



(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 274 606**

(51) Int. Cl.:

G10L 19/06 (2006.01)

G10L 13/04 (2006.01)

G10L 19/08 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Número de solicitud europea: **99309294 .9**

(86) Fecha de presentación : **22.11.1999**

(87) Número de publicación de la solicitud: **1005021**

(87) Fecha de publicación de la solicitud: **31.05.2000**

(54) Título: **Procedimiento y aparato para obtener datos de fuente y filtro basados en formantes, para codificación y síntesis, utilizando función de coste y filtrado inverso.**

(30) Prioridad: **25.11.1998 US 200335**

(45) Fecha de publicación de la mención BOPI:
16.05.2007

(45) Fecha de la publicación del folleto de la patente:
16.05.2007

(73) Titular/es:
MATSUSHITA ELECTRIC INDUSTRIAL Co., Ltd.
1006, Oaza Kadoma
Kadoma City, Osaka 571, JP

(72) Inventor/es: **Pearson, Steve**

(74) Agente: **Espiell Volart, Eduardo María**

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

Procedimiento y aparato para obtener datos de fuente y filtro basados en formantes, para codificación y síntesis, utilizando función de coste y filtrado inverso.

Antecedentes y resumen de la invención

La presente invención se refiere, en general, a síntesis del habla y de formas de onda. Además, la invención se refiere a la obtención de datos de fuente y filtro, basados en formantes, a partir de formas de onda complejas. La tecnología de la invención se puede usar para construir sintetizadores de música y de texto a habla y sistemas de codificación del habla. Además, la tecnología se puede usar para realizar un rastreo de tonos de elevada calidad y un marcado de intervalos de tonos. Las funciones de coste utilizadas por la presente invención se pueden usar como funciones discriminatorias o detectores de características en el etiquetado del habla y el reconocimiento del habla.

Un modo de analizar y sintetizar formas de onda complejas, tales como formas de onda que representan habla sintetizada o instrumentos musicales, es utilizar un modelo de fuente y filtro. Usando el modelo de fuente y filtro, se genera una señal fuente y, posteriormente, se pasa por un filtro que añade resonancias y coloración a la señal fuente. La combinación de fuente y filtro, si se selecciona adecuadamente, puede producir una forma de onda compleja que simula el habla humana o el sonido de un instrumento musical.

En los modelos de fuente y filtro, la forma de onda fuente puede ser comparativamente simple, por ejemplo, ruido blanco o un tren de impulsos simple. En tal caso, normalmente el filtro es complejo. El filtro complejo es necesario porque es el efecto acumulativo de fuente y filtro que produce la forma de onda compleja. Alternativamente, la forma de onda fuente puede ser comparativamente compleja, en cuyo caso, el filtro puede ser más simple. En términos generales, la configuración fuente y filtro ofrece varias opciones de diseño.

Se prefiere un modelo que representa más rigurosamente el grado natural de separación entre una fuente glotal humana y el filtro del conducto vocal. Cuando se analiza la forma de onda compleja vocal humana, supone un gran reto determinar qué aspectos de la forma de onda se pueden atribuir a la fuente glotal y qué aspectos se pueden atribuir al filtro del conducto vocal. Se especula, en incluso se espera, que haya una interacción acústica entre el conducto vocal y la naturaleza de la forma de onda glotal que se genera en la glotis. En muchos casos dicha interacción puede ser insignificante, por consiguiente, en la síntesis es normal ignorar esta interacción, como si la fuente y el filtro fueran independientes.

Se considera que muchos sistemas de síntesis fallan debido a un modelo de fuente y filtro con un equilibrio insuficiente entre la complejidad de la fuente y la complejidad del filtro. Con frecuencia es la facilidad de generación, en lugar de la calidad de sonido, la que establece el modelo de fuente. Por ejemplo, la codificación predictiva lineal (LPC) se puede entender en términos de un modelo de fuente y filtro en el que la fuente suele ser blanca (es decir, espectro plano). Este modelo se elimina considerablemente de la separación natural entre el conducto vocal humano y la fuente glotal y tiene como resultado cálculos insuficientes del primer formante y muchas discontinuidades en los parámetros de filtrado.

Hasta este momento, un enfoque, tomado como una alternativa de la LPC para solucionar las deficiencias de la LPC, conlleva un procedimiento denominado "análisis mediante síntesis". El análisis mediante síntesis es un enfoque paramétrico que conlleva seleccionar un conjunto de parámetros de fuente y un conjunto de parámetros de filtrado y, posteriormente, usar dichos parámetros para generar una forma de onda fuente. Posteriormente, la forma de onda fuente se pasa a través del filtro correspondiente y la forma de onda de salida se compara con la forma de onda original mediante una medición de distancia. Posteriormente, se prueban conjuntos de parámetros diferentes hasta que la distancia se reduce a un mínimo. El conjunto de parámetros que consigue el mínimo se usa entonces como una forma codificada de la señal de entrada.

Si bien el análisis mediante síntesis realiza un buen trabajo de optimización de una fuente paramétrica de voz con un filtro de modelado del conducto vocal, impone un supuesto de modelo de fuente paramétrica con el que resulta difícil trabajar.

El documento de Matsui y col, "*Improving naturalness in text-to-speech synthesis using natural glottal source*", 1991 ICASSP New York, páginas 769-772, describe un procedimiento para obtener una fuente glotal natural a partir de señales vocales mediante una técnica de filtrado inverso.

La presente invención, según se reivindica, adopta un enfoque diferente. La presente invención utiliza un filtro y un filtro inverso. El filtro tiene un conjunto asociado de parámetros de filtrado, por ejemplo, la frecuencia central y el ancho de banda de cada resonador. El filtro inverso está diseñado como el inverso del filtro (por ejemplo, los polos de uno son ceros del otro y viceversa). Por lo tanto, el filtro inverso tiene parámetros que presentan una relación respecto a los parámetros del filtro. Se suministra una señal vocal al filtro inverso para generar una señal residual. La señal residual se procesa para obtener un conjunto de puntos de datos que definen una línea o curva (por ejemplo, forma de onda) que se puede representar como segmentos múltiples.

En función de la aplicación, se pueden utilizar etapas de procesamiento diferentes para obtener y analizar los puntos de datos. Dichas etapas de procesamiento incluyen obtener datos de dominio temporal a partir de la señal residual y obtener datos de dominio frecuencial a partir de la señal residual, ya sean llevadas a cabo por separado o en combinación con otras etapas de procesamiento de señales.

Las etapas de procesamiento conllevan un cálculo de coste en función de una medida de longitud de la línea o forma de onda que en la presente memoria descriptiva se denomina "longitud de arco". La longitud de arco o su cuadrado se calcula y usa como un parámetro de coste asociado a la señal residual. Los parámetros de filtrado se ajustan selectivamente mediante iteración hasta que se reduce al mínimo el parámetro de coste. Una vez reducido al mínimo el parámetro de coste, la señal residual se usa para representar una señal fuente obtenida. Los parámetros de filtrado, asociados al parámetro de coste reducido al mínimo, también se pueden usar para construir el filtro de un sintetizador de modelos de fuente y filtro.

El uso de este procedimiento tiene como resultado una suavidad o continuidad de los parámetros de salida. Cuando dichos parámetros se usan para construir un sintetizador de modelos de fuente y filtro, la forma de onda sintetizada suena extraordinariamente natural, sin distorsiones debidas a discontinuidades. Para poner en práctica la invención se puede usar una clase de funciones de coste, basada en la medida de longitud de arco. En la siguiente memoria descriptiva se describen varios elementos de esta clase. Otros resultarán evidentes para los expertos en la materia.

