sí es posible determinar los momentos de cambios de la estadística y los intervalos en que la señal permanece aproximadamente en condiciones estacionarias. Los cambios de una sub-fuente a otra pueden clasificarse en dos tipos: cambios abruptos o cambios suaves. Los del segundo tipo implican que existe un intervalo de transición entre dos segmentos estacionarios. Como detectar los cambios tanto abruptos como suaves y cómo caracterizar los segmentos estacionarios son dos partes esenciales del problema de segmentación.

La detección y la caracterización de intervalos estacionarios según el criterio de fidelidad son problemas de interés teórico y práctico en la teoría de control, el área de comunicaciones, el análisis de series de tiempo y áreas afines. Algunos casos particulares de este tipo de problema han sido estudiados. Uno de ellos es cuando los procesos de las fuentes han sido modelados como procesos autorregresivos (AR) o autorregresivos con promedios móviles (ARMA) ([33]), en el cual se requiere determinar las características de la señal que aparentemente es ahora generada

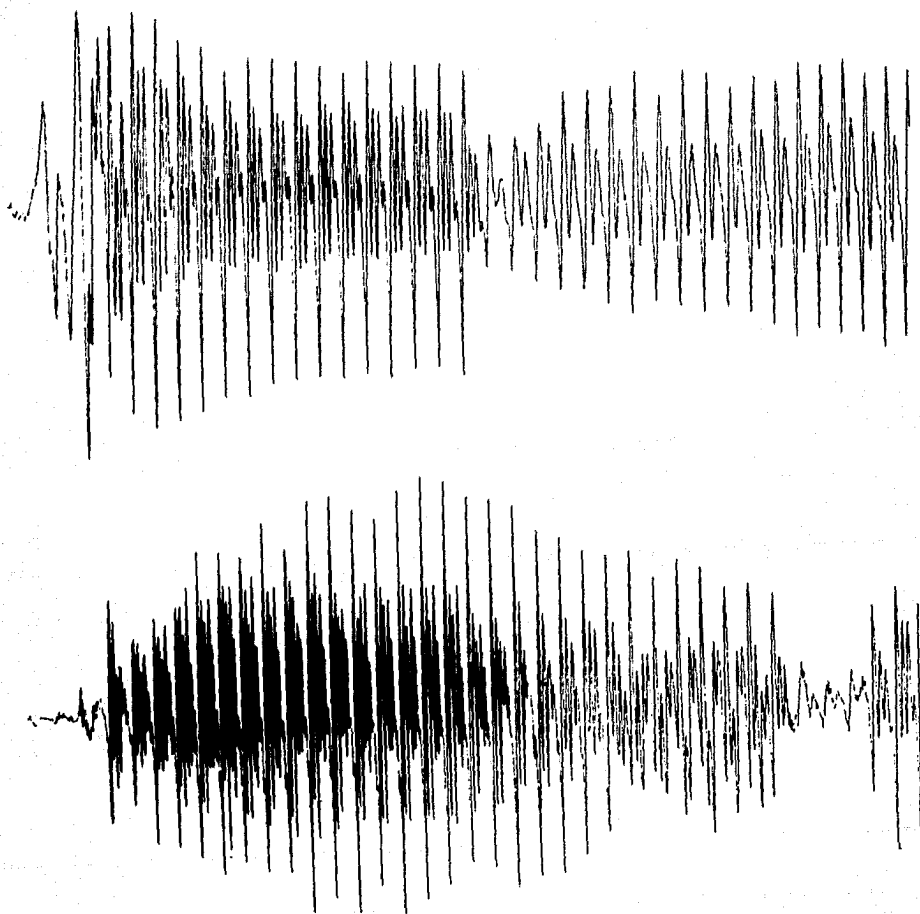


Figura V.1 Señales típicas de voz

por filtros AR o ARMA con parámetros variables. Otro enfoque usado para la solución del problema de detección de los cambios, reportado recientemente [34], presupone el conocimiento de las distribuciones de probabilidad condicional de primer orden de cada una de las sub-fuentes. Para enfoques anteriormente mencionados un problema de la realización práctica es la gran cantidad de información que se requiere sobre las subfuentes que forman a la fuente compuesta.

En este trabajo, la solución que proponemos tiene la ventaja de que no se requiere información sobre las sub-fuentes. El método proporciona buen resultado de segmentación siempre y cuando la señal observada es discreta en el tiempo y la estadística más importante es la del segundo orden, es decir la densidad espectral.

Una vez que se ha supuesto que las señales bajo estudio serán modeladas como provenientes de una fuente compuesta, es necesario cuantificar las diferencias entre dos segmentos estacionarios de la señal observada. Las medidas de distorsión establecidas en el capítulo IV nos ayudan a lograrlo. Todas éstas son aplicables de alguna forma para medir distorsión entre secuencias o bloques de muestras de longitud predeterminada. Hacemos énfasis en que el objetivo de estas medidas consiste en la detección de los tiempos en que el conmutador empleado en el modelo de la fuente compuesta ha cambiado de posición. Cabe mencionar que el tamaño de la secuencia debe ser determinado como función del mínimo número de muestras que se está dispuesto a aceptar como representativo o indicativo de la presencia de un segmento estacionario. Esto a su vez es una indicación del tiempo en que tiene que permanecer el conmutador en una subfuente antes de que el observador afirma que efectivamente el conmutador se encuentre en esa posición.

V-2. La Solución teórica

La solución del problema es basada en la siguiente proposición. Sean X y X' dos secuencias del mismo segmento y sea Y una secuencia de cualquier otro segmento, entonces $d(X, X')$ debe ser casi cero, $d(X, Y)$ y $d(X', Y)$ deben tener valores cercanos y lo mismo sucede con $d(Y, X)$ y $d(Y, X')$. Esta proposición es razonable para las medidas de distorsión tales que son sensibles a cambios de la estadística, como las propuestas anteriormente.

Consideramos dos bloques de muestras λ_n y λ_{n+k} , donde $\lambda_n = (x_n, \dots, x_{n+N-1})$. Uno se puede observar cómo evoluciona la distorsión $d(\lambda_n, \lambda_{n+k})$ como una función del desplazamiento relativo k entre ellos. Para $k=0$ obviamente $d(\lambda_n, \lambda_n) = 0$ por estar siendo comparadas de secuencias idénticas.

En los textos posteriores, los términos "segmentos estacionarios" y "subfuentes" se utilizan indistintamente, porque "segmento estacionario" es referido a la salida de la subfuente. Lo mismo sucede con "secuencia" y "bloque" de datos. Lo declaramos aquí para evitar posibles confusiones.

Al aumentar el valor de k , $d(\lambda_n, \lambda_{n+k})$ mide la distorsión de interpretar la estadística de segundo orden del bloque λ_{n+k} por la del bloque λ_n , y esta distorsión será pequeña mientras el conmutador permanezca en la misma subfuente durante el intervalo $(n, n+k+N-1)$. En el momento en que el conmutador cambia a una subfuente con estadística del segundo orden distinta a la que apuntaba originalmente, el valor de $d(\lambda_n, \lambda_{n+k})$ empezará a crecer lo cual ocurrirá hasta el momento en que permanezca en una misma subfuente durante un intervalo de tiempo mayor que N muestras: si permanece aproximadamente constante el valor de $d(\lambda_n, \lambda_{n+k})$ durante algún tiempo (dependiendo del criterio preestablecido para decir cuando un bloque es o no estacionario) diremos que se detectó un cambio de segmento y además se identificó el intervalo en el cual ocurrió el transitorio. De este

podemos decir si fue un cambio abrupto o suave dependiendo de la duración en el cual se incrementa la distorsión. Un ejemplo típico de un cambio suave (fig. V.2) ilustra el procedimiento de la segmentación.

Una vez que se ha detectado un cambio del proceso observado se reinitializa el procedimiento definiendo nuevamente un bloque o secuencia inicial dentro del último tramo estacionario detectado.

La idea básica del procedimiento de segmentación se resume como sigue:

Algoritmo de segmentación:

1) Sean $k=10$, n es la muestra inicial.

2) Para $k \leq N-1$, evalúa $d(x_k, x_{k-1})$.

3) Evalúa

$$d_k = \frac{1}{k-1} \sum_{i=1}^{k-1} d(x_i, x_{i-k+1}) \quad \forall$$

$$r_k = \frac{1}{k-1} \sum_{i=1}^{k-1} (d_k - d(x_i, x_{i-k+1}))$$

4) Si $d_k \leq a$, no hay cambio, $k=10$;

si $d_k > a$ y $r_k > 0$, empieza el cambio, $k=1$;

si $d_k > b$ y $r_k \leq c$, entonces $n=n+k-1$, $k=10$;

casos contrarios $k=1+1$.

5) regresa a (2).

En el algoritmo a , b y c son tres umbrales: uno para decidir cuando la distorsión ha crecido lo suficiente como para decir que hay un cambio, y otros dos para decidir cuando la distorsión llega a un valor estacionario. La función de los umbrales se pueden ver más clara con la ayuda de la figura V.2. De acuerdo con el algoritmo, el umbral "a" establece una cota superior de la distorsión entre dos bloques cualesquiera de un mismo segmento estacionario; por el otro lado, el umbral "b" indica el mínimo valor que puede alcanzar la distorsión entre dos segmentos estacionarios

consecutivos; y por último, el valor de "c" especifica la máxima desviación permitida de la distorsión para considerar que se detecta un segmento estacionario.

Estos umbrales dependen obviamente de la aplicación y de una evaluación subjetiva. El número de segmentos en que se segmentará un proceso será función de estos tres umbrales. El efecto de estos parámetros será discutido en el capítulo siguiente. Las variables intermedias d_L y σ_L representan el valor promedio y la desviación de la distorsión en el instante n , usando las L muestras anteriores de la señal de distorsión. La variable k que aparece en el algoritmo es el desplazamiento relativo del bloque móvil al cual estamos analizando, con respecto al bloque de referencia que es el primer bloque del segmento estacionario inmediatamente anterior. Por otra parte, la variable j desempeña el papel del contador de número de muestras de transición entre el segmento estacionario detectado y el segmento anterior.

V-3. Consideraciones Adicionales del Algoritmo

En esta sección, consideramos algunos detalles de la señal de voz, tomamos unas medidas de ajuste para un mejor funcionamiento de la segmentación.

El algoritmo de segmentación de la sección anterior, es una versión simplificada del procedimiento que presentaremos a continuación, en el sentido de que no se está tomando en cuenta algunas observaciones importantes sobre las señales en que estamos trabajando. En otras palabras, tanto el modelo de fuente compuesta como las proposiciones sobre distorsión entre segmentos son muy idealistas para un buen funcionamiento de segmentación de voz. Para lograr unos ajustes adecuados y sencillos sobre el algoritmo dado anteriormente, primero debemos analizar la señal de voz y el comportamiento de la distorsión como función del desplazamiento.

La variación de potencia es una de las características más sobresalientes de la señal de voz. Cuando uno habla, no solamente cambia la forma de onda de la señal, sino también la amplitud o la ganancia, la cual se refleja la intensidad de la voz, aun conservando la forma de onda. Para ser más tratable el efecto de la variación de potencia, reemplazamos el término "potencia" por la ganancia del filtro AK. El efecto de la variación es más notorio en los extremos que la parte central de cada segmento estacionario. Esto se debe a que los segmentos a priori y posteriori tienen generalmente ganancias distintas que la del presente. La figura 3 nos ayuda a visualizar este efecto.

Tomando en cuenta el efecto de variación de ganancia, el modelo de fuente compuesta se puede tener una ligera modificación para representar mejor la señal de voz. La modificación consiste en introducir un amplificador de ganancia variable, mientras los procesos de las subfuentes son normalizados a ganancia unitaria. Entonces, el modelo de fuente compuesta modificado será como mostrado en la figura 4.

Una restricción sobre la ganancia variable $G(n)$ es que dentro de

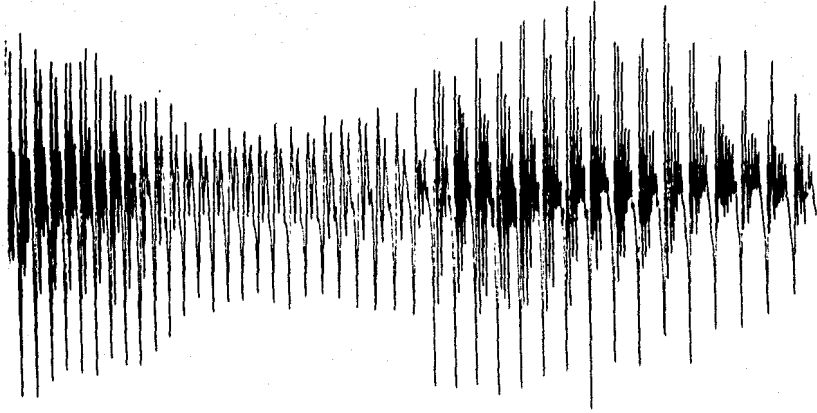


Figura V.3 Señal típica de voz con variación de potencia

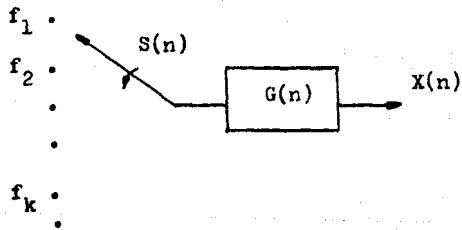


Figura V.4 Modelo de fuente compuesta modificado

un mismo segmento estacionario, $G(n)$ debe tener cambios suficientemente pequeños para que se pueda ignorar la variación de la ganancia dentro de cualquier bloque mínimo representativo. Esto es importante por que la estimación del espectro muestral de dichos bloques se requiere la estacionaridad de ellos.

Para lograr un mejor comportamiento de la segmentación de señal de voz, la compensación de la variación de ganancia representa un aspecto importante. Comparamos el modelo de fuente compuesta modificado y el modelo de fuente compuesta introducido anteriormente, nos damos cuenta de que la diferencia de estos dos modelos consiste precisamente del amplificador de ganancia variable $G(n)$. Si podemos de alguna forma conocer $G(n)$ y dividir la señal observada por $G(n)$, entonces la salida va a comportar como señal proveniente de modelos normalizados, y a ésta va es aplicable el algoritmo de la sección V.2. Recordamos que la ganancia de LPC es proporcional a la potencia promedio de la secuencia en que se obtiene el filtro AR, y la potencia promedio a la vez es proporcional a $G(n)$ cuando la secuencia es estacionario. De esa manera podemos estimar $G^2(n)$ por medio de la ganancia de predicción $E_{\hat{g}}(n)$.

Como a nosotros solamente nos interesa que la distorsión sea insensible a la variación de ganancia, la normalización de la señal se puede realizar en el momento de calcular la distorsión la cual es equivalente usar la distorsión de ganancia normalizada. La distorsión de ganancia normalizada fue introducida en el capítulo IV, que comporta como deseamos : separar los efectos de la ganancia y los modelos normalizados. Entonces el primer ajuste del algoritmo es el uso de la medida de distorsión de ganancia normalizada.

