



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 366 551**

51 Int. Cl.:
G10L 19/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **06829172 .3**

96 Fecha de presentación : **29.11.2006**

97 Número de publicación de la solicitud: **2087485**

97 Fecha de publicación de la solicitud: **12.08.2009**

54 Título: **Codificación y decodificación dependiente de una fuente de múltiples libros de códigos.**

45 Fecha de publicación de la mención BOPI:
21.10.2011

45 Fecha de la publicación del folleto de la patente:
21.10.2011

73 Titular/es: **LOQUENDO S.p.A.**
Via Arrigo Olivetti 6
10100 Torino, IT

72 Inventor/es: **Massimino, Paolo;**
Coppo, Paolo y
Vecchietti, Marco

74 Agente: **Ponti Sales, Adelaida**

ES 2 366 551 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

CAMPO TÉCNICO DE LA INVENCION

5 [0001] La presente invención se refiere en general a la codificación de señales, y en particular a la codificación de señales de voz/audio. Más detalladamente, la presente invención se refiere a la codificación y decodificación de
 10 señales de voz/audio a través del modelado de un número variable de libros de códigos, proporcionando la calidad de la señal reconstruida y la ocupación de memoria/ancho de banda de transmisión. La presente invención proporciona una aplicación ventajosa, pero no exclusiva, en la síntesis de voz, en particular en la síntesis de voz
 basada en corpus, en la que la señal de fuente se conoce a priori, y a la que la siguiente descripción hará referencia sin que implique ninguna pérdida de generalidad.

ANTECEDENTES DE LA TÉCNICA

15 [0002] En el campo de la síntesis de voz, basada en particular en la concatenación de segmentos de sonido para obtener la frase deseada, aumenta la demanda de representar el material de voz utilizado en el proceso de síntesis de forma compacta. El CELP (del inglés, *Code Excited Linear Prediction*) es una técnica muy conocida para
 20 representar una señal de voz de forma compacta, y se caracteriza por la adopción de un procedimiento, conocido como A-b-S (del inglés, *Analysis by Synthesis*), que consiste en separar la señal de voz en componentes de excitación y de secuencia vocal, codificando los coeficientes de excitación de predicción lineal (LCPs) para el componente de secuencia vocal utilizando un índice que apunta a una serie de representaciones almacenadas en un libro de códigos. La selección del mejor índice para la excitación y para la secuencia vocal se escoge comparando la señal original con la señal reconstruida. Una descripción completa de la técnica CELP se encuentra en Wai C. Chu, "Speech Coding Algorithms", ISBN 0-471-37312-5, p. 299-324. A su vez, las versiones modificadas del CELP se describen en US 2005/197833, US 2005/096901 y US 2006/206317.

25 [0003] Otros ejemplos de la técnica CELP se describen en Shigeo et Al. "Very low bit rate speech coding based on a phoneme recognition", Procedimientos del simposio internacional sobre teoría de la información (ISIT), Nueva York, IEEE, US, 19 de junio de 1988, páginas 71-72, y en Hernández-Gómez et Al., "Phonetically driven CELP coding using self-organizing maps", Procedimientos de la conferencia internacional sobre acústica, voz y procesamiento de
 30 señales (ICASSP Nueva York, IEEE, US, vol. 4, 27 de abril de 1993, páginas 628-631.

[0004] El documento WO 99/59137 describe un procedimiento para codificar voz que incluye la división de una señal de voz en una serie de tramos, convirtiendo cada uno de los tramos en una señal codificada que incluye parámetros de filtrado asignando series consecutivas de segmentos, de modo que cada segmento corresponde a una clasificación de una manifestación del lenguaje particular, y cuantificando el tramo o tramos en un segmento mediante la referencia a un libro de códigos específico de la clasificación del segmento.

35 [0005] La figura 1 muestra un diagrama de bloques de la técnica CELP para la codificación de señales de voz, en el que la secuencia vocal y la fuente glotal son modeladas mediante una fuente de impulso (excitación), indicado por F1-1, y por un filtro digital de variante temporal (filtro de síntesis), indicado por F1-2. OBJETO Y RESUMEN DE LA INVENCION.

