



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 266 843**

51 Int. Cl.:  
**G10L 19/02** (2006.01)  
**G10L 19/04** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **03745516 .9**  
86 Fecha de presentación : **14.02.2003**  
87 Número de publicación de la solicitud: **1495465**  
87 Fecha de publicación de la solicitud: **12.01.2005**

54 Título: **Métodos para modelar magnitudes de los armónicos del habla.**

30 Prioridad: **28.03.2002 US 109151**

45 Fecha de publicación de la mención BOPI:  
**01.03.2007**

45 Fecha de la publicación del folleto de la patente:  
**01.03.2007**

73 Titular/es: **Motorola, Inc., A Corporation of the  
State of Delaware  
1303 East Algonquin Road  
Schaumburg, Illinois 60196, US**

72 Inventor/es: **Ramabadran, Tenkasi V.;**  
**Smith, Aaron M. y**  
**Jasiuk, Mark A.**

74 Agente: **Ungría López, Javier**

ES 2 266 843 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Métodos para modelar magnitudes de los armónicos del habla.

5 **Campo de la invención**

Esta invención se refiere a técnicas para la codificación paramétrica o la compresión de señales de habla y, en particular, a técnicas para modelar magnitudes de los armónicos del habla.

10 **Antecedentes de la invención**

En muchos vocodificadores (codificadores de voz) paramétricos, tales como por ejemplo los Vocodificadores Sinusoidales y los Vocodificadores de Excitación Multi-Banda, las magnitudes de los armónicos de la señal del habla forman un conjunto de parámetros importante a partir del cual se puede sintetizar la señal de habla. En el caso de sonidos sonoros, éstos son las magnitudes de los armónicos de frecuencia fundamental. En el caso de sonidos sordos, éstos son típicamente las magnitudes de los armónicos de una frecuencia muy baja (menor o igual a la menor frecuencia fundamental). En el caso de sonidos de voz mixtos, éstos son las magnitudes de los armónicos fundamentales en la banda de baja frecuencia y los armónicos de una frecuencia muy baja en la banda de alta frecuencia.

Es importante una representación eficiente y precisa de las magnitudes de los armónicos para asegurar la alta calidad de la voz en los vocodificadores paramétricos. Debido a que la frecuencia fundamental cambia de persona a persona e incluso en la misma persona dependiendo de las palabras, el número de armónicos necesario para representar el habla es variable. Asumiendo un ancho de banda de la voz de 3,7 kHz, una frecuencia de muestreo de 8 kHz., y un margen para la frecuencia fundamental desde 57 Hz. hasta 420 Hz. (periodo fundamental: 19 a 139), el número de armónicos vocales puede variar entre 8 y 64. Este número variable de magnitudes de armónicos hace su representación bastante complicada.

Se han desarrollado numerosas técnicas para la representación eficiente de magnitudes de armónicos vocales. Pueden clasificarse de manera general en a) Cuantificación Directa, y b) Cuantificación Indirecta mediante un modelo. En la cuantificación directa, se utilizan técnicas de cuantificación escalar o vectorial (VQ) para cuantificar directamente las magnitudes de los armónicos. Un ejemplo es la técnica de cuantificación vectorial de Transformada No-Cuadrática descrita en "Non-Square Transform Vector Quantization for Low-Rate Speech Coding", P. Lupini and V. Cuperman, Proceedings of the 1995 IEEE Workshop on Speech Coding for Telecommunications, páginas. 87-88, Septiembre 1995. En esta técnica, el vector de dimensión variable de magnitudes (logarítmicas) de los armónicos se transforma en un vector de dimensión fija, cuantificado vectorialmente, y transformado de nuevo en un vector de dimensión variable. Otro ejemplo es la VQ de Dimensión Variable o técnica VDVQ descrita en "Variable-Dimension Vector Quantization of Speech Spectra for Low-Rate Vocoders", A. Das, A. Rao, and A. Gersho, Proceedings of the IEEE Data Compresión Conference, páginas 420-429, Abril 1994. En esta técnica, el conjunto de códigos de la VQ consiste en vectores de alta resolución cuya dimensión es como mínimo igual a la mayor dimensión de los vectores de magnitudes (logarítmicas) que hay que cuantificar. Para cualquier dimensión dada, los vectores de código primero se submuestran a la dimensión correcta y después se utilizan para cuantificar el vector de magnitud (logarítmica).

En la cuantificación indirecta, las magnitudes de los armónicos primero se modelan por otro conjunto de parámetros, y después esos parámetros del modelo son los que se cuantifican. Un ejemplo de esta aproximación puede encontrarse en el vocodificador IMBE descrito en "APCO Project 25 Vocoder Description", TIA/EIA Interim Standard, Julio 1993. En primer lugar se predicen las magnitudes (logarítmicas) de los armónicos de una trama de voz mediante las magnitudes (logarítmicas) cuantificadas correspondientes a la trama anterior. A continuación se dividen las magnitudes del error (de predicción) en seis grupos, y cada grupo se transforma mediante una DCT (Transformada de Coseno Discreto). Se toma el primer (componente continua, DC) coeficiente de cada grupo y se vuelven a transformar mediante otra DCT. Los coeficientes de esta segunda DCT junto con los coeficientes de mayor orden de las primeras seis DCTs se cuantifican escalarmenete. Dependiendo del número de magnitudes de armónicos, se cambian tanto el tamaño del grupo como los bits asignados a cada coeficiente individual de la DCT, manteniendo constante el número total de bits. Otro ejemplo se puede encontrar en el Vocodificador de Transformada Sinusoidal descrito en "Low-Rate Speech Coding Based on the Sinusoidal Model", R. J. McAulay and T. F. Quatieri, Advances in Speech Signal Processing, Eds. S. Furui and M. M. Sondhi, pp. 165-208, Marcel Dekker Inc., 1992. Primero, se obtiene la envolvente de las magnitudes de los armónicos y se calcula el (Mel) Cepstrum de esta envolvente. A continuación, la representación cepstral se trunca (a M valores) y se transforma de nuevo al dominio de la frecuencia usando una transformada del Coseno. Los M valores del dominio de la frecuencia (denominados ganancias de canal) se cuantifican utilizando técnicas DPCM (Modulación por Codificación Diferencial de Pulsos).

