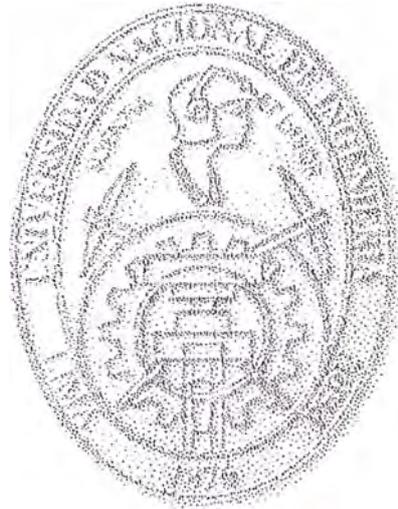


**UNIVERSIDAD NACIONAL DE INGENIERIA**

**FACULTAD DE INGENIERIA ELÉCTRICA Y ELECTRÓNICA**



**COMPRESIÓN DE LA VOZ APLICANDO LA TÉCNICA DE  
PREDICCIÓN LINEAL DE COEFICIENTES**

**INFORME DE SUFICIENCIA**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**

**INGENIERO ELECTRÓNICO**

**PRESENTADO POR:**

**HÉCTOR HUGO POZO CÁRDENAS**

**PROMOCIÓN  
1980 - I**

**LIMA – PERÚ**

**2005**

**COMPRESIÓN DE LA VOZ APLICANDO LA  
TÉCNICA DE PREDICCIÓN LINEAL DE  
COEFICIENTES**

Agradezco a mi esposa Carmen Rosa por darme el apoyo y fuerza para hacer este trabajo.  
A mi hija Nathalie por la paciencia que tuvo conmigo y quedarse los domingos en casa  
A mi hijo Renato aun pequeño pero pienso que hubiera hecho lo mismo que Nathalie

## SUMARIO

Este trabajo trata sobre el método de compresión de voz aplicando la técnica de predicción lineal de coeficientes.

Consta de una introducción, tres capítulos, conclusiones, recomendaciones anexo.

En la introducción se realiza un análisis de como a través de un estudio de la forma en que se pronuncia la voz, se puede hacer un modelo matemático y de esta forma tratar de comprimir la señal de voz.

En el primer capítulo es una introducción teórica de los distintos métodos en la que se puede representar el modelo matemático (coeficientes del filtro LPC) de la voz.

En el segundo capítulo se desarrolla las distintas técnicas para hallar los coeficientes del LPC

En tercer capítulo se ha desarrollado en Matlab, un algoritmo en el cual a partir de la captura de una frase, siguiendo la técnica del LPC, se demuestra que se puede comprimir la voz, y después recuperarla.

En las conclusiones se muestra las observaciones que uno ha realizado al correr el algoritmo y las limitaciones que se tiene para poder desarrollar un algoritmo más elaborado.

En las recomendaciones se da ideas que uno tiene acerca de cómo está la Electrónica actualmente en las Telecomunicaciones.

En el anexo se da una breve información de los distintos tipos de vocoders que existen actualmente.

## INDICE

### PROLOGO

### CAPITULO I

#### CODIFICACION DE PREDICCIÓN LINEAL DE LA PALABRA

1.1	Introducción	3
1.2	Principios Básicos de Análisis de Predicción Lineal	6
1.2.1	El método de auto-correlación	15
1.2.2	El método de covarianza	19
1.2.3	Resumen	21

### CAPITULO II

#### SOLUCION DE LAS ECUACIONES LPC

2.1	Cálculo de la ganancia para el modelo	23
2.2	Solución de descomposición de Cholesky por el método de covarianza	29
2.3	La solución recurrente de Durbin para las Ecuaciones de auto-correlación	38
2.4	Formulaciones y soluciones de Lattice	43

### CAPITULO III

#### ALGORITMO PARA HALLAR EL PERIODO PITCH Y LOS COEFICIENTES DEL LPC UTILIZANDO MATHLAB

3.1	Diagrama de bloques del codificador LPC	53
3.2	Partes del programa	53
3.2.1	Ingreso de datos generales	54
3.2.2	Grabado de voz	54

<b>3.2.3</b>	<b>Separación mediante ventanas rectangulares</b>	<b>54</b>
<b>3.2.4</b>	<b>Obtención de los coeficientes LPC</b>	<b>55</b>
<b>3.2.5</b>	<b>Obtención del pitch</b>	<b>55</b>
<b>3.2.6</b>	<b>Obtención de la ganancia</b>	<b>55</b>
<b>3.2.7</b>	<b>Compresión del pitch</b>	<b>55</b>
<b>3.2.8</b>	<b>Envío de datos</b>	<b>56</b>
<b>3.2.9</b>	<b>Recepción de datos</b>	<b>56</b>
<b>3.2.10</b>	<b>Descompresión del pitch</b>	<b>56</b>
<b>3.2.11</b>	<b>Recuperación de la señal original</b>	<b>57</b>
<b>3.2.12</b>	<b>Filtro pasabajo</b>	<b>57</b>
<b>3.2.13</b>	<b>Reproducción de la señal recuperable</b>	<b>57</b>
<b>3.3</b>	<b>Programa en Matlab para calcular: el pitch, coeficientes LPC y la energía</b>	<b>58</b>
	<b>CONCLUSIONES Y RECOMENDACIONES</b>	<b>63</b>
	<b>ANEXO</b>	<b>67</b>
	<b>BIBLIOGRAFIA</b>	<b>73</b>

## PROLOGO

La compresión de las señales de voz tiene muchas aplicaciones prácticas. Un ejemplo es en la tecnología celular digital donde varios canales de voz comparten la misma anchura de banda de la frecuencia. La compresión permite que se comparta mejor el ancho de banda del sistema de la mejor manera posible. Otro ejemplo está en el almacenaje digital de la voz (es decir, contestadores automáticos). Para un tamaño de la memoria dado, la compresión permite que mensajes más largos sean almacenados de la mejor manera.

Históricamente, las señales de voz digitalizadas se muestrean con una tasa de 8000 muestras/seg. Típicamente, cada muestra es representada por 8 dígitos binarios (que usan la ley-u o A). Esto corresponde a un índice sin comprimir de 64 kbps (kbits/seg). Con las técnicas actuales de la compresión, es posible reducir la tasa a 8 kbps con casi ninguna pérdida perceptible en calidad. La compresión adicional es posible en un costo de una calidad más baja. Basan todos los codificadores de baja tasa actuales

del discurso en el principio de la *predicción lineal de coeficientes (LPC)* que se presenta a continuación.

# **CAPITULO I**

## **CODIFICACION DE PREDICCION LINEAL DE LA PALABRA**

### **1.1 Introducción**

Una de las técnicas más eficaces del análisis de la palabra es el método de análisis de predicción lineal. Este método ha llegado a ser la técnica predominante para determinar los parámetros básicos de la palabra, por ejemplo, pitch (frecuencia fundamental), formantes, espectros, funciones del área del aparato vocal, además para representar la palabra para la transmisión o almacenamiento de baja velocidad en bits. La importancia de este método recae en dos situaciones: la primera, en la habilidad de proveer estimaciones bastante precisas de los parámetros de la palabra y la segunda, en la velocidad relativa de cálculo. En este trabajo, presentamos la formulación de ideas anteriores a la predicción lineal y discutimos algunos de los temas que se involucran en su uso en aplicaciones prácticas de la palabra.

La idea básica que se esconde detrás del análisis de predicción lineal es que una muestra de palabra puede ser tan parecida como una combinación lineal de muestras pasadas de palabra. Se puede determinar todo un conjunto de

coeficientes de predicción disminuyendo la suma de las diferencias cuadradas (sobre un intervalo finito) entre las muestras actuales de la palabra y aquellas de predicción lineal. (Los coeficientes de predicción son ponderados y utilizados en la combinación lineal.)

La filosofía de predicción lineal está estrechamente relacionada al modelo básico de síntesis de la palabra, donde se muestra que la palabra puede ser presentada como la salida de un sistema lineal con variante en el tiempo excitado ya sea por pulsos cuasi-periódicos (durante palabra sonora) o por un ruido aleatorio (durante palabra no-sonora). El método de predicción lineal provee un método preciso, confiable y resistente para calcular los parámetros que caracterizan el sistema lineal con variante en el tiempo.

La predicción lineal se podría aplicar en un esquema cuantificación diferencial para reducir la velocidad de los bits de la representación digital de la palabra en forma de onda. En realidad, la base matemática, para una predicción de adaptación de alto orden, utilizada para una codificación DPCM en forma de onda, es idéntica al análisis que debemos presentar en este capítulo. En la codificación de adaptación DPCM, el énfasis radica en encontrar un sistema de predicción que vaya a reducir la variación de la señal de diferencia para que se reduzca de igual manera la cuantificación del error. En este trabajo, nos enfocaremos en un punto de vista más general, además de mostrar cómo la idea de predicción básica lineal nos conduce hacia un conjunto de técnicas de análisis que pueden ser utilizadas para calcular parámetros de una muestra de

palabra. Con frecuencia, este conjunto general de técnicas de análisis de predicción lineal se denomina como predicción lineal de coeficientes o LPC.

Durante mucho tiempo, las técnicas y métodos de predicción lineal han estado disponibles en la literatura de la ingeniería. Las ideas de predicción lineal han estado en uso en las siguientes áreas: control y teoría de la información bajo los nombres de sistema de estimación e identificación. El término sistema de identificación describe particularmente los métodos de LPC una vez que se hayan obtenido los coeficientes extrapolados. Este sistema ha sido identificado exclusivamente hasta el punto que puede ser presentado como un sistema lineal todos-polos.

Así como se aplica para el procesamiento de la palabra, el término de predicción lineal se refiere a una variedad, principalmente de formulaciones equivalentes del problema de presentación de la palabra en forma de onda. Las diferencias entre estas formulaciones son, con frecuencia, aquellas de filosofía o de la manera de ver el problema. En otros casos, las diferencias se encuentran en los detalles de cálculo utilizados para obtener los coeficientes de predicción. De esta manera, tal y como se aplica a la palabra, las variadas formulaciones (con frecuencia, equivalentes) del análisis de predicción lineal han sido:

1. El método de covarianza.
2. La formulación de auto-correlación.
3. El método de Lattice.

4. La formulación del filtro inverso.
5. La formulación de la estimación de espectro.
6. La formulación de máxima probabilidad.
7. la formulación del producto interno

En este trabajo examinaremos, los tres primeros métodos de análisis arriba mencionados, puesto que las otras formulaciones son equivalentes a una de estas tres.

La importancia de la predicción lineal recae en la exactitud con la cual el modelo básico se aplica a la palabra. Así, la mayor parte de este trabajo está dedicada a la discusión de cómo una variedad de parámetros de la palabra puede estar bien estimada utilizando métodos de predicción lineal. Además, algunos ejemplos típicos de aplicaciones de palabra, las cuales recaen principalmente en el análisis de predicción lineal, son discutidos aquí, para mostrar el amplio margen de problemas, al cual el LPC ha sido exitosamente aplicado.

## **1.2 Principios Básicos de Análisis de Predicción Lineal**

A través de este trabajo repetidas veces nos hemos referido al modelo básico de tiempo-discreto para la producción de la palabra. La forma particular de este modelo, que es apropiada para la discusión del análisis de predicción lineal, está descrita en la figura 1. En este caso, los efectos del espectro compuesto de la radiación, aparato vocal y excitación glótica están representados por un

filtro digital variante en el tiempo cuya función del sistema de estado-estático es de la forma.

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (1)$$

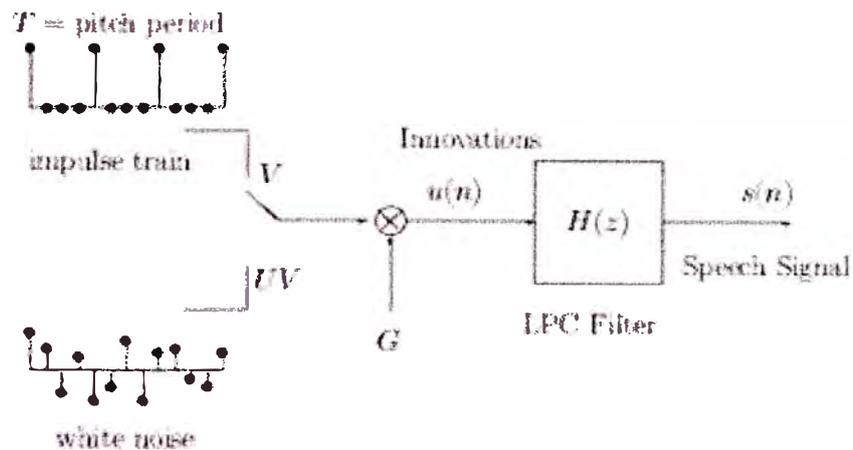


Fig. 1 Diagrama en bloques de un modelo simplificado para la producción de la palabra

Este sistema se excita por un tren de impulsos para el sonido sonoro o una secuencia de ruido blanco aleatorio para el sonido no sonoro. Así, los

parámetros de este modelo son: clasificación del sonido sonoro/no sonoro , período de pitch para palabra el sonido sonoro, parámetro de ganancia  $G$  , y los coeficientes  $\{a_k\}$  del filtro digital. Por supuesto, todos estos parámetros varían ligeramente con el tiempo.