40 [0006] El solicitante ha observado que, en general, en los procedimientos conocidos, los componentes de excitación y de secuencia vocal son modelados independientemente del hablante, lo que conduce a una codificación de la señal de voz con una ocupación de memoria reducida de la señal original. Por el contrario, el solicitante también ha observado que la aplicación de este tipo de moldeado provoca una reconstrucción imperfecta de la señal original: de hecho, cuanto menor sea la ocupación de la memoria, mayor será la degradación de la señal reconstruida respecto
 45 a la señal original. Este tipo de codificación toma el nombre de codificación con pérdidas (en el sentido de pérdida de información). En otras palabras, el solicitante ha observado que el libro de códigos del que se escoge el índice de excitación y el libro de códigos del que se escoge el mejor modelo de secuencia modular no varían sobre la base de la señal de voz que está destinada a ser un código, sino que son fijos e independientes de la señal de voz, porque los libros de códigos utilizados se construyen para trabajar para una multitud de voces y no son optimizados para las características de una sola voz.

50 [0007] El objetivo de la presente invención es por tanto proporcionar una técnica de codificación y decodificación efectiva y eficiente que dependa de la fuente, y que permita una mejor proporción entre la calidad de la señal reconstruida y el ancho de banda de ocupación/transmisión de la memoria que se quiera alcanzar con respecto a las técnicas de codificación y decodificación conocidas independientes de la fuente.

55 [0008] Este objeto se logra mediante la presente invención en que se refiere a un procedimiento de codificación, un procedimiento de decodificación, un codificador y productos de software como se define en las reivindicaciones adjuntas.

60 [0009] La presente invención consigue el objetivo arriba mencionado considerando una definiendo de un grado de aproximación en la representación de la señal original en la forma codificada basándose en la reducción deseada de la ocupación de memoria o el ancho de banda de transmisión disponible. En particular, la presente invención incluye agrupar datos en tramos; clasificar los tramos en clases; para cada clase, transformar los tramos pertenecientes a la clase en vectores de parámetros de filtro; para cada clase, calcular un libro de códigos de filtros basándose en los vectores de parámetros de filtros pertenecientes a la clase; segmentar cada tramo en sub-tramos; para cada clase, transformar los sub-tramos pertenecientes a la clase en vectores de parámetro origen, que se extraen de los sub-tramos aplicando una transformación de filtro asada en el libro de códigos de filtro computado para la clase correspondiente; para cada clase, calcular un libro de códigos de fuente basándose en los vectores de parámetros
 65

de fuente pertenecientes a la clase; y codificar los datos basándose en los libros de códigos de origen. [0010] El término clase identifica aquí una categoría de unidades o sub-unidades audibles básicas, como fonemas, demifonemas, difonemas, etc.

5 [0011] Según un primer aspecto, la invención se refiere a un procedimiento para codificar datos de audio, comprendiendo:

- agrupar datos en tramos;
- clasificar los tramos en clases:
- 10 • para cada clase, transformar los tramos pertenecientes a la clase en vectores de parámetros de filtro;
- para cada clase, calcular un libro de códigos de filtro basándose en los vectores de parámetros de filtro pertenecientes a la clase;
- segmentar cada tramo en sub-tramos; • para cada clase, transformar los sub-tramos pertenecientes a la clase en vectores de parámetros de fuente, que se extraen de los sub-tramos aplicando una transformación por filtrado basada en el libro de códigos de filtros computado para la clase correspondiente;
- 15 • para cada clase, calcular un libro de códigos de fuente basándose en los vectores de parámetros de fuente pertenecientes a la clase; y
- codificar los datos basándose en los libros de códigos de filtros y de fuente computados.

20 [0012] Preferiblemente, los datos son muestras de una señal de voz, y las clases son clases fonéticas, p ej., demifonema o fracciones de clases de demifonema.

[0013] Preferiblemente, la clasificación de los tramos en clases incluye:

- si la cardinalidad de una clase satisface cierto criterio de clasificación, asociar los tramos con la clase;
- 25 • si la cardinalidad de una clase no satisface dicho criterio de clasificación, asociar los tramos con sub-clases para conseguir una distribución uniforme de la cardinalidad de las sub-clases, en la que el criterio de clasificación está definido por una condición en la que la cardinalidad de la clase está por debajo de un umbral predeterminado y en el que los datos son muestras de una señal de voz, y en el que las clases son clases fonéticas y las sub-clases son clases de demifonemas.

