

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 701 402**

51 Int. Cl.:

G10L 19/038 (2013.01)

G10L 19/00 (2013.01)

G10L 19/107 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **31.07.2013 PCT/EP2013/066074**

87 Fecha y número de publicación internacional: **10.04.2014 WO14053261**

96 Fecha de presentación y número de la solicitud europea: **31.07.2013 E 13742646 (6)**

97 Fecha y número de publicación de la concesión europea: **19.09.2018 EP 2904612**

54 Título: **Aparato para codificar una señal de voz empleando ACELP en el dominio de autocorrelación**

30 Prioridad:

05.10.2012 US 201261710137 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
22.02.2019

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**BÄCKSTRÖM, TOM;
MULTRUS, MARKUS;
FUCHS, GUILLAUME;
HELMRICH, CHRISTIAN y
DIETZ, MARTIN**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 701 402 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

Aparato para codificar una señal de voz empleando ACELP en el dominio de autocorrelación

5 La presente invención se refiere a la codificación de señales de audio, y, en particular, a un aparato para codificar una señal de voz empleando ACELP en el dominio de autocorrelación.

10 En la codificación de voz mediante predicción lineal con excitación por código (CELP), la envolvente espectral (o de manera equivalente, estructura de tiempo a corto plazo) de la señal de voz se describe mediante un modelo predictivo lineal (LP) y el residuo de predicción se modela mediante un predictor a largo plazo (LTP, también conocido como libro de códigos adaptativo) y una señal residual representada por un libro de códigos (también conocido como libro de códigos fijo). Este último, el libro de códigos fijo, se aplica generalmente como libro de códigos algebraicos, en el que el libro de códigos se representa por una fórmula o algoritmo algebraico, mediante lo cual no hay necesidad de almacenar todo el libro de códigos, sino sólo el algoritmo, al tiempo que simultáneamente se permite un algoritmo de búsqueda rápida. Los códecs de CELP que aplican un libro de códigos algebraicos para el residuo se conocen como códecs de predicción lineal con excitación por código algebraico (ACELP) (véanse [1], [2], [3], [4]).

20 En la codificación de voz, emplear un libro de códigos de residuos algebraicos es el enfoque de elección en códecs principales tales como [17], [13], [18]. ACELP se basa en modelar la envolvente espectral mediante un filtro predictivo lineal (LP), la frecuencia fundamental de sonidos de voz mediante un predictor a largo plazo (LTP) y el residuo de predicción mediante un libro de códigos algebraicos. Los parámetros LTP y libro de códigos algebraicos se optimizan mediante un algoritmo de mínimos cuadrados en un dominio perceptual, en el que el dominio perceptual se especifica por un filtro.

25 La parte computacionalmente más compleja de algoritmos de tipo ACELP, el cuello de botella, es la optimización del libro de códigos de residuos. El único algoritmo óptimo actualmente conocido es una búsqueda exhaustiva de un espacio de tamaño N^p para cada subtrama, en la que en cada punto se requiere una evaluación de complejidad $O(N^2)$. Dado que valores típicos son una longitud de subtrama $N = 64$ (es decir 5 ms) con $p = 8$ pulsos, esto implica más de 10^{20} operaciones por segundo. Resulta evidente que esto no es una opción viable. Para permanecer dentro de los límites de complejidad establecidos por requisitos de hardware, los enfoques de optimización de libro de códigos tienen que funcionar con algoritmos iterativos no óptimos. En el pasado se han presentado muchos algoritmos y mejoras del proceso de optimización de este tipo, por ejemplo [17], [19], [20], [21], [22].

35 De manera explícita, la optimización de ACELP se basa en describir la señal de voz $x(n)$ como la salida de un modelo predictivo lineal de tal manera que la señal de voz estimada es

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

40 donde $a(k)$ son los coeficientes de LP y $\hat{e}(k)$ es la señal residual. En forma de vector, esta ecuación puede expresarse como

$$\hat{x} = H \hat{e} \tag{2}$$

45 donde la matriz H se define como la matriz de convolución de Toeplitz triangular inferior con diagonal $h(0)$ y diagonales inferiores $h(1), \dots, h(39)$ y el vector $h(k)$ es la respuesta a impulso del modelo LP. Debe observarse que en esta notación se omite el modelo perceptual (que corresponde habitualmente a un modelo LP ponderado), pero se supone que el modelo perceptual está incluido en la respuesta a impulso $h(k)$. Esta omisión no tiene ningún impacto sobre la generalidad de los resultados, pero simplifica la notación. La inclusión del modelo perceptual se aplica como en [1].

50 El ajuste del modelo se mide mediante el error cuadrático. Es decir,

$$\epsilon^2 = \sum_{k=1}^N (x(k) - \hat{x}(k))^2 = (e - \hat{e})^H H^H H (e - \hat{e}). \tag{3}$$

55 Este error cuadrático se usa para encontrar los parámetros de modelo óptimos. En este caso, se supone que se usan tanto el LTP como el libro de códigos de pulso para modelar el vector e . La aplicación práctica puede encontrarse en las publicaciones relevantes (véase [1-4]).

En la práctica, la medición anterior del ajuste puede simplificarse de la siguiente manera. Comprendiendo la matriz $B = H^T H$ las correlaciones de $h(n)$, siendo c_k el k -ésimo vector de libro de códigos fijo y estableciendo $\hat{e} = g c_k$, donde g es un factor de ganancia. Suponiendo que g se elige de manera óptima, entonces se busca en el libro de códigos maximizando el criterio de búsqueda

5

$$\frac{C_k^2}{E_k} = \frac{(x^T H c_k)^2}{c_k^T B c_k} = \frac{(d^T c_k)^2}{c_k^T B c_k} \quad (4)$$

donde $d = H^T x$ es un vector que comprende la correlación entre el vector objetivo y la respuesta a impulso $h(n)$ y el superíndice T indica la transpuesta. El vector d y la matriz B se calculan antes la búsqueda en el libro de códigos. Esta fórmula se usa habitualmente en la optimización tanto del LTP como del libro de códigos de pulso.

10

Se ha invertido mucha investigación en optimizar el uso de la fórmula anterior. Por ejemplo,

15

1) Solo se calculan los elementos de la matriz B a los que accede realmente el algoritmo de búsqueda. O:

2) Se reduce el algoritmo de ensayo y error de la búsqueda de pulso para intentar solo aquellos vectores de libro de códigos que tienen una alta probabilidad de éxito, basándose en una selección previa (véase por ejemplo [1,5]).

20

Un detalle práctico del algoritmo de ACELP se refiere al concepto de respuesta a impulso nulo (ZIR). El concepto aparece cuando se considera la señal de síntesis de dominio original en comparación con el residuo sintetizado. El residuo se codifica en bloques que corresponden al tamaño de trama o subtrama. Sin embargo, cuando se sintetiza la señal de dominio original con el modelo LP de la ecuación 1, el residuo de longitud fija tendrá una "cola" de longitud infinita, correspondiente a la respuesta a impulso del filtro LP. Es decir, aunque el vector de libro de códigos de residuo tiene una longitud finita, tendrá un efecto sobre la señal de síntesis mucho más allá de la trama o subtrama actual. El efecto de una trama en el futuro puede calcularse extendiendo el vector de libro de códigos con ceros y calculando la salida de síntesis de la ecuación 1 para esta señal extendida. Esta extensión de la señal sintetizada se conoce como la respuesta a impulso nulo. Después, para tener en cuenta el efecto de tramas anteriores en la codificación de la trama actual, se resta la ZIR de la trama anterior a partir del objetivo de la trama actual. Por tanto, en la codificación de la trama actual solo se considera la parte de la señal que no se modeló ya por la trama anterior.

25

30

En la práctica, la ZIR se tiene en cuenta de la siguiente manera: cuando se ha codificado una (sub)trama $N-1$, se extiende el residuo cuantificado con ceros hasta la longitud de la siguiente (sub)trama N . Se filtra el residuo cuantificado extendido mediante LP para obtener la ZIR de la señal cuantificada. Después se resta la ZIR de la señal cuantificada a partir de la señal original (no cuantificada) y esta señal modificada forma la señal objetivo cuando se codifica la (sub)trama N . De esta manera, todos los errores de cuantificación realizados en la (sub)trama $N-1$ se tendrán en cuenta al cuantificar la (sub)trama N . Esta práctica mejora considerablemente la calidad perceptual de la señal de salida.

35

Sin embargo, se apreciaría sumamente si pudieran proporcionarse conceptos adicionalmente mejorados para la codificación de audio.

40

El objetivo de la presente invención es proporcionar tales conceptos mejorados para la codificación de objetos de audio. El objetivo de la presente invención se resuelve mediante un aparato según la reivindicación 1, mediante un método para codificar según la reivindicación 15, mediante un sistema según la reivindicación 14 y mediante un programa informático según la reivindicación 17.