Otra característica de la señal de voz es el "pitch". La voz es una onda de presión acústica originada por movimientos psicológicos voluntarios del aparato vocal. Hay dos tipos de sonidos : sonoro y sordo. En el español, la mayor parte del sonido corresponde al primero. La señal de este tipo de voz luce localmente periódica. El pitch se caracteriza por el periodo o la frecuencia fundamental de estas señales. El mecanismo de producción de voz se

puede modelar por el esquema mostrado en la figura 5. Este esquema contiene un sintetizador el cual convierte algún tipo específico de señal de excitación imitiendo lo mejor posible a la forma de onda producida por el aparato vocal. La señal de excitación puede ser un tren de pulsos unitarios con periodo del pitch o ruido blanco de ganancia unitaria dependiendo si se produce un sonido sonoro o sordo.

Cada periodo de pitch consiste de varios ciclos, los cuales se definen como la parte de señal comprendida entre dos cruces consecutivas con cero. El primer ciclo del periodo es llamado el ciclo principal que tiene la mayor amplitud que los demás ciclos. Las amplitudes de los ciclos siguientes se caen exponencialmente. Este fenómeno se puede visualizar en la figura 6.

Ahora vamos a ver cómo se puede afectar el pitch al algoritmo de segmentación. Recordamos que la segmentación se basa en el cálculo de la distorsión entre un bloque de referencia y otro móvil usando la fórmula de Itakura-Saito. Esta distorsión se determina por los coeficientes de autocorrelación de los dos bloques. Teóricamente la distorsión se mantiene aproximadamente constante cuando el bloque móvil cae dentro de un segmento estacionario. Sin embargo, uno no puede conocer exactamente las correlaciones reales sino las estimadas. La fórmula para estimar las correlaciones fue presentada en el capítulo III (ver III.24): sea $X_n = (x(n), \dots, x(n+N-1))$ un bloque de muestras, las correlaciones se estiman por

$$R(m) = \frac{1}{N} \sum_{k=0}^{N-m-1} x(n+k)x(n+k+m) \quad m=0, \dots, M.$$

Cuando se mueve el bloque una muestra hacia adelante, tenemos el bloque X_{n+1} y las correlaciones correspondientes R_{n+1} , entonces la diferencia de las correlaciones por el corrimiento será

$$R_{n+1}(m) - R_n(m) = [(x(n+N-m)x(n+N) - x(n)x(n+m))] / N.$$

Observando la ecuación anterior, cuando le toca a alguna de las muestras el valor pico del ciclo principal de un periodo de pitch, la

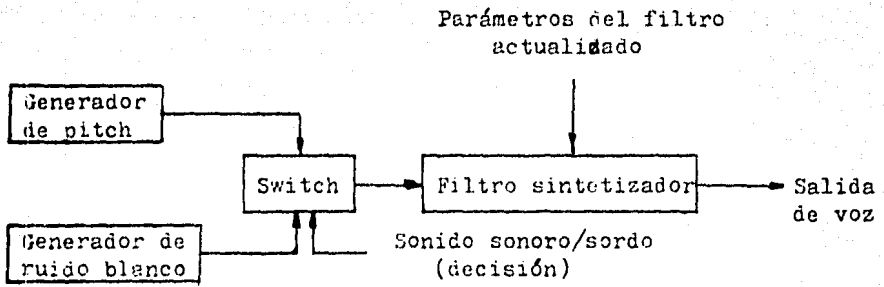


Figura. V.5 Esquema de la reproducción de voz

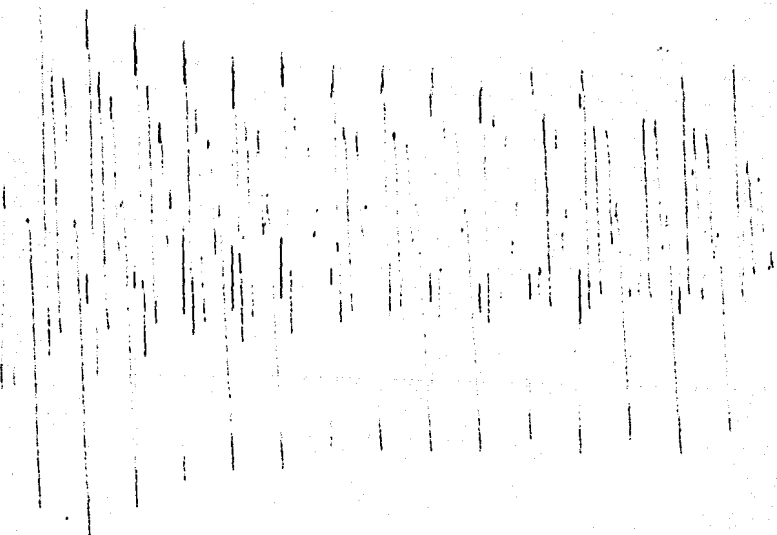


Figura V.6 Señal típica de voz con pitch

diferencia puede ser considerablemente grande. Esto se debe a que generalmente el valor pico del pitch es mucho mayor que los del resto de las muestras por un lado, y por otro el tamaño del bloque N no puede ser suficientemente grande para absorber esta diferencia grande. Esta variación grande en las correlaciones se produce en cada periodo de pitch, y su consecuencia es una oscilación del mismo periodo en la señal de distorsión. La amplitud de la oscilación depende del valor de la distorsión en ese momento. Mientras la distorsión sea mayor, la oscilación será más fuerte.

Una forma simple y efectiva de compensar esta oscilación es pasar la señal de distorsión por un promediador, que tiene función como un filtro pasa bajas. Una vez que la filtración está hecha, ya podemos aplicar el algoritmo de segmentación y tomar las decisiones sobre la señal filtrada.

Cuando cambie la señal de un segmento estacionario al otro, la detección de la estacionaridad del segundo segmento se lleva a cabo con la ayuda de dos variables auxiliares d_L y σ_L , las cuales se calculan de acuerdo con las siguientes expresiones:

$$d_L = \frac{1}{L} \sum_{k=0}^{L-1} d(x_{n-1}, x_{n-1+k})$$

$$\sigma_L = \frac{1}{L} \sum_{k=0}^{L-1} |d_L - d(x_{n-1}, x_{n-1+k})|.$$

Cuando la desviación σ_L es pequeña ($\sigma_L \leq c$) y d_L es grande ($d_L > b$), se considera que en este momento empieza otro segmento estacionario.

Como se comentaba anteriormente que la oscilación es más fuerte cuando la distorsión d_L es mayor, σ_L no es solamente función de estadísticas del segmento actual, sino también del segmento anterior en que se hace referencia. Mientras que la distorsión entre estos dos segmentos sea más grande, ésta tiene una mayor desviación σ_L . Para que esto no se afecte a la detección de estacionaridad, el umbral c debe ser proporcional al

crecimiento de σ_L causado por el crecimiento de la distorsión d_1 . Empíricamente, el factor c se puede expresar en función de d_1 como:

$$c = f(d_1) \\ = c_0 (1 + d_1)^2.$$

Cuando la estadística de segundo orden en los segmentos aquí llamados estacionarios cambia lentamente, resulta conveniente realizar un proceso iterativo como el que se describe a continuación. Una vez que se ha segmentado la señal con el algoritmo antes descrito, se repite el procedimiento pero en vez de tomar la estadística de segundo orden del primer bloque del segmento, se toma la estadística de segundo orden de todo el segmento que se considera estacionario. Al llevar a cabo una segunda pasada con esta modificación, se logran ajustes en las fronteras de los segmentos estacionarios. Lo anterior puede ser ocasionado porque posiblemente el primer bloque del segmento estacionario no sea el más representativo de todo el segmento.

V-4. Implantación Computacional

En esta sección, describimos detalladamente el procedimiento de la segmentación, tomando en cuenta los ajustes considerados en la sección V.3 tales como normalización de ganancia, compensación del efecto de pitch, de la dispersión de distorsión, etc. El criterio de fidelidad para cuantificar diferencias entre segmentos será la distorsión Itakura-Saito.

El procedimiento de segmentación se resume como sigue:

- 1) Limpia el contador de muestras de transición j : $j=0$.

El desplazamiento toma su valor inicial cero: $k=0$.

Posiciona el apuntador del bloque de referencia n en el inicio del archivo de datos.

- 2) Asigna el bloque de referencia $X_n=(x(n), \dots, x(n+N-1))$.

Calcula las primeras $p+1$ correlaciones $R_n(m)$.

$$R_n(m) = \frac{1}{N} \sum_{j=0}^{N-m-1} x(n+j)x(n+j+m) \quad m=0, \dots, p.$$

Obtiene los parámetros $a_n(m), E_n^2(n)$ del espectro

usando el algoritmo de Levinson (ver la sección III.3).

- 3) Incrementa el desplazamiento k en una unidad: $k=k+1$.

Asigna el bloque móvil $X_{n+k}=(x(n+k), \dots, x(n+k+N-1))$.

Calcula las correlaciones $R_{n+k}(m), m=0, \dots, p$

$$R_{n+k}(m) = [x(n+k+N-m-1)x(n+k+N-1) - x(n)x(n+m)] / N + R_n(m)$$

y la ganancia de predicción $E_{n+k}^2(n+k)$ con

el algoritmo Levinson.

- 4) Evalúa $d(X_n, X_{n+k})$:

$$d(X_n, X_{n+k}) = \frac{I_p(\underline{a})}{E_{n+k}^2(n+k)} - 1$$

donde

$$I_p(\underline{a}) = R_n(0)R_{n+k}(0) + 2 \sum_{m=1}^p R_n(m)R_{n+k}(m)$$

$$R_n(m) = \sum_{k=0}^{p-m} a_n(m+k) a_n(m) \quad m=0, \dots, p.$$

5) Pasa $d(x_n, x_{n+k})$ por un promediador de longitud LP, obteniendo $d(x_n, x_{n+k})$:

$$d(x_n, x_{n+k}) = d(x_n, x_{n+k-1}) + L(d(x_n, x_{n+k}) - d(x_n, x_{n+k-LP})) / LP$$

6) Evalúa

$$d_L = \frac{1}{L} \sum_{k=1}^{L-1} d(x_{n-1}, x_{n-1+k})$$

$$e_L = \frac{1}{L} \sum_{k=1}^{L-1} |d_L - d(x_{n-1}, x_{n-1+k})|$$

7) Si d_L no excede la primera cota a , es decir, $d_L < a$: se considera que no hay cambio, mantiene el contador $j=0$, regresa al paso 5):

en caso contrario, continúa al paso 8).

8) Si d_L excede la segunda cota b , y la desviación media

e_L cae dentro del rango preestablecido $c(d_L)$:

$$e_L < c(d_L) = c_0(d_L+1)^2,$$

se detecta la estacionaridad del segmento siguiente, el contenido de j en este momento indica el número de muestras en transición entre dos segmentos estacionarios. Escribe los coeficientes

$R_n, a_n, E_n^2(n)$ y j en un archivo de salida,

reinicializa el proceso con $n=n+k+N-1$, $k=0$,

limpia el contador j : $j=0$.

regresa al paso 2):

en caso contrario, incrementa el contador de transición j : $j=j+1$;

regresa al paso 3).

El sistema de segmentación es implantado con las siguientes especificaciones :

la frecuencia de muestreo = 6.4 kHz

la longitud de los bloques $N = 128$

el orden del modelo $p = 10$

y $L=5$.

Otros parámetros del algoritmo son libres de escoger por el usuario del sistema, los valores típicos son :

$LP = 20$

$b = 0.35$

$a = b / 2$

$c_n = 0.00045$.

El efecto de las variación de estos parámetros será discutido en el capítulo siguiente.

La frecuencia de muestreo de 6.4 kHz es el doble del ancho de banda de la señal de voz el cual es aproximadamente 3.2 kHz. Esta frecuencia es la mínima necesaria sin perder información de acuerdo con el teorema de Nyquist. La longitud de los bloques no es fijada arbitrariamente. Un bloque de 128 muestras dura 20 milisegundos y contiene 2-3 periodos de pitch. Esta duración es comúnmente considerada representativa de un segmento estacionario sin ser demasiado grande.

La selección del orden del filtro es basada principalmente en dos consideraciones. Por un lado, un buen modelo necesita un orden alto, pero el mejoramiento es cada vez menos significativo conforme incrementa el orden; y por el otro lado, el espectro estimado por un filtro de orden muy alto produce detalles engañosos debido a que las correlaciones son estimadas por una cantidad relativamente pequeña de muestras, a pesar de que teóricamente un filtro de alto orden permite que el espectro estimado persigue mejor a las características de espectro real. El orden típico es de 10 a 20. El orden 10 es seleccionado con mayor frecuencia en las literaturas actuales. Esto nos facilita comparar el comportamiento de nuestro sistema con los demás esquemas.

La implantación del sistema de segmentación es llevado a cabo en el Laboratorio de Cálculo Automatizado para el Diseño (CAD) de

la DEPEI (la División de Estudios de Posgrado de la Facultad de Ingeniería) de UNAM. El equipo usado consiste en una minicomputadora PDP11/40 y sus periféricos. El sistema operativo de dicha computadora es de la versión 4.0. Este sistema operativo es de un solo usuario, que permite una amplia gama de periféricos.

Los periféricos manejados por la computadora que son de nuestro interés, son los siguientes : dos unidades de disco con capacidad de almacenamiento de 2.2 Mbytes cada uno, un convertidor A/D de 10 bits, un convertidor D/A de 8 bits, una graficadora CALCOMP 907/1039 y un video de alta resolución (1024x1024 puntos). Los convertidores tanto A/D como D/A pueden trabajar a una velocidad de muestreo hasta de 10 kHz teniendo acceso directo al disco en tiempo real. Esto nos permite el fácil manejo de grandes volúmenes de datos entre la computadora y los convertidores sin ser limitado por la capacidad de la memoria principal que es moderadamente pequeña.

Antes de ser muestreada la señal de voz, ésta se pasa por un filtro paso bajas con frecuencia de corte de 3.2 kHz. El archivo de los datos muestreados es formado por registros de 128 muestras almacenadas como números reales no formateados. El acceso a tal archivo puede ser directo o secuencial. Esto permite una mayor facilidad de manejo de datos. El tamaño de los registros se fija en 128, no solamente es una simple coincidencia con la longitud de los bloques manejados por el algoritmo, sino también se optimiza la utilización de los discos debido a que los bloques físicos del disco son múltiplos de 128.

Los programas que forman el sistema de segmentación de la señal de voz se anexan en el apéndice I.

Algunos ejemplos se presentan en la figura V.7.