Para un entendimiento más completo de la invención, sus objetivos y ventajas, véase la siguiente memoria descriptiva y los dibujos adjuntos.

Breve descripción de los dibujos

Figura 1 es un diagrama de bloques del aparato que se prefiere actualmente para la práctica de la invención.

Figura 2 es un diagrama de flujo que ilustra el procedimiento de conformidad con la invención.

Figura 3 es un diagrama de formas de onda que ilustra el cálculo de longitud de arco aplicado a una señal residual de ejemplo.

Figura 4a ilustra el resultado de un función de coste de longitud elevada al cuadrado de una frase hablada de ejemplo, que ilustra frecuencias derivadas de formantes frente a tiempo.

Figura 4b ilustra el resultado conseguido usando una convencional codificación predictiva lineal (LPC) sobre la frase de ejemplo utilizada en la Figura 4a.

Figura 5 ilustra varias funciones discriminatorias de líneas etiquetadas por separado, representando la línea A la longitud de arco media de la forma de onda de dominio temporal, representando la línea B la longitud de onda media de la forma de onda de filtrado inverso, ilustrando la línea C el índice de cruce de curvas en punto cero, ilustrando la línea D la diferencia ampliada de los parámetros que se muestran en las líneas A y B.

Descripción detallada de las formas preferidas de realización

Las técnicas de la invención dan por supuesto un modelo de fuente y filtro de producción de habla (u otra forma de onda compleja, tal como una forma de onda que produce un instrumento musical). Un modelo de filtro, del tipo que tiene un conjunto asociado de parámetros de filtrado, define el modelo. Por ejemplo, el filtro puede ser una cascada de filtros de resonancia IIR (también conocido como un filtro para todo polo). En tal caso, los parámetros de filtrado pueden ser, por ejemplo, la frecuencia central y el ancho de banda de cada resonador de la cascada. Asimismo se pueden usar otros tipos de modelos de filtros.

Con frecuencia, el modelo de filtro también incluye, de manera explícita o implícita, una restricción que se puede describir fácilmente en términos matemáticos o cuantitativos. Un ejemplo de dicha restricción se produce cuando una cantidad que se puede medir se mantiene constante incluso cuando se cambian los parámetros de filtrado a cualquiera de sus valores posibles. Ejemplos específicos de dichas restricciones incluyen:

(1) se ahorra energía cuando atraviesa el filtro

(2) una señal de CC se atraviesa sin cambios (es decir, una ganancia en CC de 1) o, más en general,

(3) la función de transferencia de filtros, $H(z)$, siempre es 1 en algún punto determinado del plano Z .

La presente invención utiliza una función de coste diseñada para favorecer propiedades de una fuente real. En el caso del habla, la fuente real es una onda de presión asociada a la fuente glotal durante la entonación. Tiene propiedades de continuidad, cuasi-periodicidad y, con frecuencia, un punto de concentración (o intervalos de tonos) cuando la glotis se cierra momentáneamente entre cada abertura de la glotis. En el caso de un instrumento musical, la fuente real puede ser la onda de presión asociada a una lengüeta vibrante, por ejemplo, de un instrumento de viento.

La propiedad más importante que la función de coste de la presente invención intenta cuantificar es la presencia de resonancias inducidas por el conducto vocal o el cuerpo de instrumento musical. La función de coste se aplica a la residual del filtrado inverso de la señal musical o vocal original. Dado que el filtro inverso se ajusta de manera iterativa, se llegará a un punto en el que se habrán eliminado las resonancias y, por consiguiente, la función de coste estará a un mínimo. La función de coste debería ser sensible a las resonancias inducidas por el conducto vocal o el cuerpo de instrumento, pero debería ser insensible a las resonancias inherentes de la fuente glotal o de la fuente sonora del instrumento. Esta distinción se puede conseguir dado que sólo las resonancias inducidas dan lugar a una perturbación oscilatoria en la forma de onda residual del dominio temporal o a incursiones externas en la curva de dominio frecuencial. En cualquier caso, se detecta un aumento de la longitud de arco de la forma de onda o curva. Por el contrario, la LPC no hace esta distinción y, por consiguiente, usa partes del filtro para modelar características de la fuente glotal o de la fuente sonora del instrumento.

La Figura 1 ilustra un sistema según la invención mediante el cual se puede obtener la forma de onda fuente a partir de una señal compleja de entrada. En el proceso de obtención se usa un par de filtro inverso/filtro.

En la Figura 1, el filtro 10 está definido por su modelo de filtro 12 y por parámetros de filtrado 14. Asimismo, la presente invención utiliza un filtro inverso 16 que corresponde al inverso del filtro 10. Por ejemplo, el filtro 16 tendría los mismos parámetros de filtrado que el filtro 10, pero en cada posición en la que el filtro 10 tiene polos se reemplazarían por ceros. Por lo tanto, el filtro 10 y el filtro inverso 16 definen un sistema recíproco en el cual el efecto del filtro 10 invalida o invierte el efecto del filtro inverso 16. Por consiguiente, como se ilustra, una forma de onda vocal introducida en el filtro inverso 16 y, posteriormente, procesada por medio del filtro 10 tiene como resultado una forma de onda de salida que, en teoría, es idéntica a la forma de onda de entrada. En la práctica, ligeras variaciones en la tolerancia del filtro o ligeras diferencias entre los filtros 16 y 10 tendrían como resultado una forma de onda de salida que se desvía en cierto modo de la concordancia idéntica de la forma de onda de entrada.

Cuando una forma de onda vocal (u otra forma de onda compleja) se procesa a través del filtro inverso 16, la señal residual de salida en el nodo 20 se procesa utilizando una función de coste 22. En términos generales, dicha función de coste analiza la señal residual según una o más de una pluralidad de funciones de procesamiento, que se describen detalladamente más adelante, para producir un parámetro de coste. El parámetro de coste se usa entonces en posteriores etapas de procesamiento para ajustar los parámetros de filtrado 14 en un esfuerzo de reducir al mínimo el parámetro de coste. En la Figura 1, el bloque para reducir al mínimo el coste 24 representa esquemáticamente el procedimiento mediante el cual los parámetros de filtrado se ajustan selectivamente para producir una reducción resultante del parámetro de coste. Esto se puede llevar a cabo de manera iterativa, usando un algoritmo que ajusta de manera incremental los parámetros de filtrado a la vez que busca el coste mínimo.

Una vez conseguido el coste mínimo, la señal residual resultante en el nodo 20 se puede usar entonces para representar una señal de fuente obtenida para síntesis posterior de modelos de fuente y filtro. Los parámetros de filtrado 14 que producen el coste mínimo se usan como los parámetros de filtrado para definir el filtro 10 para uso en síntesis posterior de modelos de fuente y filtro.

La Figura 2 ilustra el procedimiento mediante el cual se obtiene la señal de formantes, y se identifican los parámetros de filtrado, para conseguir un sistema de síntesis de modelos de fuente y filtro de conformidad con la invención.

En primer lugar, en la etapa 50, se define un modelo de filtro. Se puede usar cualquier modelo de filtro adecuado que se preste a una representación parametrizada. Posteriormente, en la etapa 52, se suministra un conjunto inicial de parámetros. Se debe tener en cuenta que el conjunto inicial de parámetros se modificará de manera iterativa en etapas posteriores de procesamiento para buscar los parámetros que corresponden a una función de coste reducida al mínimo. Se pueden usar técnicas diferentes para evitar una solución sub-óptima correspondiente a mínimos locales. Por ejemplo, el conjunto inicial de parámetros que se usa en la etapa 52 se puede seleccionar de un conjunto o matriz de parámetros diseñados para suministrar varios puntos de partida diferentes a fin de evitar los mínimos locales. Por lo tanto, en la Figura 2, se debe tener en cuenta que la etapa 52 se puede llevar a cabo varias veces para diferentes conjuntos iniciales de parámetros.