El período pitch y la clasificación de sonido sonoro/no sonoro pueden ser estimados utilizando de métodos basados en el análisis de predicción lineal que serán discutidos posteriormente. Este modelo simplificado todos-polos es una representación natural de sonidos sonoros no-nasales, pero para sonidos fricativos y nasales, la teoría acústica detallada requiere ambos polos además de ceros, en la función de transferencia del aparato vocal. Sin embargo, apreciamos que si el orden  $p$  es suficientemente alto, el modelo todos-polos provee una buena representación a la mayoría de todos los sonidos de la palabra. La mayor ventaja de este modelo es que el parámetro de ganancia,  $G$  , y el filtro de coeficientes  $\{a_k\}$  pueden ser estimados de una manera de cálculo muy clara y eficaz a través del método de análisis de predicción lineal.

Para el sistema de la figura 1, las muestras de la palabra  $s(n)$  están relacionadas a la excitación  $u(n)$  a través de la ecuación de diferencia simple

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (2)$$

Un sistema de predicción lineal con coeficientes de predicción  $\alpha_k$  es definido como un sistema cuya salida es

$$s(n) = \sum_{k=1}^p \alpha_k s(n-k) \quad (3)$$

Dicho sistema es utilizado para reducir la varianza de la señal la diferencia en el esquemas de cuantificación diferencial. La función del sistema de un orden  $p^{th}$  de un sistema de predicción lineal es el polinomio

$$P(z) = \sum_{k=1}^p \alpha_k z^{-k} \quad (4)$$

El error de predicción,  $e(n)$  se define como

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k) \quad (5)$$

Desde la ecuación (5) se puede apreciar que la secuencia de error de predicción es la salida de un sistema cuya función de transferencia es

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k} \quad (6)$$

Al comparar las ecuaciones (2) y (5) se puede apreciar que si la señal de la

palabra obedece exactamente al modelo de la ecuación (2) y si  $\alpha_k = a_k$  entonces  $e(n) = Gu(n)$ . De esta manera, el filtro de error de predicción,  $A(z)$ , será un filtro inverso para el sistema,  $H(z)$ , de la Ecuación (1), es decir,

$$H(z) = \frac{G}{A(z)} \quad (7)$$

El problema básico del análisis de predicción lineal es el de determinar un conjunto de coeficientes de predicción  $\{a_k\}$  directamente de la señal de la palabra de tal manera que se pueda obtener una buena estimación de las propiedades del espectro de la señal de la palabra a través del uso de la ecuación (7). Los coeficientes de predicción deben ser estimados a partir de segmentos cortos de la señal de la palabra debido a la naturaleza variante en el tiempo de la señal de voz. El enfoque básico es encontrar un conjunto de coeficientes de predicción que vayan a minimizar el error de predicción medio-cuadrático sobre un corto segmento de forma de onda de la palabra. Entonces, los parámetros resultantes son asumidos para convertirse en los parámetros de la función del sistema,  $H(z)$  en el modelo de producción de la palabra.

Inmediatamente, no será obvio que este enfoque nos conlleve a resultados útiles, pero puede ser justificado de varias maneras. Primero, hay que recordar que si  $\alpha_k = a_k$ , entonces  $e(n) = Gu(n)$ . Para el sonido sonoro esto significa que  $e(n)$  se compone de un tren de impulsos: es decir,  $e(n)$  debe ser lo mas

pequeño en el tiempo. De esta manera, encontrando  $\alpha_k$ 's que minimiza el error de predicción, pareciera consistente de acuerdo a esta observación. Una segunda motivación para este enfoque sigue del hecho que si una señal es generada por la ecuación (2) con coeficientes no variante en el tiempo, es excitado, ya sea por un solo impulso o por una entrada del ruido blanco estacionario; luego se puede demostrar que los coeficientes de predicción, que son el resultado de minimizar el error cuadrático medio (durante todo el tiempo), son idénticos a los coeficientes de la ecuación (2). Una tercera justificación muy pragmática por utilizar el error de predicción mínima medio-cuadrático como base para calcular los modelos de los parámetros, es que este enfoque conlleve a un conjunto de ecuaciones lineales que puedan resolverse para obtener eficazmente los parámetros de predicción. Como podremos apreciar más adelante en este trabajo, los parámetros que resulten más importantes, incluyen una representación muy útil y exacta de la señal de la palabra.

El promedio del error de predicción de un tiempo corto se define como

$$E_n = \sum_m e_n^2(m) \quad (8)$$

$$= \sum_m (s_n(m) - \hat{s}_n(m))^2 \quad (9)$$

$$= \sum_m \left( s_n(m) - \sum_{k=1}^p \alpha_k s_n(m-k) \right)^2 \quad (10)$$

donde  $s_n(m)$  es un segmento de la palabra que ha sido seleccionado en la vecindad de la muestra  $n$ , es decir,

$$s_n(m) = s(n+m) \quad (11)$$

El rango de las sumatorias en las ecuaciones (8) - (10) no es especificado temporalmente, pero ya que nosotros deseamos desarrollar una técnica de análisis para tiempos cortos, la suma siempre estará sobre un intervalo finito. También notar que para obtener un promedio, debemos dividir por la longitud del segmento de la palabra. Sin embargo, esta constante es irrelevante al conjunto de Ecuaciones lineales que obtendremos y por consiguiente es omitido. Encontraremos los valores de  $\alpha_k$  que minimiza  $E_n$  en la ecuación (10) al establecer  $\partial E_n / \partial \alpha_i = 0$ ,  $i = 1, 2, \dots, p$  por lo tanto, se obtiene la Ecuación

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^p \hat{\alpha}_k \sum_m s_n(m-i)s_n(m-k) \quad (12)$$

donde  $\hat{\alpha}_k$  son los valores de  $\alpha_k$  que minimizan  $E_n$ , (Desde que  $\hat{\alpha}_k$  es único, no

utilizaremos el acento circunflejo pero sí la anotación para indicar los valores que minimizan  $E_n$ ). Si definimos

$$\phi_n(i, k) = \sum_m s_n(m-i)s_n(m-k) \quad (13)$$

entonces la ecuación (12) puede ser escrita de una manera más compacta como

$$\sum_{k=1}^p \alpha_k \phi_n(i, k) = \phi_n(i, 0) \quad (14)$$

Este conjunto de ecuaciones  $p$  en  $p$  desconocidos puede resolverse de una manera eficiente por los coeficientes de predicción desconocidos  $\{\alpha_k\}$  que minimizan el error de predicción cuadrático promedio para el segmento  $s_n(m)$ <sup>1</sup>. Utilizando las ecuaciones (10) y (12), el mínimo error cuadrático medio puede ser mostrado como

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^p \alpha_k \sum_m s_n(m)s_n(m-k) \quad (15)$$

y utilizando la ecuación (14) podemos expresar  $E_n$  como

$$E_n = \phi(0,0) - \sum_{k=1}^p \alpha_k \phi(0,k) \quad (16)$$

Así, el total del mínimo error forma parte de un componente fijo y otro componente que depende de los coeficientes de predicción.

Para optar por los coeficientes de predicción óptimos, primero, debemos calcular las cantidades  $\phi_n(i,k)$  para  $1 \leq i \leq p$  y  $0 \leq k \leq p$ . Una vez realizado este proceso, tenemos que optar por la ecuación (14) para obtener el  $\alpha_k$ 's. Así, en principio, el análisis de predicción lineal es muy directo. Sin embargo, los detalles del cálculo de  $\phi_n(i,k)$  y la subsiguiente solución de las ecuaciones es de alguna manera intrincada y se recomienda una mayor discusión.

Hasta el momento no hemos indicado explícitamente los límites de las sumas en las ecuaciones (8) - (10) y (12); sin embargo se debería enfatizar que los límites de la suma en la ecuación (12) son idénticos a los límites asumidos en el error de predicción cuadrático medio en las ecuaciones (8) – (10). Como ya hemos establecido, si deseamos desarrollar un procedimiento de análisis para tiempos cortos, los límites deben sobre un intervalo finito. Existen dos enfoques básicos a esta pregunta, y debemos ver estos dos métodos que para el análisis de predicción lineal, emergen de una consideración de los límites de la sumatoria y la definición del segmento de la forma de onda  $s_n(m)$ .

### 1.2.1 El método de auto-correlación

Para asumir que el segmento en forma de onda,  $s_n(m)$ , sea idénticamente a cero fuera del intervalo  $0 \leq m \leq N-1$ , existe un enfoque para determinar los límites de las sumas en las ecuaciones (8) - (10) y (12). Esto puede expresarse convenientemente como

$$s_n(m) = s(m+n)w(m) \quad (17)$$

donde  $w(m)$  es una ventana con una longitud finita (por ejemplo, la ventana Hamming) que es idéntica a cero fuera del intervalo  $0 \leq m \leq N-1$ .

El efecto de esta asunción, con relación a la pregunta de los límites de la sumatoria para las expresiones de  $E_n$ , puede ser apreciado teniendo en cuenta la ecuación (5). Se aprecia claramente, si  $s_n(m)$  es diferente de cero sólo para  $0 \leq m \leq N-1$ , entonces el correspondiente error de predicción,  $e_n(m)$ , para una orden  $p^{\text{th}}$  de predicción será diferente de cero sobre el intervalo  $0 \leq m \leq N-1+p$ . Así, para este caso  $E_n$  se expresa apropiadamente como

$$E_n = \sum_{m=0}^{N+p-1} e_n^2(m) \quad (18)$$

Como alternativa, hubiéramos simplemente indicado que la suma debería ser

sobre los valores diferente de cero para la sumatoria de  $-\infty$  a  $+\infty$ .

Volviendo a la ecuación (5), se puede apreciar que el error de predicción es probable que sea grande al principio del intervalo (específicamente  $0 \leq m \leq p-1$ ) porque está tratando de predecir las señales de las muestras que han sido establecidas, arbitrariamente, a cero. Del mismo modo el error puede ser grande al final del intervalo (específicamente  $N \leq m \leq N+p-1$ ) porque estamos tratando de predecir cero de las muestras que son diferente de cero. Por esta razón, una ventana que disminuye el segmento,  $s_n(m)$ , a cero es utilizado generalmente por  $w(m)$  en la ecuación (17).

Los límites en la expresión para  $\phi_n(i,k)$  en la ecuación (13) son idénticos a aquellos de la ecuación (18). Sin embargo, porque  $s_n(m)$  es idéntico a cero fuera del intervalo  $0 \leq m \leq N-1$ , es simple demostrar que

$$\phi_n(i,k) = \sum_{m=0}^{N+p-1} s_n(m-i)s_n(m-k) \quad \begin{array}{l} 1 \leq i \leq p \\ 0 \leq k \leq p \end{array} \quad (19a)$$

se puede expresar como

$$\phi_n(i,k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k) \quad \begin{array}{l} 1 \leq i \leq p \\ 0 \leq k \leq p \end{array} \quad (19b)$$

Además puede verse que en este caso  $\phi_n(i,k)$  es una la auto-correlación de

tiempo corto evaluado para  $(i - k)$ . Resultando

$$\phi(i, k) = R_n(i - k) \quad (20)$$

donde

$$R_n(k) = \sum_{m=0}^{N-1-k} s_n(m) s_n(m+k) \quad (21)$$

Puesto que  $R_n(k)$  cumple una función uniforme de ello resulta que

$$\phi_n(i, k) = R_n(|i - k|) \quad \begin{array}{l} i = 1, 2, \dots, p \\ k = 0, 1, \dots, p \end{array} \quad (22)$$

Por consiguiente la ecuación (14) se puede expresar como

$$\sum_{k=1}^p \alpha_k R_n(|i - k|) = R_n(i) \quad 1 \leq i \leq p \quad (23)$$

De igual manera, el error de predicción cuadrático medio de la ecuación (16) toma la forma de

$$E_n = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k) \quad (24)$$

El conjunto de ecuaciones otorgadas por la ecuación (23) puede ser expresado en forma matricial como

$$\begin{bmatrix} R_n(0) & R_n(1) & R_n(2) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & R_n(1) & \dots & R_n(p-2) \\ R_n(2) & R_n(1) & R_n(0) & \dots & R_n(p-3) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ R_n(p-1) & R_n(p-2) & R_n(p-3) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_1 \\ \alpha_1 \\ \dots \\ \dots \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ R_n(3) \\ \dots \\ \dots \\ R_n(p) \end{bmatrix} \quad (25)$$

La matriz  $p \times p$  de los valores de auto-correlación es una matriz Toeplitz; por ejemplo, es simétrica y todo los elementos, a lo largo de una diagonal dada, son iguales. Esta propiedad especial será explotada en la solución de ecuaciones del LPC para obtener un algoritmo eficiente para la solución de la ecuación (23).

