

01169  
2ej<sup>o</sup> |

Teoría de la Información  
con Aplicaciones al  
Procesamiento Digital de Señales

CARLOS RIVERA RIVERA

TESIS

Presentada a la División de Estudios de  
Posgrado de la  
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO  
como requisito para obtener  
el grado de  
MAESTRO EN INGENIERIA  
( CONTROL )

CIUDAD UNIVERSITARIA 10 DE DICIEMBRE DE 1986.



Universidad Nacional  
Autónoma de México



## **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 está 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

1. Agradecimientos .....	ii
2. Resumen .....	iii
3. Introducción .....	1
4. Codificadores de forma de onda y de fuente .....	5
5. Predicción Lineal .....	9
6. Descripción del sistema .....	12
7. Simulaciones .....	18
8. Comentarios .....	23
9. Bibliografía .....	27
10. Figuras .....	29

## Resumen

Se presenta un sistema de codificación de señales de voz a tasas entre 2400 y 9600 bits por segundo. El sistema está basado en la representación del espectro de la señal de voz por la transformada discreta de Fourier, en su forma polar (vector de magnitudes y fases). Se utiliza también un sistema de predicción lineal el cual es empleado como información lateral. Los parámetros de estos sistemas son codificados en forma escalar (tasas mayores a 6800 bits por segundo), y en forma vectorial (tasas menores a 6800 bits por segundo).

La información lateral está constituida por las bandas en frecuencia donde ocurren los formantes (frecuencias de resonancia del tracto vocal) de la señal de voz, así como una primera estimación de la magnitud del espectro. La codificación se hace únicamente en estas bandas.

La amplitud del espectro es representada como el producto de un filtro  $1/A(Z)$  por unos factores de corrección  $f_i$ . El filtro  $A(Z)$  es el debido al predictor lineal y es representado en su forma de vector de coeficientes de reflexión. Esta representación aprovecha la información lateral.

La codificación se realiza sobre los vectores de coeficientes de reflexión, factores  $f_i$  y fases.

Finalmente la señal sintética se obtiene al calcular la transformada discreta inversa de Fourier.

## INTRODUCCION

Los métodos de codificación de señales de voz tienen como uno de sus principales objetivos, el reducir el número de bits por segundo que se requieren para transmitir o almacenar la señal con un criterio de fidelidad especificado.

Una forma de realizar esta reducción es el quitar la mayor parte de la redundancia de la señal. Un procedimiento conocido para lograrlo es por medio de sistemas de codificación por predicción. Este tipo de codificación disminuye la redundancia al tomar únicamente la parte de la señal que no se puede predecir.

Existe otro tipo de codificación, la codificación por transformación, en la cual primero se realiza una transformación de los datos, por lo general una transformación del tiempo a la frecuencia, y posteriormente se hace la codificación con estos nuevos datos.

En este tipo de codificación la reducción no se basa en una predicción de la señal, en vez de esto, la reducción se realiza al asignar la mayor parte de los bits disponibles a las regiones donde se encuentran la energía.

En años recientes el interés por realizar sistemas de codificación que transmitan voz a tasas menores de 16000 bits por segundo ha ido en aumento.

Existen varias razones para esto, entre las cuales resaltan; Los canales disponibles para transmisiones de voz se han saturado

y el aumentarlos redundancia en un costo elevado, el avance de la electrónica ha permitido circuitos que realizan millones de operaciones por segundo (por ejemplo ; circuitos que realizan la transformada discreta de Fourier en tiempo real), nuevos modelos de la señal de voz se han propuesto, los cuales han demostrado su utilidad.

La siguiente figura muestra las diferentes tasas de transmisión, así como la calidad de los sistemas que existen en la actualidad.

### CODIFICACION DIGITAL DE VOZ

codificación de forma de onda						codificación de fuente						
						Kbps						
200	64	32	24	16	9.6	8.0	7.2	4.8	2.4	1.2	0.5	0.05
calidad de audio	calidad total			calidad de comunicaciones			calidad sintética					

En este trabajo se presenta un sistema de codificación de voz a tasas entre 2400 y 9600 bits por segundo.

El sistema se basa en codificación por predicción lineal y transformada discreta de Fourier. Los parámetros de ambas técnicas son codificados vectorialmente.

La motivación del trabajo es múltiple; al realizar estudios de la señal de voz, se ha observado que presenta la característica

de bandas en frecuencia donde existe la mayor parte de la energía de la señal. Estas bandas han sido caracterizadas por medio de los formantes, los cuales representan las zonas de resonancia del sistema vocal.

Experimentos que se realizaron, han mostrado que se puede producir una señal sintética de muy buena calidad, al codificar únicamente estas bandas.