45

Se proporciona un aparato para codificar una señal de voz determinando un vector de libro de códigos de un algoritmo de codificación de voz. El aparato comprende un determinante de matriz para determinar una matriz de autocorrelación R , y un determinante de vector de libro de códigos para determinar el vector de libro de códigos dependiendo de la matriz de autocorrelación R . El determinante de matriz está configurado para determinar la matriz de autocorrelación R determinando coeficientes de vector de un vector r , en el que la matriz de autocorrelación R comprende una pluralidad de filas y una pluralidad de columnas, en el que el vector r indica una de las columnas o una de las filas de la matriz de autocorrelación R , en el que $R(i, j) = r(|i-j|)$, en el que $R(i, j)$ indica los coeficientes de la matriz de autocorrelación R , en el que i es un primer índice que indica una de una pluralidad de filas de la matriz de autocorrelación R , y en el que j es un segundo índice que indica una de una pluralidad de columnas de la matriz de autocorrelación R .

50

55

El aparato está configurado para usar el vector de libro de códigos para codificar la señal de voz. Por ejemplo, el aparato puede generar la señal de voz codificada de tal manera que la señal de voz codificada comprende una pluralidad de coeficientes de predicción lineal, una indicación de la frecuencia fundamental de sonidos de voz (por

60

ejemplo, parámetros de altura) y una indicación del vector de libro de códigos, por ejemplo, un índice del vector de libro de códigos.

5 Además, se proporciona un decodificador para decodificar una señal de voz codificada que está codificándose mediante un aparato según la realización descrita anteriormente para obtener una señal de voz decodificada.

10 Además, se proporciona un sistema. El sistema comprende un aparato según la realización descrita anteriormente para codificar una señal de voz de entrada para obtener una señal de voz codificada. Además, el sistema comprende un decodificador según la realización descrita anteriormente para decodificar la señal de voz codificada para obtener una señal de voz decodificada.

15 Se proporcionan conceptos mejorados para la función objetivo del algoritmo de codificación de voz ACELP, que tienen en cuenta no solo el efecto de la respuesta a impulso de la trama anterior para la trama actual, sino también el efecto de la respuesta a impulso de la trama actual en la siguiente trama, cuando se optimizan parámetros de la trama actual. Algunas realizaciones implementan estas mejoras cambiando la matriz de correlación, lo cual es fundamental para la optimización de ACELP convencional por una matriz de autocorrelación, que tiene una estructura de Toeplitz hermitiana. Empleando esta estructura, es posible realizar la optimización de ACELP de manera más eficiente tanto en cuanto a complejidad computacional como en cuanto a requisitos de memoria. Simultáneamente, el modelo perceptual aplicado también se vuelve más sistemático y pueden evitarse dependencias entre tramas para mejorar el rendimiento bajo la influencia de pérdida de paquetes.

20 La codificación de voz con el paradigma ACELP se basa en un algoritmo de mínimos cuadrados en un dominio perceptual, en el que el dominio perceptual se especifica por un filtro. Según realizaciones, la complejidad computacional de la definición convencional del problema de los mínimos cuadrados puede reducirse teniendo en cuenta el impacto de la respuesta a impulso nulo en la siguiente trama. Las modificaciones proporcionadas introducen una estructura de Toeplitz en una matriz de correlación que aparece en la función objetivo, lo que simplifica la estructura y reduce los cálculos. Los conceptos propuestos reducen la complejidad computacional hasta un 17% sin reducir la calidad perceptual.

30 Realizaciones se basan en el hallazgo de que mediante una ligera modificación de la función objetivo, puede reducirse adicionalmente la complejidad en la optimización del libro de códigos de residuos. Esta reducción de la complejidad se produce sin reducción de la calidad perceptual. Como alternativa, dado que la optimización de residuos de ACELP se basa en algoritmos de búsqueda iterativos, con la modificación presentada es posible aumentar el número de iteraciones sin un aumento de la complejidad, y de esta manera obtener una calidad perceptual mejorada.

40 Tanto la función objetivo convencional como la modificada modelan la percepción e intentan minimizar la distorsión perceptual. Sin embargo, la solución óptima al enfoque convencional no es necesariamente óptima con respecto a la función objetivo modificada y viceversa. Esto por sí solo no significa que un enfoque sea mejor que el otro, pero argumentos analíticos sí que muestran que la función objetivo modificada es más sistemática. Específicamente, a diferencia de la función objetivo convencional, los conceptos proporcionados tratan todas las muestras dentro de una subtrama por igual, con modelos de señal y perceptuales sistemáticos y bien definidos.

45 En realizaciones, las modificaciones propuestas pueden aplicarse de tal manera que solo cambian la optimización del libro de códigos de residuos. Por tanto, no cambia la estructura de flujo de bits y puede aplicarse de una manera retrocompatible con códecs de ACELP existentes.

50 Además, se proporciona un método para codificar una señal de voz determinando un vector de libro de códigos de un algoritmo de codificación de voz. El método comprende:

- determinar una matriz de autocorrelación R. Y:
- determinar el vector de libro de códigos dependiendo de la matriz de autocorrelación R.

55 Determinar una matriz de autocorrelación R comprende determinar coeficientes de vector de un vector r. La matriz de autocorrelación R comprende una pluralidad de filas y una pluralidad de columnas. El vector r indica una de las columnas o una de las filas de la matriz de autocorrelación R, en el que

$$R(i, j) = r(|i - j|).$$

60 R(i, j) indica los coeficientes de la matriz de autocorrelación R, en el que i es un primer índice que indica una de una pluralidad de filas de la matriz de autocorrelación R, y en el que j es un segundo índice que indica una de una pluralidad de columnas de la matriz de autocorrelación R.

Además, se proporciona un método para decodificar una señal de voz codificada que está codificándose según el método para codificar una señal de voz según la realización descrita anteriormente para obtener una señal de voz decodificada.

5 Además, se proporciona un método. El método comprende:

- codificar una señal de voz de entrada según el método descrito anteriormente para codificar una señal de voz para obtener una señal de voz codificada. Y:

10 - decodificar la señal de voz codificada para obtener una señal de voz decodificada según el método descrito anteriormente para decodificar una señal de voz.

15 Además, se proporcionan programas informáticos para implementar los métodos descritos anteriormente cuando se ejecutan en un ordenador o procesador de señales.

En las reivindicaciones dependientes se proporcionarán realizaciones preferidas.

20 A continuación, se describen en más detalle realizaciones de la presente invención con referencia a las figuras, en las que:

la figura 1 ilustra un aparato para codificar una señal de voz determinando un vector de libro de códigos de un algoritmo de codificación de voz según una realización

25 la figura 2 ilustra un decodificador según una realización y un decodificador, y

la figura 3 ilustra un sistema que comprende un aparato para codificar una señal de voz según una realización y un decodificador.

30 La figura 1 ilustra un aparato para codificar una señal de voz determinando un vector de libro de códigos de un algoritmo de codificación de voz según una realización.

35 El aparato comprende un determinador (110) de matriz para determinar una matriz de autocorrelación R, y un determinador (120) de vector de libro de códigos para determinar el vector de libro de códigos dependiendo de la matriz de autocorrelación R.

El determinador (110) de matriz está configurado para determinar la matriz de autocorrelación R determinando coeficientes de vector de un vector r.

40 La matriz de autocorrelación R comprende una pluralidad de filas y una pluralidad de columnas, en la que el vector r indica una de las columnas o una de las filas de la matriz de autocorrelación R, en la que $R(i, j) = r(|i-j|)$.

45 $R(i, j)$ indica los coeficientes de la matriz de autocorrelación R, en la que i es un primer índice que indica una de una pluralidad de filas de la matriz de autocorrelación R, y en la que j es un segundo índice que indica una de una pluralidad de columnas de la matriz de autocorrelación R.

50 El aparato está configurado para usar el vector de libro de códigos para codificar la señal de voz. Por ejemplo, el aparato puede generar la señal de voz codificada de tal manera que la señal de voz codificada comprende una pluralidad de coeficientes de predicción lineal, una indicación de la frecuencia fundamental de sonidos de voz (por ejemplo, parámetros de altura) y una indicación del vector de libro de códigos.

55 Por ejemplo, según una realización particular para codificar una señal de voz, el aparato puede estar configurado para determinar una pluralidad de coeficientes predictivos lineales $(a(k))$ dependiendo de la señal de voz. Además, el aparato está configurado para determinar una señal residual dependiendo de la pluralidad de coeficientes predictivos lineales $(a(k))$.

Además, el determinador 110 de matriz puede estar configurado para determinar la matriz de autocorrelación R dependiendo de la señal residual.