Un modelo popular para representar la envolvente espectral del habla es el modelo todo-polos, el cual se estima típicamente utilizando métodos de predicción lineal. Es sabido en la bibliografía que el muestreo de la envolvente espectral por los armónicos de la frecuencia fundamental introduce un sesgo en la estimación de los parámetros del modelo. Se han desarrollado numerosas técnicas para minimizar el error de estimación. Un ejemplo de estas técnicas es Modelado Discreto Todo-Polos (DAP) descrito en "Discrete All-Pole Modeling". A. El-Jaroudi and Jaroudi and J. Makhoul. IEEE Trans. On Signal Processing, Vol. 39, No. 2, pp. 411-423, Febrero 1991. Dado un conjunto discreto de muestras espectrales (o magnitudes armónicas), esta técnica usa una condición mejorada de correspondencia de autocorrelación para obtener los parámetros del modelo todo-polos mediante un método iterativo. Otro ejemplo es

la técnica de Interpolación Espectral Predictiva Lineal (EILP) presentada en “*Spectral Envelope Sampling and Interpolation in Linear Predictive Analysis of Speech*”, H. Hermansky, H. Fujisaki, and Y. Sato, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, páginas. 2.2.1-2.2.4, Marzo 1984. En esta técnica, primero se interpolan las magnitudes armónicas utilizando un método de interpolado parabólico promediado. A continuación, se utiliza una Transformada Discreta de Fourier Inversa para transformar la envolvente espectral de potencia (interpolada) a una secuencia de auto-correlación. Los parámetros del modelo todo-polos, por ejemplo, los coeficientes del indicador, se calculan utilizando un método estándar LP, como una recursión Levinson-Durbin.

### Breve descripción de los dibujos

Las características novedosas que se suponen características de la invención se muestran más adelante en las reivindicaciones. Sin embargo, la invención, al igual que el modo de uso preferido, y las ventajas e inconvenientes adicionales del mismo, se entenderán mejor mediante la referencia a la descripción detallada de una realización ilustrativa junto a los dibujos adjuntos, donde:

La Fig. 1 es un diagrama de flujo de la realización preferida de un método para modelar las magnitudes de los armónicos del habla de acuerdo con la presente invención.

La Fig. 2 es una representación esquemática de la realización preferida de un sistema para modelar las magnitudes de los armónicos del habla de acuerdo con la presente invención.

La Fig. 3 es una gráfica de una forma de onda de habla de ejemplo.

La Fig. 4 es una gráfica del espectro de la forma de onda de habla ejemplar que muestra las magnitudes de los armónicos de habla.

La Fig. 5 es una gráfica de una pseudo secuencia de autocorrelación, de acuerdo con un aspecto de la presente invención.

La Fig. 6 es una gráfica de la envolvente espectral derivada de acuerdo con la presente invención.

### Descripción de la invención

Aunque esta invención es susceptible de realización de muchas formas diferentes, se muestra en los esquemas y aquí se describirá en detalle una o más realizaciones específicas, entendiéndose que la presente descripción se considerará como ejemplo de los principios de la invención y no se entenderá como limitante de la invención a las realizaciones específicas mostradas y descritas. En la descripción mostrada a continuación, se usan números de referencia para describir las partes iguales, similares o correspondientes en las diversas representaciones de los esquemas.

La presente invención proporciona un método de modelado todo-polos para representar las magnitudes armónicas del habla. El método utiliza un método iterativo para mejorar la precisión del modelado respecto a técnicas anteriores. El método de la invención puede ser referido como un método Iterativo, Interpolativo, Transformado (o IIT).

La Fig. 1 es un diagrama de flujo de una realización preferida de un método para modelar las magnitudes armónicas del habla de acuerdo con una realización de la presente invención. A continuación del bloque de comienzo 102, una trama de muestras de habla se transforma en el bloque 104 para obtener el espectro de la trama de habla. La frecuencia fundamental y las magnitudes armónicas que se desean modelar se encuentran en el bloque 106. Las K magnitudes armónicas se denotan por  $\{M_1, M_2, \dots, M_K\}$ . Claramente,  $M_k \geq 0$  para  $k = 1, 2, \dots, K$ . Similarmente, las frecuencias armónicas se denotan por  $\{\omega_1, \omega_2, \dots, \omega_K\}$ . Típicamente, las frecuencias armónicas son múltiplos de la frecuencia fundamental  $\omega_1$  para habla sonora, por ejemplo,  $\omega_k = k * \omega_1$  para  $k = 1, 2, \dots, K$ , pero el método mismo puede acomodar cualquier conjunto arbitrario de frecuencias. Para propósitos de transformación, se define un conjunto de frecuencias fijas  $\{i * \pi/N\}$  para  $i = 0, 1, \dots, N$ . El valor de N se elige para ser lo suficientemente grande para capturar la información de la envolvente espectral contenida en las magnitudes de los armónicos y para dar una resolución de muestreo adecuada, a saber,  $\pi/N$ , a la envolvente espectral. Por ejemplo, si el número de armónicos K varía entre 8 a 64, N puede elegirse como 64. Antes de introducirse al algoritmo, las frecuencias armónicas se modifican en el bloque 108. Las frecuencias armónicas modificadas se denotan por  $\{\theta_1, \theta_2, \dots, \theta_K\}$ , que se calculan de acuerdo con la fórmula de interpolación lineal

$$\theta_k = \pi/N + [(\omega_k - \omega_1)/(\omega_k - \omega_1)] * [(N - 2) * \pi/N], k = 1, 2, 3, \dots, K$$

De esta manera,  $\omega_1$  se corresponde a  $\pi/N$ , y  $\omega_k$  se corresponde a  $(N-1) * \pi/N$ . En otras palabras, las frecuencias armónicas en el rango de  $\omega_1$  a  $\omega_K$  se modifican para cubrir el rango de  $\pi/N$  a  $(N-1) * \pi/N$ . La correspondencia de arriba de las frecuencias armónicas originales a frecuencias armónicas modificadas aseguran que todas las frecuencias fijas que no sean las frecuencias D.C. (0) y de pliegue ( $\pi$ ) pueden encontrarse por interpolación. Se podrían utilizar otras correspondencias. En una realización posterior, no se utiliza ninguna correspondencia, y las magnitudes espectrales a las frecuencias fijas se encuentran por interpolación o extrapolación a partir de la original, por ejemplo, frecuencias armónicas sin modificar.

## ES 2 266 843 T3

En el bloque 110, los valores de magnitudes espectrales en las frecuencias fijas se calculan mediante interpolación (y extrapolación si fuera necesario) de las magnitudes armónicas conocidas. Las magnitudes espectrales a las frecuencias fijas se denotan por  $\{P_0, P_1, \dots, P_N\}$  correspondientes a las frecuencias  $\{i * \pi/N\}$  para  $i = 0, 1, \dots, N$ . Evidentemente, las magnitudes  $P_1$  y  $P_{N-1}$  están dadas por  $M_1$  y  $M_K$  respectivamente. Las magnitudes a las frecuencias fijas  $i * \pi/N$ ,  $i = 2, 3, \dots, N-2$  se calculan mediante interpolación de los valores conocidos a las frecuencias armónicas modificadas. Por ejemplo, si  $i * \pi/N$  cae entre  $\theta_k$  y  $\theta_{k+1}$ , la magnitud a la  $i$ -ésima frecuencia fija está dada por:

$$P_i = M_k + [(i * \pi/N) + \theta_k] / (\theta_{k+1} - \theta_k) * (M_{k+1} - M_k)$$

Aquí, se ha utilizado interpolación lineal, pero se podrían utilizar otros tipos de interpolación sin salirse de la invención. Las magnitudes  $P_0$  y  $P_N$  a las frecuencias  $0$  y  $\pi$  se calculan mediante extrapolación. Un método simple es asignar  $P_0$  igual a  $P_1$  y  $P_N$  igual a  $P_{N-1}$ . Otro método es usar extrapolación lineal. Usar  $P_1$  y  $P_2$  para calcular  $P_0$  da  $P_0 = 2 * P_1 - P_2$ . De manera similar, utilizando  $P_{N-2}$  y  $P_{N-1}$  para calcular  $P_N$ , obtenemos  $P_N = 2 * P_{N-1} - P_{N-2}$ . Por supuesto,  $P_0$  y  $P_N$  están limitados a ser mayores o iguales a cero.