El llevar a cabo un análisis teniendo en cuenta esta propiedad tiene la cualidad de que al aumentar el número de componentes en cada una de estas bandas nos lleva a una descripción de todo el espectro de voz, mientras que en el extremo contrario se tendrían las frecuencias donde ocurren los formantes de la señal de voz.

Por otra parte, durante mucho tiempo se ha conocido que el sistema de predicción lineal es un buen estimador de la amplitud del espectro de la señal de voz (por lo tanto de los formantes). La codificación por predicción lineal puede ser representada por diferentes parámetros, por ejemplo; coeficientes de correlación, coeficientes del filtro  $L(z)$ , coeficientes de reflexión. Los cuales han sido ampliamente estudiados, por lo que sus características son conocidas.

Estos parámetros muestran un buen desempeño al ser codificados vectorialmente, realizandose sistemas con tasas hasta de 800 bits por segundo.

La propiedad de ser una buena aproximación de la amplitud del espectro es importante debido a la suposición, de que el oído es muy sensible a características espectrales de la forma de onda.

En el aspecto computacional, tanto el sistema de predicción lineal como la transformada discreta de Fourier pueden ser calculados en forma eficiente.

Otra razón importante es que sistemas de codificación por transformación utilizando cuantización escalar han tenido éxito.

En [2,3], se han diseñado sistemas de codificación por transformación basados en cuantización vectorial. Las diferencias principales entre estos sistemas y el sistema descrito en este trabajo son: el tipo de transformación, la estructura del cuantizador vectorial, la forma en que es asignado el número de bits y principalmente la información lateral que se transmite al receptor.



## CODIFICADORES DE FORMA DE ONDA Y DE FUENTE

Una clase amplia de codificadores recibe el nombre de codificadores de forma de onda. La característica principal de estos codificadores es que intentan una reproducción lo más fiel posible de la señal. En principio estos sistemas son diseñados para ser independientes de la señal, por lo que pueden codificar igualmente bien una variedad grande de señales, por ejemplo; voz, música, imágenes, etc. Tienen la característica de que son robustos para diferentes personas y diferentes medios ambientes. Para conservar estas ventajas con un mínimo de complejidad, estos codificadores tienen tasas de transmisión medias. ( 20000 o más bits por segundo)

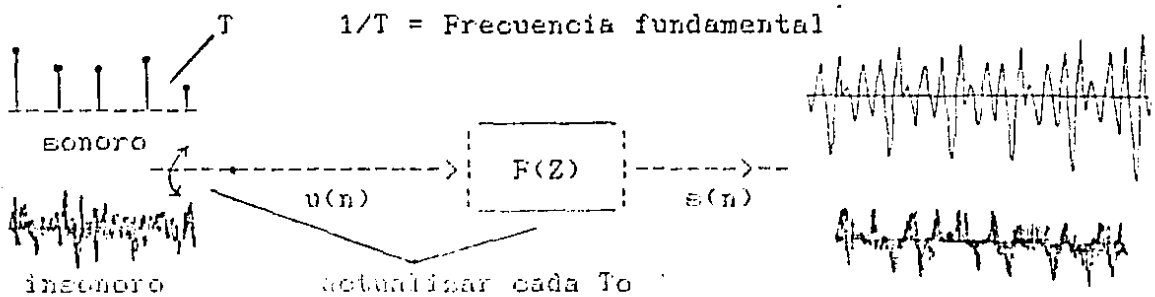
Los codificadores de onda se pueden optimizar y hacer más específicos con la finalidad de obtener un mejor rendimiento. Para conseguir lo anterior se observa la estadística de un conjunto grande de señales, de tal forma que el codificador tenga un error promedio mínimo para esta clase de señales. Estos codificadores se basan en una descripción estadística de la forma de onda de la señal, a diferencia de una parametrización de la señal de acuerdo a algún modelo físico de esta.

Una segunda clase de codificadores se basa en una descripción de la señal, utilizando un conocimiento a priori acerca de la forma en la cual fue generada en la fuente. La idea principal es que existen ciertas restricciones físicas de la

generación de la señal, las cuales pueden ser cuantificadas y por lo tanto describirla en forma eficiente.

Esto implica que la señal debe ser clasificada como perteneciente a alguna clase y parametrizada de acuerdo a esto. Estos codificadores son conocidos como codificadores de fuente, para el caso de señales de voz por lo general son referidos como vocoders.

El modelo tradicional de la señal de voz, es el mostrado en la figura.



El mecanismo de generación, se modela como una combinación lineal del modulador (la fuente) y el filtro del tracto vocal (el sistema).

Otras suposiciones acerca de la señal de voz es si son: sonoros o insonoros y si son generados por un sonido periódico producido por las cuerdas vocales o por un sonido aleatorio producido por una turbulencia de aire.

Al hacer un ajuste adecuado de los parámetros de este modelo es posible obtener una señal de reproducción de buena calidad.