60 A continuación, se describen algunas realizaciones adicionales de la presente invención.

Volviendo a las ecuaciones 3 y 4, en las que la ecuación 3 define un error cuadrático que indica un ajuste del modelo perceptual como:

$$\epsilon^2 = \sum_{k=1}^N (x(k) - \hat{x}(k))^2 = (e - \hat{e})^H H^H H (e - \hat{e}), \quad (3)$$

y en las que la ecuación 4

$$\frac{C_k^2}{E_k} = \frac{(x^T H c_k)^2}{c_k^T B c_k} = \frac{(d^T c_k)^2}{c_k^T B c_k} \quad (4)$$

indica el criterio de búsqueda, que debe maximizarse.

El algoritmo de ACELP se centra en la ecuación 4, que a su vez se basa en la ecuación 3.

Realizaciones se basan en el hallazgo de que el análisis de estas ecuaciones revela que los valores residuales cuantificados $e(k)$ tienen un efecto muy diferente sobre la energía de error ϵ^2 dependiendo del índice k . Por ejemplo, cuando se consideran los índices $k=1$ y $k=N$, si el único valor distinto de cero del libro de códigos de residuos aparece a $k=1$, entonces la energía de error ϵ^2 da como resultado:

$$\epsilon_1^2 = \sum_{k=1}^N (x(k) - e(1)h(k))^2 \quad (5)$$

mientras que para $k=N$, la energía de error ϵ^2 da como resultado:

$$\epsilon_N^2 = (x(N) - e(N)h(1))^2 + \sum_{k=1}^{N-1} (x(k))^2. \quad (6)$$

Dicho de otro modo, $e(1)$ se pondera con la respuesta a impulso $h(k)$ en el intervalo de 1 a N , mientras que $e(N)$ solo se pondera con $h(1)$. En cuanto a la ponderación espectral, esto significa que cada $e(k)$ se pondera con una función de ponderación espectral diferente, de manera que, en el extremo, $e(N)$ se pondera de manera lineal. Desde un punto de vista de modelado perceptual, tendría sentido aplicar la misma ponderación perceptual para todas las muestras dentro de una trama. Por tanto, la ecuación 3 deberá extenderse de tal manera que tenga en cuenta la ZIR en la siguiente trama. Debe constatarse en este caso que, entre otras cosas, la diferencia con respecto a la técnica anterior es que se tienen en cuenta tanto la ZIR de la trama anterior como también la ZIR en la siguiente trama.

Siendo $e(k)$ el residuo original sin cuantificar y $\hat{e}(k)$ el residuo cuantificado. Además, siendo ambos residuos distintos de cero en el intervalo de 1 a N y cero en los demás casos. Entonces

$$\begin{aligned} x(n) &= -\sum_{k=1}^m a(k)x(n-k) + e(n) = \sum_{k=1}^{\infty} e(n-k)h(k) \\ \hat{x}(n) &= -\sum_{k=1}^m a(k)\hat{x}(n-k) + \hat{e}(n) = \sum_{k=1}^{\infty} \hat{e}(n-k)h(k) \end{aligned} \quad (7)$$

De manera equivalente, las mismas relaciones en forma de matriz pueden expresarse como:

$$\begin{aligned} x &= \tilde{H} e \\ \hat{x} &= \tilde{H} \hat{e} \end{aligned} \quad (8)$$

en la que \tilde{H} es la matriz de convolución dimensional infinita correspondiente a la respuesta a impulso $h(k)$. Insertando en la ecuación 3 se proporciona

$$\epsilon^2 = \|\tilde{H} e - \tilde{H} \hat{e}\|^2 = (e - \hat{e})^T \tilde{H}^T \tilde{H} (e - \hat{e}) = (e - \hat{e})^T R (e - \hat{e}) \quad (9)$$

en la que $R = \tilde{H}^T \tilde{H}$ es la matriz de Toeplitz hermitiana de tamaño finito correspondiente a la autocorrelación de $h(n)$. Mediante una derivación similar a la de la ecuación 4, se obtiene la función objetivo:

$$\frac{(e^T R \hat{e})^2}{(\hat{e}^T R \hat{e})} = \frac{(d^T e)^2}{(\hat{e}^T R \hat{e})} \quad (10)$$

Esta función objetivo es muy similar a la ecuación 4. La principal diferencia es que en lugar de la matriz de correlación B, en este caso hay una matriz de Toeplitz hermitiana R en el denominador.

5 Tal como se explicó anteriormente, esta formulación novedosa tiene el beneficio de que todas las muestras del residuo e dentro de una trama recibirán la misma ponderación perceptual. Sin embargo, de manera importante, esta formulación también introduce beneficios considerables en la complejidad computacional y requisitos de memoria. Dado que R es una matriz de Toeplitz hermitiana, la primera columna r(0)..r(N- 1) define completamente la matriz. Dicho de otro modo, en lugar de almacenar la matriz NxN completa, es suficiente con almacenar solo el vector Nx1 r(k), proporcionando por tanto un ahorro considerable de asignación de memoria. Además, también se reduce la complejidad computacional ya que no es necesario determinar la totalidad de los NxN elementos, sino tan solo la primera columna Nx1. Además, la indexación dentro de la matriz es sencilla, ya que el elemento (i,j) puede encontrarse mediante R(i, j) = r(|i-j|).

15 Como la función objetivo en la ecuación 10 es tan similar a la ecuación 4, puede conservarse la estructura de la ACELP general. Específicamente, puede realizarse cualquiera de las siguientes operaciones con cualquier función objetivo, con tan solo modificaciones secundarias en el algoritmo:

- 20 1. Optimización del retardo de LTP (libro de códigos adaptativo)
- 2. Optimización del libro de códigos de pulso para modelar el residuo (libro de códigos fijo)
- 3. Optimización de las ganancias de LTP y pulsos, o bien por separado o bien de manera conjunta
- 25 4. Optimización de cualquier otro parámetro cuyo rendimiento pueda medirse mediante el error cuadrático de la ecuación 3.

La única parte que tiene que modificarse en aplicaciones de ACELP convencional es la gestión de la matriz de correlación B, que se sustituye por la matriz R, así como el objetivo, que debe incluir la ZIR en la siguiente trama.

Algunas realizaciones emplean los conceptos de la presente invención sustituyendo, en cualquier parte en el algoritmo de ACELP, donde aparezca la matriz de correlación B, por la matriz de autocorrelación R. Si se omiten todas las instancias de la matriz B, entonces puede evitarse calcular su valor.

35 Por ejemplo, la matriz de autocorrelación R se determina determinando los coeficientes de la primera columna r(0), ..., r(N-1) de la matriz de autocorrelación R.

La matriz R se define en la ecuación 9 mediante R=H^TH, mediante lo cual sus elementos Rij=r(i-j) pueden calcularse mediante

$$r(k) = h(k) * h(-k) = \sum_l h(l)h(l-k) \quad (9a)$$

Es decir, la secuencia r(k) es la autocorrelación de h(k).

45 Sin embargo, con frecuencia r(k) puede obtenerse mediante medios incluso más eficaces. Específicamente, en normas de codificación de voz tales como AMR y G.718, la secuencia h(k) es la respuesta a impulso de un filtro predictivo lineal A(z) filtrado mediante una función de ponderación perceptual W(z), que se interpreta que incluye el preénfasis. Dicho de otro modo, h(k) indica una respuesta a impulso perceptualmente ponderada de un modelo predictivo lineal.

El filtro A(z) se estima habitualmente a partir de la autocorrelación de la señal de voz r_x(k), es decir, r_x(k) ya se conoce. Dado que H(z) = A⁻¹(u)W(z), se desprende que la secuencia de autocorrelación r(k) puede determinarse calculando la autocorrelación de w(k) mediante

$$r_w(k) = w(k) * w(-k) = \sum_l w(l)w(l-k) \quad (9b)$$

mediante lo cual la autocorrelación de $h(k)$ es

$$r(k) = r_x(k) * r_w(k) = \sum_l r_w(l) r_x(l-k). \quad (9c)$$

5 Dependiendo del diseño del sistema global, estas ecuaciones pueden modificarse en consecuencia en algunas realizaciones.

10 Entonces puede determinarse, por ejemplo, un vector de libro de códigos de un libro de códigos basándose en la matriz de autocorrelación R . En particular, la ecuación 10 puede usarse, según algunas realizaciones, para determinar un vector de libro de códigos del libro de códigos.

15 En este contexto, la ecuación 10 define la función objetivo en la forma $f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$ que por lo demás es la misma forma que en las normas de codificación de voz AMR y G.718 pero de tal manera que la matriz R tiene ahora una estructura de Toeplitz simétrica. La función objetivo es básicamente una correlación normalizada entre el vector objetivo d y el vector de libro de códigos \hat{e} y el mejor vector de libro de códigos posible es el que proporciona el valor más alto para la correlación normalizada $f(\hat{e})$, por ejemplo, que maximiza la correlación normalizada $f(\hat{e})$.