30 [0014] Preferiblemente, los datos son muestras de una señal de voz, los vectores de parámetros de filtro obtenidos de los tramos son tales que pueden modelar una secuencia vocal de un hablante, y los vectores de parámetros de filtro son coeficientes de predicción lineales.

[0015] Preferiblemente, transformar los tramos pertenecientes a la clase en vectores de parámetros de filtro incluye llevar a cabo un algoritmo Levinson-Durbin.

35 [0016] Preferiblemente, la etapa de calcular un libro de códigos de filtro basándose en los vectores de parámetros de filtro pertenecientes a la clase incluye:

- calcular vectores de parámetros de filtro específicos que minimicen la distancia global entre ellos y los vectores de parámetros de filtro en la clase, y basándose en una distancia métrica dada; y
- 40 • calcular el libro de códigos de filtro basándose en los vectores de parámetros de filtro específicos, en el que la distancia métrica depende de la clase a la que cada vector de parámetro de filtro pertenece, más preferiblemente, la distancia métrica es la distancia euclidiana definida para un espacio de vector N dimensional.

45 [0017] Preferiblemente, los vectores de parámetro de filtro específicos son vectores de parámetro de filtro centroides computados mediante la aplicación de un algoritmo de agrupación *kmeans*, y el libro de códigos de filtros está formado por los vectores de parámetro de filtro específicos.

[0018] Preferiblemente, la etapa de segmentar cada tramo en sub-tramos incluye:

- 50 • definir una segunda ventana de análisis de muestra como un submúltiplo de la anchura de la primera ventana de análisis de muestra; y
- segmentar cada tramo en un número de sub-tramos correlacionados con el ratio entre la anchura de la primera y la segunda ventana de análisis de muestra, en el que el ratio entre la anchura de la primera y la segunda ventana de muestra de análisis va de cuatro a cinco.

55 [0019] Preferiblemente, los datos son muestras de una señal de voz, y los vectores de parámetros de fuente obtenidos de los sub-tramos son tales que pueden modelar una señal de excitación de un hablante.

[0020] Preferiblemente, la transformación por filtrado se aplica a un número de sub-tramos correlacionados con el ratio entre la anchura de la primera y la segunda ventana de análisis de muestra.

60 [0021] Preferiblemente, la etapa de calcular un libro de códigos de fuente basándose en los vectores de parámetros de fuente pertenecientes a la clase incluye:

- 65 • calcular vectores de parámetros de fuente específicos que minimicen la distancia global entre ellos y los vectores de parámetros de fuente en la clase, y basándose en una distancia métrica dada; y

- calcular el libro de códigos de fuente basándose en los vectores de parámetros de fuente específicos, en el que la distancia métrica depende de la clase a la que pertenece cada vector de parámetro de filtro.

5 [0022] Preferiblemente, la distancia métrica es la distancia euclidiana definida por un espacio de vector dimensional-N.

[0023] Preferiblemente, los vectores de parámetro de fuente específicos son vectores de parámetro de fuente centroides calculados mediante la aplicación de un algoritmo de agrupación *kmeans*, y el libro de códigos de fuente está formado por los vectores de parámetro de fuente específicos.

10 [0024] Preferiblemente, la etapa de codificar los datos basándose en los libros de códigos de filtros y de fuente calculados incluye:

- asociar a cada tramo los índices que identifican un vector de parámetro de filtro en el libro de códigos de filtro y los vectores de parámetros de fuente en el libro de códigos de fuente, que representan las muestra en el tramo en los sub-tramos respectivos.

15 [0025] Preferiblemente, la etapa de asociar a cada tramo los índices que identifican un vector de parámetro de filtro en el libro de códigos de filtro y los vectores de parámetro de fuente en el libro de códigos de fuente, que representan las muestras en el tramo y en los sub-tramos respectivos, incluye:

- 20
- definir una distancia métrica; y
 - elegir el vector de parámetro de filtro y los vectores de parámetro de fuente más cercanos basándose en la distancia métrica definida.

[0026] Preferiblemente, la etapa de elegir el vector de parámetro de filtro y los vectores de parámetro de fuente más cercanos basándose en la distancia métrica definida incluye:

- 25
- escoger el vector de parámetro de filtro y los vectores de parámetro de fuente que minimizan la distancia entre los datos originales y los datos reconstruidos, en el que los datos son muestras de una señal de voz, y en el que escoger el vector de parámetro de filtro y los vectores de parámetro de fuente más cercanos basándose en la distancia métrica definida incluye:

- 30
- escoger el vector de parámetro de filtro y los vectores de parámetro de fuente que minimizan la distancia entre los datos de voz originales y los datos reconstruidos ponderados con una función que modela la curva de percepción auditiva y la señal de voz reconstruida ponderada con la misma curva de percepción auditiva.