Sin embargo en general, en los sistemas actuales, los vocoders son muy sensibles, en términos de los parámetros ; sonoro o insonoro y frecuencia fundamental. El desempeño es frecuentemente dependiente de la persona que habla y la señal de reproducción tiene una calidad sintética(metálica). Estas características constituyen un límite en la calidad que los vocoders pueden alcanzar. Sin embargo en virtud de la parametrización que realizan de la señal de voz, los vocoders pueden alcanzar una compresión de datos muy significativa.

El límite entre los codificadores de onda y los vocoders puede verse como una clase de terreno medio, donde el criterio de diseño no es ni preservación de la forma de onda ni un modelo de la señal. En lugar de esto, el principio que se sigue es la preservación de la amplitud del espectro de duración corta de la señal de voz, la razón principal es el hecho, ya mencionado anteriormente de que el oído humano aparentemente es muy sensitivo a las características espectrales de la señal.[1]

Las técnicas de codificación por transformación han sido de las más empleadas para la compresión de datos. Aunque tradicionalmente la codificación de los coeficientes de este tipo de transformación se hace en forma escalar, puede considerarsele como un tipo de codificación vectorial, ya que cada dato en un dominio depende generalmente de todas las muestras del dominio original. Aunado a esto la codificación por transformación ha

mostrado un mejor resultado, para los casos especiales de voz e imágenes, que la codificación escalar ordinaria. Esto se debe posiblemente a que tanto el ojo como el oído tienden a ser más sensitivos a los coeficientes de la transformación que a la forma de onda original, por ejemplo a las características espectrales en el caso de la transformada de Fourier.

## FREDDICION LINEAL.

Uno de los métodos con mayor éxito en el análisis de señales de voz, es el método de predicción lineal. Este método sirve para estimar los parámetros básicos de la voz, como son : frecuencia fundamental, formantes, función de área del espectro vocal, amplitud del espectro.

La importancia de este método estriba en su habilidad para realizar una muy buena estimación de los parámetros de la voz y además en la existencia de algoritmos rápidos y eficientes en su cálculo.

En este método la señal  $s(n)$  es considerada como la salida de un sistema con entrada desconocida  $u(n)$  y tal que la siguiente relación se cumpla :

$$s(n) = \sum_{k=1}^p a(k)s(n-k) + G \sum_{i=0}^q b(i)u(n-i)$$

donde  $a(k)$ ,  $1 \leq k \leq p$ ,  $b(i)$ ,  $0 \leq i \leq q$ ,  $b(0) = 1$  y la ganancia  $G$  son los parámetros del sistema. Esta ecuación relaciona la señal  $s(n)$  como una función lineal de entradas y salidas anteriores más la entrada actual. De esta forma la señal  $s(n)$  es una predicción lineal de entradas y salidas. Este modelo se conoce con el nombre de modelo de polos y ceros.

Existen dos casos especiales que son de interés:

1. Modelo de ceros :  $a(i) = 0$   $i = 1, \dots, p$
2. Modelo de polos :  $b(i) = 0$   $i = 1, \dots, q$

tradicionalmente el modelo de polos ha sido utilizado para la

estimación de una parte del sistema vocal.

La razón principal de esto, es que el sistema de ecuaciones que resultan de la optimización de los parámetros, poseen propiedades interesantes.

El modelo de polos asume a la señal  $s(n)$  como una combinación lineal de valores anteriores más una entrada  $u(n)$  :

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

La función de transferencia está dada por :

$$S(Z)/U(Z) = G/A(Z)$$

Dada una señal  $s(n)$ , el problema es determinar los coeficientes del predictor  $a(k)$  así como la ganancia  $G$  del sistema. Existen varios métodos para el cálculo de estos parámetros, el más empleado es el método de los mínimos cuadrados. En este método se hace una predicción de la señal  $s(n)$ , de la siguiente manera:

$$\hat{s}(n) = - \sum_{i=1}^p a(i)s(n-i)$$

Se puede observar que esta predicción no depende de ninguna entrada  $u(n)$ . El error entre esta predicción y la señal real es

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i=1}^p a(i)s(n-i)$$

El problema es minimizar el valor  $e(n)$  con algún criterio de fidelidad. El método de mínimos cuadrados encuentra los parámetros del sistema al minimizar el error medio cuadrático de  $e(n)$ . Dependiendo de la forma en que sea considerada la señal  $s(n)$ , determinística o no, el rango de la sumatoria y el algoritmo de

solución se tienen las siguientes variaciones del método de mínimos cuadrados.

1. El método de la covariancia.
2. El método de autocorrelación.
3. El método del filtro de coeficientes de reflexión.
4. La formulación del filtro inverso.
5. La formulación de la estimación espectral.
6. La formulación de máxima similitud.
7. La formulación del producto interno.