20 Por tanto, pueden optimizarse vectores de libro de códigos con los mismos enfoques que en las normas mencionadas. Específicamente, por ejemplo, puede aplicarse el algoritmo muy sencillo para hallar el mejor vector de libro de códigos algebraicos (es decir, el libro de códigos fijo) \hat{e} para el residuo, tal como se describe a continuación. Sin embargo, debe observarse que se ha invertido un esfuerzo significativo en el diseño de algoritmos de búsqueda eficientes (véase AMR y G.718), y este algoritmo de búsqueda solo es un ejemplo de aplicación ilustrativo.

25 1. Definir un vector de libro de códigos inicial $\hat{e}_0 = [0, 0, \dots, 0]^T$ y establecer el número de pulsos a $p = 0$

2. Establecer la medida de calidad de libro de códigos inicial a $f_0 = 0$.

30 3. Establecer la medida de calidad de libro de códigos temporal a $\hat{f}_p = f_{p-1}$.

4. Para cada posición k en el vector de libro de códigos

(i) Aumentar p en uno.

35 (ii) Si la posición k ya contiene un pulso negativo, continuar a la etapa vii.

(iii) Crear un vector de libro de códigos temporal $\varepsilon_p^+ = \hat{e}_{p-1}$ y añadir un pulso positivo en la posición k .

40 (iv) Evaluar la calidad del vector de libro de códigos temporal mediante $f(\varepsilon_p^+)$.

(v) Si el vector de libro de códigos temporal es mejor que cualquiera de los anteriores, $f(\varepsilon_p^+) > \hat{f}_p$, entonces guardar este vector de libro de códigos, establecer $\hat{f}_p = f(\varepsilon_p^+)$ y continuar a la siguiente iteración.

45 (vi) Si la posición k ya contiene un pulso positivo, continuar a la siguiente iteración.

(vii) Crear un vector de libro de códigos temporal $\varepsilon_p^- = \hat{e}_{p-1}$ y añadir un pulso negativo en la posición k .

(viii) Evaluar la calidad del vector de libro de códigos temporal mediante $f(\varepsilon_p^-)$.

50 (ix) Si el vector de libro de códigos temporal es mejor que cualquiera de los anteriores, $f(\varepsilon_p^-) > \hat{f}_p$, entonces guardar este vector de libro de códigos, establecer $\hat{f}_p = f(\varepsilon_p^-)$ y continuar a la siguiente

iteración.

5. Definir el vector de libro de códigos \hat{e}_p para que sea el último (es decir, el mejor) de los vectores de libro de códigos guardados.

6. Si el número de pulsos p ha alcanzado el número de pulsos deseado, entonces definir el vector de salida como $\hat{e} = \hat{e}_p$, y parar. De lo contrario, continuar con la etapa 4.

Tal como ya se indicó, en comparación con aplicaciones de ACELP convencionales, en algunas realizaciones, el objetivo se modifica de tal manera que incluye la ZIR en la siguiente trama.

La ecuación 1 describe el modelo predictivo lineal usado en códecs de tipo ACELP. La respuesta a impulso nulo (ZIR, también conocida algunas veces como respuesta a entrada nula), se refiere a la salida del modelo predictivo lineal cuando el residuo de la trama actual (y todas las tramas futuras) se establece a cero. La ZIR puede calcularse fácilmente definiendo el residuo que es cero desde la posición N en adelante como

$$e_K'(n) = \begin{cases} e(n) & \text{para } n < K \\ 0 & \text{para } n \geq K \end{cases} \quad (10a)$$

mediante lo cual la ZIR puede definirse como

$$ZIR_K(n) = \sum_{k=0}^N h(k) e_K'(n-k). \quad (10b)$$

Restando esta ZIR de la señal de entrada, se obtiene una señal que depende del residuo únicamente de la trama actual en adelante.

De manera equivalente, la ZIR puede determinarse filtrando la señal de entrada pasada como

$$ZIR_K(n) = \begin{cases} x(n) & \text{para } n < K \\ -\sum_{k=1}^m \alpha(k) ZIR_K(n-k) & \text{para } n \geq K \end{cases} \quad (10c)$$

La señal de entrada en la que se ha eliminado ZIR se conoce con frecuencia como el objetivo y puede definirse para la trama que comienza en la posición K como $d(n) = x(n) - ZIR_K(n)$. Este objetivo es en principio exactamente igual al objetivo en las normas AMR y G.718. Cuando se cuantifica la señal, la señal cuantificada $\hat{d}(n)$ se compara con $d(n)$ para la duración de una trama $K \leq n < K + N$.

A la inversa, el residuo de la trama actual tiene una influencia sobre las siguientes tramas, mediante lo cual es útil considerar su influencia cuando se cuantifica la señal, es decir, por tanto puede desearse evaluar la diferencia $\hat{d}(n) - d(n)$ también más allá de la trama actual, $n > K + N$. Sin embargo, para hacer esto, puede desearse considerar la influencia del residuo de la trama actual únicamente estableciendo residuos de las siguientes tramas a cero. Por tanto, puede compararse la ZIR de $\hat{d}(n)$ en la siguiente trama. Dicho de otro modo, se obtiene el objetivo modificado:

$$d'(n) = \begin{cases} 0 & n < K \\ d(n) & K \leq n < K + N \\ -\sum_{k=1}^m \alpha(k) d'(n-k) & n > K + N. \end{cases} \quad (10d)$$

De manera equivalente, usando la respuesta a impulso $h(n)$ de $A(z)$, entonces

45

$$d'(n) = \sum_{k=K}^{K+N-1} e(k)h(n-k). \quad (10e)$$

Esta fórmula puede escribirse en una forma de matriz conveniente mediante $d' = He$ donde H y e se definen como en la ecuación 2. Puede observarse que el objetivo modificado es exactamente x de la ecuación 2.

5 En el cálculo de la matriz R , obsérvese que, en teoría, la respuesta a impulso $h(k)$ es una secuencia infinita, lo cual no es realizable en un sistema práctico.

Sin embargo, es posible o bien

10 1) truncar o acotar en intervalo la respuesta a impulso a una longitud finita y determinar la autocorrelación de la respuesta a impulso truncada, o

15 2) calcular el espectro de potencia de la respuesta a impulso usando los espectros de Fourier de los filtros perceptuales y LP asociados, y obtener la autocorrelación mediante una transformada inversa de Fourier.

Ahora se describe una extensión empleando LTP.

El predictor a largo plazo (LTP) también es en realidad un predictor lineal.

20 Según una realización, el determinante 110 de matriz puede estar configurado para determinar la matriz de autocorrelación R dependiendo de un predictor lineal perceptualmente ponderado, por ejemplo, dependiendo del predictor a largo plazo.

25 El LP y LTP pueden convolucionarse para dar un predictor conjunto, que incluye tanto la forma de envolvente espectral como la estructura de armónicos. La respuesta a impulso de un predictor de este tipo será muy larga, mediante lo cual es incluso más difícil de gestionar con la técnica anterior. Sin embargo, si ya se conoce la autocorrelación del predictor lineal, entonces la autocorrelación del predictor conjunto puede calcularse filtrando simplemente la autocorrelación con el LTP hacia delante y hacia atrás, o con un proceso similar en el dominio de frecuencia.

30 Obsérvese que los métodos anteriores que emplean LTP tienen un problema cuando el retardo de LTP es más corto que la longitud de trama, ya que el LTP provocará un bucle de retroalimentación dentro de la trama. El beneficio de incluir el LTP en la función objetivo es que cuando el retardo del LTP es más corto que la longitud de trama, entonces esta retroalimentación se tiene en cuenta de manera explícita en la optimización.

A continuación, se describe una extensión para optimización rápida en un dominio sin correlacionar.

40 Un desafío fundamental en el diseño de sistemas de ACELP ha sido la reducción de la complejidad computacional. Los sistemas de ACELP son complejos porque el filtrado mediante LP provoca correlaciones complicadas entre las muestras residuales, que se describen mediante la matriz B o en el contexto actual mediante la matriz R . Dado que las muestras de $e(n)$ están correlacionadas, no es posible cuantificar simplemente $e(n)$ con la precisión deseada, sino que tienen que intentarse muchas combinaciones de diferentes cuantificaciones con un enfoque de ensayo y error para encontrar la mejor cuantificación con respecto a la función objetivo de la ecuación 3 ó 10, respectivamente.