El modelo de filtro definido en la etapa 50 y el conjunto inicial de parámetros definido en la etapa 52 se usan entonces en la etapa 54 para construir un filtro (como en la etapa 56) y un filtro inverso (como en la etapa 58).

A continuación, la señal vocal se aplica al filtro inverso en la etapa 60 para obtener una señal residual como en la etapa 64. Como se ilustra, la forma de realización preferida usa una ventana Hanning centrada en los intervalos actuales de tonos y ajustada de modo que cubre períodos de dos tonos. También son posibles otras ventanas. En la etapa 66 se procesa la señal residual para obtener puntos de datos para uso en el cálculo de la longitud de arco.

La señal residual se puede procesar de varias maneras diferentes para obtener los puntos de datos. Como se ilustra en la etapa 68, el procedimiento puede bifurcarse a una o más de una clase seleccionada de rutinas de procesamiento. Ejemplos de tales rutinas se ilustran en la etapa 70. A continuación, en la etapa 72, se lleva a cabo el cálculo de la longitud de arco (o longitud al cuadrado). El valor resultante hace las veces de un parámetro de coste.

Tras calcular el parámetro de coste del conjunto inicial de parámetros de filtrado, en la etapa 74 se ajustan selectivamente los parámetros de filtrado y se repite el procedimiento de manera iterativa como se representa en la etapa 76 hasta que se consigue un coste mínimo.

- 5 Una vez conseguido el coste mínimo, la señal residual obtenida correspondiente a ese coste mínimo se usa en la etapa 78 como la señal fuente. Los parámetros de filtrado asociados al coste mínimo se usan como los parámetros de filtrado (etapa 80) de un modelo de fuente y filtro.

Detalles adicionales de la forma preferida de realización

10

Los datos de entrada de la forma de onda vocal se pueden analizar en tramas usando una ventana móvil para identificar entramados sucesivos. Actualmente, a tal efecto, se prefiere el uso de una ventana Hanning. La ventana Hanning se puede modificar para que sea asimétrica. Se centra en intervalos actuales de tonos y llega a cero en intervalos de tonos adyacentes, cubriendo de ese modo períodos de dos tonos. Si se desea, se puede incluir un componente multiplicador lineal adicional para compensar la mayor o menor amplitud de la señal vocal entonada.

15

El procedimiento iterativo que se usa para identificar el coste mínimo puede adoptar varios enfoques diferentes. Un enfoque es una búsqueda exhaustiva. Otro es una aproximación a una búsqueda exhaustiva utilizando un algoritmo de búsqueda de descenso más pronunciado. El algoritmo de búsqueda se debería construir de tal manera que no se eligieran los mínimos locales como el valor mínimo de coste. Para evitar el problema de los mínimos locales se pueden seleccionar diferentes puntos de partida y ejecutarlos de manera iterativa hasta que se llega a una solución. Posteriormente, se selecciona la mejor solución (valor de coste inferior). Alternativamente, o además, se pueden usar algoritmos heurísticos de suavizado para eliminar algunos de los mínimos locales. Estos algoritmos se describen detalladamente a continuación.

25

Una clase de funciones de coste

Se pueden usar uno o más elementos de una clase de funciones de coste para descubrir la señal residual que representa mejor la señal fuente. Común a la familia o clase de funciones de coste es un concepto que se denomina "longitud de arco". La longitud de arco corresponde a la longitud de la línea que se puede trazar para representar la forma de onda en un espacio multidimensional. La señal residual se puede procesar mediante una serie de técnicas diferentes (se describen más adelante) para obtener un conjunto de puntos de datos que representan una curva. Dicha representación consiste en una secuencia de puntos que definen una serie de segmentos rectilíneos que dan una aproximación lineal por segmentos rectilíneos de la curva. Esto se ilustra en la Figura 3. La curva también se puede representar usando aproximaciones por interpolación o líneas curvas. (El término longitud de arco no pretende dar a entender que los segmentos son sólo líneas curvas). El cálculo de la longitud de arco conlleva calcular la suma de las longitudes de segmentos múltiples para, de ese modo, determinar la longitud de la línea. La forma de realización que se prefiere actualmente usa un cálculo pitagórico para medir la longitud de arco. Por lo tanto, la longitud de arco se puede medir utilizando la siguiente ecuación:

40

$$\text{longitud de arco} = \sum_{n=1}^N \sqrt{(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2}$$

45

Alternativamente, el término longitud de arco, según se usa en la presente memoria descriptiva, puede incluir la longitud al cuadrado:

50

$$\text{longitud al cuadrado} = \sum_{n=1}^N \{(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2\}$$

En las ecuaciones anteriores (x_n, y_n) es una secuencia de puntos de datos.

55

Existe una clase de funciones de coste, basada en la longitud de arco, que se pueden usar para obtener una señal de formantes. Elementos de la clase incluyen:

60

- (1) longitud de arco de la forma de onda residual con ventana frente a tiempo
- (2) longitud al cuadrado de la forma de onda residual con ventana frente a tiempo,
- (3) longitud de arco de la magnitud espectral logarítmica de la residual con ventana frente a frecuencia mel,
- (4) longitud de arco en el plano z del espectro complejo de la residual con ventana, parametrizada por frecuencia,
- (5) longitud al cuadrado en el plano z del espectro complejo de la residual con ventana, parametrizada por frecuencia,

65

- (6) longitud de arco en el plano z del logaritmo complejo del espectro complejo de la residual con ventana, parametrizada por frecuencia.

Si bien en este caso se analizan explícitamente seis elementos de la clase, también se prevén otras aplicaciones que conllevan el cálculo de longitud de arco o de longitud al cuadrado.

Los últimos cuatro elementos que se han enumerado anteriormente se calculan en el dominio frecuencial usando una FFT de tamaño adecuado para calcular el espectro. Por ejemplo, respecto al elemento 6 anterior, si $Y_n = R_n \cdot \exp(j\theta_n)$ es la FFT de tamaño N.

$$\text{coste} = \sum_{n=1}^N \sqrt{\log^2 + \left(\frac{R_n}{R_{n-1}}\right) + (\theta_n - \theta_{n-1})^2}$$

En funciones de coste que incluyen el espectro de magnitud logarítmica, el suavizado puede eliminar algunos problemas relacionados con los mínimos locales, eliminando los efectos de armónicos o ceros marcados. Una función de suavizado adecuada a tal efecto puede ser un FIR de 3, 5 y 7 puntos, una LPC y un suavizado Cepstral, con suavizado heurístico para eliminar inclinaciones. La función de suavizado se puede poner en práctica como sigue: en ventanas de 3, 5 ó 7 puntos del espectro de magnitud logarítmica, los valores bajos se reemplazan por la media de dos puntos circundantes superiores o, en caso de que no existieran los puntos superiores, el punto deseado se deja sin cambios.

Los procedimientos que se han descrito anteriormente para obtener señales de formantes son intrínsecamente sincrónicos de tonos. Por consiguiente, es necesaria una estimación inicial de intervalos de tonos. En aplicaciones en las que el objetivo es la síntesis de texto a habla, puede ser aconsejable tener un marcado de los intervalos de tonos muy preciso a fin de llevar a cabo una modificación prosódica posterior. Se ha observado que los procedimientos que se han descrito anteriormente funcionan bien en la obtención de tonos y en el marcado de intervalos.