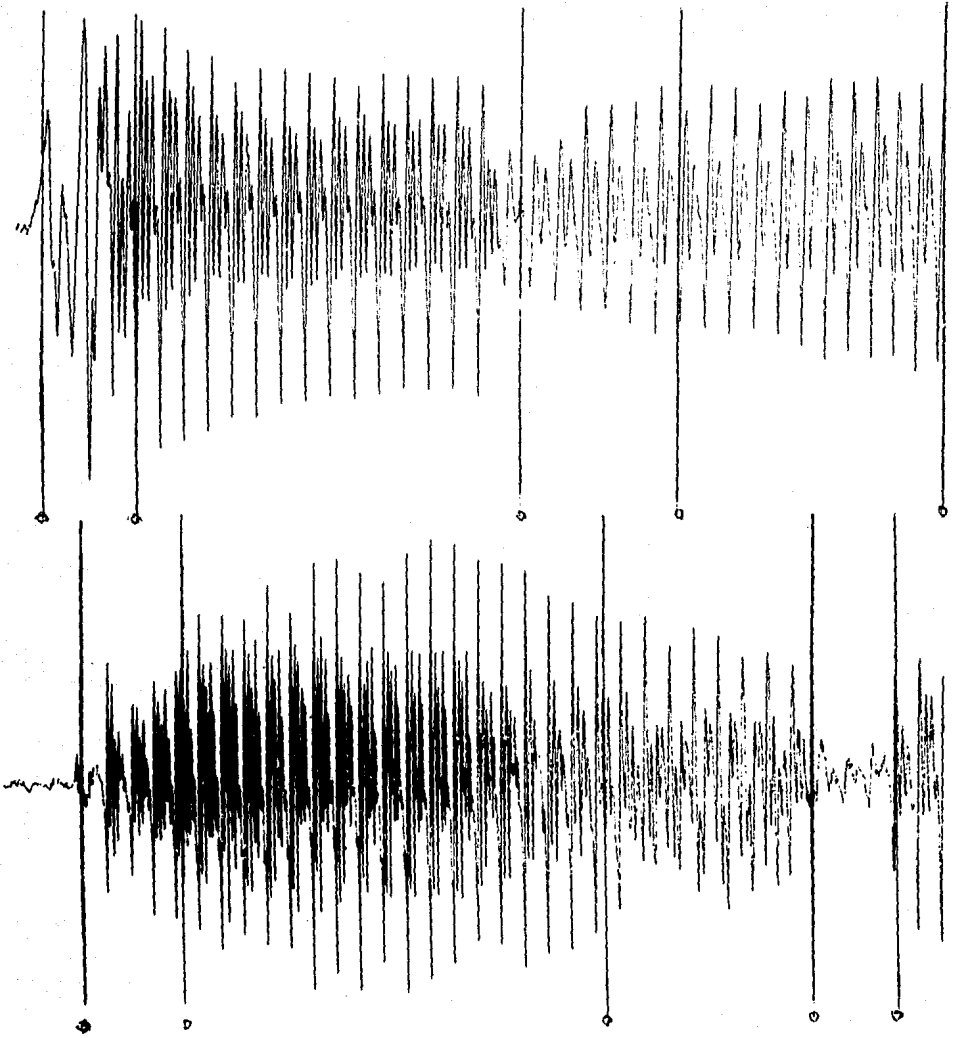


Figura V.7 Ejemplos de señales segmentadas de voz

CAPITULO VI

DISCUSION SOBRE EL ALGORITMO DE SEGMENTACION

VI-1. Efectos por Variacion de Parámetros

Hemos mencionado en el capítulo IV que una de las características principales de una medida de distorsión es la subjetividad. En el contexto de la segmentación de voz, la subjetividad consiste básicamente en la sensibilidad auditiva. Es decir, dos segmentos son subjetivamente diferentes siempre y cuando esta diferencia puede ser distinguida por el oído humano. Esta diferencia se refleja generalmente en la forma de onda de la señal, si uno grafica la señal de voz contra el tiempo. La señal de la figura VI.1 corresponde al sonido "AU" de la palabra "AURIOLA". El cambio de sonido de "A" a "U" es afirmado por el oído humano. Por el otro lado, observando esta señal, es fácil de notar el cambio de la forma de onda de "A" a "U". Sin embargo, no es nada fácil de definir las fronteras de "A", "U" y la transición en caso de considerar. En otras palabras, los criterios de decidir cuando la señal ha cambiado suficientemente son muy distintos según la aplicación y la preferencia de cada quien. Dependiendo del caso, uno puede tomar decisión estricta o relajada. La distorsión del bloque móvil con respecto al primer bloque de esta señal, el cual se considera como el bloque de referencia, se encuentra en la figura VI.2. Se observa que se presenta la misma dificultad para tomar decisión. Entonces, cómo se toma en cuenta el algoritmo de segmentación esta preferencia es el tema de discusión de esta sección.

Recordamos que la medida de distorsión de Itakura-Saito es subjetivamente sensible al oído humano. Consecuentemente, el algoritmo de segmentación presentado tiene el mismo grado de sensibilidad subjetiva que la distorsión misma. En el algoritmo, los parámetros libres de ser escogidos por el usuario son los siguientes

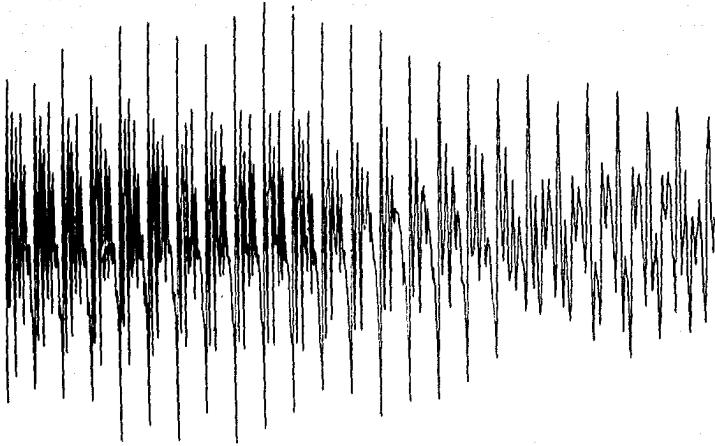


Figura VI. 1 Señal asociada al sonido "AU"

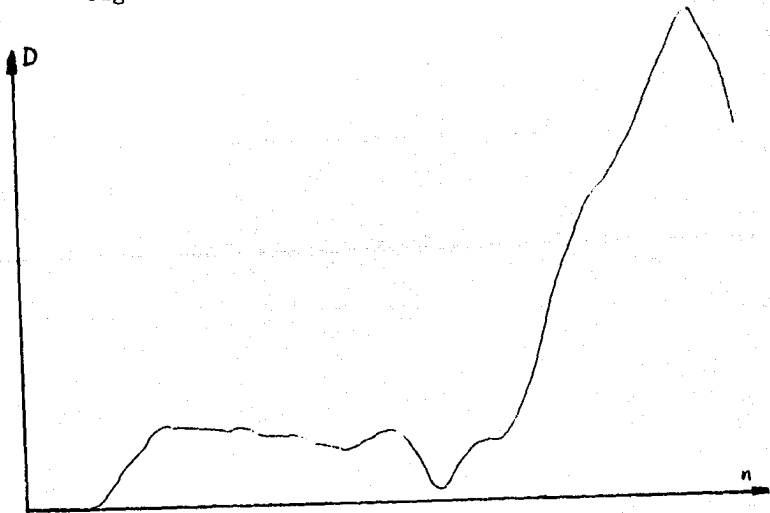


Figura VI.2 Distorsión con respecto al 1^{er} bloque

los umbrales b y c_0 , y la longitud del promediador LP. A continuación, vamos a explotar el significado y la influencia de cada uno de ellos.

El significado del umbral b es muy claro. Este umbral establece la cota inferior de la distorsión entre dos segmentos estacionarios diferentes cualesquiera de la señal en consideración. Es equivalente decir que la distorsión de ganancia normalizada entre cualquier par de bloques de diferentes segmentos estacionarios debe tener un valor por lo menos b . Un valor grande del umbral b impide la detección de cambios entre segmentos estacionarios cercanos o parecidos en el sentido subjetivo. Esto corresponde a una decisión relajada.

Por la propia conveniencia, definimos el otro umbral "a" como la mitad del umbral b . Entonces, éste no solamente afecta la mínima distorsión entre diferentes segmentos, sino también al mismo tiempo la distorsión máxima permitida entre bloques del mismo segmento estacionario. Obviamente, mientras que la cota superior de esta distorsión sea más grande, la diferencia subjetiva entre los bloques dentro del segmento será mayor. Esto implica una estacionaridad más débil de los segmentos.

Por el otro lado, un valor pequeño de b no permite que los bloques de un mismo segmento tengan distorsiones grandes entre sí, lo cual establece una decisión estricta de la segmentación. Sin embargo, un valor pequeño de b no necesariamente significa que el siguiente segmento estacionario se detecta antes que el caso de un valor grande de b , la razón es que la detección de un segmento estacionario depende principalmente del otro umbral c_0 . La otra contribución de un valor pequeño de b puede ser el permiso de una mayor duración de la transición entre dos segmentos estacionarios contiguos.

En el algoritmo, el criterio de la estacionaridad de un segmento es el siguiente : la distorsión media d_i excede el umbral b y simultáneamente la desviación σ_L cae dentro del límite preestablecido c . Como se comentaba en el capítulo anterior, la

desviación ϵ_L es una función creciente de la distorsión d_L , el umbral c debe tomarse en consideración esta relación entre ϵ_L y d_L . Experimentalmente, obtenemos una función

$$c = c_0 (1 + d_L)^2$$

la cual permite lograr un buen funcionamiento del algoritmo. En esta expresión, la constante c_0 juega un papel muy importante en la toma de decisión de la estacionaridad. Evidentemente, si c_0 toma valores extremos del rango permisible, se degenera el funcionamiento del umbral c . Cuando el valor de c_0 tiende a cero, el umbral c también tiende a cero proporcionalmente tal que ϵ_L no puede caer nunca dentro del límite establecido por c . En este caso, el algoritmo no detectará ninguna estacionaridad, la consecuencia es no segmentar nada la señal. En el otro caso extremo, si c_0 tiene un valor muy grande que hace crecer c de tal manera que ϵ_L siempre es menor que c . El resultado será como si ignorara la segunda condición del criterio de estacionaridad. En otras palabras, para detectar otro segmento estacionario solamente es necesario que la distorsión d_L excede la cota b .

Dentro del rango nominal de c_0 , el valor de éste controla tanto el momento en que se empieza un nuevo segmento como el número total de segmentos. Un valor grande permite una detección rápida del siguiente segmento que cumple las condiciones establecidas. Pero a veces el nuevo segmento detectado no es suficientemente estacionario, es decir, la referencia que es el primer bloque del segmento no es representativo del segmento. Entonces la distorsión con respecto a este bloque crece muy rápido, y el efecto es dividir la señal en un mayor número de segmentos. Por el lado contrario, con un c_0 pequeño se tiene generalmente un mayor retardo para detectar la estacionaridad. Un c_0 pequeño tiene la ventaja de que se evite tener muchos segmentos, sobre todo los segmentos que corresponden a la transición, pero al mismo tiempo se corre el riesgo de que no detectar el cambio entre segmentos que son aparentemente diferentes.

El último parámetro por fijarse es la longitud del promediador LP . Este parámetro no es tan influente como b y c_0 , pero sí

tiene la importancia que le corresponde. Un valor adecuado de LP puede regularizar el compromiso entre el retardo de detección y la compensación de la dispersión de distorsión. Una longitud grande del promediador suaviza más la distorsión, en otras palabras, la oscilación está mejor compensada; pero el promediador de LP larga absorbe las variaciones en distorsión produciendo un retardo considerable de la detección de cambios estadísticos. Reduciendo la longitud del promediador, se disminuye el retardo de la detección de cambios, pero causa problemas a la detección de estacionariedad teniendo poca suavizada la distorsión.

Estos parámetros pueden ser escogidos de acuerdo con la aplicación y la evaluación subjetiva. En nuestro caso, la segmentación es dirigida a las aplicaciones como la codificación LPC de tasa de transmisión variable y el reconocimiento de voz. Para este tipo de aplicaciones y la propia subjetividad, los valores son escogidos como:

$$b = 0.35;$$

$$c_0 = 0.00045;$$

$$LP = 20.$$

Para visualizar la influencia de cada uno de estos parámetros, se presentan algunos ejemplos de segmentación con diferentes valores de ellos. La señal de prueba es la misma para los cuatro casos que se muestran en las figuras. La figura VI.3 muestra la señal segmentada con los valores típicos, es decir, $b=0.35, c_0=0.00045$ y $LP=20$. En las figuras VI.4-VI.6, se muestran la señal segmentada con los valores $b=0.3, c_0=0.00045, LP=20$; $b=0.35, c_0=0.0004, LP=20$; y $b=0.35, c_0=0.00045, LP=50$ respectivamente.

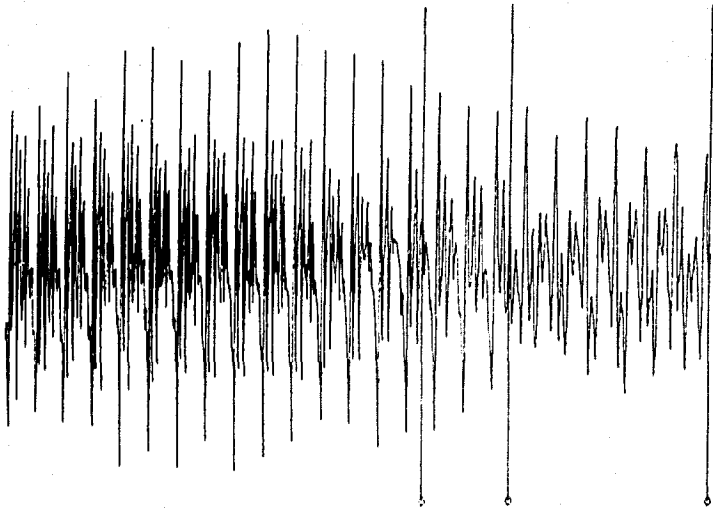


Figura. VI.3 Señal segmentada con $b=0.35, c_0=0.00045$
y $LP=20$.