En la realización descrita más arriba para bloques 108 y 110, el valor de  $N$  es fijo para diferentes  $K$  y no hay garantía de que las magnitudes armónicas que no sean  $M_1$  y  $M_K$  formen parte del conjunto de magnitudes a las frecuencias fijas, a saber,  $\{P_0, P_1, \dots, P_N\}$ . En otra realización, el valor de  $N$  se hace en función de  $K$ , a saber,  $N = (K-2) * I + 2$ , donde  $I \geq 1$  se llama el factor de interpolación. Con este valor de  $N$ , cuando las frecuencias armónicas se modifican de acuerdo con la fórmula de interpolación lineal

$$\theta_k = \pi/N + [(\omega_k - \omega_1) / (\omega_k - \omega_1)] * [(N - 2) * \pi/N], k = 1, 2, 3, \dots, K$$

en el bloque 108,  $\omega_1$  se asigna a  $\pi/N$ ,  $\omega_2$  a  $(I+1) * \pi/N$ ,  $\omega_3$  a  $(2*I+1) * \pi/N$ , y así sucesivamente hasta que  $\omega_K$  se asigne a  $((K-1)*I+1) * \pi/N = (N-1) * \pi/N$ . De esta forma, las frecuencias modificadas  $\{\theta_1, \theta_2, \dots, \theta_K\}$  a partir de un subconjunto de las frecuencias fijas  $\{i * \pi/N\}$ ,  $i = 0, 1, \dots, N$ . Correspondientemente, en el bloque 110, cuando se calculan los valores de magnitud espectral en las frecuencias fijas, las magnitudes armónicas  $\{M_1, M_2, \dots, M_K\}$  forman un subconjunto de las magnitudes espectrales en las frecuencias fijas, a saber,  $\{P_0, P_1, \dots, P_N\}$ . En la realización preferida, el valor del factor de interpolación  $I$  se elige como 4 para ( $K < 12$ ), 3 para ( $12 \leq K < 16$ ), 2 para ( $16 \leq K < 24$ ) y 1 para ( $K \geq 24$ ).

En el bloque 112, una transformada inversa se aplica a los valores de magnitud en las frecuencias fijas para obtener una (pseudo) secuencia de auto-correlación. Dadas las magnitudes en las frecuencias fijas  $\{i * \pi/N\}$ ,  $i = 0, 1, \dots, N$ , se usa una DFT (Transformada de Fourier Discreta) inversa de  $2N$  puntos para calcular una secuencia de auto-correlación asumiendo que el dominio de la frecuencia es real y par, por ejemplo,  $P_{-i} = P_i$ . Dado que la secuencia en el dominio de la frecuencia es real y par, la correspondiente secuencia en el dominio del tiempo es también real y par, como debería ser para una secuencia de auto-correlación. Sin embargo, debe resaltarse que los valores en el dominio de la frecuencia en la realización preferida son magnitudes en vez de valores de potencia (o energía), y por lo tanto la secuencia en el dominio del tiempo no es una secuencia de autocorrelación real. Por lo tanto, nos referimos a ella como una pseudo secuencia de auto-correlación. El espectro en magnitud es la raíz cuadrada del espectro de potencia y es más plano. En una realización posterior, se utiliza un espectro en magnitud logarítmica, y en otra realización el espectro en magnitud podría elevarse a un exponente diferente de 1,0.

Si  $N$  es una potencia de 2, se puede usar un algoritmo FFT (Transformada Rápida de Fourier) para calcular la DFT inversa de  $2N$ -puntos. Sin embargo, sólo se necesitan los primeros  $J+1$  valores de auto-correlación, donde  $J$  es el orden del indicador (o modelo). Dependiendo del valor de  $J$ , un cálculo directo de la DFT inversa podría ser más eficiente que una FFT. Si denotamos por  $\{R_0, R_1, \dots, R_J\}$  los primeros  $J+1$  valores de la secuencia de pseudo autocorrelación, entonces,  $R_j$  viene dada por:

$$R_j = P_0 + (-1)^j * P_N + 2 * \sum_{i=1}^{i=N-1} P_i * \cos(i * j * \pi / N).$$

En el bloque 114, los coeficientes del indicador  $\{a_1, a_2, \dots, a_J\}$  se calculan como la solución de las ecuaciones normales

$$\sum_{j=1}^{j=J} a_j * R(i - j) = R_i, \text{ para } i = 1, 2, \dots, J$$

En la realización preferida, se utiliza una recursión Levinson-Durbin para resolver esas ecuaciones, como se describe en "Discrete-Time Processing of Speech Signals", J.R. S  ller, Jr., J.G. Proakis, and J.H.L. Hansen, Macmillan, 1993.

## ES 2 266 843 T3

En el bloque de decisión 116, se realiza una comprobación para determinar si son necesarias más iteraciones. Si no, como se muestra en la rama negativa del bloque de decisión 116, el método termina en el bloque 128. Los coeficientes del indicador  $\{a_1, a_2, \dots, a_j\}$  parametrizan las magnitudes armónicas. Los coeficientes podrían estar codificados mediante técnicas de codificación conocidas para formar una representación compacta de las magnitudes armónicas. En la realización conocida, la clase de voz, la frecuencia fundamental, y un valor de ganancia se utilizan para completar la descripción de la trama de habla.

Si es necesaria una iteración adicional, como se muestra en la rama positiva del bloque de decisión 116, la envolvente espectral definida por los coeficientes del indicador se muestrea en el bloque 118 para obtener las magnitudes modeladas en las frecuencias armónicas modificadas. Denotemos por  $A(z) = 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_jz^{-j}$  el filtro de error de predicción, donde  $z$  es la variable estándar de la transformada  $Z$ . La envolvente espectral en la frecuencia  $\omega$  entonces viene dada (exacta con un factor constante de ganancia) por  $1,0 / |A(z)|^2$  con  $z = e^{j\omega}$ . Para obtener las magnitudes modeladas a las frecuencias armónicas modificadas  $\theta_k, k = 1, 2, \dots, K$ , la envolvente espectral se muestrea a estas frecuencias. Las magnitudes resultantes se denotan por  $\{M_1, M_2, \dots, M_K\}$ .