Específicamente, el seguimiento de tonos se puede llevar a cabo mejor aplicando una longitud de arco de la forma de onda residual con ventana frente a tiempo (1) con la restricción de que la salida de filtro se normaliza de manera que la magnitud máxima es constante. Esto suaviza la forma de onda residual, pero mantiene el tamaño del pico de tono. Por lo tanto, se puede aplicar la autocorrelación y es menos probable que experimente armónicos superiores.

La forma de onda residual pico con frecuencia es una aproximación coherente al intervalo de tonos, no obstante, a menudo dicho tono es estridente o desigual, provocando falta de precisión. Se ha descubierto que cuando el filtro inverso suprime satisfactoriamente los formantes, la fase de la residual se aproximaba a una fase lineal (al menos en las frecuencias más bajas). Si el original del análisis FFT se centra en el tiempo aproximado de intervalos, la fase es prácticamente plana.

Beneficiándose de esto, el punto de intervalos puede ser uno de los parámetros del espacio de reducción al mínimo cuando la función de coste incluye fase. Las funciones de coste (3), (4) y (5) que se han enumerado anteriormente incluyen fase. Por lo tanto, en estos casos, el tiempo de intervalos se puede incluir como un parámetro en la optimización. Esto produce resultados muy coherentes de marcado de intervalos, siempre que la señal vocal no sea demasiado baja. Además, la precisión de la estimación de valores de formantes correspondientes a funciones de coste de dominio frecuencial se puede mejorar considerablemente mediante la optimización simultánea del punto de intervalos de tonos y la alineación correspondiente de la ventana de análisis.

Algunas de las funciones de coste, tales como la función de coste (5), se prestan a soluciones analíticas. Por ejemplo, la función de coste (5) con una restricción lineal de los coeficientes de filtro se puede solucionar analíticamente. Asimismo, se puede encontrar una solución analítica aproximada usando la función (4). Esto puede ser importante en algunas aplicaciones para conseguir velocidad y fiabilidad.

Respecto al caso de función de coste (5) define

$$P_{i,j} = \sum_{k=0}^{N-1} X_{k-1} \cdot X_{k-j} \cdot \left(1 - \cos\left(\frac{2\pi(k - \text{cntr})}{N}\right)\right)$$

En la que X_n es la forma de onda residual, M es el orden de análisis, N es el tamaño en puntos de la ventana de análisis y cntr es el índice de puntos de muestra calculado de intervalos de tonos.

Por lo tanto, si A_i es la secuencia de coeficientes de filtro inverso y B_i es una secuencia de constantes que definen una restricción lineal de los coeficientes A_i , de tal manera que $B_0 \cdot A_0 + \dots + B_M \cdot A_M = 1$, entonces A_i se puede resolver en la siguiente ecuación matricial:

$$\begin{bmatrix} B_0 & B_1 & B_2 & \dots & B_M \\ P_{jn} & -B_j & \star & P_{an} \\ \text{para } j = 1, \dots, M \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_M \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Fijando $B_1=1$ para $i=0, \dots, M$ da una restricción (A). Fijando $B_1=1$ y $B_i=0$ para $i=1, \dots, M$ da una restricción (B).

Para hallar una solución aproximada para la función de coste (4) en la ecuación matricial anterior, se reemplaza $P_{i,j}$ por:

$$P_{i,j} = \sum_{k=1}^{N-1} \left\{ X_{k-1} \cdot X_{1-j} \cdot \cos\left(\pi \frac{k-1}{N}\right) - \cos\left(\pi \frac{k+1-2.cntr}{N}\right) \cdot S_{k-1} \right\}$$

en la que

$$S_m = \sum_{n=0}^{N/2-1} \left\{ (n+1)^a \cdot \cos\left(2\pi \frac{(n+0,5)m}{N}\right) \right\}$$

En esta ecuación, el término, $(n+1)^a$, representa una fuente idealizada. Cuando a es igual a cero, la ecuación se reduce a la de la función de coste (5). Fijando $a = 2$ da, aproximadamente, resultados equivalentes a la función de coste (4).

El procedimiento anterior se centra en el efecto de un filtro de resonancia en una fuente ideal. Una fuente ideal tiene fase lineal y una envolvente espectral de caída suave. Cuando una fuente ideal de este tipo se aplica a un filtro de resonancia el filtro produce un rodeo circular en el corto recorrido del espectro complejo. La técnica de reducción al mínimo de la longitud de arco pretende eliminar el rodeo usando tanto información de fase como de magnitud. Es por esto por lo que las funciones de coste de dominio frecuencial funcionan bien. En comparación, la LPC convencional da por supuesto una fuente blanca e intenta alisar el espectro de magnitud. No obstante, no tiene en cuenta la fase y, por lo tanto, predice resonancias para modelar las características de fuente.

Quizás una de las funciones de coste más sólida es utilizar simultáneamente tanto información de fase como de magnitud. Para utilizar información de fase y de magnitud simultánea en una función de coste de dominio frecuencial se tienen en cuenta otros supuestos acerca del filtro. Se da por supuesto que el filtro es una cascada de polos y ceros (resonancias de segundo orden y anti-resonancias). Se trata de un supuesto razonable dado que un tubo ideal tiene la acústica de una cascada de polos, mientras que un tubo con una abertura lateral (tal como la cavidad nasal) se puede modelar añadiendo ceros a la cascada.

Diseñar la función de coste para utilizar tanto información de fase como de magnitud conlleva tener en cuenta cómo un único polo afectará al espectro complejo (transformada de Fourier) de una fuente ideal que se da por supuesto que tiene una fase casi plana, casi lineal y una magnitud de caída lenta y suave con una frecuencia básica muy por debajo de la frecuencia del polo. La función de coste debería rechazar los efectos del polo.

Si se tiene en cuenta la trayectoria del espectro complejo, continuando desde la frecuencia cero hasta el ancho de banda límite, se observa que adopta un recorrido sinuoso que depende de la forma de onda. Si la forma de onda es de una fuente ideal, el recorrido es bastante simple. Se inicia cerca del origen del acceso real y se mueve rápidamente, en una línea recta, hacia un punto cuya distancia refleja la intensidad de la frecuencia básica. Posteriormente, retorna muy lentamente, en una línea recta de vuelta hacia el origen. Cuando se aplica a la fuente un único polo, la trayectoria toma un rodeo en un recorrido circular en el sentido de las agujas del reloj y luego continua. Dicho rodeo es conforme a la respuesta de frecuencia conocida de un polo. Según aumenta la intensidad del polo (es decir, ancho de banda más estrecho) aumenta el tamaño del rodeo circular. Nuevamente, se puede aplicar la longitud de arco para reducir al mínimo el rodeo y mejorar, de ese modo, el rendimiento de la función de coste. Una función de coste basada en la longitud de arco del espectro complejo en el plano Z, parametrizada por frecuencia, hace las veces de una función de coste especialmente ventajosa para analizar formantes.

Asimismo, se ha observado que otras dos funciones de coste del mismo tipo tienen excelentes resultados. La primera se define añadiendo la distancia al cuadrado de cada etapa cuando se cruza el recorrido del espectro. Esta es realmente más sencilla, desde el punto de vista del cálculo, que algunas otras técnicas, dado que no exige sacar una raíz cuadrada. La segunda de dichas funciones de coste se define tomando el logaritmo del espectro complejo y calculando la longitud de arco de esa trayectoria en el plano Z. Dicha función de coste es más equilibrada en su sensibilidad a polos y ceros.