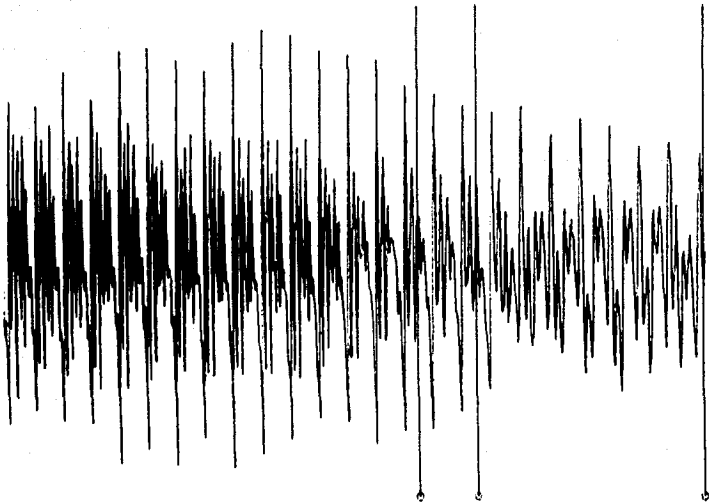


Figura VI.4 Señal Segmentada con $b=0.3, c_0=0.00045$ y $LP=20$

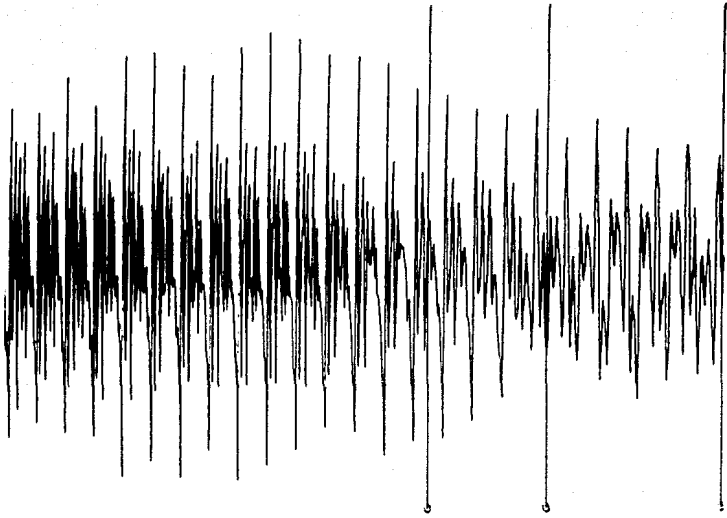


Figura VI.5 Señal segmentada con $b=0.380$, $c_0=0.0004$, y $LP=20$

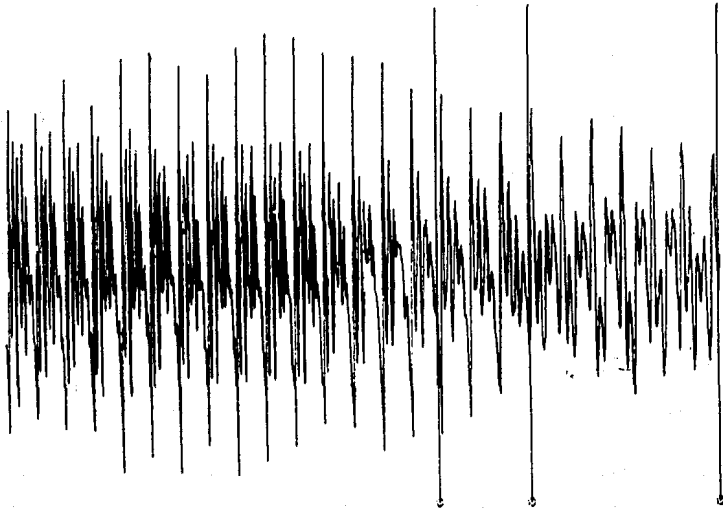


Figura VI.6 Señal segmentada con $b=0.35$, $c_0=0.00045$ y $LP=50$

VI-2. Comparación con Otros Algoritmos

En esta sección, se describen algunos otros esquemas de segmentación, y se comparan con el algoritmo desarrollado en el capítulo anterior.

Uno de los esquemas está enfocado en la detección secuencial de cambios en características espectrales de señales digitales ([35]). Bajo este esquema, la segmentación convierte simplemente en la detección de cambios. Una forma efectiva para detectar cambios en modelos estadísticos consta de

- a) filtrar la señal observada $x(n)$ por uno o varios filtros conocidos, y
- b) construir una prueba estadística de tipo acumulativo la cual es afectada por el cambio de modelo.

Se supone que la señal observada $\{x(n)\}$ es descrita por un modelo AR cuyos parámetros pueden variar en algún momento desconocido q , es decir,

$$x(n) = - \sum_{k=1}^p a^n(k)x(n-k) + e(n)$$

$$\text{Var}(e(n)) = E^2(n)$$

donde

$$a^n(k) = a^0(k) \quad 1 \leq k \leq p, n < q;$$

$$E^2(n) = E_0^2 \quad n < q;$$

$$a^n(k) = a^1(k) \quad 1 \leq k \leq p, n \geq q;$$

$$E^2(n) = E_1^2 \quad n \geq q;$$

y $e(n)$ es una secuencia de ruido blanco. Los parámetros AR se denotan por

$$A^j = (a^j(1), \dots, a^j(p)), \quad j=0,1;$$

y las observaciones pasadas por

$$X(n-1) = (x(n-1), x(n-2), \dots, x(1))^T.$$

Los casos básicos se distinguen en lo siguiente: el caso ideal de modelos conocidos y la situación real de modelos desconocidos. En el caso ideal, se supone que los parámetros son conocidos antes y

después del cambio, y la única cantidad desconocida es el tiempo q en que ocurre el cambio. Entonces el problema consiste en la detección secuencial (con un posible retraso) del cambio y la estimación del tiempo q . Este caso ideal es considerado principalmente para la derivación del algoritmo de interés junto con su interpretación estadística.

En la práctica, frecuentemente se desconocen los parámetros reales de los modelos AR antes y después del cambio. Además, la estructura de los modelos reales es desconocida, y los modelos AR son usados solamente como una herramienta para la segmentación. En estas situaciones, el problema de la identificación secuencial de modelos AR surge aparte del problema de detección del cambio y estimación.

En situaciones tanto ideal como real, las principales propiedades deseables de algoritmos son las siguientes :

- a) pocas falsas alarmas;
- b) poco retraso para la detección;
- c) detección simétrica.

Hay dos posibles maneras para llevarse a cabo estas propiedades, de acuerdo con el número de modelos AR usados. Una manera posible consiste en filtrar la señal observada $\{x(n)\}$ por un filtro AR conocido y buscar cambios en la señal del error residual $\{e(n)\}$, es decir, probar qué tan diferente es $\{e(n)\}$ de un ruido blanco. Actualmente el uso de las técnicas de suma acumulada basadas en el error residual $e(n)$ es una manera estándar para la detección de cambios. La otra alternativa es precisamente la que usamos en el capítulo anterior que emplea dos modelos AR y una medida de distorsión adecuada.

En el artículo [35], M. Basseville et. al. analizan tres pruebas estadísticas diferentes.

La primera prueba usa solamente un modelo AR:

Sea la señal observada $x(n)$ descrita por una densidad de probabilidad condicional $g^0(x(n);X(n-1))$ y $g^1(x(n);X(n-1))$ antes y después del cambio respectivamente. Se consideran las siguientes estadísticas :

$$U(n) = \sum_{k=1}^n I(k)$$

donde

$$I(n) = g^q(x; X(n-1)) \log [g^q(x; X(n-1))] dx \\ - \log [g^q(x(n); X(n-1))]$$

Esta estadística tiene un valor esperado condicional nulo bajo la hipótesis H_0 , i.e., antes de la ocurrencia del cambio.

Además, se puede demostrar que el valor esperado condicional de $I(n)$ después del cambio es

$$E_1(I(n); X(n-1)) = H(q^1) - H(q^0) + I(q^1; q^0)$$

donde H e I son entropías condicional y cruzada respectivamente.

Una condición suficiente para la positividad de la variación de $U(n)$ después del cambio es obviamente

$$H(q^1) \geq H(q^0)$$

es decir, la entropía condicional no decrece.

Cuando el proceso es gaussiano, se obtiene la suma acumulada

$$U(n) = \frac{1}{2} \sum_{k=1}^n \left(\frac{e_0^2(k)}{E_0} - 1 \right)$$

donde $e_0(k)$ es el error residual en el instante k con respecto al modelo AR con parámetros A^0 y E_0 . En este caso, la condición suficiente de la positividad es

$$E_1 \geq E_0.$$

Cuando $E_1 < E_0$, $E_1(I(n); X(n-1))$ puede ser positivo o negativo. Debido a este comportamiento imordecible de esta prueba para una gran parte de situaciones en las cuales la energía de la señal decrece después del cambio.

Otra desventaja de esta prueba se debe a que en la condición de positividad, no se considera la entropía mutua, sino solamente las entropías diferenciales de las densidades condicionales g^0 y g^1 .

Otras dos pruebas propuestas de suma acumulada usan dos modelos AR:

a) La prueba es la razón de similitud. En este caso, la prueba de suma acumulada $U'(n)$ tiene forma

$$U'(n) = \sum_{k=1}^n T'(k)$$

donde

$$T'(n) = \log \frac{g^1(x(n); X(n-1))}{g^0(x(n); X(n-1))}$$

Obviamente, la estadística de la prueba no es más que la razón de similitud entre dos probabilidades conjuntas $p^0(x(n), x(n-1), \dots, x(1))$ y $p^1(x(n), x(n-1), \dots, x(1))$. En otras palabras, ésta es una prueba de la hipótesis "todas las observaciones obedecen la densidad condicional g^0 " contra la hipótesis "todas las observaciones obedecen g^1 ", la cual no es adecuada para el problema planteado en que la hipótesis alternativa es "las observaciones $x(k)$ obedecen a g^0 en el intervalo $1 \leq k \leq q$, y obedece a g^1 en $q \leq k \leq n$ ". Sin embargo, se puede ver que esta es adecuada para implantación real.

b) La prueba se toma en cuenta tanto la entropía diferencial como la entropía mutua.

Se propone la estadística de suma acumulada como

$$U''(n) = \sum_{k=1}^n T''(k)$$

donde

$$T''(n) = g^0(x; X(n-1)) \log \frac{g^1(x; X(n-1))}{g^0(x; X(n-1))} dx \\ - \log \frac{g^1(x(n); X(n-1))}{g^0(x(n); X(n-1))}$$

La única diferencia entre $T'(n)$ y $T''(n)$ es la corrección por la media condicional antes del cambio.

En el caso de que el proceso AR sea gaussiano, se definen $e_0(n)$ y $e_1(n)$ como errores residuales de modelos 0 y 1 respectivamente. Entonces las dos pruebas $T'(n)$ y $T''(n)$ tienen la siguiente forma:

$$T'(n) = \frac{1}{2} \log \frac{E\tilde{\epsilon}}{E\tilde{\eta}} + \frac{e\tilde{\epsilon}(n)}{2E\tilde{\epsilon}} - \frac{e\tilde{\eta}(n)}{2E\tilde{\eta}}$$

$$y \quad T''(n) = \frac{e_0(n)e_1(n)}{E\tilde{\eta}} - \frac{E\tilde{\epsilon}}{2E\tilde{\eta}} \left(1 + \frac{e\tilde{\epsilon}(n)}{E\tilde{\epsilon}} \right) - \frac{E\tilde{\epsilon}}{2E\tilde{\eta}} + \frac{1}{2}$$

Se puede demostrar ([35]) que

$$E_0(T'(n) | X(n-1)) < 0,$$

$$E_1(T'(n) | X(n-1)) > 0;$$

$$E_0(T''(n) | X(n-1)) = 0,$$

$$E_1(T''(n) | X(n-1)) < 0.$$

Es razonable pensar que las sumas acumuladas de estas pruebas descritas tengan comportamientos similares que los valores esperados de las mismas. Entonces, las sumas acumuladas pueden ser utilizadas para la segmentación de señal por detección de cambios de modelos. El algoritmo de segmentación bajo este enfoque es relativamente sencillo: si la suma acumulada excede alguna cota preestablecida, se considera un cambio y reinicializa el proceso. En situaciones reales, los parámetros de los filtros AR son estimados con los métodos convencionales.

En la figura VI.7, se ilustra el comportamiento de $U(n)$, $-U''(n)$ y $U'(n)$ para una señal AR con parámetros:

$$a^0(1) = 0.85, \quad a^0(2) = -0.25, \quad a^0(3) = 0.06;$$

$$a^1(1) = 1.33, \quad a^1(2) = -0.45, \quad a^1(3) = -0.04;$$

y ganancias unitarias para los dos filtros.

Comparando las gráficas anteriores con los resultados obtenidos por M. Baseville et. al., se nota que los comportamientos resultantes

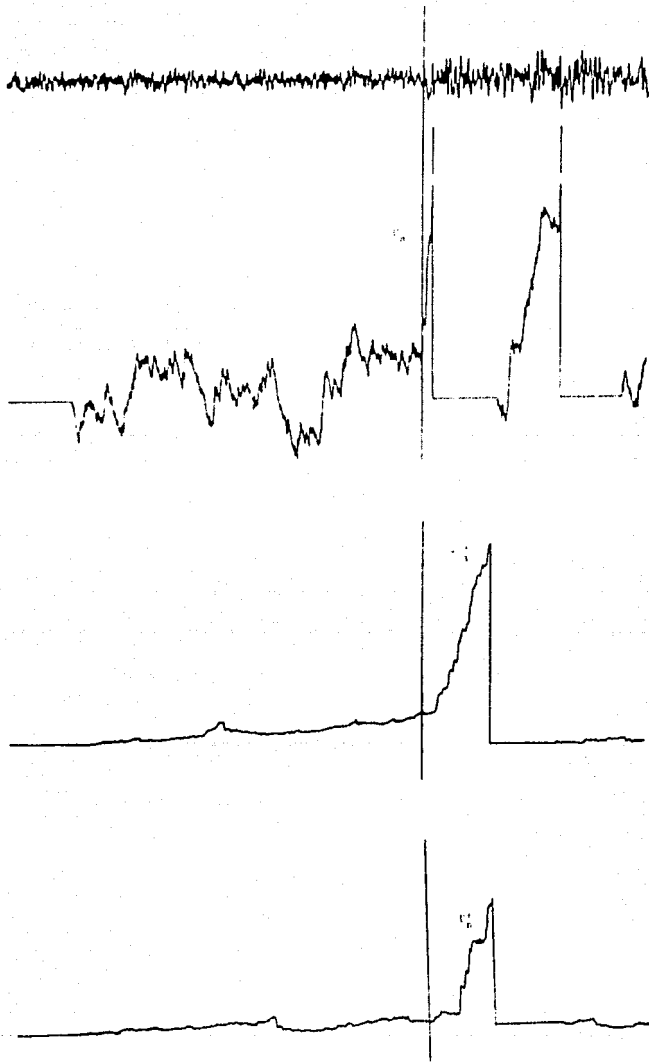


Figura VI.7 La señal de simulación y el comportamiento de las pruebas U_n , U'_n y U''_n .

de distorsión y el retardo de la detección del cambio por los diferentes algoritmos son bastante similares para la señal sintética por filtros AR de bajo orden. Sin embargo, para lograr el objetivo de segmentar alguna clase particular de señales, entre ellas, la señal de voz, los algoritmos presentados por [35] llevan ciertas desventajas obvias sobre el nuestro. Como se comentaba en el capítulo anterior, la variación de ganancia y el pitch de una señal de voz influyen mucho en la segmentación de la misma. Los algoritmos que son sensibles a estos efectos, por ejemplo los presentados por [35], generalmente llegan a resultados lejanos de los deseados por la subjetividad. Además, los métodos que emplean medidas de distorsión son más tratables analíticamente, tienen una interpretación física y un significado subjetivo mejor. Particularmente la medida que usamos está íntimamente relacionada con el concepto de información mutua, la cual juega un papel muy importante en la teoría de información.