Si las variables en el dominio de la frecuencia que se usaron para obtener la secuencia de pseudo auto-correlación no son magnitudes armónicas sino alguna función de las magnitudes, son necesarias operaciones adicionales para obtener las magnitudes modeladas después de muestrear la envolvente espectral.

En el bloque 120 se calculan factores de escala en las frecuencias armónicas modificadas para ajustar las magnitudes modeladas y las magnitudes armónicas conocidas a esas frecuencias. Antes de calcular los factores de escala, es necesario asegurarse de que las magnitudes conocidas y las magnitudes modeladas en las frecuencias armónicas modificadas están normalizadas de una manera adecuada. Una forma simple es usar normalización de energía, por ejemplo,  $\sum |M_k|^2 = \sum |M_k|^2$ . Otra aproximación simple es forzar que el valor de pico sea el mismo, por ejemplo,  $\max(\{M_k\}) = \max(\{M_k\})$ . Cualquiera que sea el método de normalización utilizado, la misma normalización debe aplicarse a las magnitudes modeladas a las frecuencias fijas.

Los  $K$  factores de escala son calculados como  $S_k = M_k / M_k, k = 1, 2, \dots, K$ . Si, para algún  $k, M_k = 0$ , entonces el correspondiente  $S_k$  se toma como 1,0.

En el bloque 122 los factores de escala en las frecuencias armónicas modificadas se interpolan para obtener los factores de escala en las frecuencias fijas. Los factores de escala en las frecuencias fijas ( $i * \pi/N$ ),  $i = 0, 1, \dots, N$  se denotan como  $\{T_0, T_1, \dots, T_N\}$ . Los valores  $T_0$  y  $T_N$  se asignan a 1,0. Los otros valores se calculan mediante interpolación de los valores conocidos en las frecuencias armónicas modificadas. Por ejemplo, si  $i * \pi/N$  cae entre  $\theta_k$  y  $\theta_{k+1}$ , el factor de escala en la  $i$ -ésima frecuencia fija está dada por

$$T_i = S_k + \left[ \frac{(i * \pi / N) - \theta_k}{(\theta_{k+1} - \theta_k)} * (S_{k+1} - S_k) \right], \text{ para } i = 1, 2, \dots, N-1.$$

En el bloque 124 la envolvente espectral se muestrea para obtener las magnitudes modeladas a las frecuencias fijas ( $i * \pi/N$ ),  $i = 0, 1, \dots, N$ . Las magnitudes modeladas en las frecuencias fijas se denotan por  $\{P_0, P_1, \dots, P_N\}$ .

En el bloque 126, un nuevo conjunto de magnitudes en las frecuencias fijas se calcula multiplicando las magnitudes modeladas (y normalizadas) en esas frecuencias por los factores de escala correspondientes, por ejemplo,  $P_i = \underline{P}_i * T_i, i = 0, 1, \dots, N$ .

El flujo vuelve al bloque 112, donde se aplica una transformada inversa al nuevo conjunto de magnitudes en las frecuencias fijas y se encuentran los coeficientes del indicador en el bloque 114.

Cuando se completa el proceso iterativo, los coeficientes del indicador obtenidos en el bloque 114 son los parámetros del modelo todo-polos que se necesitaban. En el decodificador correspondiente, las magnitudes armónicas modeladas se calculan muestreando la envolvente espectral en las frecuencias armónicas modificadas.

Para un orden dado del modelo, la precisión de modelado generalmente mejora con el número de iteraciones llevadas a cabo. La mayoría de la ganancia, sin embargo, se realiza después de una única iteración. La invención proporciona un método de modelado todo-polos para representar un conjunto de magnitudes armónicas del habla. A través de un procedimiento iterativo, el método mejora la curva de interpolación que se usa en el dominio de la frecuencia. Medido en términos de distorsión espectral, la exactitud de modelado de este método ha demostrado ser mejor que métodos conocidos anteriores.

En la realización descrita más arriba, se asume que  $N > J+1$ , lo cual se cumple normalmente. Los  $J$  coeficientes del indicador  $\{a_1, a_2, \dots, a_j\}$  modelan las  $N+1$  magnitudes espectrales en las frecuencias fijas, a saber,  $\{P_0, P_1, \dots, P_N\}$ , y por lo tanto, las  $K$  magnitudes armónicas  $\{M_0, M_1, \dots, M_K\}$  con cierto error de modelado. Una posterior realización utiliza un valor de  $J$  tal que  $K \leq J+1$ . En esta realización es posible modelar las magnitudes armónicas exactamente (a excepción de un factor de ganancia) como sigue. Si  $K < J+1$ , se añaden algunos valores ( $\geq 0$ ) de magnitudes

## ES 2 266 843 T3

armónicas de relleno, de tal forma que  $K = J+1$ .  $N$  se elige como  $N = K-1 = J$ , y las frecuencias armónicas se asignan de tal forma que  $\omega_1$  corresponda a  $0*\pi/N$ ,  $\omega_2$  a  $1*\pi/N$ ,  $\omega_3$  a  $2*\pi/N$ , y así sucesivamente, y finalmente  $\omega_K$  a  $(K-1)*\pi/N = \pi$ . De esta manera, las magnitudes armónicas  $\{M_1, M_2, \dots, M_K\}$  corresponden exactamente al conjunto  $\{P_0, P_1, \dots, P_N\}$ . En el bloque 112, el conjunto  $\{P_0, P_1, \dots, P_N\}$  se transforma en el conjunto  $\{R_0, R_1, \dots, R_J\}$  mediante la DFT inversa que es invertible. En el bloque 114, el conjunto  $\{R_0, R_1, \dots, R_J\}$  se transforma en el conjunto  $\{a_1, a_2, \dots, a_J\}$  mediante una recursión Levinson-Durbin que es también invertible a excepción de un factor constante. De esta forma los coeficientes del indicador  $\{a_1, a_2, \dots, a_J\}$  modelan las magnitudes armónicas  $\{M_1, M_2, \dots, M_K\}$  exactamente a excepción de un factor constante. No es necesaria ninguna iteración adicional. No hay error de modelado en este caso. Cualquier codificación, por ejemplo, cuantificación, de los coeficientes del indicador podría introducir algún error de codificación. Para obtener las magnitudes armónicas a partir de los coeficientes del indicador, los coeficientes del indicador  $\{a_1, a_2, \dots, a_J\}$  se transforman a  $\{R_0, R_1, \dots, R_J\}$  y entonces  $\{R_0, R_1, \dots, R_J\}$  se transforman a  $\{P_0, P_1, \dots, P_N\}$ , que son iguales a  $\{M_1, M_2, \dots, M_K\}$  mediante las transformaciones inversas apropiadas.