Todas las funciones de coste de “recorrido del espectro” anteriores parecen funcionar muy bien. Dado que tienen características diferentes, una u otra pueden ser más útiles para una aplicación específica. Las que son susceptibles de una solución analítica matemática pueden representar la mejor elección cuando es necesario fiabilidad y velocidad de cálculo.

La Figura 4a muestra el resultado de función de coste de la longitud elevada al cuadrado de la frase “coming up”. Se trata de un gráfico de frecuencias derivadas de formantes frente a tiempo. Asimismo, se incluye el ancho de banda como la longitud de las pequeñas líneas transversales. Obsérvese que no existen señales deformadas ni desplazamientos de filtro, tal como normalmente aparecen en el análisis LPC.

En la Figura 4b se muestra la misma frase, analizada usando LPC. En cada gráfico, la forma de onda se muestra en la parte superior y el gráfico encima de la forma de onda es el tono que se obtiene usando el filtro inverso con autocorrelación.

La Figura 5 muestra varias funciones discriminatorias. La función (A) es la longitud de arco media de la forma de onda de dominio temporal. La función (B) es la longitud de arco media de la forma de onda de filtrado inverso. La función (C) ilustra el índice de cruce de curvas en punto cero (una propiedad no directamente aplicable en este caso, pero que se muestra a efectos de perfección). La función (D) es la diferencia ampliada de los parámetros (A) y (B). La función de diferencia (D) parece tomar un valor bajo o negativo, dependiendo de cómo están constreñidos los articuladores. En particular, se debe tener en cuenta que durante la “m” contenida dentro de la frase “coming up” los articuladores están constreñidos. Esta característica se puede usar para detectar nasales y los límites entre nasales y vocales.

Se desarrolló un tipo de filtrado previo para análisis que aumentaba considerablemente la precisión, especialmente de marcado de intervalos de tonos. Esto se aplica cuando el análisis usa una función de coste no logarítmica del dominio frecuencial. En ese caso, el análisis es muy sensible a las bajas frecuencias y, por lo tanto, se encontraron alteraciones derivadas de un soplo de aire u otras fuentes de baja frecuencia. Un simple filtrado de paso alto con filtros FIR parecía empeorar las cosas.

Se puso en práctica la siguiente solución: Durante la optimización de una función de coste, la forma de onda vocal original, con ventana en dos impulsos glotales, se filtra inversamente de manera repetida. La forma de onda de entrada, $x[n]$, se modifica restando un polinomio en n , $A \cdot n^2 + B \cdot n + C$, en la que $n=0$ es el punto de intervalos, así como el origen de la FFT que se usa en la función de coste. Esto significa que se da por supuesto que la distorsión a baja frecuencia se aproxima mediante una forma de onda polinomial aditiva sobre la ventana de dos periodos. Para hallar A, B, C, los mismos se incluyen en la optimización con el objetivo de reducir al mínimo la función de coste. Se observó que un modo no incurría en demasiado cálculo adicional. El resultado fue un efecto de paso elevado que mejoró el análisis y el marcado de intervalos en partes de baja amplitud de la forma de onda.

Evaluación del rendimiento

Para evaluar la precisión se pusieron en práctica dos medidas de distancia espectral y se realizó una prueba de comparación del habla sintética. La primera medida se basa en la distancia, en el plano Z, entre el polo deseado y el polo que se calculó mediante el procedimiento de análisis. La distancia se calculó de manera independiente para los formantes uno a cuatro, así como también para la suma de los cuatro, y se acumuló a toda la expresión de prueba.

La segunda medida es la medida de distorsión de sumas de la potencia de la raíz (RPS) (sensibilidad espectral pico), definida por

$$\text{dist} = \sum_{k=1}^N (k \cdot (c_{1k} - c_{2k}))^2$$

en la que c_{1k} y c_{2k} son el coeficiente cepstral kth del espectro deseado y del espectro analizado, respectivamente, y N se eligió lo suficientemente grande como para que representara de manera adecuada el espectro logarítmico.

El análisis se llevó a cabo en una oración completamente entonada, “Where were you a year ago?” que se produjo mediante un sintetizador de formantes basado en reglas. Se enfatizaron varias palabras para provocar un patrón de entonación bastante extremo. El sintetizador de formantes produjo seis formantes, y cada procedimiento de análisis localizó seis, no obstante, en las medidas de distancia sólo se tuvieron en cuenta los cuatro primeros formantes. Los parámetros de formantes conocidos del sintetizador hicieron las veces de valores deseados.

Como referencia, la oración se analizó mediante LPC estándar de orden 16, usando el procedimiento de estimación de autocorrelación. La LPC se realizó sincrónicamente por tonos, similar a los otros procedimientos y la ventana fue una ventana Hanning centrada en periodos de dos tonos. Los polos de modelado de formantes se separaron de los polos de modelado de fuentes seleccionando las resonancias más intensas (es decir, los anchos de banda más estrechos). El análisis LPC cometió varios errores de discontinuidad, sin embargo para las mediciones de precisión, dichos errores se corrigieron manualmente mediante la reasignación de formantes.

ES 2 274 606 T3

Para el análisis se puede usar cualquier combinación de función de coste y restricción de filtro, no obstante, algunas de dichas combinaciones dan resultados muy insuficientes. Las combinaciones no productivas no se tuvieron en cuenta. Combinaciones con un rendimiento bastante bueno, como se enumera en la Tabla 1, para compararlas con ellas mismas y con LPC. La escala o unidades asociadas a estos números es arbitraria, pero los valores relativos dentro de una columna son comparables.

Distancia de polos del plano Z correspondiente a formantes.

TABLA 1

Medición de errores de los procedimientos de análisis. Los procedimientos están designados por número de función de coste y letra de constreñida

	1	2	3	4	suma	RPS
LPC	3,57	3,24	2,93	3,63	13,4	17,6
1C	9,32	5,45	4,73	5,07	24,6	81,1
1A	4,51	5,86	5,63	7,03	23,0	38,7
2A	11,80	11,08	6,56	9,54	39,0	115,0
3A	2,12	2,43	1,81	2,07	8,4	12,2
4A	1,26	2,37	2,32	2,83	8,8	11,1
4B	3,22	7,82	4,98	4,13	20,2	46,7
5A	1,57	4,13	4,27	8,30	18,3	24,8
6A	1,23	2,88	2,51	2,84	9,5	7,6

Suponiendo que dichas medidas de distancia son válidas, en general, se llega a la conclusión de que las funciones de coste a partir del dominio frecuencial y el uso del constreñimiento de ganancia unidad en CC superan la precisión de la LPC. Especialmente perceptible es su mejora respecto a la precisión del primer formante.

Se puede llegar a la conclusión de que los procedimientos (3A), (4A) y (6A) son, igualmente, posibles candidatos para una aplicación de análisis, no obstante, se deben tener en cuenta factores adicionales. Los mismos se refieren a mínimos locales y convergencia. Los procedimientos (3A) y (6A), que conllevan el logaritmo, tienen muchas más probabilidades de encontrar mínimos locales y converger más lentamente. Esto es poco acertado dado también son los que tienen más posibilidades de rastrear ceros.

Los procedimientos (4A) y (5A) casi nunca encuentran mínimos locales, de hecho, respecto al procedimiento (5A) todavía no se han observado mínimos locales. Por otro lado, dichos procedimientos suelen calcular anchos de banda demasiado estrechos. Por lo tanto, respecto a los mismos, se añadía una pequeña desventaja a la función de coste para rechazar anchos de banda demasiado estrechos. Si bien, en general, el procedimiento (5A) es inferior, puede ser muy útil dado que rastrea de manera precisa el formante uno con convergencia más rápida y sin mínimos locales.