Una variante del algoritmo presentado utiliza la medida de distorsión Itakura-Saito versión discreta, en lugar de la versión continua. En este caso la distorsión se calcula de acuerdo con la siguiente expresión:

$$D(f,g) = \frac{1}{2nN} \sum_{k=0}^{N-1} \left(\frac{f(k)}{g(k)} - \log \frac{f(k)}{g(k)} - 1 \right)$$

Las muestras de los espectros f y g son obtenidas por sus coeficientes de correlación usando la técnica FFT. La ventaja de este algoritmo es su buen comportamiento de distorsión y retardo el cual se puede apreciar en la figura VI.10. Pero la desventaja del mismo es el gran consumo del tiempo de procesador debido a que se requiere realizar un FFT de N puntos para cada muestra de la señal de voz.

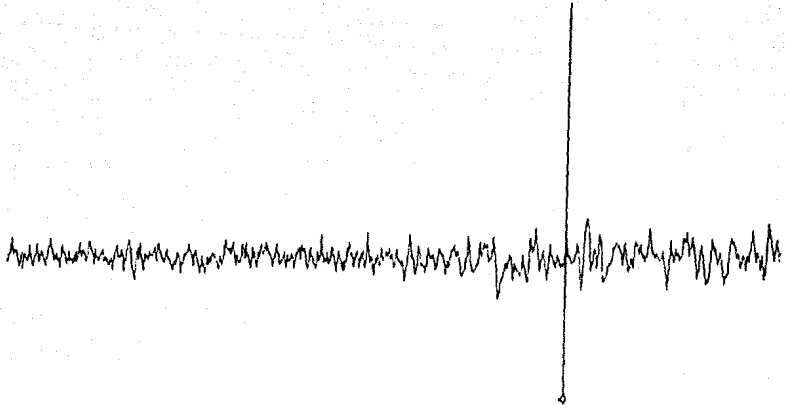


Figura VI.8 Señal sintética con los parámetros especificados en el texto

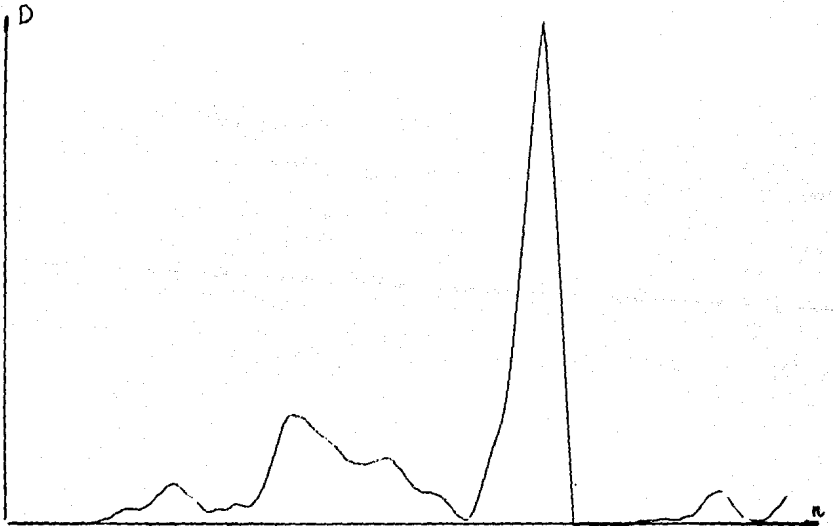


Figura VI.9 El comportamiento de distorsión usando la medida Itakura-Saito con ganancia normalizada

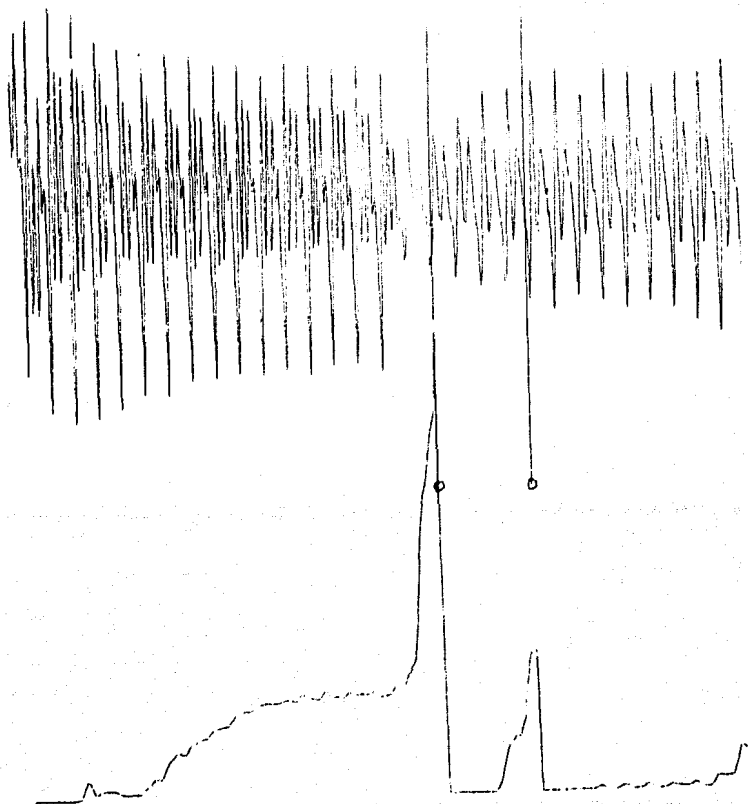


Figura VI.10 Ejemplo de segmentación con la medida de distorsión Itkura-Saito discreta

CAPITULO VII

APLICACIONES DE SEGMENTACION

En este capítulo, vamos a dar dos aplicaciones de la segmentación. Una de ellas va a ser la codificación LPC de razón de transmisión variable. La otra es el reconocimiento de patrones basado en segmentación. Estos dos sistemas han sido implantados experimentalmente en el Laboratorio de Cálculo Automatizado para el Diseño de la DEPTI.

VII-1. Codificación LPC de Tasa de Transmisión Variable

Debido a los recientes avances tecnológicos, el problema de la codificación, transmisión y almacenaje eficiente de señal de voz ha sido de gran interés. Una cantidad considerable de investigaciones se ha llevado a cabo en el desarrollo de nuevos algoritmos de codificación que aseguran una reducción aceptable de calidad con tasa de transmisión baja. Estas tasas bajas garantizan una adecuada utilización de canales de comunicación o un almacenaje eficiente a costo bajo de la información contenida en la señal.

Hoy en día, existen básicamente dos categorías de codificadores de señal de voz. La primera de ellas, en la cual el esquema de codificación trata de reproducir exactamente la forma de onda de la señal de voz, es llamada codificadores de forma de onda. El miembro más obvio de esta clase, es muestrear directamente la señal a frecuencia 6.4-8 kHz con un cuantizador de 8 bits requiriendo una tasa de transmisión de orden de 30-60 kbits/seg, con una muy buena calidad. Aceptando una calidad más baja e incrementando la complejidad del sistema, se puede reducir la tasa de bits a 16 kbits/seg aproximadamente. Estos esquemas que incluyen las técnicas clásicas como PCM, DPCM, DM y ADM, son descritos en [36] y [37]. La segunda categoría de codificadores de voz está basada en modelos del proceso de generación de voz. Como la modelación de

este proceso es equivalente a la modelación de la fuente de voz, esta clase se denomina codificadores de fuente o "vocoder". Estos sistemas son tratados también en [36] y [37]. Con un grado aceptable de degradación de calidad, estos sistemas se pueden reducir la tasa de bits a menos de 2.4 kbits/seg.

Una técnica ampliamente usada para la compresión de ancho de banda es basada en la codificación con predicción lineal (LPC) ([13], [14]), la misma se ha discutido en el capítulo II. En este esquema, los parámetros del codificador LPC son extraídos en una tasa fija de la señal de voz. Después los parámetros extraídos se cuantizan, codifican y transmiten al receptor a la misma tasa fija.

Como comentábamos en los capítulos anteriores, la señal de voz se puede modelar como proveniente de una fuente compuesta. En otras palabras, se supone que la señal consiste en la concatenación de segmentos estacionarios. Explotando el efecto de que los periodos de tiempo de los segmentos estacionarios no siempre tienen la misma duración, es posible desarrollar esquemas de codificación de tasa de transmisión variable con una calidad comparable que un sistema LPC estandar o de tasa fija. La idea fundamental de estos sistemas es transmitir los parámetros codificados solamente cuando la señal de voz ha tenido cambios suficientemente grandes en algún sentido predeterminado con respecto al segmento anterior.

La idea de codificar señal de voz con una tasa variable no es nueva. En un artículo reciente, Viswanata ([38]) describió un tal esquema de codificación que es basado en la medida de distorsión que es función de los coeficientes LAR's (Log Area Ratios) producidos recursivamente de los coeficientes de la predicción lineal. Su algoritmo calcula la distorsión entre dos bloques de señal desplazados e interpola para bloques intermedios. Después de calcular los valores reales de los parámetros de los bloques intermedios, se los comparan con los interpolados. Si la diferencia entre los valores reales y los interpolados excede a una cota dada, el proceso se reinicializa en el último bloque usado para la interpolación; en el caso contrario, el proceso se repite para el

siguiente bloque el cual será usado para interpolar.

Este algoritmo usa un número variable de bloques de señal de longitud constante para la codificación de voz. Sin embargo, uno puede darse cuenta inmediatamente que la señal dentro de cada uno de los bloques no siempre corresponde a la salida de una subfuente estacionaria. Por otro lado, la codificación usando nuestro modelo de fuente compuesta y nuestro algoritmo de segmentación, es basada en modificar el modelo de fuente sólo cuando un cambio de salida de subfuente ha sido detectado ([39]). Este esquema es capaz de representar transición entre segmentos estacionarios adyacentes como una interpolación de las salidas de aquellos segmentos, produciendo una mejor aproximación no solamente durante intervalos estacionarios sino también en transición entre ellos.

Los pasos de codificación basada en segmentación se puede resumir en el siguiente:

Algoritmo de codificación:

- 1) Sean $X_n = (x_n, \dots, x_{n+N-1})$ una secuencia de señal de N muestras y X_j la secuencia de datos consistida en x_n y los j muestras siguientes de la señal.
- 2) Determine la longitud (número de muestras) del segmento estacionario usando el algoritmo de segmentación (es decir $N+k$).
- 3) Calcule los parámetros LPC y el valor de pitch para el segmento estacionario X_j usando un número variable de bloques de longitud fija (ellos pueden tener traslapes) a calcular correlaciones, promediando sus valores sobre el número de bloques y estime el mejor filtro AR con los coeficientes promedios de la correlación.
- 4) Reemplace n por $n+N+k+j-1$ y regrese a 1).

Para cada segmento estacionario, tienen que transmitir los siguientes datos: los p coeficientes de transmisión a_k , $k=1, \dots, p$, la ganancia $E\hat{z}$, el pitch P , el número de muestras del segmento estacionario y el tamaño de la transición posterior del segmento. El indicador del pitch está incluido en P .

Cuando $P=0$, significa que el segmento es de sonido agudo.

En el receptor, la señal se reproduce de acuerdo con la información recibida. Los segmentos estacionarios se sintetiza de la siguiente forma:

$$x(n) = - \sum_{k=1}^P a_k x(n-k) + E \tilde{z}(n)$$

donde a_k y $E \tilde{z}$ son los coeficientes de predicción y la ganancia del filtro AR del segmento, $v u(n)$ es la señal de entrada la cual puede ser un tren de pulsos unitarios de periodo P o ruido blanco de ganancia unitaria.

Las transiciones se generan de forma similar, la diferencia es que el filtro AR es variante con el tiempo, obtenido por algún tipo de interpolación de los parámetros de los segmentos en que comprende la transición. La interpolación que usamos es la interpolación lineal de los primeros $p+1$ coeficientes de correlación correspondientes a los segmentos inmediatamente anterior y posterior. El valor de pitch se interpola también linealmente, excepto cuando el valor interpolado está fuera del rango permisible se considera como sonido agudo.

El resultado de esta técnica se ilustra en la figura VII.1, donde la señal original se encuentra en la figura VII.1a mientras la señal mostrada en la figura VII.1b es la sintética. Uno se puede ver que subjetivamente existe una correspondencia razonablemente buena entre las dos señales. Pruebas de oído confirma esta conclusión y muestra que el resultado es de calidad no más baja que un sistema LPC de tasa fija. Sin embargo, la ganancia de usar esta técnica es significativa: los bits requeridos de este enfoque es aproximadamente 1/3 en promedio de los que se necesita un LPC estandar con una misma calidad subjetiva.

En el experimento, la señal de voz fue muestreada a una frecuencia de 6.4 kHz, el tamaño de los bloques fue de 128 muestras (20 mseg), el orden del filtro fue 10. El programa y las subrutinas asociadas que realizan la síntesis se encuentran en el apéndice 2,

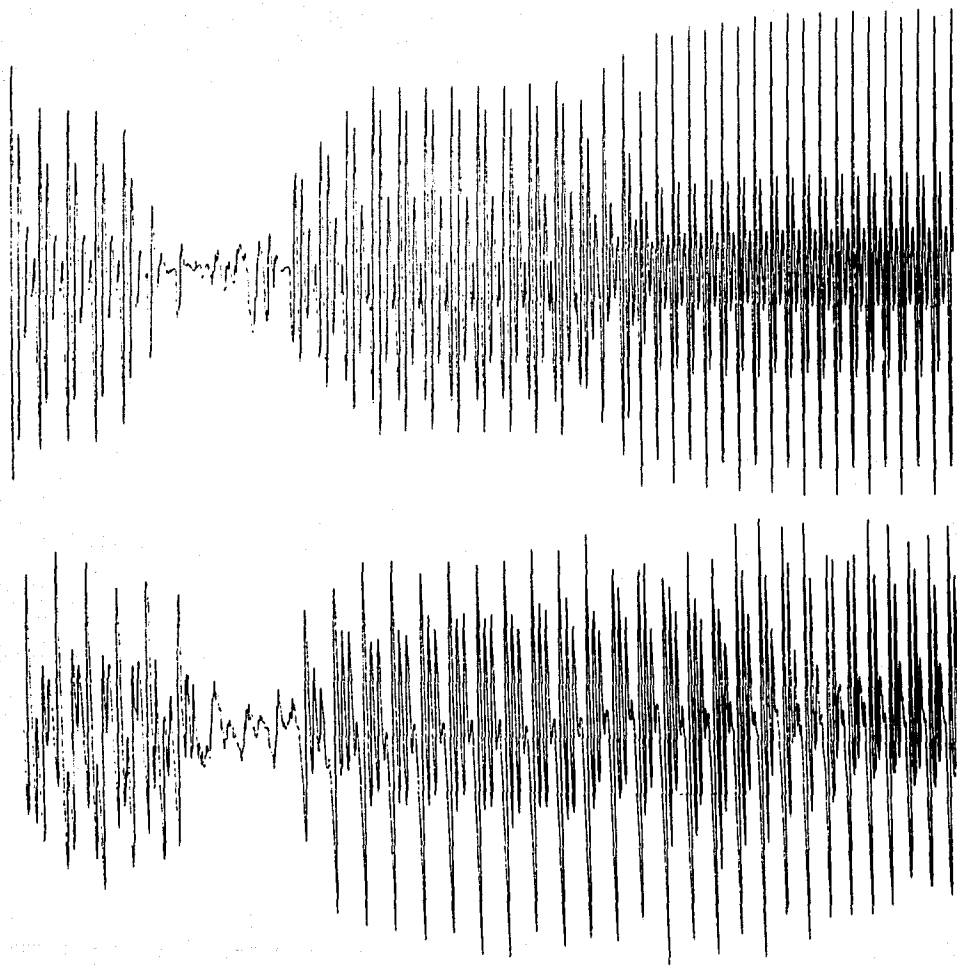


Figura VII.1 La señal sintonizada y la señal original

mientras la parte de codificación está incluida en el de la segmentación.