La Fig. 2 muestra una realización preferida de un sistema para modelar las magnitudes armónicas del habla de acuerdo con una realización de la presente invención. Haciendo referencia a la Fig. 2, el sistema tiene una entrada 202 para una recibir una trama de habla, y un analizador armónico 204 para calcular las magnitudes armónicas 206 y frecuencias armónicas 208 del habla. Las frecuencias armónicas se transforman en el modificador de frecuencia 210 para obtener frecuencias armónicas modificadas 212. Las magnitudes armónicas 206 y las frecuencias armónicas modificadas 212 se pasan al interpolador 214, donde se calculan las magnitudes espectrales a las frecuencias fijas  $F = \{0, \pi/N, 2\pi/N, \dots, \pi\}$  (216). Las magnitudes espectrales 218 en las frecuencias fijas se pasan al transformador inverso de Fourier 220, donde se aplica una transformada inversa para obtener una pseudo secuencia de autocorrelación 222. Un análisis LP de la pseudo secuencia de autocorrelación se realiza con un analizador LP 224 para dar como resultado los coeficientes del indicador 225. Los coeficientes de predicción 225 se pasan a un cuantificador de coeficientes o codificador 226. Esto produce los coeficientes cuantificados 228 como salida. Los coeficientes de predicción cuantizados 228 (o los coeficientes de predicción 225) y las frecuencias armónicas modificadas 212 se suministran al calculador de espectro 230 que calcula las magnitudes modeladas 232 en las frecuencias armónicas modificadas muestreando la envolvente espectral correspondiente a los coeficientes de predicción.

Los coeficientes de predicción finales podrían cuantificarse o codificarse antes de ser guardados o transmitidos. Cuando la señal de habla se recupera mediante síntesis, se utilizan los coeficientes cuantificados o codificados. Por consiguiente, un cuantificador o codificador/decodificador se aplica a los coeficientes 225 en una posterior realización. Esto asegura que el modelo producido por los coeficientes cuantificados es tan exacto como sea posible.

A partir de las magnitudes armónicas modeladas 232 y las magnitudes armónicas reales 206, el calculador de escala 234 calcula un conjunto de factores de escala 236. El calculador de escala también calcula un valor de ganancia o valor de normalización como se ha descrito más arriba en referencia a Fig. 1. Los factores de escala 235 se interpolan por el interpolador 238 en las frecuencias fijas 216 para dar los factores de escala interpolados 240.

Los coeficientes de predicción cuantificados 228 (o los coeficientes de predicción 225) y las frecuencias fijas 216 también se proporcionan al calculador de espectro 242 que calcula las magnitudes modeladas 244 en las frecuencias fijas muestreando la envolvente espectral.

Las magnitudes modeladas 244 en las frecuencias fijas y los factores de escala interpolados 240 se multiplican en el multiplicador 246 para dar como resultado el producto  $P \cdot T$ , 248. El producto  $P \cdot T$  se devuelve al transformador inverso 220 de manera que una iteración pudiera ser realizada.

Cuando el proceso de iteración se haya completado, los coeficientes del indicador cuantificados 238 se sacan como parámetros del modelo, junto con la clase de habla, la frecuencia fundamental, y el valor de ganancia.

Las Figs. 3-6 muestran resultados de ejemplo producidos por una realización del método de la invención. Fig. 3 es una gráfica de una forma de onda de habla muestreada a 8 kHz. El habla es sonora. Fig. 4 es una gráfica de la magnitud espectral de la forma de onda de habla. La magnitud se muestra en decibelios. Las magnitudes armónicas se denotan por los círculos en los picos del espectro. Los valores marcados con círculos son las magnitudes armónicas,  $M$ . La frecuencia fundamental es 102.5 Hz. Fig. 5 es una gráfica de la pseudo secuencia de autocorrelación,  $R$ .  $N = 64$  en este ejemplo. Los coeficientes del indicador se calculan a partir de  $R$ . Fig. 6 es una gráfica de la envolvente espectral en las frecuencias fijas, derivadas a partir de los coeficientes del indicador después de varias iteraciones. El orden del indicador es 14. También se muestran en la Fig. 6 los círculos que denotan las magnitudes de los armónicos,  $M$ . Puede verse que la envolvente espectral provee una buena aproximación a las magnitudes de los armónicos en las frecuencias armónicas.

La Tabla 1 muestra resultados de ejemplo calculados utilizando una base de datos de 3 minutos de habla de 32 pares de frases. La base de datos consta de 4 hablantes masculinos y 4 femeninos con 4 pares de frases cada uno. Solamente se han incluido tramas sonoras en los resultados, ya que son la clave para una buena calidad del habla de salida. En este ejemplo 4258 tramas eran sonoras de un total de 8726 tramas. Cada trama tenía una longitud de 22.5 ms. En la tabla, la presente invención (método ITT) se compara con el modelado discreto todo-polos (DAP) para varios órdenes del modelo diferentes.

# ES 2 266 843 T3

TABLA 1

*Orden del modelo vs. Distorsión media (dB)*

	Orden del Modelo	DAP	IIT			
		15 iteraciones	Ninguna iteración	1 iteración	2 iteraciones	3 iteraciones
5						
10	10	3.71	3.54	3.41	3.39	3.38
	12	3.34	3.27	3.10	3.06	3.03
	14	2.95	2.98	2.75	2.68	2.65
15	16	2.60	2.74	2.43	2.33	2.28

La distorsión D en dB se calcula como

$$D = \frac{1}{N} \sum_{i=1}^{i=N} D_i$$

donde

$$D_i = \sqrt{\frac{1}{K_i} \sum_{k=1}^{k=K_i} [20 * \log_{10}(M_{k,i}) - 20 * \log_{10}(\underline{M}_{k,i})]^2}$$

$M_{k,i}$  es la magnitud del armónico k-ésimo de la i-ésima trama, y  $\underline{M}_{k,i}$  es la magnitud modelada k-ésima de la trama i-ésima. Tanto la magnitud real como la modelada de cada trama primero se normalizan de manera que su media logarítmica sea cero.

La distorsión media se reduce por un método iterativo de la presente invención. Mucho de la mejora se obtiene después de una única iteración.

Aquellos de habilidad ordinaria en el arte reconocerán que la presente invención podría ser implementada en *software* ejecutándose en un procesador o utilizando componentes de *hardware* equivalentes tal y como *hardware* de propósito especial y/o procesadores dedicados, los cuales son equivalentes a la invención descrita y reivindicada. De manera similar, se podría utilizar ordenadores de propósito general, ordenadores basados en microprocesador, procesadores digitales de señal, microcontroladores, procesadores dedicados, circuitos *custom* (de diseño específico), ASICS y/o lógica dedicada implementada en hardware para construir implementaciones alternativas equivalentes de la presente invención.