### 1.2.2 El método de covarianza

El segundo enfoque básico para definir el segmento de la palabra  $s_n(m)$  y los límites de las sumas, es fijar el intervalo sobre el cual el error cuadrático medio es calculado y luego considera el efecto en el cálculo de  $\phi_n(i, k)$ . Esto es, si podemos definir

$$E_n = \sum_{m=0}^{N-1} e_n^2(m) \quad (26)$$

luego,  $\phi_n(i, k)$  se convierte

$$\phi_n(i, k) = \sum_{m=0}^{N-1} s_n(m-i)s_n(m-k) \quad \begin{array}{l} 1 \leq i \leq p \\ 0 \leq k \leq p \end{array} \quad (27)$$

En este caso, si cambiamos el índice de la suma, podemos expresar  $\phi_n(i, k)$  como esto

$$\phi_n(i, k) = \sum_{m=-i}^{N-i-1} s_n(m)s_n(m+i-k) \quad \begin{array}{l} 1 \leq i \leq p \\ 0 \leq k \leq p \end{array} \quad (28a)$$

o

$$\phi_n(i, k) = \sum_{m=-k}^{N-k-1} s_n(m) s_n(m+k-i) \quad \begin{array}{l} 1 \leq i \leq p \\ 0 \leq k \leq p \end{array} \quad (28b)$$

A pesar que estas ecuaciones parecen ser similares a la ecuación (19b), apreciamos que los límites de la suma no son los mismos. La ecuación (28), busca los valores de  $s_n(m)$  fuera del intervalo  $0 \leq m \leq N-1$ . De hecho, para evaluar  $\phi_n(i, k)$  todos los valores requeridos de  $i$  y de  $k$ , se requiere el uso de valores de  $s_n(m)$  en el intervalo  $-p \leq m \leq N-1$ . Si tuviéramos que ser consecuentes con los límites con en la ecuación (26), entonces no tendríamos alternativa sino la de proporcionar los valores requeridos. En este caso no tiene sentido el alineamiento del segmento de la palabra a cero, al final como fuera del intervalo  $0 \leq m \leq N-1$ . Este enfoque nos conduce hacia la función que no es una verdadera función de auto-correlación, sino, una inter.-correlación entre dos segmentos de longitud finita, muy similares mas no idénticos, de la onda de la palabra. A pesar que las diferencias entre las ecuaciones (28) y (19b) puedan parecer detalles de cálculo menores, el conjunto de ecuaciones

$$\sum_{k=1} \alpha_k \phi_n(i, k) = \phi_n(i, 0) \quad (29a)$$

tiene propiedades significativamente diferentes que afectan fuertemente el método de solución y las propiedades del resultado óptimo de predicción. Estas

ecuaciones se convierten en la forma matriz

$$\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & \phi_n(1,3) & \dots & \phi_n(1,p) \\ \phi_n(2,1) & \phi_n(2,2) & \phi_n(2,3) & \dots & \phi_n(2,p) \\ \phi_n(3,1) & \phi_n(3,2) & \phi_n(3,3) & \dots & \phi_n(3,p) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \phi_n(p,1) & \phi_n(p,2) & \phi_n(p,3) & \dots & \phi_n(p,p) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \dots \\ \dots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \phi_n(3,0) \\ \dots \\ \dots \\ \phi_n(p,0) \end{bmatrix} \quad (29b)$$

En este caso, ya que  $\phi_n(i,k) = \phi_n(k,i)$  (ver la ecuación (28)), la matriz  $p \times p$  de valores de correlación-tipo es simétrica mas no Toeplitz. De hecho, se puede apreciar que los elementos diagonales están relacionados con la Ecuación.

$$\phi_n(i+1,k+1) = \phi_n(i,k) + s_n(-i-1)s_n(-k-1) - s_n(N-1-i)s_n(N-1-k) \quad (30)$$

El método de análisis basado en este método de cálculo de  $\phi_n(i,k)$  ha llegado a conocerse como el método de covarianza porque la matriz de los valores  $[\phi_n(i,k)]$  tiene las propiedades de una matriz de covarianza<sup>2</sup>.

### 1.2.3 Resumen

Se ha demostrado que al utilizar diferentes definiciones de los segmentos de la señal a ser analizada, puede obtenerse dos conjuntos distintos de ecuaciones

de análisis. Para el método de auto-correlación, la señal es una selección de ventana dada por una ventana de  $N$  punto, y las cantidades  $\phi_n(i, k)$  se obtienen utilizando una función de auto-correlación en tiempos cortos. La matriz, obtenida de las correlaciones es Toeplitz, dirigiéndose hacia un tipo de solución para los coeficientes de predicción. Por el método de covarianza, se asume que la señal es conocida por los valores del conjunto  $-p \leq n \leq N-1$ . Fuera de este intervalo no se necesita hacer suposiciones con respecto a la señal, puesto que estos son los únicos valores necesitados en el cálculo. En este caso, la matriz de correlación obtenida es simétrica pero no Toeplitz. El resultado es que los dos métodos de cálculo de las correlaciones conduce a diferentes métodos de solución de las ecuaciones de análisis y a los conjuntos de coeficientes de predicción con algún tipo de propiedades diferentes.

## CAPITULO II

### SOLUCIONES DE LAS ECUACIONES LPC

En secciones posteriores, vamos a comparar y contrastar detalles de cálculo y resultados para estas técnicas así como para otro método todavía a discutir. Sin embargo, primero mostraremos, cómo obtener la ganancia,  $G$ , en la figura 1 pudiendo determinarse con la expresión de error de predicción.

#### 2.1 Cálculo de la ganancia para el modelo

Es razonable esperar que la ganancia,  $G$ , podría determinarse al combinar la energía en la señal con la energía de las muestras de predicción lineal. De hecho, esto es verdad cuando se realizan suposiciones apropiadas sobre la señal de excitación al sistema LPC

Es posible relacionar la ganancia constante  $G$  a la excitación de la señal y el error en predicción al referirnos a las ecuaciones (2) y (5)<sup>3</sup>. La señal de excitación,  $Gu(n)$ , se puede expresar como

$$Gu(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k) \quad (31a)$$

donde la señal de error de predicción  $e(n)$  se expresa como

$$e(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k) \quad (31b)$$

En el caso donde  $a_k = \alpha_k$  por ejemplo, los coeficientes actuales de predicción y aquellos del modelo son idénticos, entonces

$$e(n) = Gu(n) \quad (32)$$

es decir, la señal de salida es proporcional a la señal de error con la constante de proporcionalidad, siendo la ganancia,  $G$ , constante.

Puesto que la ecuación (32) es sólo aproximada (es decir, es válido para la extensión en que los parámetros de predicción lineal actuales e ideales, sean idénticos) por lo general no es posible resolver directamente  $G$  de una manera fiable de la misma señal de error. En cambio, la suposición más razonable es que la energía en la señal de error sea igual a la energía de la excitación de

salida, es decir,

$$G^2 \sum_{m=0}^{N-1} u^2(m) = \sum_{m=0}^{N-1} e^2(m) = E_n \quad (33)$$

A esta altura debemos realizar algunas suposiciones acerca  $u(n)$  para poder relacionar  $G$  a cantidades conocidas, es decir, a los  $\alpha_k$ 's y los coeficientes de correlación. Hay dos casos de interés para la excitación. Para el sonido sonoro es razonable asumir  $u(n) = \delta(n)$ , es decir, la excitación es una muestra de unidad en  $n = 0$ <sup>4</sup>. Para que esta suposición sea válida requiere que los efectos de la forma de impulso glótico utilizados en la excitación actual para el sonido sonoro, sean agrupados junto a la función de transferencia del aparato vocal y por consiguiente ambos efectos son esencialmente moldeados por el sistema de predicción lineal variante en el tiempo. Esto requiere que el orden de predicción,  $p$ , sea lo suficientemente amplia para considerar ambos, el aparato vocal y los efectos del impulso glótico. Es más razonable para el sonido no sonoro asumir que  $u(n)$  es un proceso de ruido blanco, estacionario, y valor medio cero.

Basados en estas suposiciones, ahora podemos determinar la ganancia constante  $G$  utilizando la ecuación (33). Para el sonido sonoro, tenemos como entrada  $G\delta(n)$ . Si llamamos resultado de salida a esta entrada particular  $h(n)$  (puesto que en realidad es la respuesta de impulso del sistema con la

función de transferencia  $H(z)$  como en la ecuación (1) obtenemos la relación

$$h(n) = \sum_{k=1}^p \alpha_k h(n-k) + G\delta(n) \quad (34)$$

La función de auto-correlación de  $h(n)$ , definida como

$$R(m) = \sum_{n=0}^{\infty} h(n)h(m+n) \quad (35)$$

satisface las relaciones

$$R(m) = \sum_{k=1}^p \alpha_k R(|m-k|) \quad m = 1, 2, \dots, p \quad (36a)$$

y

$$R(0) = \sum_{k=1}^p \alpha_k R(k) + G^2 \quad (36b)$$

Puesto que la ecuación (36) es idéntica a la ecuación (23), de ello resulta que

$$R(m) = R_n(m) \quad 1 \leq m \leq p \quad (37)$$

Ya que la energía total en la señal ( $R(0)$ ) y la respuesta del impulso ( $\bar{R}(0)$ ) deberían ser iguales, podemos utilizar las ecuaciones (24), (33) y (36b) para obtener

$$G^2 = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k) = E_n \quad (38)$$

Es interesante notar que la ecuación (37) y el requerimiento de la energía de la respuesta impulso sea igual a la energía de la señal, juntos requieren que los primeros coeficientes de la función de auto-correlación  $p+1$  de la respuesta del impulso del modelo sean idénticos a los primeros coeficientes de la función de auto-correlación  $p+1$  de la señal de la palabra.

Para el caso del sonido no sonoro, las correlaciones se definen como promedios estadísticos. Se asume que la entrada es ruido blanco con valor medio cero, es decir,

$$E[u(n)u(n-m)] = \delta(m) \quad (39)$$

Si excitamos al sistema con señales de entrada aleatoria  $Gu(n)$  y llamamos a la salida  $g(n)$  entonces

$$g(n) = \sum_{k=0}^p \alpha_k g(n-k) + Gu(n) \quad (40)$$

Si ahora colocamos  $R(m)$  que indica la función de auto-correlación  $g(n)$ , entonces

$$\begin{aligned} R(m) = E[g(n)g(n-m)] &= \sum_{k=1}^p \alpha_k E[g(n-k)g(n-m)] + E[Gu(n)g(n-m)] \\ &= \sum_{k=1}^p \alpha_k R(m-k) \quad m \neq 0 \end{aligned} \quad (41)$$

ya que  $E[u(n)g(n-m)] = 0$  para  $m > 0$  porque  $u(n)$  no está en correlación con ninguna señal anterior a  $u(n)$ . Para  $m = 0$  obtenemos

$$\begin{aligned} R(0) &= \sum_{k=1}^p \alpha_k R(k) + GE[u(n)g(n)] \\ &= \sum_{k=1}^p \alpha_k R(k) + G^2 \end{aligned} \quad (42)$$

puesto que  $E[u(n)g(n)] = E[u(n)(Gu(n) + \text{terminos posteriores a } n)] = G$ . Dado que la energía en la respuesta a  $Gu(n)$  debería ser igual a la energía de la señal, obtenemos

$$R(m) = R_n(m) \quad 0 \leq m \leq p \quad (43)$$

o

$$G^2 = R_n(0) - \sum_{k=1}^p \alpha_k R_n(k) \quad (44)$$

como fue el caso para la excitación del impulso para la palabra sonora.

Para poner en práctica, con eficiencia, un sistema de análisis de predicción lineal, es necesario resolver las ecuaciones lineales de igual manera. A pesar que una variedad de técnicas puedan ser aplicadas para resolver un sistema de ecuaciones lineales de  $p$  en  $p$  desconocidas, estas técnicas no son de igual eficiencia. Debido a las propiedades especiales de las matrices de coeficiente es posible resolver las ecuaciones con mayor eficiencia que lo es posible en general. En esta sección vamos a discutir al detalle dos métodos para obtener los coeficientes de predicción.

## 2.2 Solución de descomposición de Cholesky por el método de covarianza

Por el método de covarianza, el conjunto de ecuaciones que debe ser resuelto

es de la forma:

$$\sum_{k=1}^p \alpha_k \phi_n(i, k) = \phi_n(i, 0) \quad i = 1, 2, \dots, p \quad (45)$$

o en notación de una matriz

$$\Phi \alpha = \psi \quad (46)$$

donde  $\Phi$  es una matriz simétrica definida positiva con  $(i, j)^{th}$  elemento  $\phi_n(i, j)$  y  $\alpha$  y  $\psi$  son vectores columnas con elementos  $\alpha_p$  y  $\phi_n(i, 0)$  respectivamente.