50 Mediante la introducción de la matriz R , se obtiene una nueva perspectiva con respecto a estas correlaciones. Concretamente, dado que R tiene una estructura de Toeplitz hermitiana, pueden aplicarse varias descomposiciones de matriz eficientes, tales como la descomposición de valor individual, descomposición de Cholesky o descomposición de Vandermonde de matrices de Hankel (las matrices de Hankel son matrices de Toeplitz dadas la vuelta, mediante lo cual pueden aplicarse las mismas descomposiciones a las matrices de Toeplitz y de Hankel) (véanse [6] y [7]). Siendo $R = E D E^H$ una descomposición de R de tal manera que D es una matriz diagonal del mismo tamaño y rango que R . Entonces la ecuación 9 puede modificarse de la siguiente manera:

$$\epsilon^2 = (e - \hat{e})^H R (e - \hat{e}) = (e - \hat{e})^H E D E^H (e - \hat{e}) = (f - \hat{f}) D (f - \hat{f}) \quad (11)$$

55 donde $\hat{f} = E^H \hat{e}$. Dado que D es diagonal, el error para cada muestra de $f(k)$ es independiente de otras muestras $f(i)$. En la ecuación 10, se supone que el vector de libro de códigos se ajusta a escala mediante la ganancia óptima, mediante lo cual la nueva función objetivo es

$$\frac{(f^H D \hat{f})^2}{\hat{f}^H D \hat{f}} \quad (12)$$

5 En este caso, las muestras están correlacionadas de nuevo (dado que cambiar la cuantificación de una línea cambia la ganancia óptima para todas las líneas), pero en comparación con la ecuación 10, el efecto de la correlación en este caso es limitado. Sin embargo, aunque se tenga en cuenta la correlación, la optimización de esta función objetivo es mucho más sencilla que la optimización de las ecuaciones 3 ó 10.

10 Usando este enfoque de descomposición, es posible

1. aplicar cualquier técnica de cuantificación escalar o vectorial convencional con la precisión deseada, o
2. usar la ecuación 12 como función objetivo con cualquier algoritmo de búsqueda de pulsos de ACELP convencional.

15 Ambos enfoques proporcionan una cuantificación casi óptima con respecto a la ecuación 12. Dado que las técnicas de cuantificación convencionales generalmente no requieren ningún método de fuerza bruta (con la excepción de un posible bucle de velocidad), y dado que la matriz D es más sencilla que cualquiera de B o R, ambos métodos de cuantificación son menos complejos que los algoritmos de búsqueda de pulsos de ACELP convencionales. La principal fuente de complejidad computacional en este enfoque es, por tanto, el cálculo de la descomposición de matriz.

Algunas realizaciones emplean la ecuación 12 para determinar un vector de libro de códigos del libro de códigos.

25 Por ejemplo, existen varias factorizaciones de matriz para R de la forma $R = E^H D E$. Por ejemplo,

(a) Puede calcularse la descomposición de valores propios, por ejemplo, usando la GNU Scientific Library (http://www.gnu.org/software/gsl/manual/html_node/Real-Symmetric-Matrices.html). La matriz R es real y simétrica (así como la de Toeplitz), mediante lo cual puede usarse la función "gsl_eigen_symm()" para determinar las matrices E y D. Otras implementaciones de la misma descomposición de valores propios están fácilmente disponibles en la bibliografía [6].

35 (b) Puede usarse la factorización de Vandermonde de matrices de Toeplitz [7] usando el algoritmo descrito en [8]. Este algoritmo devuelve las matrices E y D de tal manera que E es una matriz de Vandermonde, que es equivalente a una transformada discreta de Fourier con una distribución de frecuencia no uniforme.

Usando tales factorizaciones, el vector residual e puede transformarse en el dominio de transformada mediante $f = E^H e$ o $f = D^{1/2} E^H e$. Puede aplicarse cualquier método de cuantificación común en estos dominios, por ejemplo,

40 1. El vector f puede cuantificarse mediante un libro de códigos algebraicos exactamente como en implementaciones habituales de ACELP. Sin embargo, dado que los elementos de f no están correlacionados, no se necesita una función de búsqueda complicada como en ACELP, sino que puede aplicarse un algoritmo sencillo, tal como

- (a) Establecer la ganancia inicial a $g=1$
- 45 (b) Cuantificar f mediante $\hat{f} = \text{round}(gf)$.
- (c) Si el número de pulsos en f es mayor que una cantidad predefinida p, $\|\hat{f}\|_1 > p$, entonces aumentar la ganancia g y volver a la etapa b.
- 50 (d) De lo contrario, si el número de pulsos en \hat{f} es menor que una cantidad predefinida p, $\|\hat{f}\|_1 < p$, entonces reducir la ganancia g y volver a la etapa b.
- (e) De lo contrario, el número de pulsos en \hat{f} es igual a la cantidad predefinida p, $\|\hat{f}\|_1 = p$, y puede detenerse el procesamiento.
- 55

2. Puede usarse un codificador aritmético similar al usado en la cuantificación de líneas espectrales en TCX en las normas AMR-WB+ o MPEG USAC.

5 Debe observarse que, dado que los elementos de f son ortogonales (tal como puede observarse a partir de la ecuación 12) y tienen el mismo peso en la función objetivo de la ecuación 12, pueden cuantificarse por separado, y con el mismo tamaño de etapa de cuantificación. Esta cuantificación encontrará automáticamente el valor óptimo (el más grande) de la función objetivo en la ecuación 12, lo cual es posible con esa precisión de cuantificación. Dicho de otro modo, los algoritmos de cuantificación presentados anteriormente devolverán, ambos, la cuantificación óptima con respecto a la ecuación 12.

10 Esta ventaja de optimalidad está vinculada al hecho de que los elementos de f pueden tratarse por separado. Si se usa un enfoque de libro de códigos, en el que los vectores de libro de códigos c_k no son triviales (tienen más de un elemento distinto de cero), entonces estos vectores de libro de códigos ya no tendrán elementos independientes y se pierde la ventaja de la factorización de matriz.

15 Obsérvese que la factorización de Vandermonde de una matriz de Toeplitz puede elegirse de tal manera que la matriz de Vandermonde es una matriz de transformada de Fourier pero con frecuencias uniformemente distribuidas. Dicho de otro modo, la matriz de Vandermonde corresponde a una transformada de Fourier con alteración de frecuencia. En este caso se desprende que el vector f corresponde a una representación en el dominio de frecuencia de la señal residual en una escala de frecuencia alterada (véase la "propiedad de intercambio de raíz" en [8]).

20 Resulta importante observar que esta consecuencia no se conoce bien. En la práctica, este resultado establece que si se filtra una señal x con una matriz de convolución C , entonces

$$\|C x\|^2 = \|D V x\|^2 \quad (13)$$

25 donde V es una transformada de Fourier (por ejemplo, alterada) (que es una matriz de Vandermonde con elementos en el círculo unitario) y D una matriz diagonal. Es decir, si se desea medir la energía de una señal filtrada, puede medirse de manera equivalente la energía de la señal con alteración de frecuencia. A la inversa, cualquier evaluación que se realice en un dominio de Fourier alterado puede realizarse de manera equivalente en un dominio de tiempo filtrado. Debido a la dualidad del tiempo y la frecuencia, también existe una equivalencia entre acotamiento de intervalos en el dominio de tiempo y la alteración de tiempo. Sin embargo, una cuestión práctica es que hallar una matriz de convolución C que satisfaga la relación anterior supone un problema numéricamente sensible, mediante lo cual con frecuencia resulta más fácil hallar en su lugar soluciones aproximadas C .

Puede emplearse la relación $\|C x\|^2 = \|D V x\|^2$ para determinar un vector de libro de códigos de un libro de códigos.

35 Para ello, en primer lugar, debe observarse que, en este caso, mediante H se indicará una matriz de convolución tal como en la ecuación 2 en lugar de C . Entonces, si se desea minimizar el ruido de cuantificación $e = Hx - H\hat{x}$, puede medirse su energía:

$$\begin{aligned} \varepsilon^2 &= \|Hx - H\hat{x}\|^2 = \|H(x - \hat{x})\|^2 = (x - \hat{x})^T H^T H (x - \hat{x}) = (x - \hat{x})^T R (x - \hat{x}) \\ &= (x - \hat{x})^T V^H D V (x - \hat{x}) = \|D^{1/2} V (x - \hat{x})\|^2 = \|D^{1/2} V (x - \hat{x})\|^2 = \\ &= \|D^{1/2} (f - \hat{f})\|^2 = \|f' - \hat{f}'\|^2. \end{aligned} \quad (13a)$$

40 Ahora, se describe una extensión para la independencia de tramas.

45 Cuando la señal de voz codificada se transmite por líneas de transmisión imperfectas tales como ondas de radio, de manera invariable algunas veces se pierden paquetes de datos. Si las tramas dependen unas de otras, de tal manera que se necesita el paquete N para decodificar $N-1$ de manera perfecta, entonces la pérdida del paquete $N-1$ corromperá la síntesis de ambos paquetes $N-1$ y N . Por otro lado, si las tramas son independientes, entonces la pérdida del paquete $N-1$ solo corromperá la síntesis del paquete $N-1$. Por tanto, resulta importante diseñar métodos que estén libres de dependencias entre tramas.