Si bien la invención se ha descrito en su forma preferida de realización en la actualidad, se entenderá que la invención permite ciertas modificaciones sin apartarse del alcance de la invención según se expone en las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un procedimiento para obtener una señal fuente basada en formantes y en parámetros de filtrado a partir de una
5 señal vocal, que comprende:

- a. definir (50) un modelo de filtro del tipo que tiene un conjunto asociado de parámetros de filtrado,
- b. proporcionar (54) un primer filtro basado en dicho modelo de filtro (12),
- 10 c. suministrar (60) dicha señal vocal a dicho primer filtro para generar una señal residual,
- d. procesar (66) dicha señal residual para obtener un conjunto de puntos de datos que definen una línea de
15 segmentos múltiples y calcular una medida de longitud de dicha línea con el fin de determinar un parámetro
de coste asociado a dicha señal residual,
- e. ajustar selectivamente (74) dichos parámetros de filtrado para producir una reducción resultante de dicho
parámetro de coste,
- 20 f. repetir de manera iterativa (76) las etapas c - e hasta que dicho parámetro de coste se reduzca al mínimo
y, posteriormente, utilizar dicha señal residual, para representar una señal fuente obtenida y los parámetros
de filtrado.

2. El procedimiento de la reivindicación 1 que comprende además proporcionar un segundo filtro correspondiente
25 al inverso de dicho primer filtro para uso en el procesamiento de dicha señal fuente obtenida para generar habla
sintetizada.

3. El procedimiento de la reivindicación 1 en el que dicha etapa d se lleva a cabo obteniendo (70) datos de dominio
temporal a partir de dicha señal residual.

30 4. El procedimiento de la reivindicación 1 en el que dicha etapa d se lleva a cabo obteniendo (70) datos de dominio
temporal a partir de dicha señal residual y calculando la longitud al cuadrado de la distancia a través de dichos datos
de dominio temporal.

35 5. El procedimiento de la reivindicación 1 en el que dicha etapa d se lleva a cabo obteniendo (70) la magnitud
espectral logarítmica de dicha señal residual en el dominio frecuencial.

6. El procedimiento de la reivindicación 1 en el que dicha etapa d se lleva a cabo obteniendo (70) el espectro
complejo del plano z de dicha señal residual parametrizada por frecuencia.

40 7. El procedimiento de la reivindicación 1 en el que dicha etapa d se lleva a cabo obteniendo (70) el logaritmo
complejo del plano z del espectro complejo de dicha señal residual parametrizada por frecuencia.