El sistema de ecuaciones dado por la ecuación (45) puede resolverse de una manera eficaz puesto que la matriz  $\Phi$  es una matriz definida, positiva y simétrica. El método de solución obtenido se llama descomposición Cholesky (o algunas veces se llama el método de la raíz cuadrada). Para este método la matriz  $\Phi$  se expresa de la forma

$$\Phi = VDV^t \quad (47)$$

donde  $V$  es una matriz triangular inferior (cuyos principales elementos diagonales todos son 1's), y  $D$  es una matriz diagonal. El superíndice  $t$  indica la matriz transpuesta. Inmediatamente, los elementos de las matrices  $V$  y  $D$  son determinados por la ecuación (47) por resolución de los elementos  $(i, j)^{th}$

de ambos lados de la ecuación (47) dando

$$\phi_n(i, j) = \sum_{k=1}^j V_{ik} d_k V_{jk} \quad 1 \leq j \leq i-1 \quad (48)$$

o

$$V_{ij} d_j = \phi_n(i, j) - \sum_{k=1}^{j-1} V_{ik} d_k V_{jk} \quad 1 \leq j \leq i-1 \quad (49)$$

y para los elementos diagonales

$$\phi_n(i, i) = \sum_{k=1}^i V_{ik} d_k V_{ik} \quad (50)$$

o

$$d_i = \phi_n(i, i) - \sum_{k=1}^{i-1} V_{ik}^2 d_k \quad i \geq 2 \quad (51)$$

con

$$d_1 = \phi_n(1, 1) \quad (52)$$

Para ilustrar el uso de las ecuaciones (47) - (52) hay que considerar un ejemplo con  $p = 4$ , y los elementos matrices  $\phi_n(i, j) = \phi_{ij}$ . Entonces la ecuación (47) es de la forma.

$$\begin{bmatrix} \phi_{11} & \phi_{21} & \phi_{31} & \phi_{41} \\ \phi_{21} & \phi_{22} & \phi_{32} & \phi_{42} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{43} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ V_{21} & 1 & 0 & 0 \\ V_{31} & V_{32} & 1 & 0 \\ V_{41} & V_{42} & V_{43} & 1 \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix} \begin{bmatrix} 1 & V_{21} & V_{31} & V_{41} \\ 0 & 1 & V_{32} & V_{42} \\ 0 & 0 & 1 & V_{43} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Para resolver de  $d_1$  a  $d_4$  y el  $V_{ij}$  comenzamos con la ecuación (52) para  $i = 1$

dando

$$d_1 = \phi_{11}$$

Utilizando la ecuación (49) para  $i = 2, 3, 4$  resolvemos  $V_{21}, V_{31}$  y  $V_{41}$  como

$$V_{21}d_1 = \phi_{21} \quad , \quad V_{31}d_1 = \phi_{31} \quad , \quad V_{41}d_1 = \phi_{41}$$

$$V_{21} = \phi_{21} / d_1 \quad , \quad V_{31} = \phi_{31} / d_1 \quad , \quad V_{41} = \phi_{41} / d_1$$

Utilizando la ecuación (51) para  $i = 2$  da

$$d_2 = \phi_{22} - V_{21}^2 d_1$$

Utilizando la ecuación (49) para  $i = 3$  y  $4$  da

$$V_{32}d_2 = \phi_{32} - V_{31}d_1V_{21}$$

$$V_{42}d_2 = \phi_{42} - V_{41}d_1V_{21}$$

o

$$V_{32} = (\phi_{32} - V_{31}d_1V_{21}) / d_2$$

$$V_{42} = (\phi_{42} - V_{41}d_1V_{21}) / d_2$$

Ahora se utiliza la ecuación (51) para  $i = 3$  para resolver  $d_3$ , luego también se utiliza la ecuación (49) para  $i = 4$  para resolver  $V_{43}$  y finalmente la ecuación (51) para  $i = 4$  para resolver  $d_4$ .

Una vez determinadas las matrices  $V$  y  $D$  es relativamente simple resolver el vector columna ( en un procedimiento de dos pasos). Se obtiene de las ecuaciones (46) y (47)

$$VDV^t\alpha = \psi \quad (53)$$

ecuación que puede estar escrita como

$$VY = \psi \quad (54)$$

y

$$DV^t\alpha = Y \quad (55)$$

o

$$V^t\alpha = D^{-1}Y \quad (56)$$

Por eso, a partir de la matriz  $V$ , se puede resolver la ecuación (54) para el vector columna  $Y$  utilizando una simple recurrencia de la forma

$$Y_i = \psi_i - \sum_{j=1}^{i-1} V_{ij}Y_j \quad p \geq i \geq 2 \quad (57)$$

con una condición inicial

$$Y_1 = \psi_1 \quad (58)$$

De igual manera habiendo resuelto  $Y$ , la ecuación (56) se puede resolver recurrentemente para  $\alpha$  utilizando la relación

$$\alpha_i = Y_i / d_i - \sum_{j=i+1}^p V_{ji} \alpha_j \quad 1 \leq i \leq p-1 \quad (59)$$

con una condición inicial

$$\alpha_p = Y_p / d_p \quad (60)$$

Se debería tomar en cuenta que el subíndice  $i$  en la ecuación (59) procede al revés desde  $i = p-1$  hasta  $i = 1$ .

Para ilustrar el uso de las ecuaciones (57) - (60) continuamos con nuestro ejemplo anterior y primero resolvemos  $Y_i$ 's asumiendo que  $V$  y  $D$  son conocidos. En la forma de matriz obtenemos la ecuación

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ V_{21} & 1 & 0 & 0 \\ V_{31} & V_{32} & 1 & 0 \\ V_{41} & V_{42} & V_{43} & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{bmatrix}$$

De las ecuaciones (57) y (58) obtenemos

$$\begin{aligned} Y_1 &= \psi_1 \\ Y_2 &= \psi_2 - V_{21}Y_1 \\ Y_3 &= \psi_3 - V_{31}Y_1 - V_{32}Y_2 \\ Y_4 &= \psi_4 - V_{41}Y_1 - V_{42}Y_2 - V_{43}Y_3 \end{aligned}$$

De  $Y_i$ 's, resolvemos la ecuación (56) la cual es de la forma

$$\begin{bmatrix} 1 & V_{21} & V_{31} & V_{41} \\ 0 & 1 & V_{32} & V_{42} \\ 0 & 0 & 1 & V_{43} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = \begin{bmatrix} 1/d_1 & 0 & 0 & 0 \\ 0 & 1/d_2 & 0 & 0 \\ 0 & 0 & 1/d_3 & 0 \\ 0 & 0 & 0 & 1/d_4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} Y_1/d_1 \\ Y_2/d_2 \\ Y_3/d_3 \\ Y_4/d_4 \end{bmatrix}$$

De las ecuaciones (59) y (60) obtenemos

$$\begin{aligned} \alpha_4 &= Y_4 / d_4 \\ \alpha_3 &= Y_3 / d_3 - V_{43}\alpha_4 \\ \alpha_2 &= Y_2 / d_2 - V_{32}\alpha_3 - V_{42}\alpha_4 \\ \alpha_1 &= Y_1 / d_1 - V_{21}\alpha_2 - V_{31}\alpha_3 - V_{41}\alpha_4 \end{aligned}$$

así completamos la solución de las ecuaciones de covarianza.

El uso del procedimiento de descomposición de Cholesky nos lleva a una expresión muy simple del mínimo error del método de covarianza en términos del vector columna  $Y$  y de la matriz  $D$ . Recordemos que para el método de covarianza, el error de predicción  $E_n$  fue de forma.

$$E_n = \phi_n(0,0) - \sum_{k=1}^p \alpha_k \phi_n(0,k) \quad (61)$$

o en anotación matriz

$$E_n = \phi_n(0,0) - \alpha^t \psi \quad (62)$$

De la ecuación (56) podemos sustituir la expresión  $\alpha^t$  por  $Y^t D^{-1} V^{-1}$  dando

$$E_n = \phi_n(0,0) - Y^t D^{-1} V^{-1} \psi \quad (63)$$

Utilizando la ecuación (54) obtenemos

$$E_n = \phi_n(0,0) - Y^t D^{-1} Y \quad (64)$$

o

$$E_n = \phi_n(0,0) - \sum_{k=1}^p Y_k^2 / d_k \quad (65)$$

De esta manera el error de predicción del cuadrático-medio  $E_n$  puede determinarse directamente del vector en columna  $Y$  y de la matriz  $D$ . Además, la ecuación (65) puede utilizarse para conseguir el valor de  $E_n$  para cualquier valor de  $p$  hasta el valor de  $p$  utilizado para resolver las ecuaciones de matrices. Así uno puede tener la idea de cómo el error de predicción cuadrático-medio varía con el número de coeficientes de predicción utilizados en la solución.

### **2.3 La solución recurrente de Durbin para las ecuaciones de auto-correlación**

Para el método de auto-correlación la matriz de la ecuación para resolver los coeficientes de predicción es de la forma.

$$\sum_{k=1}^p \alpha_k R_n(|i-k|) = R_n(i) \quad 1 \leq i \leq p \quad (66)$$

Al explotar la naturaleza Toeplitz de los coeficientes de la matriz, se han inventado varios procedimientos recurrentes y eficientes para resolver este sistema de ecuaciones. A pesar que los más populares y conocidos de estos métodos son los algoritmos de Levinson y Robinson, el método más conocido y eficiente para resolver este particular sistema de ecuaciones, es el procedimiento recurrente de Durbin que puede establecerse como sigue (por conveniencia de anotación, nosotros omitiremos el subíndice en la función de auto-correlación):

$$E^{(0)} = R(0) \quad (67)$$

$$k_i = \left( R(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} R(i-j) \right) / E^{(i-1)} \quad 1 \leq i \leq p \quad (68)$$

$$\alpha_1^{(i)} = k_i \quad (69)$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{1-j}^{(i-1)} \quad 1 \leq j \leq i-1 \quad (70)$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad (71)$$

Las ecuaciones (68) - (71) se resuelven recurrentemente para  $i = 1, 2, \dots, p$  y la solución final es dada como

$$\alpha_j = \alpha_j^{(p)} \quad 1 \leq j \leq p \quad (72)$$

Tome nota que en el proceso de solución de los coeficientes extrapolados del orden de predicción  $p$ , las soluciones para los coeficientes de predicción de todas las ordenes menores que  $p$ , también se han obtenido – es decir,  $\alpha_j^{(i)}$  es el coeficiente de predicción  $j^{\text{th}}$  para una predicción de orden  $i$ .

Para ilustrar el procedimiento arriba mencionado, se considera un ejemplo al obtener los coeficientes extrapolados para un extrapolado de orden 2. La ecuación original de la matriz es de la forma

$$\begin{bmatrix} R(0) & R(1) \\ R(1) & R(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \end{bmatrix}$$

Utilizando las ecuaciones (67) - (72), obtenemos

$$\begin{aligned}
E^{(0)} &= R(0) \\
k_1 &= R(1)/R(0) \\
\alpha_1^{(1)} &= R(1)/R(0) \\
E^{(1)} &= \frac{R^2(0) - R^2(1)}{R(0)} \\
k_2 &= \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)} \\
\alpha_2^{(2)} &= \frac{R(2)R(0) - R^2(1)}{R^2(0) - R^2(1)} \\
\alpha_1^{(2)} &= \frac{R(1)R(0) - R(1)R(2)}{R^2(0) - R^2(1)} \\
\alpha_1 &= \alpha_1^{(2)} \\
\alpha_2 &= \alpha_2^{(2)}
\end{aligned}$$

Se debería tomar en cuenta que la cantidad  $E^{(i)}$  en la ecuación (71) es el error de predicción para una predicción de orden  $i$ . Por eso, se puede determinar, en cada etapa del cálculo del error de predicción, una predicción de orden  $i$ . Además, si los coeficientes de auto-correlación  $R(i)$  son reemplazados por un conjunto de coeficientes normales de auto-correlación, es decir,  $r(k) = R(k)/R(0)$ , luego la solución de la matriz de la ecuación se mantiene inalterable. Sin embargo, el error  $E^{(i)}$  ahora es interpretado como un error normalizado. Si llamamos a esto error normalizado  $V^{(i)}$ , entonces

$$V^{(i)} = \frac{E^{(i)}}{R(0)} = 1 - \sum_{k=1}^i \alpha_k r(k) \quad (73)$$

con

$$0 \leq V^{(i)} \leq 1 \quad i \geq 0 \quad (74)$$

Se puede demostrar que el error normalizado para  $i = p$  (es decir  $V^{(p)}$ ) puede escribirse en la forma

$$V^{(p)} = \prod_{i=1}^p (1 - k_i^2) \quad (75)$$

donde las cantidades  $k_i$ , están en el rango

$$-1 \leq k_i \leq 1 \quad (76)$$

Esta condición es importante en los parámetros  $k_i$ , puesto que se puede demostrar que es una condición necesaria y suficiente para todas las raíces del polinomio  $A(z)$  para estar dentro de la unidad del círculo, por ende garantizando la estabilidad del sistema  $H(z)$ . Desafortunadamente, una prueba de este resultado nos llevaría bastante lejos; sin embargo, es posible mostrar que no hay tal garantía de estabilidad disponible en el método de la covarianza.