Mientras que la invención se ha mostrado y descrito particularmente con referencia a una realización preferida, se entenderá por aquellos con experiencia en el arte que allí se podrían hacer varios cambios en la forma y detalle sin salirse del espíritu y ámbito de la invención. En particular, la invención podría usarse para modelar señales tonales de fuentes diferentes que no sean de habla. Las componentes frecuenciales de las señales tonales no necesitan estar relacionadas armónicamente, sino que pueden estar espaciadas irregularmente espaciadas.

Mientras que la invención ha sido descrita en conjunto con implementaciones específicas, es evidente que muchas alternativas, modificaciones, permutaciones y variaciones se harán aparentes a aquellos de experiencia en el arte a la luz de la descripción a continuación. Por consiguiente, se espera que la presente invención abarque todas esas alternativas, modificaciones y variaciones que caigan dentro del ámbito de las reivindicaciones añadidas.

REIVINDICACIONES

1. Un método para modelar una señal representada por una trama de muestras que comprenda las etapas de:

5 a. Identificar (106) una pluralidad de frecuencias armónicas de la señal;

b. Identificar (106) una pluralidad de magnitudes de armónicos correspondientes a las magnitudes espectrales de la señal en la pluralidad de frecuencias armónicas;

10 c. Interpolar (110) la pluralidad de magnitudes de los armónicos para obtener una pluralidad de magnitudes espectrales en un conjunto de frecuencias fijas;

15 d. Transformar inversamente (112) la pluralidad de magnitudes espectrales para obtener una pseudo secuencia de auto-correlación;

e. Calcular (114) los coeficientes de predicción lineal a partir de la pseudo secuencia de auto-correlación;

20 f. Calcular (118) las magnitudes de los armónicos del modelo mediante el muestreo de una envolvente espectral definida por los coeficientes de predicción lineal;

g. Calcular (120) un primer conjunto de factores de escala como el cociente de las magnitudes de los armónicos y las magnitudes de los armónicos del modelo;

25 h. Interpolar (122) el primer conjunto de factores de escala para obtener un segundo conjunto de factores de escala en el conjunto de frecuencias fijas;

i. Calcular (124) las magnitudes espectrales del modelo en el conjunto de frecuencias fijas muestreando la envolvente espectral definida por los coeficientes de predicción lineal en el conjunto de frecuencias fijas;

30 j. Multiplicar (126) las magnitudes espectrales del modelo en el conjunto de frecuencias fijas por el segundo conjunto de factores de escala para obtener una nueva pluralidad de magnitudes espectrales;

35 k. Transformar de manera inversa (112) la nueva pluralidad de magnitudes espectrales para obtener una nueva pseudo secuencia de auto-correlación; y

l. Calcular (114) nuevos coeficientes de predicción lineal a partir de la nueva pseudo secuencia de autocorrelación, donde la señal se modela por los nuevos coeficientes de predicción lineal.

40 2. Un método de acuerdo con la reivindicación 1, que además comprende:

Modificar la pluralidad de frecuencias armónicas para obtener una pluralidad de frecuencias armónicas modificadas,

45 donde la pluralidad de magnitudes espectrales en un conjunto de frecuencias fijas se calculan interpolando a partir de la pluralidad de las frecuencias armónicas modificadas al conjunto de frecuencias fijas.

50 3. Un método de acuerdo con la reivindicación 1, en donde el conjunto de frecuencias fijas incluye frecuencias fuera de la pluralidad de frecuencias armónicas, comprendiendo además:

Calcular magnitudes espectrales fuera de la pluralidad de frecuencias armónicas extrapolando a partir de la pluralidad de frecuencias armónicas.

55 4. Un método de acuerdo con la reivindicación 1, en donde la transformada inversa es una de una transformada inversa rápida de Fourier y una transformada inversa discreta de Fourier.

5. Un método de acuerdo con la reivindicación 1, en donde los coeficientes de predicción lineal se calculan mediante recursión de Levinson-Durbin.

60 6. Un método de acuerdo con la reivindicación 1, en donde la señal se modela posteriormente por una clase de habla, una frecuencia fundamental y un valor de ganancia.

65 7. Un método de acuerdo con la reivindicación 1, en donde los coeficientes de predicción lineal se cuantifican para obtener coeficientes de predicción lineal cuantificados, y en donde las magnitudes armónicas del modelo y las magnitudes espectrales del modelo se calculan a partir de los coeficientes de predicción lineal cuantificados.

## ES 2 266 843 T3

8. Un método de acuerdo con la reivindicación 1, en donde las magnitudes armónicas del modelo se normalizan para tener una de 1) la misma suma de cuadrados que la pluralidad de magnitudes armónicas y 2) el mismo valor de pico que la pluralidad de magnitudes armónicas.

5 9. Un método de acuerdo con la reivindicación 1, en donde en la interpolación de la pluralidad de magnitudes armónicas para obtener una pluralidad de magnitudes espectrales en un conjunto de frecuencias fijas se utiliza una interpolación lineal y no-lineal.

10 10. Un método de acuerdo con la reivindicación 1, en donde en la interpolación del primer conjunto de factores de escala para obtener un segundo conjunto de factores de escala en el conjunto de frecuencias fijas se utiliza una interpolación lineal y no-lineal.

15 11. El método de modelar una señal de acuerdo con la reivindicación 1, en donde la transformación inversa de la pluralidad de las magnitudes espectrales comprende:

i) calcular una pluralidad modificada de magnitudes espectrales en un conjunto de frecuencias fijas aplicando una función de modificación a la pluralidad de magnitudes espectrales en un conjunto de frecuencias fijas;

20 ii) transformar de forma inversa la pluralidad modificada de magnitudes espectrales para obtener la pseudo secuencia de autocorrelación.

12. Un método de acuerdo con la reivindicación 11, en donde la función de modificación es una de una función logarítmica y una función de potenciación.

25 13. Un sistema adaptado para modelar una señal de acuerdo al método de acuerdo con cualquiera de las reivindicaciones 1 a 12, que comprende:

Una entrada para recibir la señal;

30 Un medio con función de procesado que realiza cada una de las funciones de identificar la pluralidad de magnitudes de los armónicos, identificar la pluralidad de frecuencias armónicas, interpolar la pluralidad de magnitudes de los armónicos, transformar de manera inversa la pluralidad de magnitudes espectrales, calcular las magnitudes armónicas del modelo, calcular un primer conjunto de factores de escala, interpolar el primer conjunto de factores de escala, calcular las magnitudes espectrales del modelo, multiplicar las magnitudes espectrales del modelo, transformar de  
35 manera inversa la nueva pluralidad de magnitudes espectrales, y calcular los nuevos coeficientes de predicción lineal, y

Una salida para sacar los nuevos coeficientes de predicción lineal.