35 BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0027] Para comprender mejor la presente invención, una realización preferida, destinada solo a modo de ejemplo y no construida como limitativa, se describirá ahora en detalle, en referencia a los dibujos adjuntos, en los que:

- 40
- La figura 1 muestra un diagrama de bloque que representa la técnica CELP para la codificación de señal de voz;
 - La figura 2 muestra un diagrama de flujo del procedimiento según la presente invención;
 - Las figuras 3 y 4 muestran una señal de voz y las cantidades implicadas en el procedimiento de la presente invención;
 - La figura 5 muestra un diagrama de bloque de una transformación de tramos en vectores de código;
 - La figura 6 muestra otra señal de voz y las cantidades implicadas en el procedimiento de la presente invención;
 - 45 La figura 7 muestra un diagrama de bloque de una transformación de sub-tramos en parámetros de fuente;
 - La figura 8 muestra un diagrama de bloque de una fase de codificación; y la figura 9 muestra un diagrama de bloque de una fase decodificación.

50 DESCRIPCIÓN DETALLADA DE LAS REALIZACIONES PREFERIDAS DE LA INVENCION

[0028] La siguiente descripción se presenta para permitir a un experto en la técnica hacer y utilizar la invención. Varias modificaciones a las realizaciones serán aparentes a los expertos en la técnica, y los principios genéricos se puede aplicar aquí a otras realizaciones y aplicaciones sin apartarse del espíritu y el ámbito de la presente invención. Por tanto, la presente invención no está destinada a limitarse a las realizaciones mostradas, sino que concuerda con el ámbito más amplio compatible con los principios y características aquí descritos y definidos en las reivindicaciones adjuntas.

55 [0029] Además, la presente invención es implementada por medio de un producto de programa informático que incluye porciones de código de software para implementar, cuando el producto de programa informático se carga en la memoria del sistema de procesamiento y se ejecuta en el sistema de procesamiento, un procedimiento de codificación y decodificación, como se describe a continuación en referencia a las figuras 2 a 9.

60 [0030] Adicionalmente, un procedimiento se describirá ahora para representar y compactar un conjunto de datos, no necesariamente del mismo tipo (por ejemplo, la compresión con pérdidas de una señal de voz originada a partir de múltiples fuentes y/o una señal musical). El procedimiento tiene una aplicación ventajosa pero no exclusiva en los

datos que contiene información relacionada con las señales de voz digitales y/o musicales, donde las unidades de datos individuales se corresponden con una sola muestra digital.

[0031] En referencia al diagrama de flujo mostrado en la figura 2, el procedimiento según la presente invención proporciona ocho etapas de procesamiento de datos para conseguir la representación codificada y una etapa para reconstruir los datos iniciales, y en particular:

1. Clasificación y agrupamiento de datos en clases (bloque 1);
2. Selección de una primera ventana de análisis de datos, es decir, el número de elementos de datos consecutivos que se den considerar como una sola unidad de información, referida a continuación como tramo, en la siguiente etapa (bloque 2);
3. transformación, para cada clase identificada, de los tramos identificados en la etapa previa y pertenecientes a la clase bajo consideración, en parámetros de filtro (bloque 3);
4. Cálculo, para cada clase identificada, de un grupo de N parámetros que representa globalmente unidades de información de filtro de análisis pertenecientes a la clase bajo consideración, y almacenar los parámetros extraídos a continuación referidos como Libro de Códigos de Filtro (bloque 4);
5. Selección de una segunda ventana de análisis de datos, es decir, el número de elementos de datos consecutivos considerados como una sola unidad de información en la siguiente etapa (bloque 5);
6. Extracción, para cada clase identificada, de parámetros de fuente utilizando el Libro de Códigos de Filtro como modelo; esta descomposición difiere de la transformación de la etapa previa 3 en que depende del Libro de Códigos de Filtro, que no está presente en la etapa 3, y en la diferente definición de la ventana de análisis (bloque 6);
7. Cálculo, para cada clase identificada, de un grupo de N parámetros que representan globalmente los datos fuente pertenecientes a la clase bajo consideración, y almacenar los valores extraídos en un libro de códigos referidos a continuación como Libro de Códigos de Fuente (bloque 7);
8. Codificación de datos (bloque 8); y
9. Decodificación de datos (bloque 9).