Para este trabajo fueron considerados, el método de la autocorrelación y el algoritmo de Levinson para la solución del sistema.

## DESCRIPCION DEL SISTEMA

La señal de voz es representada como la concatenación de un conjunto de vectores  $\{ x(n), n=0, \dots \}$ , donde el vector  $x(n)$  consiste de  $K$  muestras consecutivas de la señal de voz.

Este vector a su vez es representado en el dominio de la frecuencia, por medio de la transformada discreta de Fourier, por otro vector  $X(n)$ .

La relación entre los vectores ;  
 $x(n) = \{ x[k], k=0, K-1 \}$  y  $X(n) = \{ X[i], i=0, K-1 \}$  ( se puede suprimir el índice  $n$  sin que exista confusión ), está dada por el siguiente par de ecuaciones :

$$X[i] = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} x[k] \exp( -j2\pi i k / K ) \quad i = 0, 1, \dots, K-1$$

$$x[k] = \frac{1}{\sqrt{K}} \sum_{i=0}^{K-1} X[i] \exp( j2\pi i k / K ) \quad k = 0, 1, \dots, K-1$$

o en forma matricial

$$X = F x \quad x = F^{-1} X$$

donde  $F$  es una matriz cuyo elemento en el  $i$  renglón y la columna  $k$  es  $\exp( -j2\pi i k / K )$ .

El vector  $X(n)$  tiene coeficientes que son números complejos y en general requiere para su representación de  $2K$  números de punto flotante.



Sin embargo como  $x(n)$  es un vector con componentes reales, se tiene de la propiedad de simetría de la TDF

$$\operatorname{Re}\{X[i]\} = \operatorname{Re}\{X[K-i]\} \quad 1 \leq i \leq K/2$$

$$\operatorname{Im}\{X[i]\} = -\operatorname{Im}\{X[K-i]\} \quad 1 \leq i \leq K/2$$

donde  $\operatorname{Re}$  y  $\operatorname{Im}$  son la parte real e imaginaria respectivamente.

El vector  $X(n)$  puede ser también representado por medio de un vector de magnitudes  $X_m(n)$  y un vector de fases  $X_f(n)$  :

$$X_m(n) = ( |X[i]|, i=0, K-1 ), \quad X_f(n) = ( \exp\{j\theta_x(i)\}, i=0, K-1 )$$

$$\text{donde } \theta_x = \tan^{-1} ( \operatorname{Im}\{X[i]\} / \operatorname{Re}\{X[i]\} )$$

también por la propiedad de simetría, se tiene que

$$|X[i]| = |X[K-i]| \quad 1 \leq i \leq K/2$$

$$\exp\{j\theta_x(i)\} = \exp\{-j\theta_x(K-i)\} \quad 1 \leq i \leq K/2$$

Puesto que el sistema de predicción lineal es una estimación del espectro de la señal, la amplitud de los coeficientes de la TDF se pueden expresar en función del filtro  $A(z)$  como:

$$|X[k]| = G f(k) / |A(\exp\{j\omega(k)\})|$$

donde  $A(\exp\{j\omega(k)\})$  es la respuesta en frecuencia del filtro debido al sistema de predicción lineal,  $G$  es la energía del vector  $X[k]$  y  $f(k)$  son factores de corrección de amplitud. El vector de magnitudes  $X_m$  se puede expresar como :

$$X_m = G f(k) / |A(\exp\{j\omega(k)\})|$$

Finalmente se tiene una relación entre un vector  $x(n)$  de señal de voz en el tiempo y una pareja de vectores en frecuencia;

$$\{x(n)\} \longleftrightarrow \{X_m(n), X_f(n)\}$$

Como se puede obtener una señal sintética de buena calidad al codificar únicamente las bandas que están centradas en los formantes, y el filtro  $A(Z)$  proporciona una manera sencilla de encontrar las frecuencias de los formantes junto con una aproximación del espectro, representar la amplitud de la TDF en función del sistema de predicción lineal aprovecha en forma interesante esas propiedades.

Para realizar la codificación del espectro y teniendo en cuenta que se deben codificar con más cuidado las bandas donde se concentra la energía, primero los vectores de magnitud y fase son subdivididos en varios subvectores, esta división se lleva a cabo de la siguiente manera;

primero se encuentran  $N$  formantes y alrededor de estas frecuencias se toman  $N_b$  coeficientes de la TDF, tanto de magnitud como de fase, si existe transiapa de las bandas se hace un corrimiento con el fin de que siempre se consideren  $N \times N_b$  frecuencias del espectro.

La longitud de cada uno de estos subvectores varía dependiendo del orden de la transformada. Para el caso de una TDF de orden 8 (256 frecuencias), se tienen subvectores de 13 a 19 componentes. Estas longitudes se encontraron al realizar experimentos y ver la calidad de la señal sintética.