VII-2. Reconocimiento de Voz Basado en Segmentación

En esta sección, describimos un esquema de reconocimiento de patrones para la señal de voz basado en la identificación de las subfuentes. El sistema contempla las técnicas como codificación por predicción lineal (LPC), cuantización vectorial ([40]) y segmentación usando la medida de distorsión espectral Itakura-Saito.

El sistema consiste en dos partes fundamentales: la construcción de una base de datos (archivos) de los patrones y la identificación de los segmentos de datos por el criterio de vecino más cercano con respecto a los patrones preestablecidos.

La secuencia de entrenamiento es cuidadosamente seleccionada para que esta contenga todos los sonidos que se poseen las palabras de algún diccionario específico. Esta secuencia de entrenamiento se segmenta con el algoritmo presentado en el capítulo V. Cada segmento tanto estacionario como transitorio, se divide en bloques (translapados en general), los cuales proporcionan los parámetros indispensables de identificación de los mismos como los coeficientes de autocorrelación, de predicción o de reflexión, y la ganancia de LPC, pitch, etc. El número de bloques en cada segmento depende de la longitud del mismo. Además, los parámetros del filtro promedio de cada segmento son disponibles también.

A cada segmento de señal se le asigna un símbolo que es representativo del sonido que lleva. La asignación de símbolo es realizada actualmente por el medio del oído humano. Una vez que los segmentos estén etiquetados, los clasifican de acuerdo con los símbolos asignados. Dentro de cada grupo identificado por un mismo símbolo, se forman unos subgrupos con los bloques que tengan distorsiones pequeñas entre sí. Finalmente a cada uno de los subgrupos se calcula el centroide o el filtro promedio, el último va a ser un código del codificador ([40]).

De esta manera, se obtiene el codificador como una lista de códigos (codebook). Los códigos están formados por el símbolo

asignado y los parámetros de identificación. Generalmente a un símbolo le corresponden varios códigos, los cuales forman un cuantizador vectorial para este símbolo. Mientras que la lista de códigos sea más grande, el reconocimiento tendrá un mayor porcentaje de aciertos, claro que el precio que paga es el tiempo de procesamiento y la memoria que ocupan los códigos.

A continuación, presentamos la parte complementaria del sistema de reconocimiento : identificación de los segmentos.

Supongamos que tenemos una señal de voz que deseamos reconocer. Obviamente esta señal debe ser proveniente del mismo diccionario que generó los códigos. A esta secuencia de voz le aplicamos un procedimiento similar que usó en la primera parte. Se segmenta la señal, se dividen los segmentos tanto estacionarios como transitorios en bloques y se estiman los coeficientes de autocorrelación de cada uno de los bloques.

La idea de identificación de los segmentos es buscar en la lista de códigos un símbolo que mejor representa cada segmento en el sentido de vecino más cercano. Sea $S(j)$ el j -ésimo segmento de la secuencia de datos, el cual está formado por los bloques $s(j,k)$, $k=1, \dots, K(j)$; sea $C(n)$ el n -ésimo cuantizador vectorial en la lista de símbolos, el cual está formada por los códigos $c(n,m)$, $m=1, \dots, M(n)$. Entonces, el cuantizador $C(n_c)$, que se va a representar al segmento $S(j)$ de acuerdo con la regla de vecino más cercano, se determina por:

$$D(S(j), C(n_c)) = \min_n \frac{1}{K(j)} \sum_{k=1}^{K(j)} \min_m d(s(j,k), c(n,m))$$

La medida de distorsión usada es la de Itakura-Saito. Realmente estamos utilizando una doble minimización : calcular la distorsión de cada bloque del segmento a todos los códigos que pertenece a un mismo cuantizador, seleccionar aquel que da el menor valor de distorsión como la distorsión entre dicho bloque y el cuantizador; la distorsión entre el segmento y un cuantizador se define como el valor promedio de las distorsiones entre sus bloques y el cuantizador,

entonces el segmento se identifica por el símbolo del cuantizador que proporciona la menor distorsión.

Este esquema también es realizado en el Laboratorio CAD de la DEFFI, pero el principal contribuyente de este labor es el Ing. Carlos Rivera R. Lo presentamos en este trabajo para enseñar una aplicación más del procedimiento de la segmentación para el procesamiento digital de señal de voz.

CAPITULO VIII

CONCLUSIONES Y SUGERENCIAS

Se ha presentado un algoritmo para segmentar una señal discreta cuasi-estacionaria, y se muestran los resultados que se obtienen al segmentar señales de voz.

Para éstas, la segmentación puede ser usada para ayudar a resolver otros problemas entre los que se encuentran la compresión, el reconocimiento, el análisis y la síntesis de estas señales. Aunque el algoritmo sólo ha sido empleado para segmentar este tipo de señales, creemos que puede ser usado para segmentar otras señales cuasi-estacionarias, siempre y cuando sea la estadística de segundo orden el criterio más importante para la detección de los cambios.

El algoritmo está basado en el modelado de la señal como una fuente compuesta y en una medida de distorsión empleada ampliamente en el área de procesamiento digital de señales.

La complejidad del algoritmo es comparable con la de otros existentes en la literatura, con la ventaja de que tiene una interpretación interesante, lo cual puede convertirlo en un procedimiento importante para llevar a cabo la segmentación de señales.

La complejidad del algoritmo puede ser reducida considerablemente si se emplean técnicas recursivas para calcular los parámetros que intervienen en la medida de distorsión, además de que para algunos de ellos, en particular el valor de $E\hat{\xi}(n)$ (que es el que más tiempo consume) hemos observado que no es necesario calcularlo para cada valor de n , sino tan sólo se requiere actualizarlo de vez en cuando durante el proceso. Sin embargo esto puede ser función del tipo de señal que se esté analizando.

Resulta ser bastante prometedora esta técnica y los experimentos realizados hasta ahora indican que puede ser una magnífica herramienta para procesar señales de voz en tiempo real.

La aplicabilidad del algoritmo de segmentación es ilustrada por

dos sistemas de procesamiento de voz. El primero se trata del sistema de codificación LPC de tasa de transmisión variable. El sistema es basado en modificar el modelo de fuente compuesta sólo cuando un cambio de salida de subfuente ha sido detectado. Este esquema es capaz de representar transición entre segmentos estacionarios adyacentes como una interpolación de las salidas de aquellos segmentos, produciendo una mejor aproximación no solamente durante intervalos estacionarios sino también en la transición entre ellos. Esta técnica ofrece una ganancia significativa: los bits requeridos son aproximadamente una tercera parte en promedio de los que se necesita un LPC estandar con la misma calidad subjetiva.

El segundo sistema, en su modo de operación, básicamente segmenta la señal de voz y selecciona para cada uno de los segmentos el cuantizador que mejor produce (según el criterio de Itakura-Saito) al segmento de entrada, los resultados que se obtienen son correctos en un 90% en las vocales que aparecen en el texto y tiene un 70% de aciertos en las consonantes. Es importante notar que no se utilizó ninguna gramática que pudiera ayudar a la corrección de errores.

APENDICE I
PROGRAMAS DE SEGMENTACION

PROGRAMA DE SEGMENTACION CON DISTORSION DE ITAKURA-SAITO
POR
ZHENYU WU
Y
JIANYU YOU

DESCRIPCION GENERAL :

EL PROGRAMA TOMA UN ARCHIVO DE DATOS PROVENIENTES DE UNA SEÑAL DE VOZ COMO ENTRADA , DETECTA LOS CAMBIOS ESTADISTICOS DE LA SEÑAL , IDENTIFICA LOS INTERVALOS ESTACIONARIOS Y TRANSITORIOS, Y CARACTERIZA LOS SEGMENTOS ESTACIONARIOS POR UN FILTRO AR.

ENTRADAS :

ARCHIVO DE LAS MUESTRAS DE SEÑAL DE VOZ, EL ACCESO AL ARCHIVO ES DIRECTO, Y CADA REGISTRO CONTIENE 128 MUESTRAS COMO NUMEROS REALES NO-FORMATEADOS.

SALIDAS :

ARCHIVO DE LOS COEFICIENTES DE PREDICCION A'S;
ARCHIVO DE LOS COEFICIENTES DE CORRELACION R'S
ARCHIVO DE LAS MARCAS DE SEGMENTACION.
LOS ARCHIVOS SON NO-FORMATEADOS, DE ACCESO SECUENCIAL.

DIMENSION A(11),X1(266),X(128),D(128),CRRLAT(11),E(128)
DIMENSION COEF(11),CORRAS(11),CRMEAN(11),D1(128),D2(128)
DATA M/128/
DATA M10/10/
TYPE 1000
1000 FORMAT(10X,'INICIO DE LA EJECUCION',/)
TYPE 1001
1001 FORMAT(10X,'EL NOMBRE DEL ARCHIVO DE ENTRADA',/)
CALL ASSIGN(1,'',-1,'OLD')
TYPE 1105
1105 FORMAT(10X,'EL BLOQUE INICIAL Y FINAL')
ACCEPT 1106,ICTINI,ICTFIN
1106 FORMAT(2I5)
DEFINE FILE 1(ICTFIN,256,U,ICT)
TYPE 1003

```

1003  FORMAT(10X,'EL NOMBRE DE ARCHIVO DE AS',Z)
      CALL ASSIGN(3,'',-1,'NEW')
      TYPE 2003
2003  FORMAT(10X,'EL NOMBRE DE ARCHIVO DE RL',Z)
      CALL ASSIGN(4,'',-1,'NEW')
      TYPE 1111
1111  FORMAT(10X,'EL NOMBRE DEL ARCHIVO DE MARCAS',Z)
      CALL ASSIGN(8,'',-1,'NEW')
      TYPE 1101
1101  FORMAT(5X,'EL RANGO DE PREFERENCIA Y NIVEL',Z,Z,15X,
*      '( LOS VALORES POR DEFAULT : 0.00045,0.35,0.05 )')
      ACCEPT 1102,CR1,CR2,CR4
1102  FORMAT(3F10.7)
      IF(CR1.EQ.0.0)CR1=0.00045
      IF(CR2.EQ.0.0)CR2=0.35
      IF(CR4.EQ.0.0)CR4=0.05
      TYPE 1103
1103  FORMAT(5X,'EL PERIODO DE ACTUALIZACION',Z,Z,20X,
*      '( EL VALOR POR DEFAULT : T = 10 ; )')
      ACCEPT 1104,KPR
1104  FORMAT(I4)
      IF(KPR.EQ.0)KPR=10
      TYPE 1107
1107  FORMAT(5X,'LA LONGITUD DE VENTANA',Z,Z,20X,
*      '( EL VALOR POR DEFAULT : L = 20 ; )')
      ACCEPT 1104,LOGVTN
      IF(LOGVTN.EQ.0)LOGVTN=20
      TYPE 1108
1108  FORMAT(5X,'EL NUMERO MAXIMO DE ITERACION',Z,Z,20X,
*      '( EL VALOR POR DEFAULT : NI = 4 ; )')
      ACCEPT 1109,NUMITR
1109  FORMAT(I3)
      M1=M10+1
      M2=M+M10
      MM=M*2
      MM1=M+M1
      MM2=MM+M10
      M3=3*M
      ICT=ICTINI
      INICIO=1
      CR5=0.6
      CR6=0.75*CR2
      READ(1'ICT')(X1(K),K=M1,M2)
      DO 5 I=1,M10
      IMM=12+I-1
      X1(I)=X1(IMM)
      READ(1'ICT')(X1(K),K=MM1,MM2)
      ICTEMP=ICT
      DO 10 I=1,M
      IM10=I+M10

```

```

D(I)=0.0
10 X(I)=X1(IM10)
WRITE(8)ICTINI,1
IBLQ1=ICTINI
MST1=1
ICT2IN=ICT2
MARCAF=0
ITERAC=0
60 CONTINUE
MARCAM=0
IF(MARCAF,EQ,1)GO TO 62
CALL DBTAS(X,CRRLAT,A,SIGMA,EPROCS)
SIGMA2=SIGMA
GO TO 63
62 CALL DBTAS(X,CRRLAT,COEF,SIGMA1,EPROCS)
CALL CRMED(IBM1,MST1,IBLQ2,MST2,CRMEAN,ICTFIN)
CALL ALGLEV(CRMEAN,A,SIGMA)
SIGMA2=SIGMA
ICT=ICTEMP
SIGMA=SIGMA1
63 NI=0
CALL DBTRAS(A,CORRAS)
CALL DBTERE(CORRAS,CRRLAT,ET)
ET=ET/SIGMA-1.0
IF(ET,LT,CR6)GO TO 64
NUM1=(IBLQ2-IBLQ1)*M+MST2-MST1
IBLQ2=IBLQ2-1
IF(NUM1,LT,M)IBLQ2=IBLQ1+1
IF(NUM1,LT,M)MST2=MST1
GO TO 60
66 MINIC1=M+INICIO-1
CRIPIC=0.025
IF((SIGMA*128.0),LT,1.5)CRIPIC=0.25
IPITCH=0
NP=0
NCONTP=25
DO 15 I=INICIO,MINIC1
II=I
IF(I,GT,M)II=I-M
E(II)=0.0
DO 20 J=1,11
IJ1=I-J+1+M10
20 E(II)=E(II)+A(J)*X1(IJ1)
E(II)=E(II)**2/SIGMA/128.
IF((E(II),GT,CRIPIC),AND,(NCONTP,GT,24))GO TO 16
NCONTP=NCONTP+1
GO TO 15
16 NP=NP+1
IF(NP,EQ,2)IPITCH=NCONTP
IF((NP,GT,2),AND,(ABS(NCONTP-IPITCH),LT,10))GO TO 17