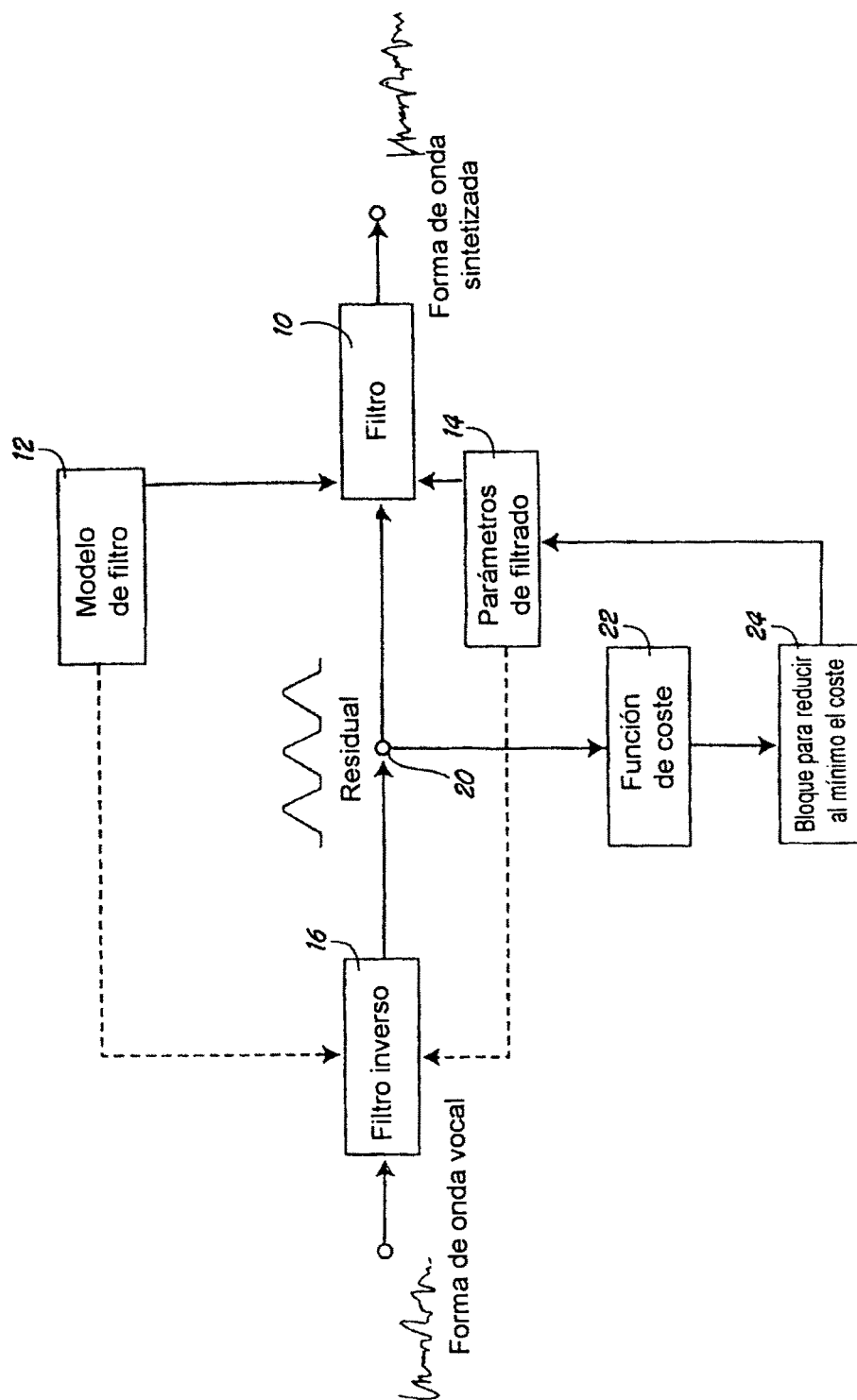


Figura 1

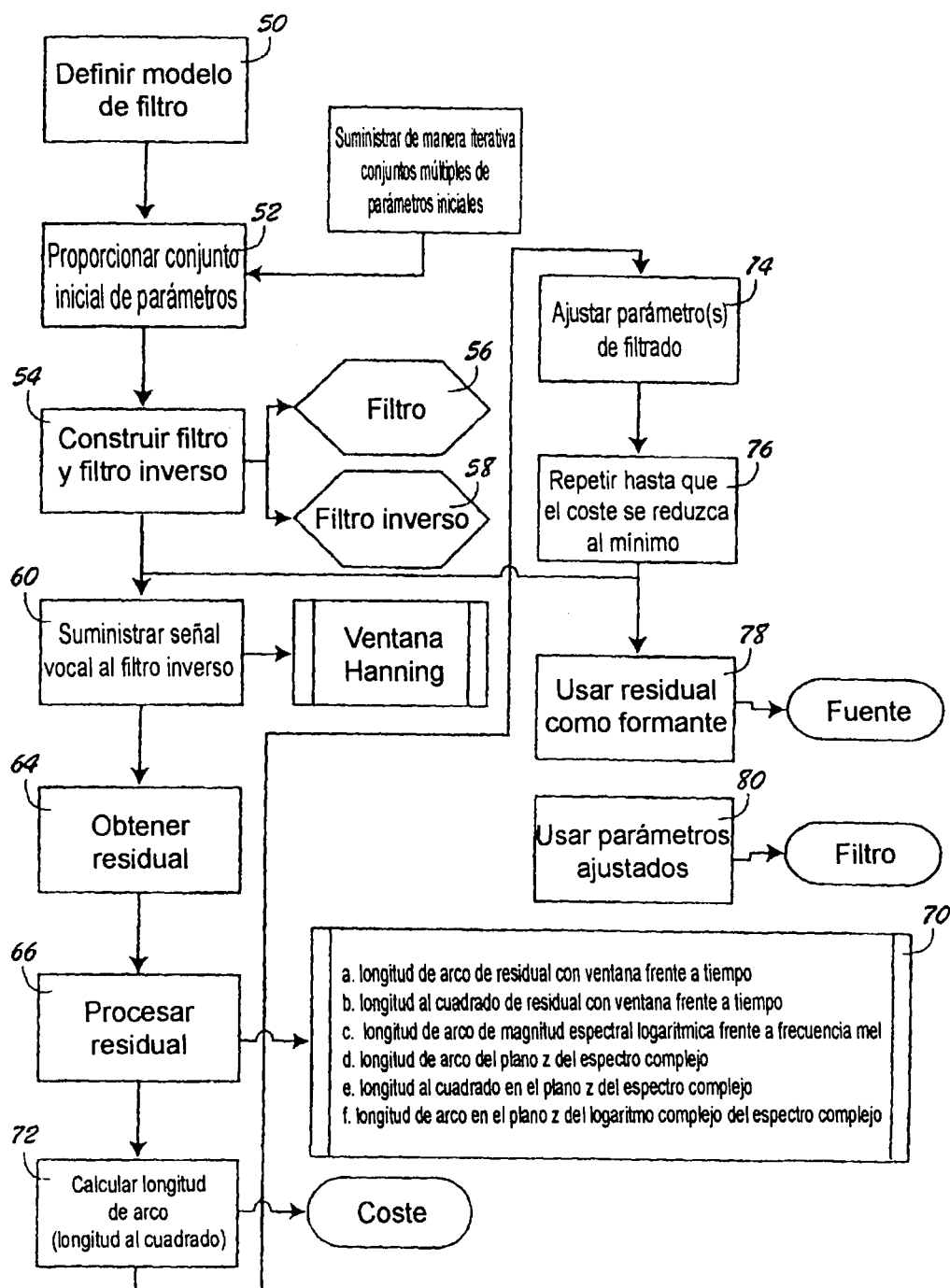


Figura 2

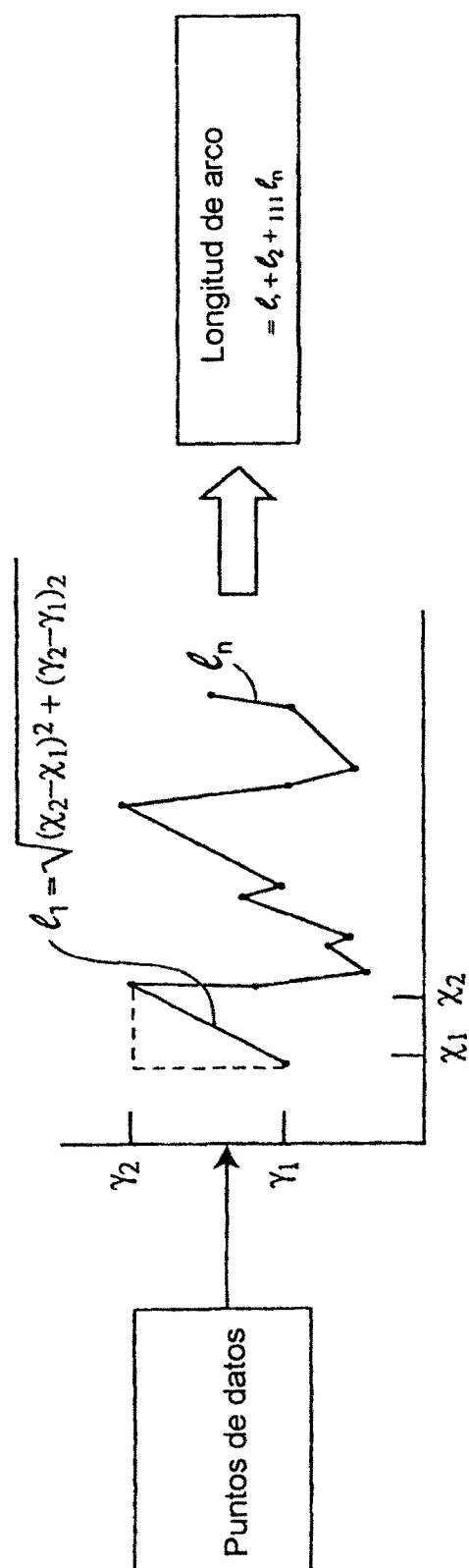
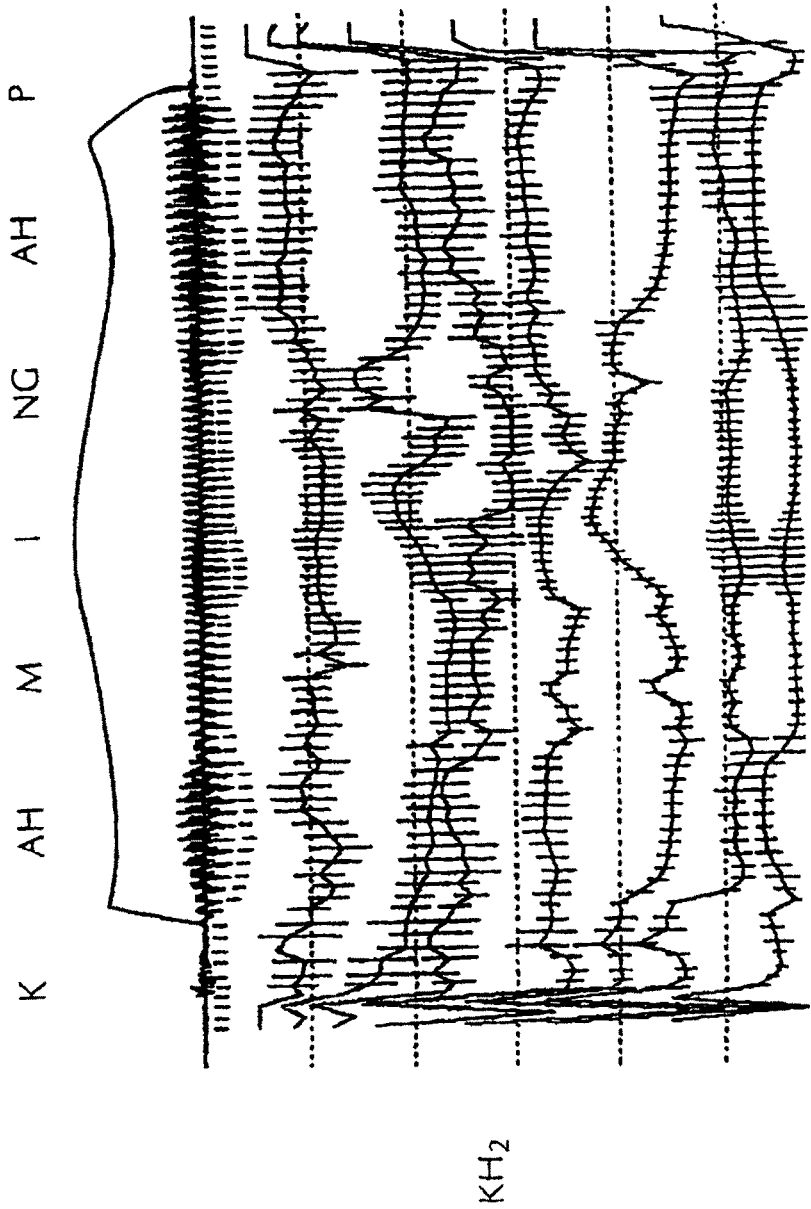
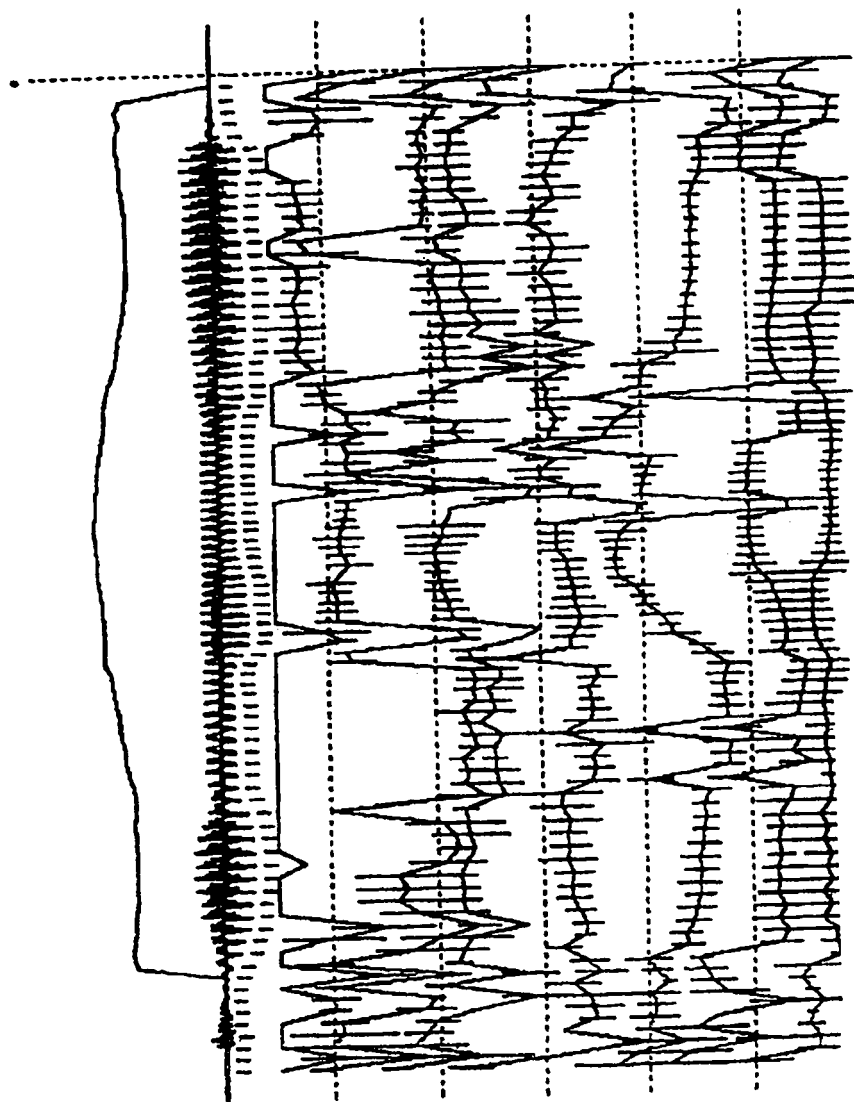