50 En sistemas de ACELP convencionales, la fuente principal de dependencia entre tramas es el LTP y en cierta medida también el LP. Específicamente, dado que ambos son filtros de respuesta a impulso infinita (IIR), una trama corrompida provocará una cola "infinita" de muestras corrompidas. En la práctica, esa cola puede tener una longitud de varias tramas, lo cual es perceptualmente molesto.

55 Usando el entramado de la presente invención, la trayectoria a través de la cual se genera dependencia entre tramas puede cuantificarse mediante la ZIR desde la trama actual a la siguiente que se realiza. Para evitar esta dependencia entre tramas, necesitan realizarse tres modificaciones a la ACELP convencional.

1. Cuando se calcula la ZIR desde la trama anterior a la (sub)trama actual, debe calcularse a partir del residuo original (no cuantificado) extendido con ceros, no a partir del residuo cuantificado. De esta manera, los errores de cuantificación procedentes de la (sub)trama anterior no se propagarán a la (sub)trama actual.

5 2. Cuando se cuantifica la trama actual, debe tenerse en cuenta el error en la ZIR en la siguiente trama entre las señales original y cuantificada. Esto puede realizarse sustituyendo la matriz de correlación B por la matriz de autocorrelación R, tal como se explicó anteriormente. Esto garantiza que el error en la ZIR en la siguiente trama se minimiza junto con el error dentro de la trama actual.

10 3. Dado que la propagación de error se debe tanto al LP como al LTP, deben incluirse ambos componentes en la ZIR. Esto difiere del enfoque convencional en el que la ZIR se calcula únicamente para el LP.

Si no se tienen en cuenta los errores de cuantificación de la trama anterior cuando se cuantifica la trama actual, se pierde eficiencia de la calidad perceptual de la salida. Por tanto, es posible elegir tener en cuenta los errores anteriores cuando no existe riesgo de propagación de error. Por ejemplo, un sistema de ACELP convencional aplica una delimitación de tramas en la que cada trama de 20 ms se subdivide en 4 ó 5 subtramas. El LTP y el residuo se cuantifican y codifican por separado para cada subtrama, pero la trama completa se transmite como un bloque de datos. Por tanto, no pueden perderse subtramas individuales, sino solo tramas completas. Se desprende que solo se requiere usar ZIR independientes de trama en los límites de trama, pero pueden usarse ZIR con dependencias entre tramas entre las subtramas restantes.

Realizaciones modifican los algoritmos de ACELP convencionales mediante inclusión del efecto de la respuesta a impulso de la trama actual en la siguiente trama, en la función objetivo de la trama actual. En la función objetivo del problema de optimización, esta modificación corresponde a sustituir una matriz de correlación por una matriz de autocorrelación que tiene una estructura de Toeplitz hermitiana. Esta modificación tiene los siguientes beneficios:

1. Se reducen la complejidad computacional y los requisitos de memoria debido a la estructura de Toeplitz hermitiana añadida de la matriz de autocorrelación.

30 2. Se aplicará el mismo modelo perceptual a todas las muestras, haciendo que el diseño y ajuste del modelo perceptual sean más sencillos, y su aplicación más eficiente y sistemática.

35 3. Pueden evitarse completamente correlaciones entre tramas en la cuantificación de la trama actual, teniendo en cuenta únicamente la respuesta a impulso sin cuantificar desde la trama anterior y la respuesta a impulso cuantificada en la siguiente trama. Esto mejora la robustez de sistemas en los que se espera pérdida de paquetes.

La figura 2 ilustra un decodificador 220 para decodificar una señal de voz codificada que está codificándose por un aparato según la realización descrita anteriormente para obtener una señal de voz decodificada. El decodificador 220 está configurado para recibir la señal de voz codificada, en el que la señal de voz codificada comprende una indicación del vector de libro de códigos, que se determina por un aparato para codificar una señal de voz según una de las realizaciones descritas anteriormente, por ejemplo, un índice del vector de libro de códigos determinado. Además, el decodificador 220 está configurado para decodificar la señal de voz codificada para obtener una señal de voz decodificada dependiendo del vector de libro de códigos.

45 La figura 3 ilustra un sistema según una realización. El sistema comprende un aparato 210 según una de las realizaciones descritas anteriormente para codificar una señal de voz de entrada para obtener una señal de voz codificada. La señal de voz codificada comprende una indicación del vector de libro de códigos determinado, determinado por el aparato 210 para codificar una señal de voz, por ejemplo, comprende un índice del vector de libro de códigos. Además, el sistema comprende un decodificador 220 según la realización descrita anteriormente para decodificar la señal de voz codificada para obtener una señal de voz decodificada. El decodificador 220 está configurado para recibir la señal de voz codificada. Además, el decodificador 220 está configurado para decodificar la señal de voz codificada para obtener una señal de voz decodificada dependiendo del vector de libro de códigos determinado.

55 Aunque algunos aspectos se han descrito en el contexto de un aparato, estos aspectos también representan una descripción del método correspondiente, en el que un bloque o dispositivo corresponde a una etapa de método o a una característica de una etapa de método. De manera análoga, aspectos descritos en el contexto de una etapa de método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente.

60 La señal descompuesta de la invención puede almacenarse en un medio de almacenamiento digital o puede transmitirse por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión cableado tal como Internet.

5 Dependiendo de determinados requisitos de implementación, pueden implementarse realizaciones de la invención en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, por ejemplo, un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control electrónicamente legibles almacenadas en el mismo, que actúan conjuntamente (o que pueden actuar conjuntamente) con un sistema informático programable de tal manera que se realiza el método respectivo.

10 Algunas realizaciones según la invención comprenden un soporte de datos no transitorio que tiene señales de control electrónicamente legibles, que pueden actuar conjuntamente con un sistema informático programable, de tal manera que se realiza uno de los métodos descritos en el presente documento.

15 Generalmente, realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, pudiendo funcionar el código de programa para realizar uno de los métodos cuando se ejecuta el producto de programa informático en un ordenador. El código de programa puede estar almacenado, por ejemplo, en un soporte legible por máquina.

Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en el presente documento, almacenado en un soporte legible por máquina.

20 Dicho de otro modo, una realización del método de la invención es, por tanto, un programa informático que tiene un código de programa para realizar uno de los métodos descritos en el presente documento, cuando se ejecuta el programa informático en un ordenador.

25 Una realización adicional de los métodos de la invención es, por tanto, un soporte de datos (o un medio de almacenamiento digital o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los métodos descritos en el presente documento.

30 Una realización adicional del método de la invención es, por tanto, un flujo de datos o una secuencia de señales que representa el programa informático para realizar uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales pueden estar configurados, por ejemplo, para transferirse a través de una conexión de comunicación de datos, por ejemplo a través de Internet.

35 Una realización adicional comprende unos medios de procesamiento, por ejemplo, un ordenador, o un dispositivo lógico programable, configurados o adaptados para realizar uno de los métodos descritos en el presente documento.

Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para realizar uno de los métodos descritos en el presente documento.

40 En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo, una matriz de puertas programable en campo) para realizar algunas o todas de las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, una matriz de puertas programable en campo puede actuar conjuntamente con un microprocesador con el fin de realizar uno de los métodos descritos en el presente documento. Generalmente, los métodos se realizan preferiblemente mediante cualquier aparato de hardware.

45 Las realizaciones descritas anteriormente son simplemente ilustrativas de los principios de la presente invención. Se entiende que modificaciones y variaciones de las disposiciones y los detalles descritos en el presente documento resultarán evidentes para otros expertos en la técnica. Por tanto, la intención es limitarse únicamente por el alcance de las reivindicaciones de patente a continuación y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones en el presente documento.

50 Bibliografía

[1] Salami, R. y Laflamme, C. y Bessette, B. y Adoul, J.P., "ITU-T G. 729 Annex A: reduced complexity 8 kb/s CS-ACELP codec for digital simultaneous voice and data", Communications Magazine, IEEE, vol. 35, n.º 9, págs. 56-63, 1997.

[2] 3GPP TS 26.190 V7.0.0, "Adaptive Multi-Rate (AMR-WB) speech codec", 2007.

60 [3] ITU-T G.718, "Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s", 2008.

[4] Schroeder, M. y Atal, B., "Code-excited linear prediction (CELP): High-quality speech at very low bit rates", Acoustics, Speech, and Signal Processing, IEEE Int Conf, págs. 937-940, 1985.