La codificación se realiza únicamente dentro de las bandas donde existen los formantes, asignandosele a los coeficientes de las frecuencias fuera de estas bandas una magnitud cero. La razón

de esto, es que a tasas bajas es muy importante la asignación de bits y como ya se mencionó es posible obtener una señal sintética de buena calidad al codificar únicamente estas bandas. Si la tasa de transmisión aumenta, la asignación de bits se sigue haciendo en estas bandas, solamente que el número de componentes puede aumentar, manteniéndose siempre centrado alrededor del formante. Es importante notar que a los formantes con más energía se les debe asignar mayor número de bits.

Debido a que codificadores vectoriales de estas longitudes son para fines prácticos difíciles de realizar, la codificación se hace por medio de códigos producto. Este tipo de codificación divide el bloque o vector en  $m$  subvectores y realiza la codificación de cada uno de estos subvectores en forma separada.

Al tener por un lado un sistema de predicción lineal y por otro la transformada discreta de Fourier el sistema completo es un sistema de codificación por medio de código producto, ya que tanto el sistema de predicción lineal, el cual puede ser representado por un vector de coeficientes de reflexión, como los vectores de fase y corrección de amplitud son codificados en forma separada, aunque primero se debe codificar el vector de coeficientes de reflexión ya que de este vector dependen tanto las bandas donde se realiza la codificación como los factores de corrección, además se le debe agregar a lo anterior que la energía del vector  $X(n)$  es codificado en forma escalar.

Debido a que la codificación de los diferentes parámetros puede ser realizada en forma separada, no es necesario el tener una sola medida de distorsión, sino que es posible tener para cada subvector una distorsión diferente.

Al llevar a cabo las simulaciones, cada uno de los codificadores tiene un número arbitrario de bits y posteriormente se examina el efecto en la señal sintética.

Para la codificación de los coeficientes de reflexión se utilizaron las distorsiones de Itakura-Saito y de error cuadrático medio. Para la codificación de los factores de corrección  $\phi_i$  y las fases la distorsión que se empleó fue error cuadrático medio.

La codificación de la fase no se realiza en forma directa ya que es difícil el calcular los centroides en forma adecuada debido a la propiedad de modulo  $2\pi$  que tiene la fase. En lugar de esto el número complejo de magnitud unitaria y la misma fase es el que se codifica, esto tiene el inconveniente de que cada fase está representada por dos números.

La codificación por transformación genera un ruido debido a la discontinuidad que se presenta entre los bloques. Una manera de reducir este efecto es extender la longitud del bloque de transformación, con esto el espectro del ruido se traslada a la región de frecuencias donde es enmascarado por la señal de voz, por ejemplo entre 300 y 3400 Hz. Otro método es el de emplear ventanas en las cuales exista un traslape y reduzcan el efecto de

discontinuidad entre los bloques, el efecto de las ventanas es que permiten que la señal luzca continua entre bloques. Las ventanas son colocadas antes y después de la transformada discreta de Fourier tanto en el codificador como en el decodificador. No es necesario que estas ventanas sean iguales en el análisis y síntesis, es necesario solamente el tener en cuenta que si no existe cuantización la señal de reproducción sea igual a la original.

En el sistema descrito en este trabajo se utilizó una ventana trapezoidal en el codificador y una ventana rectangular en el decodificador. Al tener un traslape, la tasa del sistema aumenta pero este aumento es compensado por la mejora en la calidad de la señal de reproducción.

## SIMULACIONES.

La secuencia de entrenamiento con la que se calcularon los códigos, consiste de grabaciones de 9 diferentes personas. Cada sesión tiene un segmento de 49.152 segundos de voz ( 1536 bloques, cada bloque de 256 muestras). La señal se digitalizó con una frecuencia de 8000 muestras por segundo en un convertidor de 8 bits, el cual tiene la ley de compresión  $\mu$ .

Para hacer las pruebas se tuvieron 2 grabaciones más, una de ellas pertenece a un individuo que estuvo dentro de la secuencia de entrenamiento, la otra grabación fue tomada del radio y en ella intervienen 2 personas ( no al mismo tiempo), las cuales están fuera del entrenamiento. Cada una de estas grabaciones tiene una duración de 16.364 segundos ( 512 bloques de 256 muestras cada uno).

El sistema consiste de las siguientes etapas:

Sistema de predicción lineal:

1. N correlaciones de un bloque de K muestras.
2. solución del sistema de predicción lineal.
3. codificación del vector de coeficientes de reflexión.
4. filtro  $A(Z)$  correspondiente a los coeficientes de reflexión codificados.
5. espectro de  $A(Z)$ .
6. formantes del espectro y bandas de codificación de la TDF.