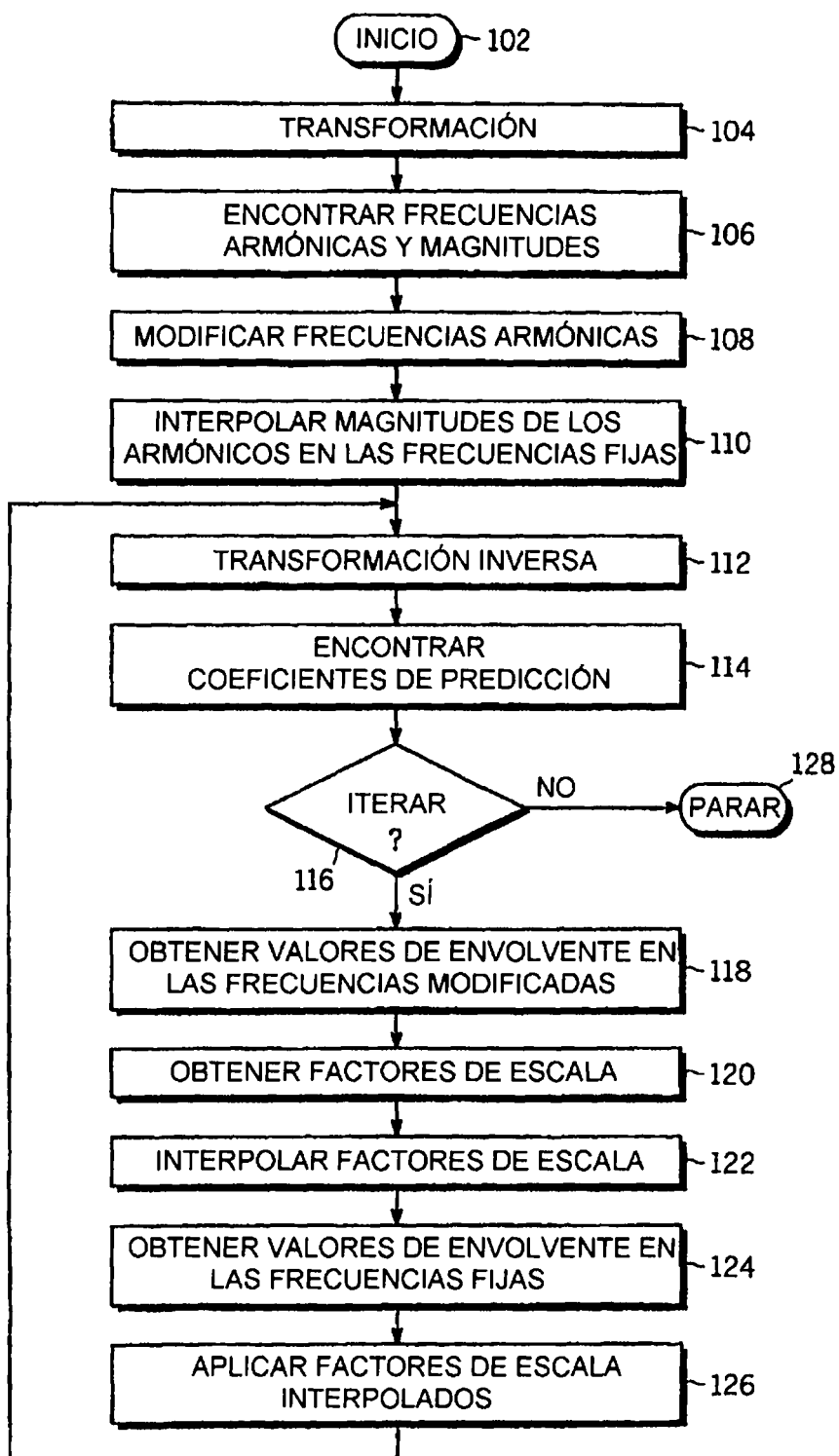
40 14. Un dispositivo adaptado para modelar una señal de acuerdo con el método de cualquiera de las reivindicaciones 1 a 12, en donde el dispositivo está dirigido por un programa de ordenador guardado en como mínimo uno de una memoria, un circuito integrado específico a la aplicación, un procesador digital de señal, y una FPGA (Matriz de puertas programable), en donde el programa de ordenador es operable para realizar cada una de las funciones de identificar la pluralidad de magnitudes de los armónicos, identificar la pluralidad de frecuencias armónicas, interpolar la pluralidad  
45 de magnitudes de los armónicos, transformar de manera inversa la pluralidad de magnitudes espectrales, calcular las magnitudes armónicas del modelo, calcular un primer conjunto de factores de escala, interpolar el primer conjunto de factores de escala, calcular las magnitudes espectrales del modelo, multiplicar las magnitudes espectrales del modelo, transformar de manera inversa la nueva pluralidad de magnitudes espectrales, y calcular los nuevos coeficientes de predicción lineal.

50 15. Un medio que pueda ser leído por un ordenador que contenga instrucciones que, cuando se manejen en un ordenador, lleven a cabo un proceso de modelar una pluralidad de magnitudes armónicas en una pluralidad de frecuencias armónicas de acuerdo a cualquiera de las reivindicaciones 1 a 12.