Figura 3



Formante=2 conjunto=0 fichero= e1754

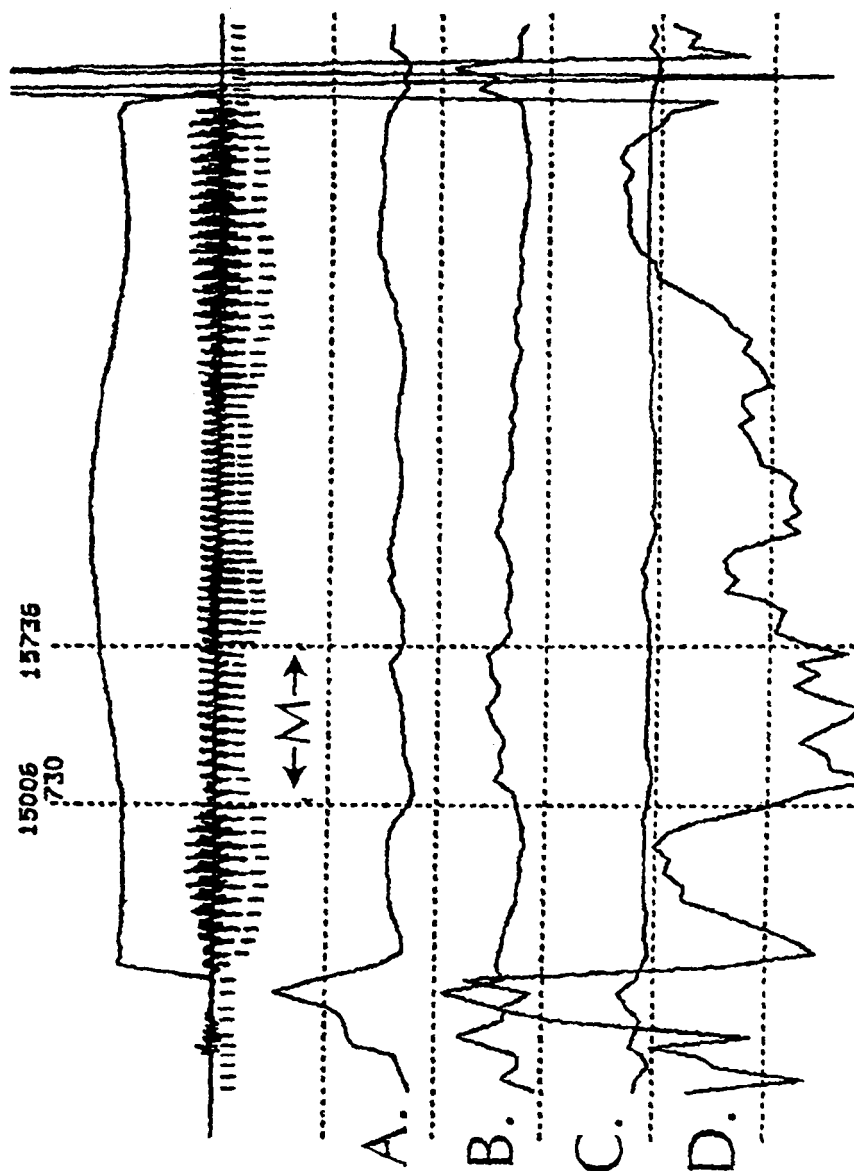
Figura 4a



Formante=7 conjunto=0 fichero= e1754

Figura 4b

(Técnica anterior)



Formante=2 conjunto=0 fichero= e1754

Figura 5