## 2.4 Formulaciones y soluciones de Lattice

Como ya hemos podido apreciar, ambos métodos de covarianza y auto-correlación conllevan dos pasos:

1. Cálculo de una matriz de los valores de auto-correlación.
2. Solución de un conjunto de ecuaciones lineales.

Estos métodos han sido ampliamente utilizados y con gran éxito en las aplicaciones del procesamiento de la palabra. Sin embargo, otros tipos de métodos, llamados métodos Lattice, ha evolucionado donde los dos pasos anteriores se han combinado en cierta manera en un algoritmo recurrente para determinar los parámetros de predicción linealmente. Para apreciar cómo estos métodos están relacionados, es de gran ayuda empezar con el algoritmo de Durbin. Primero, permítanos recordar que en la etapa  $i^{\text{th}}$  de este procedimiento, el conjunto de coeficientes  $\{\alpha_j^{(i)} \ j = 1, 2, \dots, i\}$  son los coeficientes de predicción linealmente de orden  $i^{\text{th}}$  óptimo. Utilizando estos coeficientes podemos definir

$$A^{(i)}(z) = 1 - \sum_{k=1}^i \alpha_k^{(i)} z^{-k} \quad (77)$$

para ser la función del sistema de orden  $i^{\text{th}}$  de filtro inverso (o filtro de error de predicción). Si la entrada a este filtro es el segmento de la señal,

$s_n(m) = s(n+m)w(m)$ , entonces la salida sería el error de predicción,  
 $e_n^{(i)}(m) = e^{(i)}(n+m)$ , donde

$$e^{(i)}(m) = s(m) - \sum_{k=1}^i \alpha_k^{(i)} s(m-k) \quad (78)$$

Tome en cuenta que por simplicidad debemos de aquí en adelante dejar de lado el subscrito  $n$  que indica el hecho que estamos considerando un segmento de la señal localizada en la muestra  $n$ . En términos de transformada  $z$ , la ecuación (78) es

$$E^{(i)}(z) = A^{(i)}(z)S(z) \quad (79)$$

Sustituyendo la ecuación (70) por la (77) obtenemos una fórmula de recurrencia para  $A^{(i)}(z)$  en términos de  $A^{(i-1)}(z)$ , por ejemplo,

$$A^{(i)}(z) = A^{(i-1)}(z) - k_i z^{-i} A^{(i-1)}(z^{-1}) \quad (80)$$

Sustituyendo la ecuación (80) en la (79) obtenemos

$$E^{(i)}(z) = A^{(i-1)}(z)S(z) - k_i z^{-i} A^{(i-1)}(z^{-1})S(z) \quad (81)$$

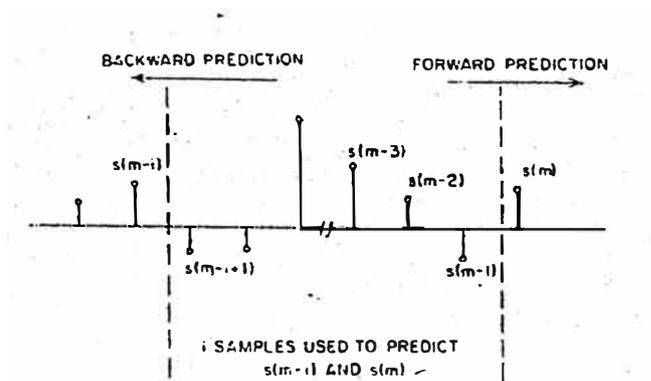


Fig. 2 Ilustración de la predicción delantera y posterior utilizando un orden  $i^{th}$  de predicción

El primer término en la ecuación (81) es obviamente la transformada  $z$  del error de predicción para un orden  $(i-1)^{th}$ . Al segundo término se le puede otorgar una interpretación similar si definimos

$$B^{(i)}(z) = z^{-i} A^{(i)}(z^{-1}) S(z) \quad (82)$$

Se muestra fácilmente que la transformada de  $B^{(i)}(z)$  es

$$b^{(i)}(m) = s(m-i) - \sum_{k=1}^i \alpha_k^{(i)} s(m+k-i) \quad (83)$$

Esta ecuación sugiere que estamos intentando predecir  $s(m-i)$  de las muestras  $i$  de entrada  $\{s(m-j+k) \mid k=1,2,\dots,i\}$  continuando con  $s(m-i)$ . De esta manera  $b^{(i)}(m)$  se llama la secuencia de error de predicción posterior. En la figura 2 se observa que las muestras  $i$  involucradas en la predicción son las mismas utilizadas para predecir  $s(m)$  en términos de muestras  $i$  pasadas en la ecuación (78). Ahora regresando a la ecuación (81) apreciamos que la secuencia del error de predicción  $e^{(i)}(m)$  puede ser expresada como

$$e^{(i)}(m) = e^{(i-1)}(m) - k_i b^{(i-1)}(m) \quad (84)$$

Al sustituir la ecuación (81) por la (82) obtenemos

$$B^{(i)}(z) = z^{-1} A^{(i-1)}(z^{-1}) S(z) - k_i A^{(i-1)}(z) S(z) \quad (85)$$

o

$$B^{(i)}(z) = z^{-1} B^{(i-1)}(z) - k_i E^{(i-1)}(m) \quad (86)$$

Así el escenario del error de predicción posterior  $i^{\text{th}}$  es

$$b^{(i)}(m) = b^{(i-1)}(m-1) - k_i e^{(i-1)}(m) \quad (87)$$

Ahora las ecuaciones (84) y (87) definen las secuencias del error de predicción delantero y posterior para un extrapolado de orden  $i^{\text{th}}$  en términos de los errores de predicción correspondientes para los coeficientes de predicción de orden  $(i-1)^{\text{th}}$ . Utilizando una predicción de orden cero es similar al uso de la no predicción del todo, así que

$$e^{(0)}(m) = b^{(0)}(m) = s(m) \quad (88)$$

Así podemos delinear las ecuaciones (84) y (87) por el gráfico de circulación de la figura 3. Tal estructura se llama una red Lattice. Es claro si ampliamos la Lattice a las secciones  $p$  la salida de la rama superior será el error de predicción posterior como lo demuestra la figura 3. Así, dicha figura es una implementación de red digital del filtro de error de predicción con una función de transferencia  $A(z)$ .

A esta altura debemos enfatizar que esta estructura es una consecuencia directa del algoritmo Durbin, y los parámetros  $k_i$ , se pueden obtener de las ecuaciones (67) – (72). Observe también que el coeficiente de predicción no

aparece explícitamente en la figura 3. Itakura ha mostrado que los parámetros  $k_i$  pueden estar directamente relacionados con los errores de predicción delanteros y posteriores y debido a la naturaleza de la estructura Lattice todo el conjunto de coeficientes  $\{k_i, i = 1, 2, \dots, p\}$  se pueden calcular sin calcular el coeficiente de predicción. La relación es

$$k_i = \frac{\sum_{m=0}^{N-1} e^{(i-1)}(m)b^{(i-1)}(m-1)}{\left\{ \sum_{m=0}^{N-1} (e^{(i-1)}(m))^2 \sum_{m=0}^{N-1} (b^{(i-1)}(m-1))^2 \right\}^{1/2}} \quad (89)$$

Esta expresión se presenta en la forma de una función normal de correlación transversal; es decir, es indicativo del grado de correlación entre los errores de predicción delanteros y posteriores. Por esta razón los parámetros  $k_i$  se llaman coeficientes de correlación parciales o coeficientes PARCOR (Partial correlation coefficients). Es relativamente claro verificar que la ecuación (89) es idéntica a la ecuación (68) al sustituir las Ecuaciones las ecuaciones (78) y (83) por la ecuación (89).

Se puede apreciar que si la ecuación (89) es reemplazada por la ecuación (68) en el algoritmo Durbin, el coeficiente de predicción puede ser calculado recurrentemente como antes. Así el análisis PARCOR nos lleva hacia una alternativa de la inversión de una matriz y da resultados idénticos al método de auto-correlación; por ejemplo, el conjunto de coeficientes es equivalente al

conjunto de coeficientes de predicción que minimiza el error de predicción posterior de cuadrático-medio. Algo más importante es que este acercamiento abre todo un tipo de procedimientos basados en la configuración Lattice de la figura 3

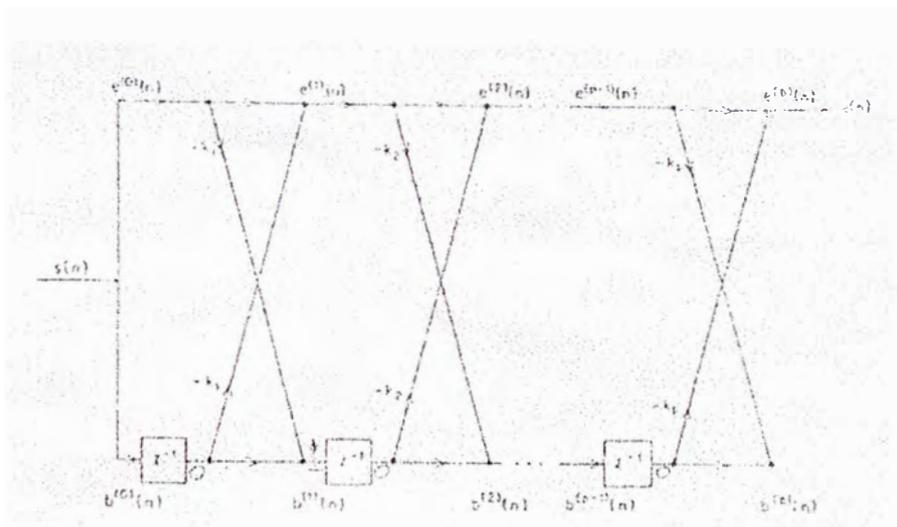


Fig. 3 Diagrama de bloques de la implementación por el método Lattice

En particular, Burg ha desarrollado un procedimiento basado en minimizar la suma de los errores de predicción delantera y posterior cuadrático-medio en la figura 3, es decir,

$$\bar{E}^{(i)} = \sum_{m=0}^{N-1} [(e^{(i)}(m))^2 + (b^{(i)}(m))^2] \quad (90)$$

Sustituyendo las ecuaciones (84) y (87) en la ecuación (90) y diferenciando  $E^{(i)}$

con respecto a  $k_i$ , obtenemos

$$\begin{aligned} \frac{\partial E^{(i)}}{\partial k_i} &= -2 \sum_{m=0}^{N-1} [e^{(i-1)}(m) - k_i b^{(i-1)}(m-1)] b^{(i-1)}(m-1) \\ &= -2 \sum_{m=0}^{N-1} [b^{(i-1)}(m-1) - k_i e^{(i-1)}(m)] e^{(i-1)}(m) \end{aligned} \quad (91)$$

Estableciendo la derivada igual a cero y resolviendo  $k_i$  obtenemos

$$k_i = \frac{2 \sum_{m=0}^{N-1} [e^{(i-1)}(m) b^{(i-1)}(m-1)]}{\sum_{m=0}^{N-1} [e^{(i-1)}(m)]^2 + \sum_{m=0}^{N-1} [b^{(i-1)}(m-1)]^2} \quad (92)$$

Se puede demostrar que si  $k_i$  es estimado utilizando la ecuación (92) entonces

$$-1 \leq k_i \leq 1 \quad (93)$$

Sin embargo, debería estar claro que los estimados  $k_i$ 's, utilizando la ecuación (92), en general, va a diferir de aquellos estimados utilizando la ecuación (89) o equivalente, del método de auto-correlación.

En resumen, los pasos involucrados para determinar los coeficientes de predicción y los parámetros  $k$  son los siguientes:

1. Primeramente establecer  $e^{(0)}(m) = s(m) = b^{(0)}(m)$ .
2. Calcular  $k_1 = \alpha_1^{(1)}$  de la ecuación (92).
3. Determinar los errores de predicción delanteros y posteriores  $e^{(1)}(m)$  y  $b^{(1)}(m)$  de las ecuaciones (84) y (87).
4. Establecer  $i = 2$ .
5. Determinar  $k_i = \alpha_i^{(i)}$  de la ecuación (92).
6. Determinar  $\alpha_j^{(i)}$  para  $j = 1, 2, \dots, i-1$  de la ecuación (70).
7. Determinar  $e^{(i)}(m)$  y  $b^{(i)}(m)$  de las ecuaciones (84) y (87).
8. Establecer  $i = i + 1$ .
9. Si  $i$  es menor o igual que  $p$ , ir al paso 5.
10. El procedimiento ha finalizado.

Claramente se puede apreciar varias diferencias en la aplicación entre el método Lattice y las aplicaciones de covarianza y auto-correlación ya discutidos. Una diferencia mayor es que en el método Lattice los coeficientes extrapolados se obtienen directamente de las muestras de la palabra sin un cálculo intermedio de una función de auto-correlación. Al mismo tiempo el método es garantizado para obtener un filtro estable sin tener necesidad de utilizar una ventana. Por estas razones la formulación Lattice se ha convertido en un aporte importante y viable para la aplicación del análisis de predicción lineal.