```



```

IF(NCONTP.GT.IPITCH)NCONTP=IPITCH+9
IF(NCONTP.LT.IPITCH)NCONTN=IPITCH-9
17 IF(NP.GT.2)IFPITCH=(IPITCH*(NP-2)+NCONTP)/(NP-1)
NCONTP=1
15 CONTINUE
INIC1=INICIO-1
DO 22 I=INICIO,M
D(I)=ET
D1(I)=ET
D2(I)=ABS(D1(I))
22 CONTINUE
DO 25 I=1,INIC1
D1(I)=ET
D2(I)=ABS(D1(I))
25 D(I)=ET
28 DO 30 I=INICIO,M
KI=KI+1
IM10=I+M10
IM2=I+M2
DO 310 J=1,11
IMJ1=IM10+J-1
IMJ2=IM2-J+1
CORREC=X1(IM2)*X1(IMJ2)-X1(IM10)*X1(IMJ1)
CRRLAT(J)=CRRLAT(J)+CORREC/128.
310 CONTINUE
IF(KI.LT.KPR)GO TO 31
CALL ALGLEV(CRRLAT,COEF,SIGMA)
KI=0
31 E2=0.0
DO 35 J=1,11
IJ1=I-J+1+M2
35 E2=E2+A(J)*X1(IJ1)
E2=E2**2/SIGMA/128.0
I11=I-1
IF(I11.LT.1)I11=M+I11
D(I)=D(I11)+E2-E(I)
E(I)=E2
IF((E(I).GT.CRIPIC).AND.(NCONTP.GT.24))GO TO 36
NCONTP=NCONTP+1
GO TO 39
36 NP=NP+1
IF(NP.EQ.2)IPITCH=NCONTP
IF((NP.GT.2).AND.(ABS(NCONTP-IPITCH).LT.10))GO TO 47
IF(NCONTP.GT.IPITCH)NCONTP=IPITCH+9
IF(NCONTP.LT.IPITCH)NCONTP=IPITCH-9
47 IF(NP.GT.2)IFPITCH=(IPITCH*(NP-2)+NCONTP)/(NP-1)
NCONTP=1
39 IPOS=:
CALL DETCAM(D,D1,IPOS,LOGVTN,EPR,EDIF)
D2(I)=ABS(D1(I))

```

```

ABSD=ABS(EPR)
CR3=CR1*(ABSD+1.0)**3
IF((EDIF.LT.0.001).AND.(ABSD.LT.0.15))GO TO 37
IF(MARCAM.EQ.1)GO TO 38
MARCAM=1
ICTCAM=ICT-1
MSTCAM=I
GO TO 38
37 IF(ABSD.LE.CR2)MARCAM=0
38 VDIFF=0.0
IF(ABSD.GT.CR5)VDIFF=EDIF
IF(((EDIF.LT.CR3).OR.(VDIFF.GT.CR4)).AND.(ABSD.GT.CR2))GO TO 45
30 CONTINUE
DO 40 K=1,M2
KM=K+M
40 X1(K)=X1(KM)
IF(ICT.GT.ICTFIN)GO TO 98
READ(1,ICT)(X1(K),K=MM1,MM2)
ICTEMP=ICT
INICIC=1
GO TO 28
45 INICIO=I
IBLOCK=ICT-1
IF(MARCAI.EQ.0)GO TO 44
NUMSTS=(IBLQ21-IBLOCK)*128+MST21-INICIO
IF(NUMSTS.LT.0)NUMSTS=-1*NUMSTS
IF((NUMSTS.LT.5).OR.(ITERAC.GE.NUMITR))GO TO 43
44 IF(NUMITR.EQ.0)GO TO 43
IBLQ21=IBLOCK
MST21=INICIO
IBLQ2=IBLOCK
MST2=INICIO
MARCAI=1
IF(MARCAM.EQ.1)IBLQ2=ICTCAM
IF(MARCAM.EQ.1)MST2=MSTCAM
ITERAC=ITERAC+1
ICT=IBLQ1
INICIO=MST1
READ(1,ICT)(X1(KKK),KKK=M1,M2)
READ(1,ICT)(X1(KKK),KKK=MM1,MM2)
ICTEMP=ICT
DO 49 KKK=1,M
KKKINI=KKK+INICIO-1+M10
49 X(KKK)=X1(KKKINI)
GO TO 60
43 MARCAI=0
ITERAC=0
IF(ICT.GE.ICTFIN)ICTCAM=ICTFIN
IF(ICT.GE.ICTFIN)MSTCAM=M
CALL CRRMED(IBLQ1,MST1,ICTCAM,MSTCAM,CRMEAN,ICTFIN)

```

```

CALL ALGLEV(CRMEAN,A,SIGMAN)
IBLQ1=IBLOCK
MST1=INICIO
IP=IPITCH
IF(IP.GE.80)IP=0
IF(ICT.LT.ICTFIN)GO TO 53
WRITE(4)(CRMEAN(K),K=1,11),0.0,SIGMAN,ICTFIN,M,ICTFIN,M,IP,1
WRITE(3)(A(K),K=1,11),0.0,SIGMAN,ICTFIN,M,ICTFIN,M,IP,1
WRITE(8)ICTFIN,M
GO TO 99
53 WRITE(4)(CRMEAN(K),K=1,11),0.0,SIGMAN,
* ICTCAM,MSTCAM,IBLQ1,MST1,IP,1
WRITE(3)(A(K),K=1,11),0.0,SIGMAN,ICTCAM,MSTCAM,IBLQ1,MST1,IP,1
WRITE(8)IBLQ1,MST1
DO 50 K=1,M2
KINIC=K+M
50 X1(K)=X1(KINIC)
IF(ICT.GT.ICTFIN)GO TO 98
READ(1'ICT')(X1(K),K=MM1,MM2)
DO 55 K=1,M
KINIC1=K+INICIO-1+M10
55 X(K)=X1(KINIC1)
GO TO 60
98 WRITE(8)ICTFIN,M
SIGMAN=SIGMA2
WRITE(4)(CRRLAT(K),K=1,11),0.0,SIGMAN,ICTFIN,M,ICTFIN,M,0,1
WRITE(3)(A(K),K=1,11),0.0,SIGMAN,ICTFIN,M,ICTFIN,M,0,1
99 CALL CLOSE(1)
CALL CLOSE(4)
CALL CLOSE(3)
CALL CLOSE(8)
CALL EXIT
END

```

```

C      (* *)
C      (* SUBROUTINA QUE CALCULA LAS CORRELACIONES. *)
C      (* *)
SUBROUTINE CRRLT(PROCS,CRRLAT,EPROCS)
DIMENSION CRRLAT(11),PROCS(128)
N=128
NN=11
DO 105 I=1,NN
CRRLAT(I)=0.0
IN=N-I+1
DO 103 J=1,IN
JI=J+I-1
CRRLAT(I)=CRRLAT(I)+PROCS(JI)*PROCS(J)
103 CONTINUE
CRRLAT(I)=CRRLAT(I)/128.
105 CONTINUE
EPROCS=CRRLAT(1)
RETURN
END

C      (* *)
C      (* SUBROUTINA QUE REALIZA EL ALGORITMO DE LEVINSON *)
C      (* *)
SUBROUTINE ALGLEV(CRRLAT,COEFA,SIGMA)
DIMENSION CRRLAT(11),COEFA(11),COEFB(11),VECTRK(10)
COEFA(1)=1.0
COEFB(1)=0.0
SIGMA=CRRLAT(1)
DO 705 I=1,10
PR1=0.0
DO 710 J=1,I
JI2=I+2-J
COEFB(JI2)=COEFA(J)
PR1=PR1+COEFA(J)*CRRLAT(JI2)
710 CONTINUE
VECTRK(I)=-1.0*PR1/SIGMA
SIGMA=SIGMA*(1.0-VECTRK(I)*VECTRK(I))
DO 715 J=2,I
715 COEFA(J)=COEFA(J)+VECTRK(I)*COEFB(J)
I1=I+1
COEFA(I1)=VECTRK(I)
705 CONTINUE
RETURN
END

C
SUBROUTINE OBTAS(DATOS,CRRLAT,COEF,P,EPROCS)
DIMENSION DATOS(128),COEF(11),CRRLAT(11),CRAS(11)
CALL CRRLT(DATOS,CRRLAT,EPROCS)
CALL ALGLEV(CRRLAT,COEF,P)
RETURN
END

```

```

C      (* *)
C      (* SUBROUTINA QUE CALCULA LAS CORRELACIONES PROMEDIAS. *)
C      (* *)
SUBROUTINE CRMEAN(BLQ1,MST1,BLQ2,MST2,CRMEAN,ICTFIN)
DIMENSION Y(128),Y1(256),CRMEAN(11),CRRLA1(11)
INTEGER BLQ1,BLQ2
NUMBLQ=BLQ2-BLQ1+1
NUMB1=NUMBLQ-1
NUMTLP=128-MST2+MST1
IF(NUMB1.GT.0)NUMTLP=NUMTLP/NUMB1
ICT=BLQ1
TYPE 401,BLQ1,MST1,BLQ2,MST2
401  FORMAT('X',MUESTRAS INICIAL Y FINAL :',2(X,'(',I3,',',I3,')')')
READ(1,ICT)(Y(I),I=1,128)
ICT=ICT+1
READ(1,ICT)(Y1(I),I=129,256)
ICT=ICT+1
MSTINI=MST1
DO 420 J=1,NUMBLQ
J1=J-1
DO 405 I=1,128
405  IMST1=I+MSTINI-1
Y(I)=Y1(IMST1)
CALL CRRL1(Y,CRRLA1,EP)
DO 410 I=1,11
410  IF(J.EQ.1)CRMEAN(I)=CRRLA1(I)
IF(J.GT.1)CRMEAN(I)=(CRMEAN(I)*J1+CRRLA1(I))/J
CONTINUE
MSTINI=MSTINI+NUMTLP+128
IF((MSTINI.LT.129).OR.(J.EQ.NUMBLQ))GO TO 420
MSTINI=MSTINI-128
DO 415 I=1,128
415  IM=I+128
Y1(I)=Y1(IM)
IF(ICT.GT.ICTFIN) GO TO 421
READ(1,ICT)(Y1(I),I=129,256)
ICT=ICT+1
420  CONTINUE
421  CONTINUE
RETURN
END
C      (* *)
C      (* SUBROUTINA QUE CALCULA LA DISTORSION PROMEDIA Y DESVIACION *)
C      (* *)
SUBROUTINE DETCAM(D,B1,IPOS,LOGVTN,EPR,EDTF)
DIMENSION D(128),B1(128)
IPOS2=IPOS-LOGVTN
IPOS1=IPOS-1
IF(IPOS1.LT.0)IPOS1=128
IF(IPOS2.LT.0)IPOS2=IPOS2+128

```

```

D1(IPOS)=D1(IPOS1)+(D(IPOS)-D(IPOS2))/LOGVTN
EPR=0.0
DO 21 J=1,5
IPOSJ=IPOS-J+1
IF(IPOSJ.LE.0)IPOSJ=IPOSJ+128
21 EPR=EPR+D1(IPOSJ)
EPR=EPR/5.0
EDIF=0.0
DO 22 J=1,5
IPOSJ=IPOS-J+1
IF(IPOSJ.LE.0)IPOSJ=IPOSJ+128
22 EDIF=EDIF+ABS(D1(IPOSJ)-EPR)
EDIF=EDIF/5.0
RETURN
END
C (* *)
C (* SUBROUTINA QUE CALCULA LA CORRELACION DE LOS A'S *)
C (* *)
SUBROUTINE DBTRAS(A,CORRAS)
DIMENSION A(11),CORRAS(11)
DO 11 I=1,11
CORRAS(I)=0.0
I11=11-I+1
DO 12 J=1,I11
JI=J+I-1
CORRAS(I)=CORRAS(I)+A(JI)*A(J)
12 CONTINUE
11 CONTINUE
RETURN
END
C (* *)
C (* SUBROUTINA QUE CALCULA LA DISTORSION ITAKURA-SAITO *)
C (* *)
SUBROUTINE DBTERR(CORRAS,CRRLAT,ERRT)
DIMENSION CORRAS(11),CRRLAT(11)
ERRT=CORRAS(1)*CRRLAT(1)
DO 13 I=2,11
ERRT=ERRT+CORRAS(I)*CRRLAT(I)*2
13 CONTINUE
RETURN
END

```

APENDICE II
PROGRAMAS DE SINTESIS DE
TASA VARIABLE

```

C...
C      PROGRAMA PRINCIPAL
C...
      TYPE 3
3      FORMAT(4X,'# DE BLOQUES DE 128')
      ACCEPT 4,INB
4      FORMAT(I5)
      ITR=128
      N=ITR
      ITB=ITR*2
      TYPE 66
66     FORMAT(2X,'EL NOMBRE DEL ARCHIVO DE SE&AL SINTETICA:',/)
      CALL ASSIGN(2,'',-1,'NEW')
      DEFINE FILE 2(INB,ITR,U,ICT2)
      TYPE 93
93     FORMAT(2X,'EL NOMBRE DEL ARCHIVO DE AS:',/)
      CALL ASSIGN(6,'',-1,'OLB')
      TYPE 96
96     FORMAT(10X,'QUIERES TRANSICION ? (0:SI; 1:NO.)',/)
      ACCEPT 95,ITRNS
95     FORMAT(I2)
      ICT2=1
      CALL DTSINT(INB,ICT2,ITRNS)
      CALL CLOSE(2)
      CALL CLOSE(6)
      CALL EXIT
      END
      SUBROUTINE DTSINT(IBLQFI,ICT2,ITRNS)
      DIMENSION COEF(60),DATGEN(188),A1(60)
      IA=0
      IB=0
      DO 501 ICOEF=1,60
      DATGEN(ICOEF)=0.0
      COEF(ICOEF)=0.0
301    CONTINUE
      IB1=1
      MST1=1
      READ(6)(COEF(I),I=1,12),SIGMA,IB2,MST2,IBB,MST,NP,NR
      NCONT=NP
      SIGMAM=SIGMA
      IF(NR.EQ.1)GO TO 514
      DO 513,I=2,NR
      I1=I*16-19
      I2=I1+15
      READ(6)(COEF(K),K=I1,I2)