- [5] Byun, K.J. y Jung, H.B. y Hahn, M. y Kim, K.S., "A fast ACELP codebook search method", Signal Processing, 2002 6th International Conference on, vol. 1, págs. 422-425, 2002.
- 5 [6] G. H. Golub y C. F. van Loan, "Matrix Computations", 3rd Edition, John Hopkins University Press, 1996.
- [7] Boley, D.L. y Luk, F.T. y Vandevoorde, D., "Vandermonde factorization of a Hankel matrix", Scientific computing, págs. 27-39, 1997.
- 10 [8] Bäckström, T. y Magi, C., "Properties of line spectrum pair polynomials - A review", Signal processing, vol. 86, n.º 11, págs. 3286-3298, 2006.
- [9] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. Laine, y J. Huopaniemi, "Frequencywarped signal processing for audio applications," J. Audio Eng. Soc, vol. 48, n.º 11, págs. 1011-1031, 2000.
- 15 [10] T. Laakso, V. Välimäki, M. Karjalainen, y U. Laine, "Splitting the unit delay [FIR/all pass filters design]," IEEE Signal Process. Mag., vol. 13, n.º 1, págs. 30-60, 1996.
- [11] J. Smith III y J. Abel, "Bark and ERB bilinear transforms," IEEE Trans. Speech Audio Process., vol. 7, n.º 6, págs. 697-708, 1999.
- 20 [12] R. Schappelle, "The inverse of the confluent Vandermonde matrix," IEEE Trans. Autom. Control, vol. 17, n.º 5, págs. 724-725, 1972.
- [13] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, y K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," Speech and Audio Processing, IEEE Transactions on, vol. 10, n.º 8, págs. 620-636, 2002.
- 25 [14] M. Bosi y R. E. Goldberg, Introduction to Digital Audio Coding and Standards. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2003.
- 30 [15] B. Edler, S. Disch, S. Bayer, G. Fuchs, y R. Geiger, "A time-warped MDCT approach to speech transform coding," in Proc 126th AES Convention, Múnich, Alemania, mayo de 2009.
- [16] J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE, vol. 63, n.º 4, págs. 561-580, abril de 1975.
- 35 [17] J.-P. Adoul, P. Mabillean, M. Delprat, y S. Morissette, "Fast CELP coding based on algebraic codes," en Acoustics, Speech, and Signal Processing, IEEE Int Conf (ICASSP'87), abril de 1987, págs. 1957-1960.
- [18] ISO/IEC 23003-3:2012, "MPEG-D (MPEG audio technologies), Part 3: Unified speech and audio coding," 2012.
- 40 [19] F.-K. Chen y J.-F. Yang, "Maximum-take-precedence ACELP: a low complexity search method," en Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on, vol. 2. IEEE, 2001, págs. 693-696.
- 45 [20] R. P. Kumar, "High computational performance in code exited linear prediction speech model using faster codebook search techniques," in Proceedings of the International Conference on Computing: Theory and Applications. IEEE Computer Society, 2007, págs. 458-462.
- [21] N. K. Ha, "A fast search method of algebraic codebook by reordering search sequence," en Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, vol. 1. IEEE, 1999, págs. 21-24.
- 50 [22] M. A. Ramirez y M. Gerken, "Efficient algebraic multipulse search," in Telecommunications Symposium, 1998. ITS'98 Proceedings. SBT/IEEE International. IEEE, 1998, págs. 231-236.
- [23] ITU-T Recommendation G.191, "Software tool library 2009 user's manual," 2009.
- 55 [24] ITU-T Recommendation P.863, "Perceptual objective listening quality assessment," 2011.
- [25] T. Thiede, W. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg et al., "PEAQ - the ITU standard for objective measurement of perceived audio quality," Journal of the Audio Engineering Society, vol. 48, 2012.
- 60 [26] ITU-R Recommendation BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems," 2003.
- 65

REIVINDICACIONES

1. Aparato para codificar una señal de voz determinando un vector de libro de códigos de un algoritmo de codificación de voz, en el que el aparato comprende:

5 un determinador (110) de matriz para determinar una matriz de autocorrelación R, y
un determinador (120) de vector de libro de códigos para determinar el vector de libro de códigos dependiendo de la matriz de autocorrelación R,

10 en el que el determinador (110) de matriz está configurado para determinar la matriz de autocorrelación R determinando coeficientes de vector de un vector r, en el que la matriz de autocorrelación R comprende una pluralidad de filas y una pluralidad de columnas, en el que el vector r indica una de las columnas o una de las filas de la matriz de autocorrelación R, en el que

$$15 \quad R(i, j) = r(|i - j|),$$

en el que R(i, j) indica los coeficientes de la matriz de autocorrelación R, en el que i es un primer índice que indica una de una pluralidad de filas de la matriz de autocorrelación R, y en el que j es un segundo índice que indica una de una pluralidad de columnas de la matriz de autocorrelación R,

20 en el que el determinador (120) de vector de libro de códigos está configurado para determinar el vector de libro de códigos aplicando la fórmula

$$25 \quad f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

en el que R es la matriz de autocorrelación, y en el que \hat{e} es uno de los vectores de libro de códigos del algoritmo de codificación de voz, en el que $f(\hat{e})$ es una correlación normalizada, y en el que d^T se define según

$$30 \quad \frac{(e^T R \hat{e})^2}{(\hat{e}^T R \hat{e})} = \frac{(d^T e)^2}{(e^T R e)},$$

en el que e es una señal residual original sin cuantificar.

35 2. Aparato según la reivindicación 1,

en el que el determinador (120) de vector de libro de códigos está configurado para determinar el vector de libro de códigos \hat{e} del algoritmo de codificación de voz que maximiza la correlación normalizada.

$$40 \quad f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}.$$

3. Aparato según la reivindicación 1 ó 2,

45 en el que el determinador (110) de matriz está configurado para determinar los coeficientes de vector del vector r aplicando la fórmula:

$$r(k) = h(k) * h(-k) = \sum_l h(l)h(l - k)$$

50 en el que h(k) indica una respuesta a impulso perceptualmente ponderada de un modelo predictivo lineal, y en el que k es un índice que es un número entero, y en el que l es un índice que es un número entero.

4. Aparato según una de las reivindicaciones anteriores,

en el que el determinante (110) de matriz está configurado para determinar la matriz de autocorrelación R dependiendo de un predictor lineal perceptualmente ponderado.

5. Aparato según una de las reivindicaciones anteriores, en el que el determinante (120) de vector de libro de códigos está configurado para descomponer la matriz de autocorrelación R realizando una descomposición de matriz.

6. Aparato según la reivindicación 5, en el que el determinante (120) de vector de libro de códigos está configurado para realizar la descomposición de matriz para determinar una matriz diagonal D para determinar el vector de libro de códigos.

7. Aparato según la reivindicación 6, en el que el determinante (120) de vector de libro de códigos está configurado para determinar el vector de libro de códigos empleando

$$\frac{(f^H D \hat{f})^2}{\hat{f}^H D \hat{f}},$$

en el que D es la matriz diagonal, en el que f es un primer vector, y en el que \hat{f} es un segundo vector.

8. Aparato según la reivindicación 6 ó 7, en el que el determinante (120) de vector de libro de códigos está configurado para realizar una factorización de Vandermonde en la matriz de autocorrelación R para descomponer la matriz de autocorrelación R para realizar la descomposición de matriz para determinar la matriz diagonal D para determinar el vector de libro de códigos.

9. Aparato según una de las reivindicaciones 6 a 8, en el que el determinante (120) de vector de libro de códigos está configurado para emplear la ecuación

$$\| Cx \|^2 = \| DVx \|^2$$

para determinar el vector de libro de códigos, en el que C indica una matriz de convolución, en el que V indica una transformada de Fourier, y en el que x indica la señal de voz.

10. Aparato según una de las reivindicaciones 6 a 9, en el que el determinante (120) de vector de libro de códigos está configurado para realizar una descomposición de valor individual en la matriz de autocorrelación R para descomponer la matriz de autocorrelación R para realizar la descomposición de matriz para determinar la matriz diagonal D para determinar el vector de libro de códigos.

11. Aparato según una de las reivindicaciones 6 a 9, en el que el determinante (120) de vector de libro de códigos está configurado para realizar una descomposición de Cholesky en la matriz de autocorrelación R para descomponer la matriz de autocorrelación R para realizar la descomposición de matriz para determinar la matriz diagonal D para determinar el vector de libro de códigos.

12. Aparato según una de las reivindicaciones anteriores, en el que el determinante (120) de vector de libro de códigos está configurado para determinar el vector de libro de códigos dependiendo de una respuesta a impulso nulo de la señal de voz.