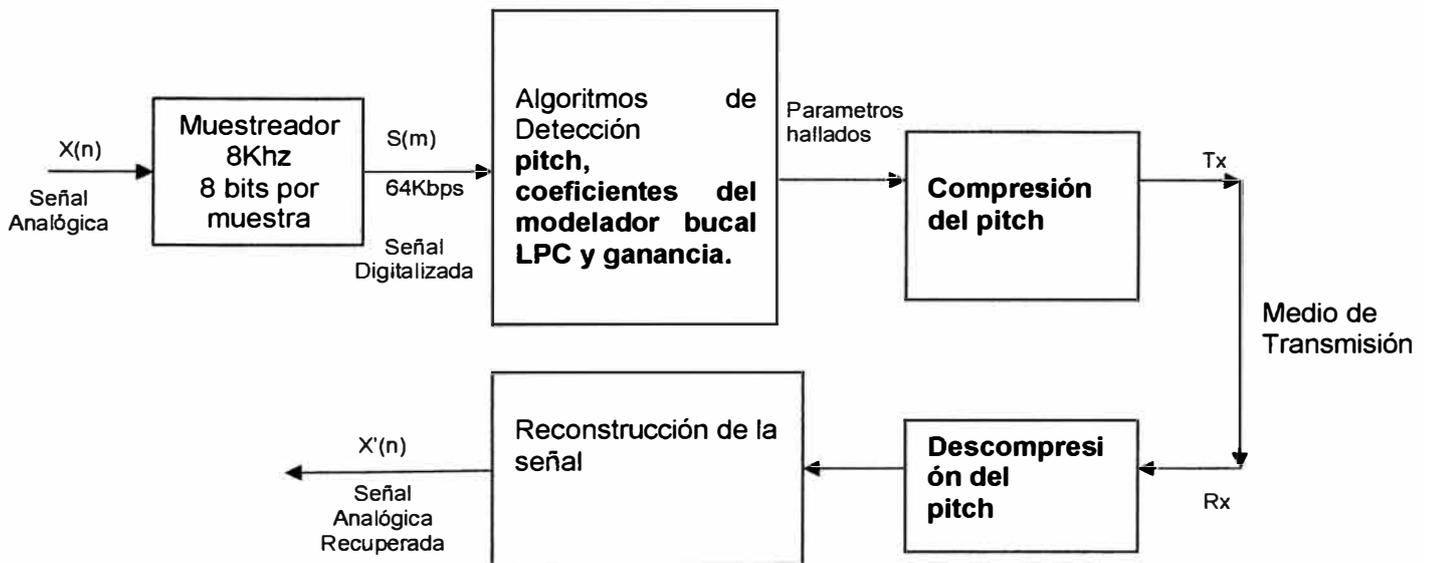
## NOTAS

1. Es claro que  $\alpha_k$ 's son funciones de  $n$  (tiempo index, por el cual son calculados) a pesar que esta dependencia no será mostrada. Además, debemos encontrar ventajoso, el no mencionar los subscritos  $n$  en  $E_n, S_n(m)$  y  $\phi_n(i, k)$ , sin confusión alguna.
2. Esta terminología, es algo confusa, ya que la palabra covarianza, por lo general se refiere a la correlación de una señal con su media trasladada.
3. Notar que la ganancia también es una función de tiempo.
4. Notar que para este aporte sea válido, requiere que el intervalo del análisis tenga la misma amplitud que un período pitch.

## CAPITULO III

### ALGORITMO PARA HALLAR EL PERIODO PITCH Y LOS COEFICIENTES LPC UTILIZANDO MATHLAB

#### 3.1 Diagrama de bloques del codificador LPC:



#### 3.2 Partes del programa:

Los siguientes 13 pasos, explicados a continuación, son los pasos generales más importantes del programa. Los cuales se explicarán en forma detallada.

### 3.2.1 Ingreso de datos generales.

En este primer paso se define los valores deseados de para el muestreo de la señal de voz. Estos valores son los siguientes:

**duración:** que es el tiempo de duración de grabado deseado.

**bpm:** nos indican la cuantización que se desea tener en la señal original, es decir la cantidad de bits por muestra que pueden ser 8 bpm o 16 bpm.

**fmuestreo:** que es la frecuencia de muestreo que puede ser 8000Hz, 11025Hz, 22050Hz ó 44100Hz.

Estos tres parámetros nos determinaran:

La cantidad de muestras que se tomaran como parte de la señal original:

**duracion\*fmuestreo.**

Tasa de velocidad de voz: **fmuestreo\*bpm** que nos da la cantidad de bits por segundo.

Con estos parámetros establecidos por el usuario final pasamos a la siguiente fase.

### 3.2.2 Grabado de voz.

Mediante la subrutina realizada en el programa se accede a la tarjeta de sonido indicándole los parámetros antes mencionados que son necesarios para la adquisición de datos.

### 3.2.3 Separación mediante ventanas rectangulares.

Una vez obtenidos todos los datos, estos se proceden a separar por tramas para ser analizados. Para dicha separación se utiliza una ventana rectangular cuya. El único factor que utilizamos para la construcción de la ventana es el tiempo ingresado por el usuario. Este tiempo es de 15ms pero puede ser variado. Este tiempo es en otros términos el número de muestras analizadas por ventana.

### **3.2.4 Obtención de los coeficientes LPC.**

Una vez hecha la separación de los datos se procede a analizar cada trama. A partir del subprograma creado se obtienen los coeficientes LPC pertenecientes a ese tramo de muestras. La técnica recursiva de Durvin es usada para el programa de obtención de los coeficientes LPC

Se definen por default la obtención de 12 coeficientes LPC, y como veremos la cantidad de coeficientes influyen en la respuesta del filtro. Después cada coeficiente LPC es quantizado con 10 bits, los cuales son suficientes para obtener un mínimo error de cuantización.

### **3.2.5 Obtención del pitch.**

En este paso se procede a calcular el pitch de la trama analizada. La técnica utilizada es por medio de la obtención de la primera derivada para saber las concavidades y con la segunda derivada se obtiene las inflexiones. Este paso es muy importante ya que es el pitch el que nos permitirá el éxito del codificador. El pitch es un tren de impulsos es decir una secuencia de unos y ceros que necesita el filtro LPC reconstructor.

### **3.2.6 Obtención de la ganancia.**

La ganancia juega un papel muy importante en el análisis ya que nos determina la intensidad de voz del hablante. Se determina el promedio del valor absoluto de las muestras por número de muestras que se está analizando.

### **3.2.7 Compresión del pitch.**

La técnica utilizada para la compresión depende del tamaño de muestras de la trama analizada, como solo es necesario saber la posición en la parte de la trama donde van los unos, entonces esa es la información que determinamos. De acuerdo a nuestro análisis en 15 milisegundos se tiene 120 muestras,

entonces con  $2^7 = 128$  posiciones se puede determinar en que posición de la trama están los unos.

### **3.2.8 Envío de datos.**

En este punto definimos la tasa de compresión. Los bits que enviamos por trama analizada son los siguientes:

12 coeficientes LPC cuantizados con 10 bits cada uno nos dan 120bits

La ganancia cuantizada con 15 bits

Y el tren de impulsos comprimido que son aproximadamente 49bits.

Es decir por cada trama analizada que tiene un total de 960bits solo transmitimos 184bits (120+15+49).

Lo que nos da un factor de compresión de: **5.2174** = (960/184). Que equivale a una tasa de transmisión de **12Kbps**.

Cabe señalar que estos cálculos efectuados son para una tasa de 64Kbps (8000Hz y 8 bits por muestra).

Finalmente los datos son enviados al medio de transmisión y dependiendo del sistema se utilizarán diferentes técnicas de modulación y encriptación.

### **3.2.9 Recepción de datos.**

Se reciben los datos del medio de transmisión y son analizados, cada trama es enviada a su respectivo descompresor o decuantizador.

### **3.2.10 Descompresión del pitch.**

El descompresor debe tener las mismas características del compresor vistas en el anexo 5.

### **3.2.11 Recuperación de la señal original.**

Una vez descomprimido el pitch, decuantizada la ganancia y decuantizados los coeficientes LPC se procede a la reconstrucción de la voz utilizando los coeficientes del filtro LPC.

### **3.2.12 Filtro pasabajo.**

La finalidad del filtro pasabajo es eliminar el ruido "clic" introducido por la separación del total de muestras por la ventana rectangular. Este es un filtro FIR de 42 coeficientes. Y sus características están definidas en el anexo 6.

### **3.2.13 Reproducción de la señal recuperada.**

El reproductor es seteado con los mismos valores ingresados por el usuario al inicio del programa. Es decir reproduce a la misma tasa de velocidad que la señal original.

Pruebas de Calidad de Voz y performance del sistema:

Para definir las pruebas de calidad de voz recibida tenemos varios factores, como son:

1. Frecuencia de muestreo (fmuestreo).
2. Bits por muestras (bpm).
3. Ventana utilizada para la separación de la señal original.
4. Cantidad de coeficientes LPC utilizados.
5. Filtros utilizados para la recuperación y mejoramiento de la señal.

Después de implementar el sistema y haber escuchado la señal reconstruida notamos que ésta sale robotizada.

### 3.3 Programa en Matlab para calcular: el pitch, coeficientes LPC, la energía de una señal vocal y su recuperación.

```

%CODEC LPC
close all;
clear all;
clc;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%
%%% Adquisición de datos de voz

duracion=input('Ingrese la duración de grabado : ');
bpm=input('Ingrese los bits por muestra del grabado (8 16) : ');
fmuestreo=8000;%input('Ingrese la frecuencia de muestreo (8000, 11025,
22050, 44100) : ');

ai=analoginput('winsound');
chan=addchannel(ai,1);
set(ai,'SampleRate',fmuestreo);
set(ai,'BitsPerSample',bpm);
ActualRate=get(ai,'SampleRate');
set(ai,'SamplesPerTrigger',duracion*ActualRate);
set(ai,'TriggerType','Manual');
disp('Inicio')
pause;
start(ai);
trigger(ai);
data=getdata(ai);
disp('Final')
delete(ai);
clear ai;

%%%%%salida
ao=analogoutput('winsound');
chan=addchannel(ao,1);
set(ao,'SampleRate',ActualRate);
set(ao,'BitsPerSample',bpm);
set(ao,'TriggerType','manual');
set(ao,'RepeatOutput',0)
putdata(ao,data*2);
pause;
start(ao);
trigger(ao);

figure(2)
subplot(2,1,1)
plot(data);
title('Señal de entrada');

%%%definimos un vector salidatotal

```

```

salidatotal=[];
%separacion de la data en bloques de 15 mseg.

ldata=length(data);
dt=1/fmuestreo;
qms=floor(0.015*fmuestreo);
nqms=round(ldata/qms);
if ldata<nqms*qms
    ndata=[data;zeros(nqms*qms-ldata,1)];
else
    ndata=[data];
end

media=[zeros(nqms,1)];
figure(1)
sdata=ndata(1:qms);
for N=1:(nqms-1)
    subplot(5,1,1);
    plot(sdata);
    title('Señal original');
    grid
    %%Scamos la media del Conjunto de Muestras (CM)
    media(N)=median(sdata);

    %%creamos el filtro LPC para el CM
    coelpc=olpcc(sdata,12);
    coelpc

    %Calculamos el pitch respectivo para cada CM
    autocorrelacion=xcorr(sdata);
    aucol=autocorrelacion(floor(qms/2):floor(qms/2)+qms+1);
    eraderiv=diff(aucol); %primera derivada para
localizar los puntos de inflexion %segunda derivada para localizar la
segderiv=diff(eraderiv);
concavidad de la señal
eraderiv1=eraderiv;
signo=sign(eraderiv);
for nn=1:qms
    if signo(nn)==signo(nn+1)
        eraderiv1(nn)=0;
    else
        if sign(segderiv(nn))<0
            eraderiv1(nn)=1;
        else
            eraderiv1(nn)=0;
        end
    end
end
eraderiv1(qms+2)=0;%%%pitch puntos maximos

%%%%%%%%%%
%Para comprimir el periodo pitch

```

```

contador=1;
envio=[];
for comp=1:qms
    if eraderiv1(comp)==1
        envio(contador)=comp;
        contador=contador+1;
    end
end

subplot(5,1,2);
plot(eraderiv);
subplot(5,1,3);
plot(segderiv);
subplot(5,1,4);
plot(eraderiv1);

%%Sacamos la potencia media
abs1=abs(sdata);
mediana1=median(abs1);
mediana1
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%recuperamos la trama de envio
repcion=zeros(qms,1);
for comp=1:length(envio)
    recepcion(envio(comp))=1;
end

%%Se transmiten los siguientes parametros: coeficientes LPC,
Potencia y Pitch
%%Se genera un filtro con los coeficientes creados
filtrol=filter(mediana1,coelpc,sdata);%%eraderiv1);

%se genera un filtro pasabajas
salida1=filtra(filtrol) '*1.5;

salidatotal=[salidatotal;salida1];
subplot(5,1,5);
plot(salida1);

%tomamos la siguiente muestra de la señal....
sdata=ndata(qms*N+1:qms*(N+1));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
pause;
end;
figure(2)
subplot(2,1,2);
plot(salidatotal);
disp('Escuchar señal reconstruida');
%%%%salida
ao=analogoutput('winsound');
chan=addchannel(ao,1);

```

```

set(ao, 'SampleRate', ActualRate);
set(ao, 'BitsPerSample', bpm);
set(ao, 'TriggerType', 'manual');
set(ao, 'RepeatOutput', 1);
putdata(ao, salidatotal*2);
pause;
start(ao);
trigger(ao);

```

### Subrutina para calcular los coeficientes del filtro LPC

```

function [alfas]=OLPC1(x,ncoef)
RRRR=autocore(x');
EE=[];
KK=[];
dlpsm=0;
alpha=[];
RR=RRRR(2:length(RRRR));
ate=RRRR(1);
KK(1)=RR(1)/ate;
alpha(1,1)=KK(1);
EE(1)=(1-(KK(1))^2)*ate;

for i=2:ncoef
    dlpsm=0;
    for yyy=1:i-1
        dlpsm=dlpsm+[alpha(i-1,yyy)*RR(i-yyy)];
    end
    KK(i)=(RR(i)-dlpsm)/EE(i-1);
    alpha(i,i)=KK(i);
    for yyy=1:i-1
        alpha(i,yyy)=alpha(i-1,yyy)-KK(i)*alpha(i-1,i-yyy);
    end
    alpha;
    EE(i)=(1-(KK(i))^2)*EE(i-1);
end
alfas=alpha(ncoef,:);
alfas=[1 -alfas];

```

### Subrutina para hallar la autocorrelación

```

function [ww]= autoco(x);
k=length(x);
a=0;
for i=1:k
    ww(i)=x(1+a:k)*x(1:k-a)';
    a=a+1;
end