55

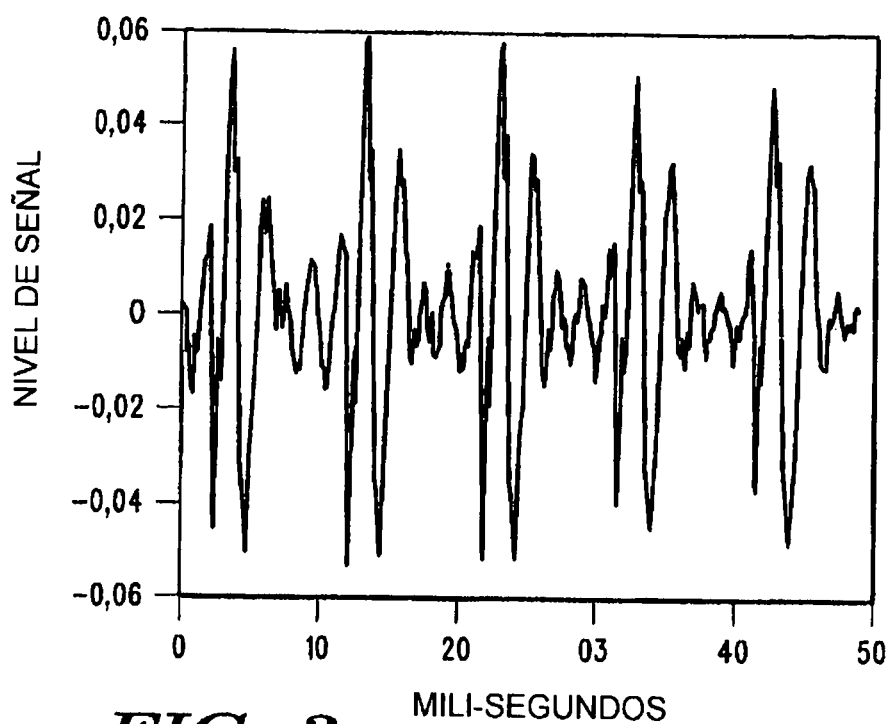
60

65

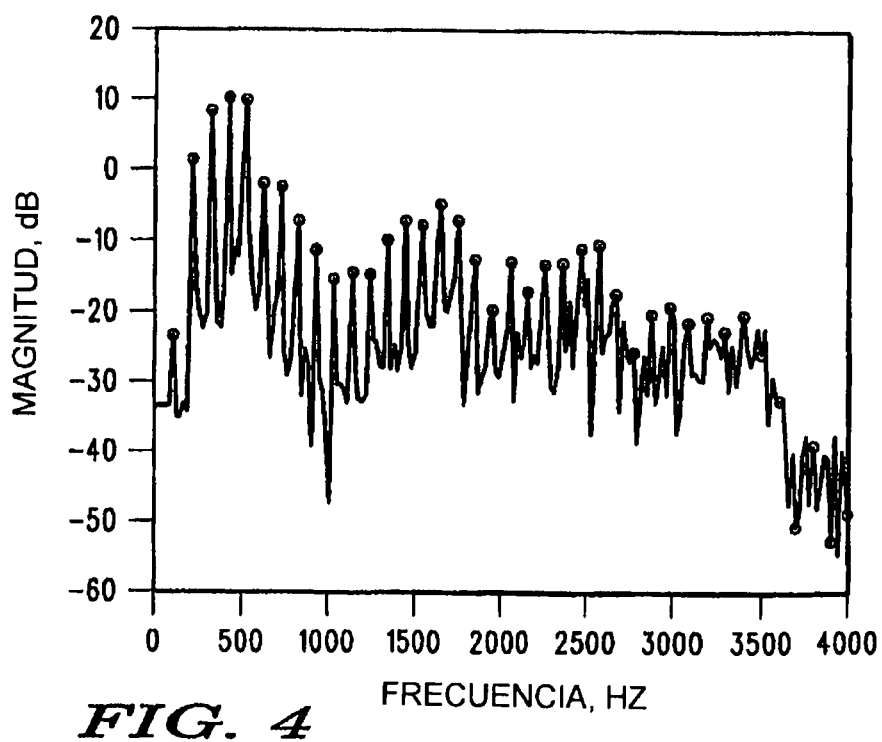


**FIG. 1**

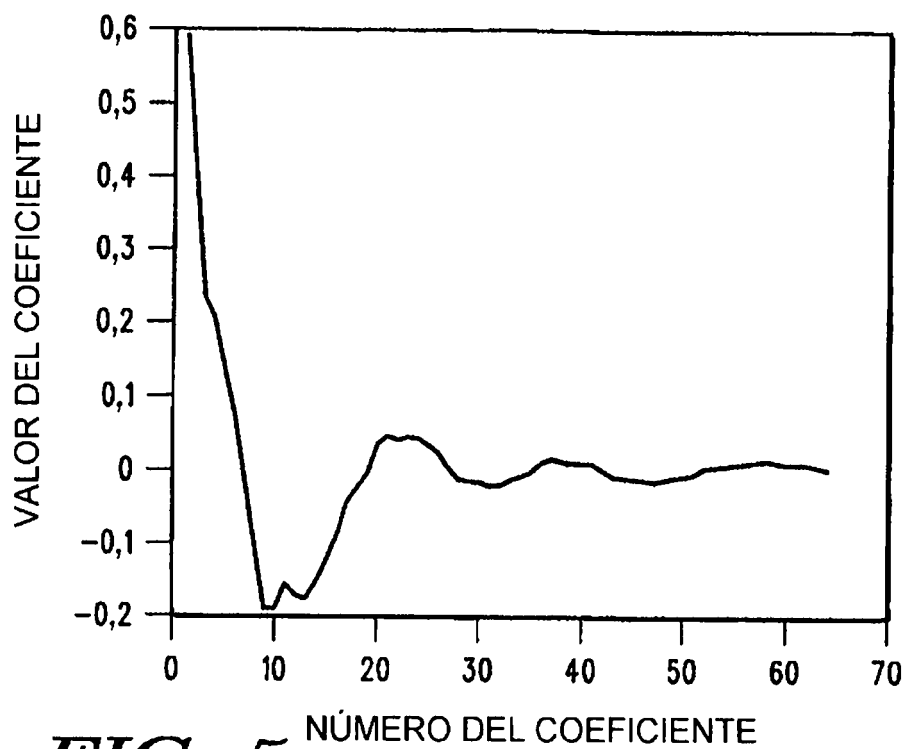




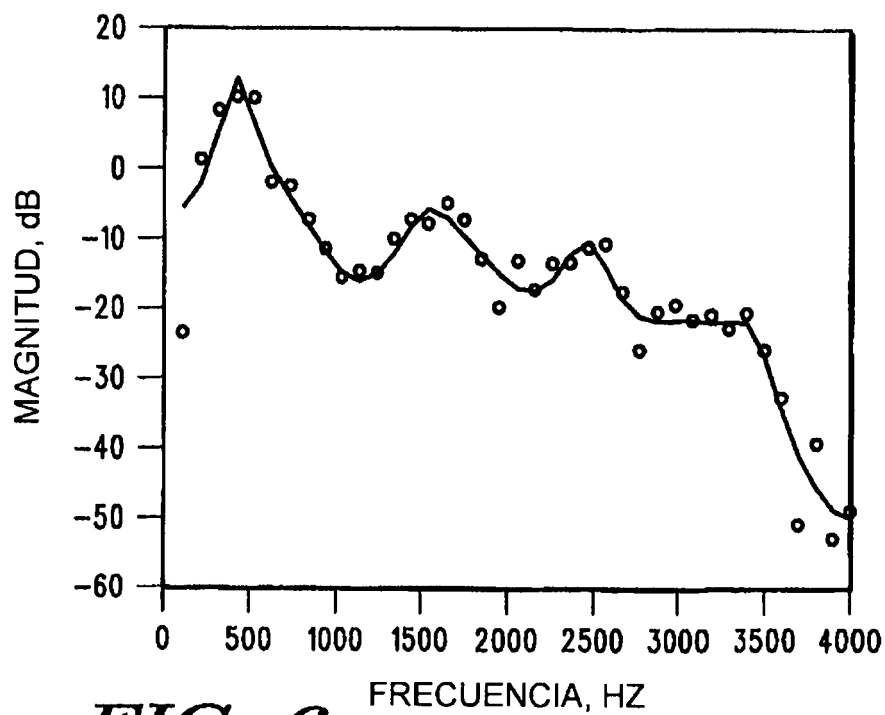
**FIG. 3**



**FIG. 4**



**FIG. 5**



**FIG. 6**