13. Aparato según una de las reivindicaciones anteriores, en el que el aparato es un codificador para codificar la señal de voz empleando codificación de voz mediante predicción lineal con excitación por código algebraico, y

en el que el determinante (120) de vector de libro de códigos está configurado para determinar el vector de libro de códigos basándose en la matriz de autocorrelación R como vector de libro de códigos de un libro de códigos algebraicos.

14. Sistema, que comprende: un aparato (210) según una de las reivindicaciones 1 a 13 para codificar una señal de voz de entrada para obtener una señal de voz codificada, y

un decodificador (220) para decodificar la señal de voz codificada para obtener una señal de voz decodificada,

5 en el que el decodificador (220) está configurado para recibir la señal de voz codificada, en el que la señal de voz codificada comprende una indicación de un vector de libro de códigos, que se determina por el aparato según una de las reivindicaciones 1 a 13,

10 en el que el decodificador (220) está configurado para decodificar la señal de voz codificada para obtener la señal de voz decodificada dependiendo del vector de libro de códigos.

15. Método para codificar una señal de voz determinando un vector de libro de códigos de un algoritmo de codificación de voz, en el que el método comprende:

15 determinar una matriz de autocorrelación R, y

determinar el vector de libro de códigos dependiendo de la matriz de autocorrelación R,

20 en el que determinar una matriz de autocorrelación R comprende determinar coeficientes de vector de un vector r, en el que la matriz de autocorrelación R comprende una pluralidad de filas y una pluralidad de columnas, en el que el vector r indica una de las columnas o una de las filas de la matriz de autocorrelación R, en el que

$$R(i, j) = r(|i-j|),$$

25 en el que R(i, j) indica los coeficientes de la matriz de autocorrelación R, en el que i es un primer índice que indica una de una pluralidad de filas de la matriz de autocorrelación R, y en el que j es un segundo índice que indica una de una pluralidad de columnas de la matriz de autocorrelación R,

30 en el que la determinación del vector de libro de códigos se realiza aplicando la fórmula

$$f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

35 en el que R es la matriz de autocorrelación, y en el que \hat{e} es uno de los vectores de libro de códigos del algoritmo de codificación de voz, en el que $f(\hat{e})$ es una correlación normalizada, y en el que d^T se define según

$$\frac{(e^T R \hat{e})^2}{(\hat{e}^T R \hat{e})} = \frac{(d^T e)^2}{(\hat{e}^T R \hat{e})},$$

40 en el que e es una señal residual original sin cuantificar.

16. Método que comprende:

45 codificar una señal de voz de entrada según el método según la reivindicación 15 para obtener una señal de voz codificada, en el que la señal de voz codificada comprende una indicación de un vector de libro de códigos, y

decodificar la señal de voz codificada para obtener la señal de voz decodificada dependiendo del vector de libro de códigos.

50 17. Programa informático para implementar el método según la reivindicación 15 ó 16, cuando se ejecuta en un ordenador o procesador de señales.

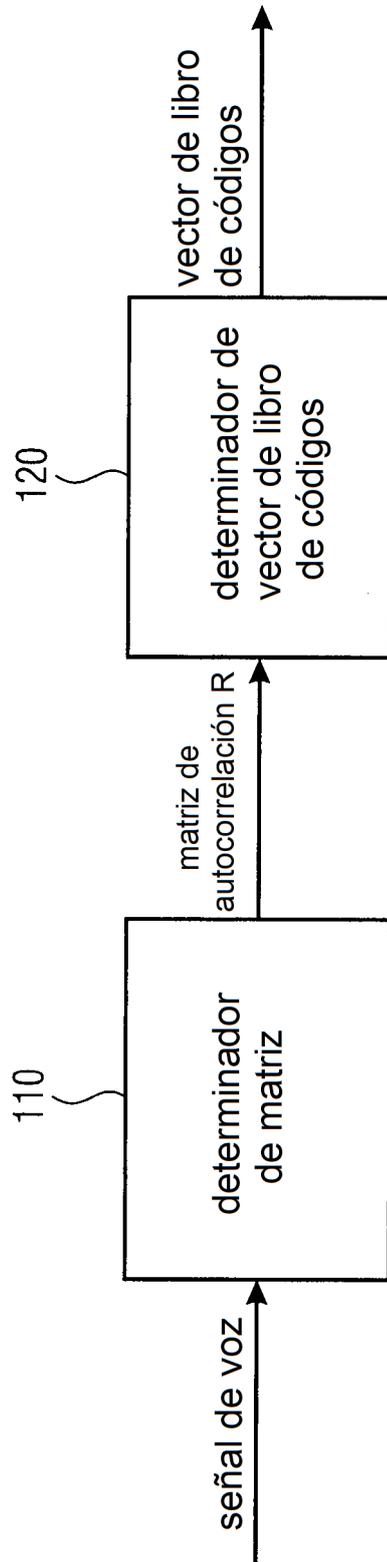


FIG 1

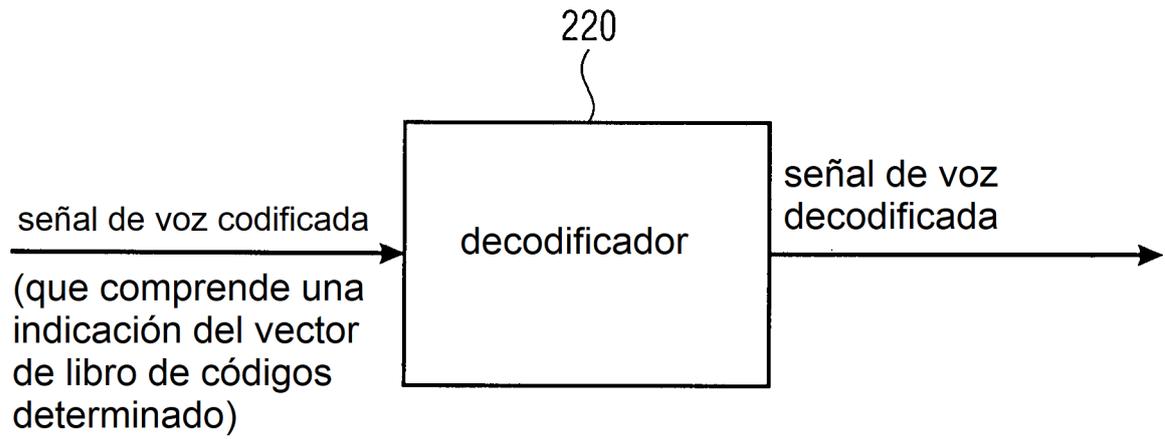


FIG 2

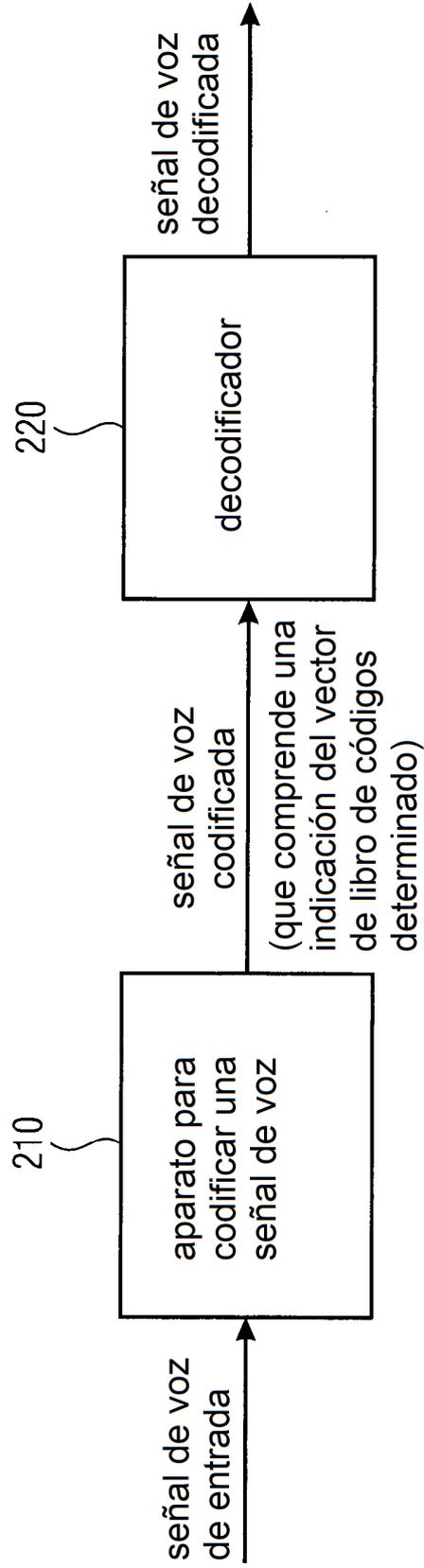


FIG 3