```

### Subrutina del filtro basabajos para recuperar la señal de voz

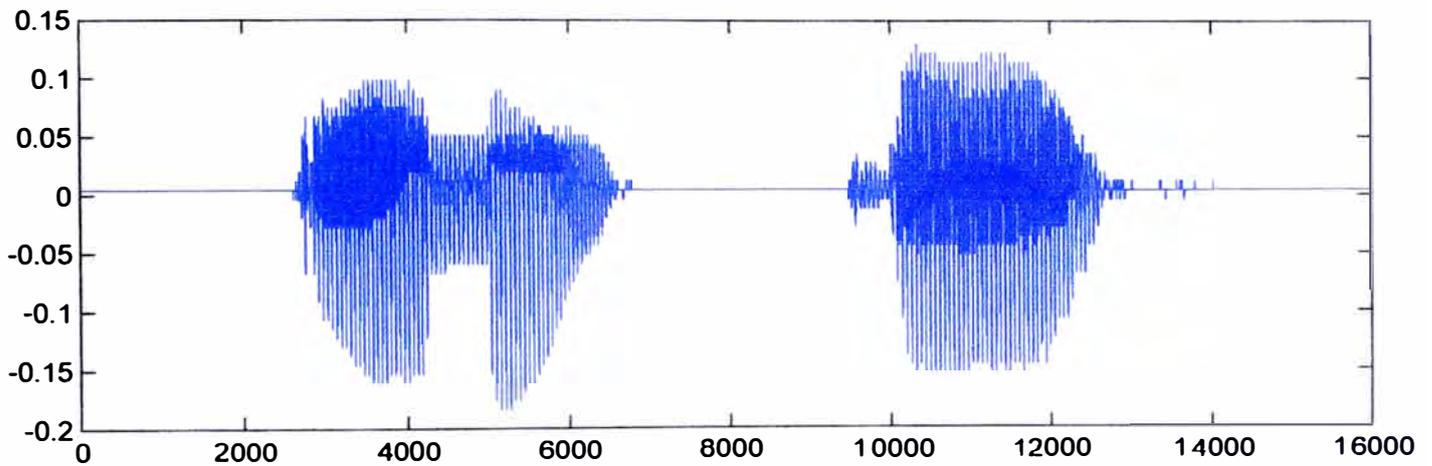
```
function [RR]=filtra(us)

uss=[us us us];
nus=length(us);
NN=41;
FFF=fir1(41,0.3);
A3=2*conv(FFF,uss);
RR=A3([(nus+floor(NN/2)+4):(2*nus+floor(NN/2)+3)]);
%RR=A3([(nus+NN):(2*nus+NN-1)]);
```

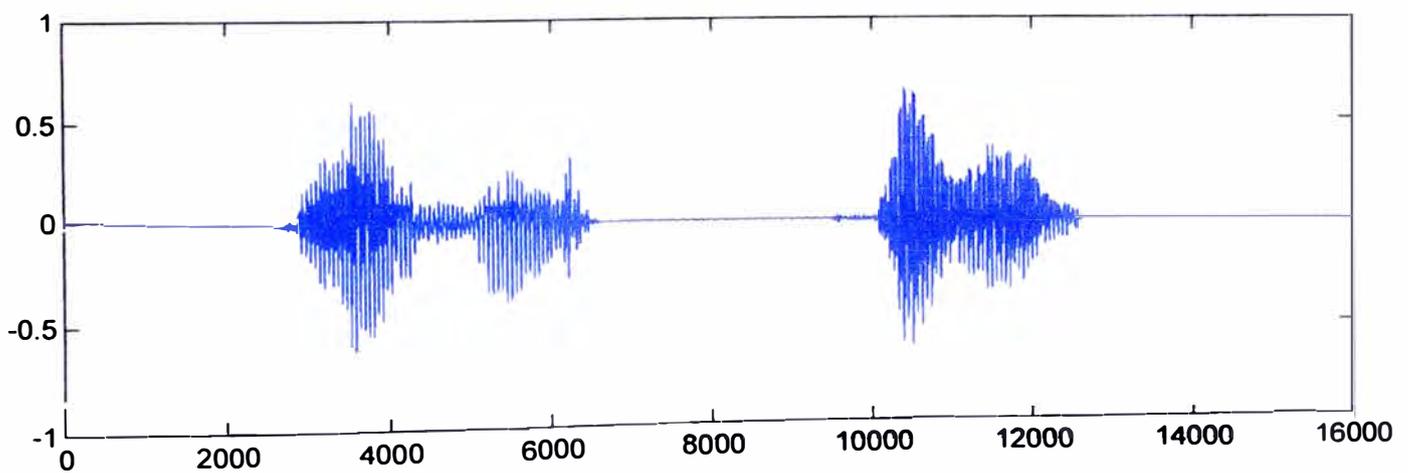
SE MUESTRAN LAS FIGURAS CON LAS DISTINTAS ETAPAS DE LA COMPRESIÓN DE VOZ, DESPUES DE CORRER EL ALGORITMO DE LA TÉCNICA LPC.

La frase es: "Uno Dos"

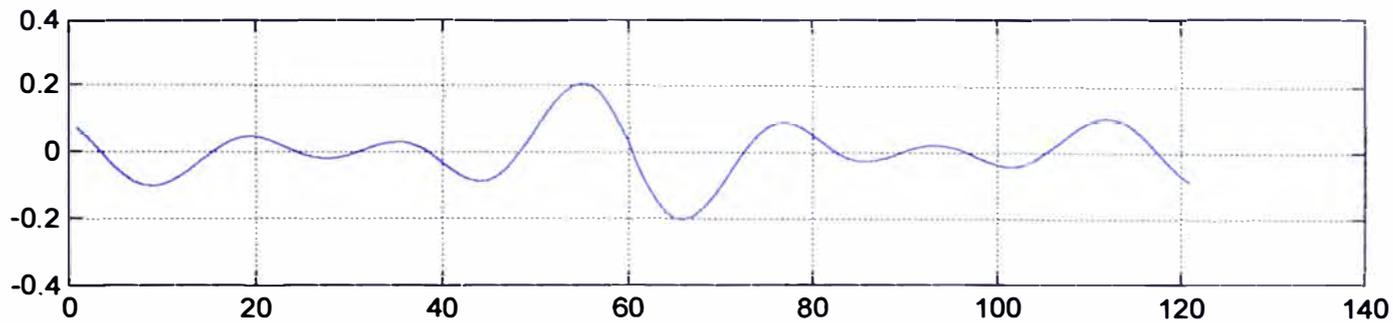
Señal de Entrada



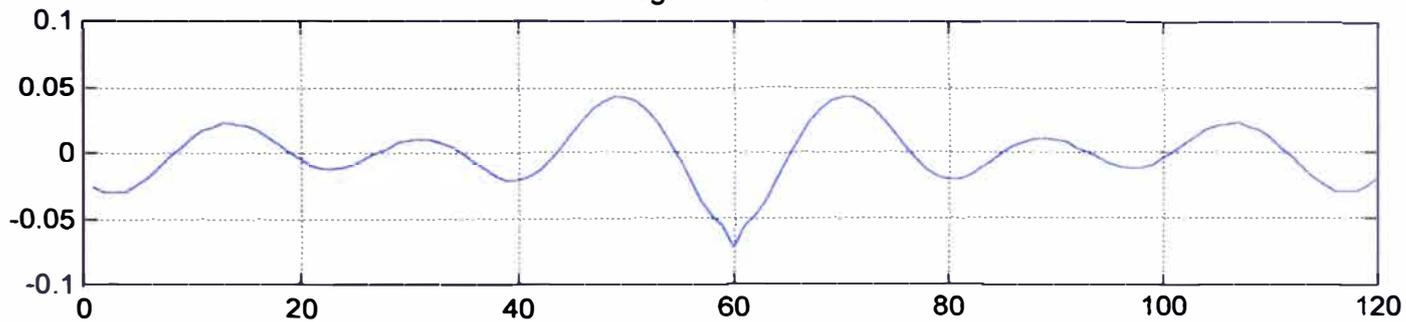
Señal de Salida



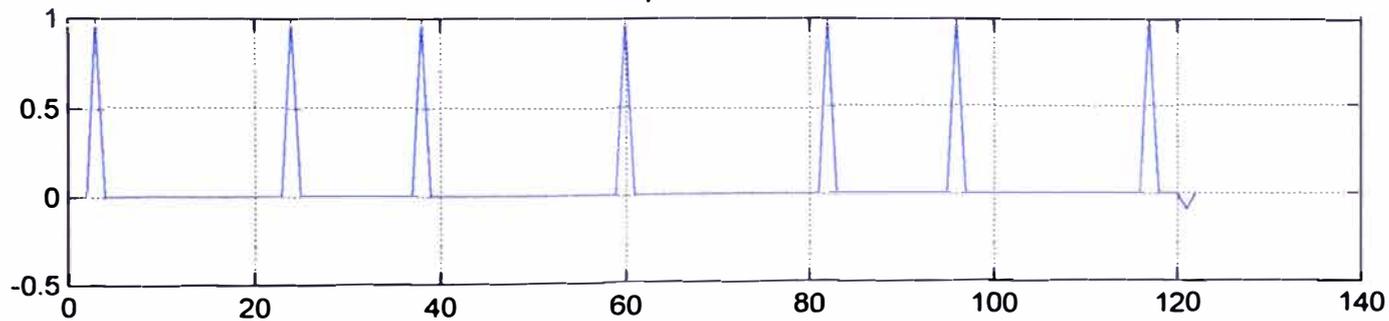
Primera Derivada



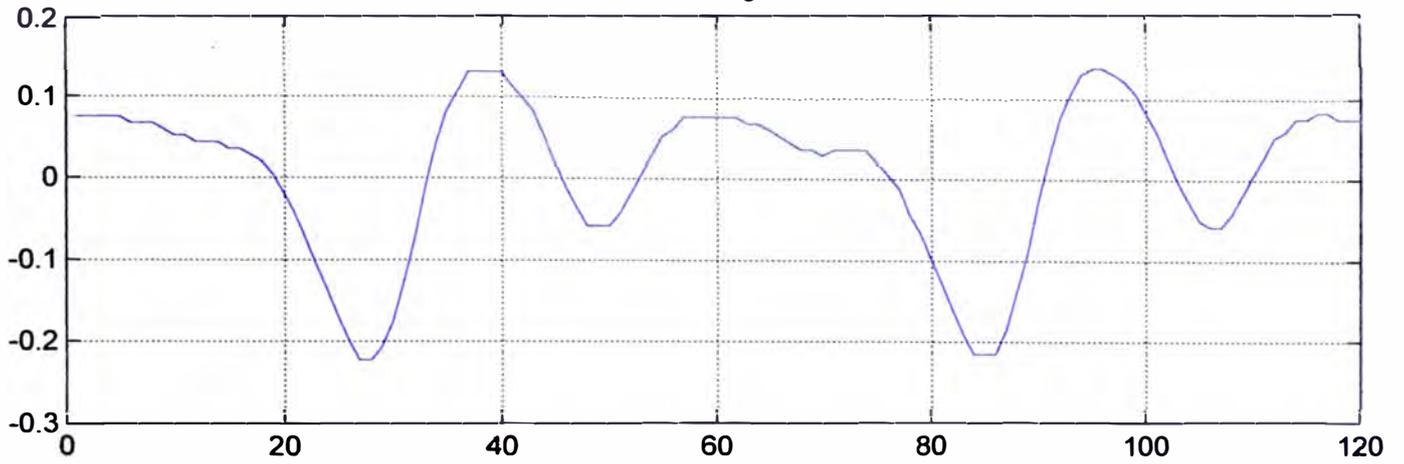
Segunda Derivada



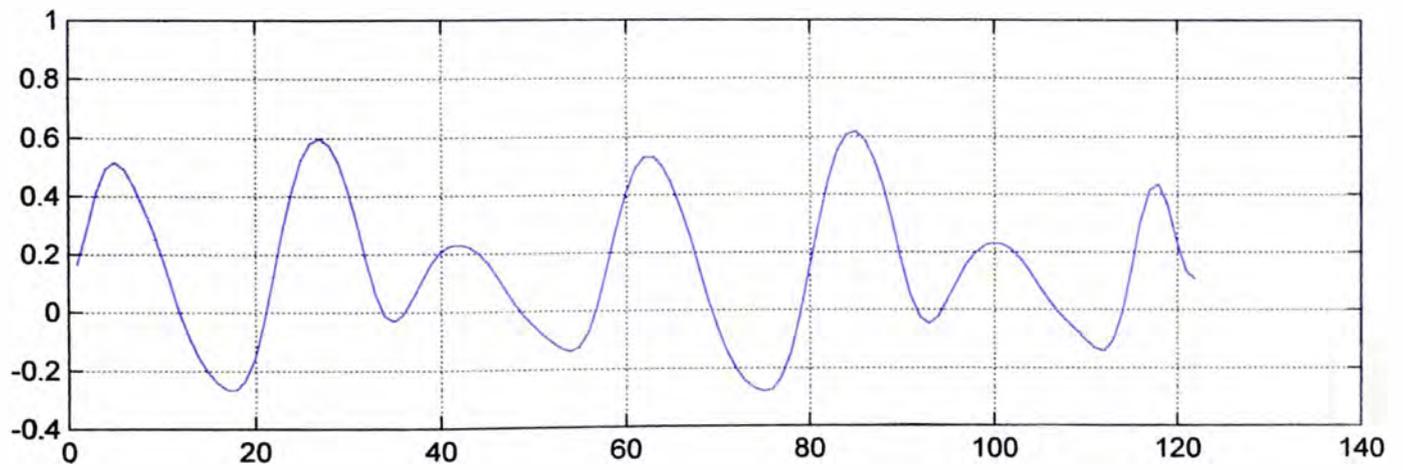
Tren de Impulsos o Periodo Pitch



Señal original



Señal Reconstruida



## CONCLUSIONES

Después de realizado este trabajo se ha llegado a las siguientes conclusiones:

1. Se comprobó que bajo la técnica de la autocorrelación es posible determinar los coeficientes del LPC.
2. Es muy difícil determinar el periodo (pitch), a pesar de que se tuvo en cuenta tres pasos para determinarlo (cálculo de energía, cruce por cero y la autocorrelación) se necesita mayor información al respecto.
3. Se ha podido demostrar que a mayor número de tramas se comprime más la señal.
4. Experimentalmente se ha visto que el número muestras por trama ideal es de 256, en nuestro caso.
5. Hay muchas técnicas de compresión de voz, pero la más utilizada es la del LPC.
6. Actualmente se han hecho modificaciones a esta técnica lográndose una muy buena recuperación de voz con una alta compresión.