Transformada discreta de Fourier.

1. TDF de un bloque de  $K_0$  muestras.
2. vectores de amplitudes y fases correspondientes a las bandas de codificación.
3. vector de factores de corrección  $f_i$ .
4. energía del bloque  $G$ .
5. codificación de vectores de fases,  $f_i$  y energía del sistema.

Para calcular el sistema de predicción lineal se toma un bloque de  $K$  muestras, por lo general este bloque consiste de 3 subbloques el primero corresponde al bloque anterior, el segundo al bloque del cual se quiere encontrar el predictor y el último bloque es un segmento de la señal que va a ser procesada. Esto tiene como finalidad el disminuir lo más posible cambios abruptos entre bloques. Por lo general el bloque se multiplica por una ventana que tiene ciertas características deseables en frecuencia. Notese que aun cuando esta multiplicación no se haga en forma explícita, siempre está presente en la forma de una ventana rectangular.

Con este vector de señal, se calcula un vector de correlaciones y posteriormente se resuelve el sistema de predicción lineal. La solución da como resultado varios parámetros, pero solo el vector de coeficientes de reflexión se utiliza.

Con el objeto de tener codificadores vectoriales de facil

realización, el vector de coeficientes se divide en varios subvectores. Para su codificación se tuvieron dos criterios de distorsión, en el primer subvector se empleó la distorsión de Itakura-Saito, mientras que en los restantes el criterio fue el error medio cuadrático.

Si bien con las longitudes de estos subvectores es posible diseñar códigos de hasta 10 bits (1024 patrones), los experimentos mostraron que códigos de 9 bits tienen muy buen desempeño.

Este nuevo vector de coeficientes de reflexión es el que se transmite al receptor. Teniendo este vector, se calculan los coeficientes del filtro  $A(Z)$  y su espectro correspondiente. Con este espectro se encuentran de manera sencilla los formantes, con estos sus frecuencias y las bandas de codificación de la TDF.

La transformada discreta de Fourier se calcula tomando un bloque de  $K_0$  muestras. De manera similar al predictor lineal, se tienen 3 subbloques que corresponden; al bloque anterior, actual y siguiente respectivamente. El bloque se multiplica por una ventana, para evitar las discontinuidades que se puedan presentar debido a los efectos de la cuantización. A diferencia de la ventana del predictor, aquí las características deseables están expresadas en el tiempo. Teniendo este vector se calcula la transformada discreta de Fourier, puesto que una manera óptima de hacerlo es con la transformada rápida de Fourier, existen ciertas restricciones acerca de la longitud del vector  $K_0$ , por lo que



generalmente se toma una longitud de  $2^{11}M_0$  donde  $M_0$  es un valor entero.

Teniendo los vectores de magnitud y fase, y junto con la información de las bandas donde se debe hacer la codificación, se forman varios subvectores, uno por banda.

Para la magnitud se toman los  $N_f$  subvectores que tengan los formantes con más energía. De estos  $N_f$  subvectores, solamente se toman  $M_f$  para calcular los factores  $f_i$ . Los vectores  $f_i$  a su vez son divididos, con el fin de tener codificadores vectoriales de hasta 10 bits, pero que son razonablemente realizables debido a las longitudes de cada subvector. Aquí de la misma forma que el caso del predictor, se encontró que con 9 bits se obtienen buenos resultados.

Para los vectores de fases, dependiendo de la tasa de transmisión que se quiera tener, se toman los primeros formantes (sin importar la energía que tengan). Para tasas bajas, se toma únicamente el primer vector, en ambos casos se hace una subdivisión. Debe notarse que al no hacer la codificación de la fase en forma directa, sino como un número complejo de magnitud unitaria, las longitudes de los vectores se duplican.

El criterio de distorsión fue también el error medio cuadrático. En la codificación vectorial de estos datos se debe tener cuidado en que los centroides obtenidos tengan magnitud unitaria, de otra forma influirían en la codificación de la

magnitud.

Para realizar la síntesis, el receptor debe calcular primero el filtro  $A(Z)$  correspondiente a los coeficientes de reflexión que recibió. Con el espectro de este filtro debe buscar las  $N_f$  crestas con más energía, las cuales corresponden a los formantes, finalmente obtiene las bandas donde se debe sintetizar el espectro.

Para calcular la magnitud de la TDF en estas bandas, simplemente se multiplica el correspondiente factor  $f_i$  por la magnitud de  $A(Z)$ . En las bandas donde los factores  $f_i$  no fueron codificados su magnitud es la del filtro  $A(z)$  (se puede suponer que los factores son iguales a uno en esa banda). Fuera de estas la magnitud se hace cero.

Para formar el vector de fases, las que se recibieron se repiten en todas las bandas, fuera de estas no importa su valor.