```

```

513 CONTINUE
514 NUMSTS=(IR2-IR1)*128+MST2-MST1
    NUMINT=NUMSTS/4
    NINT3=NUMINT*3
    IF((IRB.EQ.IBLOF1).AND.(MST.EQ.128))GO TO 515
    READ(4)(A1(I),I=1,12),SIGMA2,IR3,MST3,IRB0,MST0,NP0,NR0
    DO 601 I=13,60
601  A1(I)=0.0
    GO TO 510
515  SIGMA2=SIGMA
    IR3=IRB
    MST3=MST
518  IF(ITRNS.EQ.1)IR2=IRB
    IF(ITRNS.EQ.1)MST2=MST
    * IF(ITRNS.EQ.1)CALL FUENTE(IG,MST1,
    IR2,MST2,SIGMA,NP,COEF,DATGEN,ICT2,IA,IR,NCONT)
    NUMT0=(((IR3-IR1)*128+MST3-MST1)*3*((IRB-IR2)*128+MST-MST2))/4
    P0=SQRT(SIGMA)
    FMO=P0+(SQRT(SIGMA2)-P0)*NUMINT/NUMT0
    IDELAY=NP/8
    INICIO=MST1
    MSTFIN=128
    MSTCT1=0
    MSTCT2=NUMINT
    PSQRTM=SQRT(SIGMA)
517  CONTINUE
    IF(ICT2.GE.IR2)MSTFIN=MST2
    DO 537 ICOEF=INICIO,MSTFIN
    MSTCT1=MSTCT1+1
    MSTCT2=MSTCT2-1
    PSQRT=P0
    IF(ITRNS.EQ.1)GO TO 700
    IF(MSTCT1.LE.NUMINT)PSQRT=(P0*MSTCT1+PSQRTM*MSTCT2)/NUMINT
    IF(MSTCT1.GT.NINT3)PSQRT=P0+(FMO-P0)*(MSTCT1-NINT3)/NUMINT
700  IF(P0.LT.0.005)PSQRT=PSQRT/10.0
    FN=PSQRT*5.0
    ICOEFM=ICOEF+60
    R1=RAN(IA,IB)
    R2=RAN(IA,IB)
    RR=SQRT(-2.0*ALOG(R1)/ALOG(2.718282))*SIN(6.283185*R2)
    DATGEN(ICOEFM)=RR*PSQRT
    IF(NP.EQ.0)GO TO 535
    DATGEN(ICOEFM)=0.0
    IF(NCONT.EQ.IDELAY*3)DATGEN(ICOEFM)=0.35*FN
    IF(NCONT.EQ.IDELAY*2)DATGEN(ICOEFM)=-0.45*FN
    IF(NCONT.EQ.IDELAY)DATGEN(ICOEFM)=FN*0.82
    IF(NCONT.EQ.NP)DATGEN(ICOEFM)=-FN
    IF(NCONT.GE.NP)NCONT=NCONT-NP
    NCONT=NCONT+1
535  DO 536 ITERA=2,60

```



```

ICOEFR=ICOEFM-TTERA+1
DATGEN(ICOEFM)=DATGEN(ICOEFM)-DATGEN(ICOEFR)*COEF(ITERA)
536 CONTINUE
537 CONTINUE
IF(MSTFIN.NE.128)GO TO 524
TYPE 555,ICT2,IB2
555 FORMAT(3X,'EL ICT2 EN RCSYAG : ',I3,3X,'IB2 = ',I3)
WRITE(2,ICT2)(DATGEN(JK),JK=51,188)
IF(ICT2.GT.182)GO TO 524
DO 523 I=1,60
I128=I+128
523 DATGEN(I)=DATGEN(I128)
INICIO=1
GO TO 517
TYPE 525,SIGMA
524 FORMAT(2X,'LA POTENCIA DE ERROR ES : ',F7.5,/)
IF((IBB.EQ.IBLOFI).AND.(MST.EQ.128))GO TO 526
IB1=IBB
MST1=MST+1
IF(MST1.GT.128)IB1=IB1+1
IF(MST1.GT.128)MST1=MST1-128
SIGMAM=FM0**2
IF(ITRNS.EQ.1)SIGMAM=SIGMA2
CALL INTFAS(COEF,SIGMA,IB2,MST2,IBB,MST,NF,NR,ICT2,
* DATGEN,IA,IB,NCONT,SIGMAM,A1,SIGMA2,IB3,MST3,IBB0,MST0,NP0,NR0)
GO TO 514
526 RETURN
END
SUBROUTINE INTFAS(A1,F1,IB1,M1,IB2,M2,NF1,NR1,ICT2,DATGEN,IA,IB,
* NCONT,PM,A2,P2,IB33,MST33,IBB,MST,NP2,NR2)
DIMENSION A1(60),A2(60),AT(60),DATGEN(188)
DIMENSION RX1(60),RX2(60),RXT(60),RC1(60),RC2(60),RCT(60)
IF(NR2.EQ.1)GO TO 414
DO 413,I=2,NR2
I1=I*16-19
I2=I1+15
413 READ(6)(A2(K),K=I1,I2)
CONTINUE
414 PSQRT1=SQRT(PM)
NUMSTS=(IB2-IB1)*128+M2-M1
IF(NUMSTS.EQ.0)GO TO 424
NUMTOT=((IB33-IB2)*128+MST33-M2)/4+NUMSTS
PSQRT2=PSQRT1+(SQRT(P2)-PSQRT1)*NUMSTS/NUMTOT
PM=PSQRT2**2
TYPE 444,NUMSTS,IB1,M1,IB2,M2
444 FORMAT(10X,I4,' MUESTRAS PARA INTERPOLAR :',2(2X,'(',2I4,')'))
M60=60
CALL STEPDI(A1,RC1,M60)
CALL STEPDI(A2,RC2,M60)
CALL DRC2AU(PSQRT1,RC1,M60,RX1)

```

```

CALL DRC2AU(PSQRT2,RC2,M60,RX2)
DO 415 I=1,M60
415 AT(I)=A1(I)
CALL STEPUP(AT,RC1,M60)
TYPE 999,(AT(KK),KK=1,60)
999 FORMAT(10(2X,F10.6))
MSTCT1=NUMSTS
MSTCT2=0
INICIO=M1
MSTFIN=128
417 CONTINUE
IF(ICT2.GE.IB2)MSTFIN=M2
DO 437 ICDEF=INICIO,MSTFIN
ICDEFM=ICDEF+60
R1=RAN(IA,IB)
R2=RAN(IA,IB)
RR=SQRT(-2.0*ALOG(R1)/ALOG(2.718282))*SIN(6.283185*R2)
NF=(NF1*MSTCT1+NF2*MSTCT2)/NUMSTS
NF12=NF1*NF2
IF((NF12.EQ.0).AND.(P1.GE.P2))NF=NF1
IF((NF12.EQ.0).AND.(P1.LT.P2))NF=NF2
IDELAY=NF/8
IF(PSQRTP**2.LT.0.005)PSQRTP=PSQRTP/10.0
PN=PSQRTP*5.0
RXT(K)=(RX1(K)*MSTCT1+RX2(K)*MSTCT2)/NUMSTS
CALL ALGLEV(RXT,AT,SIGMAT)
PSQRTP=SQRT(SIGMAT)
MSTCT1=MSTCT1-1
MSTCT2=MSTCT2+1
DATGEN(ICDEFM)=RR*PSQRTP
IF(NF.EQ.0)GO TO 435
DATGEN(ICDEFM)=0.0
IF(NCONT.EQ.IDELAY*3)DATGEN(ICDEFM)=0.35*PN
IF(NCONT.EQ.IDELAY*2)DATGEN(ICDEFM)=-0.45*PN
IF(NCONT.EQ.IDELAY)DATGEN(ICDEFM)=PN*0.82
IF(NCONT.EQ.NF)DATGEN(ICDEFM)=-PN
IF(NCONT.GE.NF)NCONT=NCONT-NF
NCONT=NCONT+1
435 DO 436 ITERA=2,60
ICOEFR=ICDEFM-ITERA+1
DATGEN(ICDEFM)=DATGEN(ICDEFM)-DATGEN(ICOEFR)*AT(ITERA)
436 CONTINUE
437 CONTINUE
IF(MSTFIN.NE.128)GO TO 424
TYPE 455,ICT2
455 FORMAT(3X,'EL ICT2 EN INTPAS : ',I3)
WRITE(2,ICT2)(DATGEN(JK),JK=61,188)
IF(ICT2.GT.IB2)GO TO 424
DO 423 I=1,60
I128=I+128

```

423 DATGEN(I)=DATGEN(I128)

INICIO=1
GO TO 417

424 IB1=IB33
IB2=IBB
M1=MST33

M2=MST

P1=P2

NP1=NP2

NR1=NR2

DO 427 K=1,60

427 A1(K)=A2(K)

RETURN

END

SUBROUTINE DRC2AU(RMSS,RC,N,R)

C..
C.. CONVERT GAIN AND REFLECTION COEFFICIENTS TO AUTOCORRELATIONS

C..
C.. INPUTS:

C.. RMSS - SQRT OF LPC GAIN

C.. RC - REFLECTION COEFFICIENT ARRAY

C.. N - NUMBRER OF REFLECTION COEFFICIENTS

C..
C.. OUTPUTS:

C.. R - ARRAY OF N+1 AUTOCORRELATIONS

C..

DIMENSION RC(1),R(1),A(21),A1(21)

CALL DRC2E(RC,N,RMSS,E)

R(1)=E

DO 50 M=1,N

R(M+1)=-E*RC(M)

E=E*(1-RC(M)**2)

A1(M)=-RC(M)

IF(M.EQ.1) GO TO 20

DO 10 J=1,M-1

R(M+1)=R(M+1)+A(J)*R(M-J+1)

10 A1(J)=A1(J)+RC(M)*A(M-J)

20 DO 40 J=1,M

40 A(J)=A1(J)

50 CONTINUE

RETURN

END

SUBROUTINE DRC2E(RC,M,RMSS,E)

DIMENSION RC(60)

CF=1.0

IF(M.EQ.0)GO TO 20

DO 10 I=1,M

10 CF=CF*(1.0-RC(I)**2)

20 E=RMSS**2/CF
RETURN
END

END OF PROGRAM

REFERENCIAS

1. A. Papoulis
Probability, Random Variables and Stochastics Processes
McGraw-Hill, 1965.
2. E. Parzen
Stochastics Processes
Holden-Day, 1962
3. E. Cinlar
Introduction to Stochastic Processes
Prentice-Hall, 1975
4. A. M. Yaglom
Introduction to the Theory of Stationary Random Functions
Prentice-Hall, 1962
5. G. R. Cooper & C. D. McGillem
Probabilistic Method of Signal and System Analysis
Holt Rinehart Winston, 1971
6. J. Lamperti
Stochastic Processes
Springer-Verlag, 1977
7. Van Trees & L. Harry
Detection, Estimation and Modulation Theory
Wiley, 1968
8. A. Papoulis
Signal Analysis
McGraw-Hill, 1977

9. J. P. Burg
 Maximun Entropy Spectral Analysis
 Ph. D. Dissertation, Dep. Geophysics
 Stanford University, CA, May, 1975
10. A. Papoulis
 Maximun Entropy and Spectral Estimation : A Review
 IEEE Trans. Acoustics, Speech and Signal Processing
 Vol. ASSP-29, No.6, December 1981
11. T. Berger
 Rate Distortion Theory
 Prentice Hall, Englewood Cliffs, N. J. 1971
12. A. Buzo, F. Kuhlmann, Zhenyu Wu & Jianyu You
 Segmentación Estadística de Series de Tiempo
 in Proc. IX Congreso de la Academia Nacional de Ingeniería,
 León, México, septiembre 1983
13. J. Makhoul
 Linear Prediction: A Tutorial Review
 Proc. IEEE, VOL.63, No.4, April 1975
14. J. D. Markel & A. H. Gray
 Linear Prediction of Speech
 New York : Springer-Verlag, 1976
15. G. M. Jenkins & D. G. Watts
 Spectral Analysis and its Applications
 Hilden-Day, 1968
16. L. H. Koopmans
 The Spectral Analysis of Time Series
 Academic Press, 1974

17. S. M. Kay & S. L. Marple
Spectrum Analysis - A Modern Perspective
Proc. IEEE, Vol.69, No.11, November 1981
18. J.W. Cooley and J.W. Tukey
An Algorithm for the Machine Computation of Complex Fourier Series
Math. Comp.,19, Apr.,1965
19. A. V. Oppenheim & R. W. Schaffer
Digital Signal Processing
Prentice-Hall, 1975
20. L. R. Rabiner & B. Gold
Theory and Applications of Digital Signal Processing
Prentice-Hall, 1975
21. F. J. Harris
On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform
Proc. IEEE, Vol.66, No.1, January 1978
22. J. S. Bendat & A. G. Piersol
Random Data: Analysis and Measurement Procedures
Wiley-Interscience, 1971
23. A. VonDenBus
Alternative Interpretation of Maximum Entropy Spectral Analysis
IEEE Trans. Inform. Theory, Vol.IT-17, July 1971
24. F. Itakura & S. Saito
Analysis Synthesis Telephony Based on the Maximum

Likelihood Method

Proc. 6th. Int. Congr. Acoust. Tokyo Japan 1968

25. F. Itakura & S. Saito
A Statistical Method for Estimation of Speech Spectral
Density and Formant Frequencies
Electro. Commun. Jan. Vol. 53-A, 1970
26. R. M. Gray, A. Buzo, A. H. Gray & Y. Matsuyama
Distortion Measures for Speech Processing
IEEE Trans. Acoustics, Speech and Signal Processing
Vol. ASSP-28, No 4, August 1980
27. N. Levinson
The Wiener (Root Mean Square) Error Criterion in
Filter Design and Prediction
J. Math. Phys. Vol. 25, 1947
28. J. Durbin
The Fitting of Time Series Models
Rev. Inst. Int. de Stat. Vol. 28, 1960
29. R. B. Ash
Real Analysis and Probability
New York : Academic 1972
30. R. M. Gray, A. H. Gray, G. Rebolledo & J. E. Shore
Rate-Distortion Speech Coding with a Minimum Discrimination
Information distortion measure
IEEE Trans. Inform. Theory, Vol. IT-27, No. 6, November 1981
31. S. Kullback
Information Theory and Statistics
New York : Dover, 1968

32. M. Finskev
Information and Information Stability of Random
Variables and Processes
San Francisco, CA : Holden-Day, 1964
33. I. B. MacNeil
Properties of Sequences of Partial Sum of Polynomial
Regression Residuals with Application to Test for Change of
Regression at Unknown Times
Ann. stat. Vol. 6, pp. 422-433, 1978
34. J. Seyer & A. C. Sanderson
Detecting Changes in Time-Series
IEEE Trans. Inform. Theory Vol. IT-26, No.2, March 1980
35. M. Basseville & A. Benveniste
Sequential Detection of Abrupt Changes in Spectral
Characteristics of Digital Signals
IEEE Trans. Inform. Theory, Vol. IT-29, No.5, September 1983
36. R. E. Crochiere & J. L. Flanagan
Current Perspectives in Speech Processing
IEEE Communications Magazine, Jan. 1983
37. J. L. Flanagan, M. R. Schroeder, B. S. Atal, R. E. Crochiere,
N. S. Jayant & J. M. Tribolet
Speech Coding
IEEE Trans. Comm. Vol. Com-27, No.4, April 1979
38. V. R. Viswanathan, J. Makhoul, R. M. Schwartz & W. F. Huggins
Variable Frame Rate Transmission : a Review of Methodology and
Applications to Narrow-Band LPC Speech Coding
IEEE Tran. Inform. Theory, Vol. IT-29, No.4, July 1983

39. F. Kuhlmann, A. Buzo & Zhenyu Wu
Variable Bit-Rate LPC Coding of Speech Signals
in Proc. the 26-th Midwest Symposium on Circuits and Systems
Puebla, México. August 1983
40. A. Buzo, A. H. Gray, R. M. Gray & J. D. Markel
Speech Coding Based Upon Vector Quantization
IEEE Trans. Acoustics, Speech and Signal Processing
Vol. ASSP-28, No.5, October 1980