7. Una continuación del desarrollo de este trabajo sería mejorar la forma de recuperar la voz, y en base a esto, enviarlo por internet para saber como se comporta, para esto se tendrá que implementar un circuito.
  
8. La tendencia actual de comprimir la información es sobre todo para bajar los costos de transmisión de la información.

## RECOMENDACIONES

1. Este trabajo realizado es de actualidad ya que como dijo en alguna parte del trabajo la tendencia actual es tratar de comprimir la voz para poder utilizar el menor ancho de banda.
2. Actualmente la tendencia en lo que se refiere a Telecomunicaciones se està dirigido a la informática, debido a que ahora se habla de voz sobre IP, digitalización de imagen, mensajes de textos, mensajes de voz, etc., toda esta información es manejada por procesadores
3. En lo que respecta a mantenimiento de equipos de comunicación ya no se repara una tarjeta, debido a que todo viene en circuitos integrados, lo que se hace actualmente es cambiarlo y enviarlo a la fábrica para que lo reparen, o en su defecto sea reemplazado por otra.
4. En el futuro no muy lejano van a desaparecer las Centrales Telefónicas, lo que va haber son grandes procesadores, en la que uno ya no va ha tener un número telefónico si no un número IP, por lo que hay que preparar a los alumnos para enfrentar estos avances.
5. La orientación actual de la electrónica está mas hacia informática, las nuevas tecnologías, es decir Internet, voz sobre IP, imagen sobre IP, que eso es lo que se va imponer en el futuro.
6. Con respecto a lo académico en los últimos ciclos sería bueno tener algunos cursos de administración porque actualmente el Ingeniero

aunque no lo quiera va ha tener a cargo técnicos y debe estar preparado para manejar personal, otro tema es la redacción para hacer los informes.

7. En la actualidad en las empresas de Telecomunicaciones la mayor parte de ingenieros titulados son de otras Universidades, los que son de la UNI son muy pocos y es una gran desventaja para la competencia, para que no suceda esto, en los dos últimos ciclos se debe incluir un curso de tesis, con el cual el alumno podría elegir un tema de tesis y desarrollarlo y al concluir la universidad ya tendría el 80% de su Tesis hecha, y de esta manera no va ha pasar lo que nos está sucediendo a nosotros, al menos a no muchos.

## **ANEXO**

### **Tipos de vocoders**

Los vocoders son una clase de sistemas de codificación que analizan la señal de voz en el transmisor, transmiten los parámetros derivados del análisis y sintetizan la voz en el receptor usando esos parámetros. Los vocoders son en general mucho más complejos que los codificadores de forma de onda y desarrollan mucho más economía en la tasa de bits de transmisión, aunque son menos robustos y su desempeño tiende a ser muy dependiente de la voz del usuario. El vocoder más popular es el codificador predictor lineal (LPC), otros vocoders son: channel vocoder, formant vocoder, espectro vocoder y voice excited vocoder.

Los vocoders caracterizan el proceso del "speechmaking" que es un modelo dinámico. Al hablar, el aire que viene de los pulmones pasa por la traquea hacia las cuerdas vocales y se modula por la lengua, boca, labios y dientes. Hay básicamente dos categorías del habla: sonido sonoro y sonido no sonoro. La primera de éstas (sonido sonoro) es cualquier sonido de voz que provoque excitación de pulso quasi-periódico de las cuerdas vocales. El segundo tipo de habla (sonido no sonoro) no toma en cuenta a las cuerdas vocales y no es periódica, más bien se consideraría como aleatorio. Los vocoders analizan la voz periódicamente y determinan los coeficientes de filtrado, ganancia del sistema, periodo de tonos y modelar la excitación como pulsos o como ruido.

Hay muchas diferentes formas para determinar y transmitir estas características.

La voz se divide en tramas las cuales van del orden de 10-30 ms. Cada trama se analiza y un conjunto de parámetros se extraen, estos parámetros son después transmitidos al receptor dónde se usan para sintetizar la voz.

### **Channel Vocoders (vocoder de canal)**

El vocoder de canal fue el primer sistema de análisis y síntesis probado prácticamente. Los vocoders de canal son vocoders en el dominio del tiempo que determinan la envolvente de la señal de voz para un número de bandas de frecuencias y luego las muestrean, codifican, multiplexan estas muestras con la salidas codificadas de los otros filtros. La muestra se realiza sincronizadamente de cada 10ms a 30ms.

### **Formant vocoders**

El formant vocoder es similar al vocoder de canal. Teóricamente el formant vocoder puede operar con tasas de bits más bajas que el vocoder de canal debido a que usa menos señales de control. En lugar de mandar muestras de la envolvente del espectro de potencia, el formant vocoder transmite las posiciones de los picos (los que dan la forma, de ahí el nombre) de la envolvente del espectro.

Los formant vocoders pueden reproducir voz a razones menores de 1200Kbps. Aunque debido a dificultades en la precisión para localizar los picos formadores de las transiciones de la voz humana, los formant vocoders no han tenido un desempeño muy satisfactorio.

### **Vocoders de espectro**

Los vocoders de espectro separan la excitación y el espectro del tracto vocal mediante la transformada inversa de Fourier del espectro de la magnitud logarítmica para producir el "espectro" de la señal. Los coeficientes de baja

frecuencia del espectro resultante corresponden a la envolvente del espectro del tracto vocal y con los coeficientes de excitación de alta frecuencia se forma un tren de pulsos de múltiplos del periodo de muestreo. En el receptor los coeficientes del tracto vocal provenientes del espectro son sometidos a una transformación de Fourier para producir la respuesta al impulso del tracto vocal.

### **Voice-Excited Vocoder (vocoders de excitación de voz)**

Los voiced-excited vocoders eliminan la necesidad de la extracción de tonos y operaciones de detección de voz. Estos sistemas utilizan una combinación híbrida de transmisión PCM para la banda de bajas frecuencias de la voz, combinado con un vocoder de canal para la banda de altas frecuencias.

Los voice-excited vocoders han sido diseñados para operar en el rango de los 7.2Kbps a los 9.6Kbps y su calidad es normalmente superior a aquella obtenida por los tradicionales vocoders excitados por tonos.

### **Enhanced Variable Rate Coder (EVRC)**

El codificador de razón variable extendido "The Enhanced Variable Rate Coder (EVRC)", es un codificador de voz en CDMA compatible con la tecnología utilizada en la actualidad, que promete la optimización necesaria para permitir una alta calidad de voz y mayores niveles de capacidad. EVRC, estandarizado recientemente como IS-127, se apoya de la tecnología actual de software y hardware para ofrecer comunicación de voz de alta calidad a bajas tasas de bits. Las transmisiones de voz de alta calidad prometen mejorar la satisfacción del cliente, mientras que las bajas tasas de bits reducen el ruido del sistema, incrementa la capacidad del sistema y últimamente han bajado los costos de operación por suscriptor. Ahora, como un estándar internacional aprobado,

EVRC también ofrece estabilidad a proveedores en servicio de comunicaciones, interoperabilidad y desarrollo de tiempo de prueba.

El codificador original de voz en CDMA, IS-96A, era un diseño excitado por un Code-Excited Linear Predictive (CELP). IS-96A transmite 8Kbps de información de voz a 9.6Kbps cuando se le agrega la corrección de error. El deseo por mejorar la calidad de voz hizo que la TIA "Telecommunications Industry Association" empezara a trabajar en un vocoder (CELP) a 13Kbps el cual provee una más alta calidad en la transmisión de voz. El vocoder en CDMA de 13Kbps utiliza una tasa de bits más alta para mejorar dicha calidad de voz, sin embargo, la mejora de la misma reduce la capacidad del sistema y el área de cobertura de la célula .

Últimamente se desarrollo el sistema EVRC de 8Kbps. Un sistema EVRC de 8kbps, usando el mismo número de sitios de célula, ofrece el doble de la capacidad que los sistemas de 13Kbps o alternativamente la misma capacidad para la mitad de células. El hecho de tener menos bits implica menor ruido en el sistema y mejor aprovechamiento del ancho de banda, dando como resultado el poder ofrecer a los suscriptores calidad de línea alambrada a un menor costo.

El fuerte del EVRC es un algoritmo llamado "Relaxed Code- Excited Linear Predictive coding(RCELP)", que se desarrolló en 1994 por Bell Labs y Lucent's research. Una generalización del algoritmo de la codificación de voz CELP es RCELP el cual es idóneo para operaciones de tasa variable y fortalecer el ambiente de CDMA. CELP utiliza tramas de voz de 20 ms para la codificación y decodificación. Cada intervalo de tiempo de 20 ms, el codificador procesa 160 muestras de voz. Los codificadores de tasa variable examinan el contenido de la trama y determinan la tasa de codificación necesaria. De acuerdo al nivel de la complejidad de la señal detectada, el codificador analiza la voz a cuatro

diferentes tasas de bits: 8Kbps, 4Kbps, 2Kbps o 1Kbps. Esto difiere de otros codificadores (half rate y multi rate coders), donde la tasa de bits se determina una vez por llamada. Los codificadores de tasa variable emplean un algoritmo conocido como detección de actividad de voz. El vocoder codifica a tasas de bits altas al detectar actividad de voz, generando con esto una mejor calidad de la misma y al no haber detección de voz, el vocoder reduce la tasa de codificación y la tasa de transmisión efectiva disminuye de 9.6Kbps a 1.2Kbps. El transmisor genera un octavo de la energía de interferencia y utiliza un octavo del espectro de una transmisión al 100%.

Los vocoders también incrementan la calidad en la voz mediante la supresión del ruido de fondo. Cualquier ruido que sea constante en la naturaleza, tal como el ruido del camino, es atenuado. El constante sonido de fondo se interpreta por el vocoder como ruido que no contiene información relevante, por lo que se remueve tanto como sea posible. Esto mejora la claridad de la voz en ambientes ruidosos, tal como en el interior de los coches o en lugares donde hay mucha gente.

Para cada trama de 20 ms, el algoritmo CELP examina los datos y genera 10 coeficientes lineales del filtro de codificación de predicción. Los coeficientes se representan con la precisión 4, 2, o 1 bit, dependiendo de la tasa variable determinada por el factor de complejidad. Estos coeficientes describen la forma de onda de excitación, un predictor derivado de un modelo matemático de los pulmones humanos y de las cuerdas vocales, y otro predictor de corto plazo con factor de periodicidad y ganancia derivado de un modelo similar a la zona vocal. La forma de onda de excitación para cada 5 ms se selecciona de un código que consiste en una señal formada por una gran cantidad de vectores candidatos. El vector de código seleccionado para diseñar los filtros minimiza el error entre la voz original y la sintetizada.

A diferencia con el CELP, el RCELP no pretende generar la voz original exactamente, acopla una versión en el tiempo de la señal residual para

conformar un contorno de tono simplificado el cual se obtiene mediante la estimación de retraso de tono en cada trama y linealmente usando interpolación de tono de trama a trama resultando en una mayor calidad de voz por bit transmitido. La representación del tono simplificado deja más bits disponibles en cada paquete para la excitación estocástica y para la protección de perjuicios al canal que sería posible si se usara una aproximación de tono fraccional tradicional. El resultado es la prevención de errores mejorado sin la reducción de la calidad de la voz.

## **BIBLIOGRAFIA**

### **1. DIGITAL PROCESSING OF SPEECH SIGNALS**

**AUTOR:** Laurence R. Rabiner

Ronald W. Schafer

**EDITORIAL:** Prentice-Hall Inc – 1978

### **2. DIGITAL CODING OF WAVEFORMS**

**Principles and applications to speech and video**

**AUTOR:** Ns Jayant

Peter Noll

**EDITORIAL:** Prentice-Hall – 1984

### **3. TRATAMIENTO DIGITAL DE SEÑALES**

**AUTOR:** John G. Proakis

Dimitris G. Manolakis

**EDITORIAL:** Prentice-Hall – 1998