Teniendo los vectores de magnitud y fase se calcula la transformada inversa de Fourier, después se aplica la ventana con lo cual se tiene la señal sintética de ese bloque.

Es importante notar que la longitud del bloque del predictor  $K$  y la del bloque de transformada  $K_0$  pueden ser diferentes, únicamente se debe cumplir que la segunda parte de cada uno de ellos sea igual, por lo que la diferencia de longitudes correspondería al traslape que cada uno de ellos tenga con los bloques adyacentes.

## COMENTARIOS.

En este trabajo se describe un sistema de codificación por transformación. con un predictor lineal como información lateral.

La codificación de estos parámetros se hace por medio de cuantización vectorial y con una estructura de código producto. Al utilizar el predictor lineal como información lateral, se aprovecha la característica de que es un buen estimador de la amplitud del espectro, por lo que la transformada discreta de Fourier se representa en forma polar. En la codificación de la amplitud se pudo notar que la distorsión promedio al codificar vectorialmente los factores de corrección  $f_i$  es pequeña, y la señal que se obtiene es de buena calidad.

Debido a que; por un lado el predictor lineal, principalmente en su representación de coeficientes de reflexión, no sufre una distorsión significativa al ser codificado vectorialmente y por el otro los factores  $f_i$  tampoco se degradan, hacen que la formulación de la amplitud de los coeficientes de la TDF sea interesante.

El sistema utiliza también la peculiaridad de que tiene zonas en el espectro donde se concentra la energía, por lo que sólo estas bandas son codificadas. Esto es consistente con el hecho de que la aproximación del espectro por el predictor lineal se ajusta mejor en las crestas (en los formantes) que en los valles (en las regiones con menos energía).

Es una suposición común el considerar que la fase es en gran parte ignorada por el oído, este atributo se refleja en el sistema, al tomar únicamente una parte de las fases y repetir las en las bandas adecuadas.

Para el sistema de orden 8, 127 fases por bloque, se toman 19 fases con las cuales se logra una señal sintética de buena calidad. Se tiene además que si se codifican todas las bandas, la calidad de esa señal no aumenta notablemente, a diferencia de la tasa que si lo hace.

Aquí se debe hacer la siguiente observación; si bien el número de componentes que se deben codificar son pocas, en los experimentos se notó que esta codificación es muy sensible y por lo general la señal se degrada al presentarse un ruido de bajas frecuencias.

La codificación de la fase se hizo de manera tanto vectorial como escalar, en el primero de los casos la longitud de los vectores fue máximo de 4 componentes y su tasa de 10 bits. Para el caso escalar los experimentos mostraron, suponiendo que la fase está uniformemente distribuida, que con 5 bits por coeficiente la señal sintética es de buena calidad.

La asignación de bits que se hizo fue el resultado de una serie de experimentos y no como la minimización de la distorsión total del sistema.

El no hacer lo anterior tiene las siguientes justificaciones;

i) al tener tantos parámetros involucrados en el diseño del sistema, un algoritmo que minimize la distorsión total sería bastante complejo y difícil de hacer.

ii) puesto que en la actualidad no se conoce una distorsión que refleje todas las propiedades que tiene el oído humano, hacer una minimización sobre una distorsión dada no es indicación de que la señal sintética sea de buena calidad.

Si esta distorsión fuera conocida, cualquier tipo de codificación podría lograr un buen desempeño, además que por medio de la teoría de la información se podría calcular la tasa mínima del sistema con una distorsión dada. Es por esto que se debe juzgar la calidad de la señal en forma subjetiva.

De manera experimental se encontró que una asignación de 2/3 partes a la magnitud ( predictor lineal, factores de corrección y energía del bloque) y 1/3 parte a la fase, daba como resultado una señal de buena calidad.

Si bien ya una vez encontradas las frecuencias de los formantes y con ello las bandas de codificación, la asignación de bits está fija, se debe hacer hincapié en que considerando todo el

espectro, la asignación se realiza en forma dinámica.

Finalmente debo mencionar las propiedades novedosas del sistema;

i) La manera de utilizar la información contenida en el LPC, en especial la información acerca de los formantes, con lo cual se hace la asignación dinámica de bits.

ii) Formular la amplitud de la TDF como el producto del filtro de predicción lineal por unos factores de corrección, esto es igual a considerar solamente la amplitud del residuo del predictor.

iii) La manera en la cual la fase de la señal es codificada y posteriormente reconstruida.

## BIBLIOGRAFIA

1. J. BELLAMY, Digital Telephony, Wiley, New York, 1982
2. N. HE., A. BUZO, AND F. KUHLMANN, "A frequency domain waveform speech compression system based on product vector quantizers," Proceeding of ICASSP, Tokyo, Japan, April 1986
3. PAO-CHI CHANG, ROBERT M. GRAY, AND JACK MAY, "Fourier Transform Vector Quantization for Speech Coding"
4. R. ZELINSKI and P. NOLL, "Adaptive transform coding of speech signals," IEEE Transactions on Acoustic, Speech, and Signal Processing, vol. ASSP-25, pp.299-309, August 1977.
5. J.M. TRIBOLET and R.E. CROCHIERE, "Frequency domain coding of speech," IEEE Transactions on Acoustic, Speech, and Signal Processing, vol. ASSP-27, pp.512-530, October 1979.
6. P. A. WINTZ, "Transform picture coding," Proceedings of the IEEE, vol. 60, pp.809-820, July 1972
7. T. BERGER, Rate Distortion Theory, Prentice-hall Inc., Englewood Cliffs, N. J., 1971.
8. A. GERSHO, T. RAMSTAD, and I. VERSVIK, "Fully vector-quantized subband coding with adaptive codebook allocation," Proceedings ICASSP, vol. 1, pp. 10.7.1-4, March 1984.
9. H. ABUT and S. LUSE, "Vector quantizers for subband coded waveforms," Proceedings ICASSP, vol. 1, pp. 10.6.1-4, March 1984.
10. J. PEARL, "On coding and filtering stationary signals by discrete Fourier transforms," IEEE Transactions on Information Theory, pp. 229-232, March 1973.
11. R.M.GRAY, "Vector Quantization," IEEE ASSP Magazine, vol. 1, No.2, pp.4-29, April 1984.
12. Y. LINDE, A. BUZO, AND R.M. GRAY, "An algorithm for vector quantizer design," IEEE Transactions on Communications, vol. COM-28, pp. 84-95, January 1980.
13. A. BUZO, A.H. GRAY Jr., R. GRAY and J.D. MARKEL, "Speech Coding Based Upon Vector Quantization", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-28, pp.562-574, Oct. 1980.

14. R.M GRAY, A. EUZO, A.H. GRAY and Y. MATSUYAMA, " Distortion Measures for Speech Processing", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-28, pp.367-376, August 1980.
15. R. Bracewell, The Fourier Transform and its Applications , McGraw-Hill, New York, 1965.
16. J.D. MARKEL, A.H. GRAY, Jr., Linear Prediction of Speech, Springer Verlag, Berlin, Heidelberg, New York, 1976.
17. J. Makhoul, " Linear Prediction : a tutorial review," IEEE Proc., Vol.63, pp.551-580, April 1975.
18. M.J. SABIN and R.M. GRAY, " Product code vector quantizers for waveform and voice coding," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-32, pp.474-488, June 1984.
19. W.A. PEARLMAN and R.M. GRAY, "Source coding of the discrete Fourier transform," IEEE Transactions on Information Theory, vol. IT-24, pp. 683-692, November 1978.
20. A.H. GRAY, Jr. and J.D. MARKEL, " Distance Measures for Speech Processing," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-24, pp.380-391, October 1976.
21. F. ITAKURA and SAITO, " Analytic synthesis telephony based upon the maximum likelihood method," Conf. Rec., 6th Int. Congr. Acoust., Y. Yonasi, Ed., Tokyo, Japan, 1968.
22. L.R. RABINER and R.W. SCHAFER, Digital Processing of Speech Signals, Prentice-Hall, New Jersey 07632.



### Parámetros de la TDF.

1. Orden de la transformada .....	8
2. Ventana de análisis .....	256
3. Ventana de corrimiento .....	240
4. Translape .....	16
5. Máximo número de formantes .....	6
6. Componentes por banda .....	17
7. Tipo de ventana .....	Trapezoidal

### Parámetros del predictor lineal.

1. Orden del predictor .....	16
2. Ventana de análisis .....	320
3. Ventana de corrimiento .....	240
4. Translape .....	80
5. Tipo de ventana .....	Hanning

Bandas del vector de coeficientes de reflexión.

Coeficientes

	1 - 3	4 - 8	9 - 12	13 - 17
Bits	10	10	9	6

total de bits = 35

Bandas de factores  $f_1$ .

Primer formante.

Coeficientes

	1 - 4	5 - 7	8 - 10	11 - 13	14 - 17
Bits	9	9	9	9	9

total de bits = 45

Coeficientes

	1 - 4	5 - 8	9 - 12	13 - 17
Bits	9	9	9	9

total de bits = 36

Coeficientes

	1 - 6	7 - 12	13 - 17
Bits	9	9	9

total de bits = 27

Segundo formante

Coeficientes

	1 - 6	7 - 12	13 - 17
Bits	9	9	9

total de bits = 27

Bandas de fases

Primer formante

Coeficientes

	1 - 3	4 - 6	7 - 9	10 - 13	14 - 17
Bits	10	10	10	10	10

total de bits = 50