[0032] A continuación se describirá más detalladamente cada etapa de procesamiento de datos.

1. Clasificación y agrupamiento de datos

[0033] En esta etapa, los datos disponibles se agrupan en clases para su análisis subsiguiente. Las clases que representan el contenido fonético de la señal se pueden identificar en la señal de voz. En general, se identifican los grupos de datos que cumplen con una métrica determinada. Una posible elección puede ser la subdivisión de los datos disponibles en clases fonéticas predefinidas. Una elección diferente puede ser la subdivisión de los datos disponibles en clases demifonéticas predefinidas. La estrategia elegida es una mezcla de estas dos estrategias. Esta etapa proporciona la subdivisión de los datos disponibles en fonemas si el número de elementos de datos pertenecientes a la clase está por debajo del límite determinado. Si en su lugar se sobrepasa el límite, se realiza una subdivisión sucesiva e subclases demifonéticas en las clases que sobrepasan el límite. Este procedimiento de subdivisión se puede repetir un número de veces en las subclases que tienen un número de elementos mayor que el límite, que puede variar en cada repetición y se puede definir para lograr una distribución uniforme de la cardinalidad de las clases. Para conseguir este objetivo, los demifonemas derecho e izquierdo, o en general fracciones de demifonemas, pueden por ejemplo identificarse y se puede realizar otra clasificación basándose en estas dos clases. La figura 3 muestra una señal de voz y la clasificación y el agrupamiento descritos anteriormente, en el que las clases identificadas se indican como C_i con $1 < i < N$, donde N es el número total de clases.

2. Selección de la primera ventana de análisis de datos.

[0034] En esta etapa, una ventana de análisis de muestra WF se define para su análisis subsiguiente. Para una señal de voz, se puede elegir una ventana que se corresponde con 10-30 milisegundos. Las muestras son segmentadas en tramos que contienen un número de muestras igual a la anchura de la ventana. Cada tramo pertenece a solo una clase. En casos en los que un tramo coincide con diferentes clases, se puede definir una distancia métrica y el tramo asignado a la clase más cercana. El criterio de selección para determinar la anchura óptima de la ventana de análisis depende del detalle de la representación de muestra deseado. Cuanto menor sea la anchura de la ventana de análisis, mayor será el detalle de la representación de la muestra y mayor la ocupación de memoria, y viceversa. La figura 4 muestra una señal de voz con la ventana de análisis de muestra WF, los tramos F_i , y las clases C_i , donde cada tramo pertenece a solo una clase.

3. Transformación de los tramos en vectores de parámetros de filtro

[0035] En esta etapa, la transformación de cada tramo en un vector de parámetro de filtro correspondiente, conocida generalmente como vector de código, se realiza a través de la aplicación de una transformación matemática T1. En el caso de una señal de voz, la transformación se aplica a cada tramo para extraer de la señal de voz contenida en el tramo un vector de código que modela la secuencia vocal y obtiene los LPCs o parámetros equivalentes. Un algoritmo para conseguir esta descomposición es el algoritmo Levinson-Durbin descrito en el arriba mencionado Wai C. Chu, "Speech Coding Algorithms", ISBN 0-471-37312-5, p. 107-114. En particular, en la etapa previa 2, cada

tramo se ha etiquetado como perteneciente a una clase. En particular, el resultado de la transformación de un solo tramo perteneciente a una clase es un conjunto de parámetros de filtro de síntesis que forman un vector de código F_{si} ($1 < i < N$), que pertenece a la misma clase que el tramo correspondiente. Para cada clase, un conjunto de vectores de código FS se genera aquí con los valores obtenidos aplicando la transformación a los tramos correspondientes F. El número de vectores de código FS generalmente no es el mismo en todas las clases, debido a que el número de tramos es diferente en cada clase. La transformación aplicada a las muestras en los tramos puede variar como una función de la clase a la que pertenecen, para maximizar la coincidencia del modelo creado con los datos reales, y como una función del contenido de información de cada tramo. La figura 5 muestra un diagrama de bloque que representa la transformación T1 de tramos F en vectores de código FS.

4. Generación de los libros de códigos de filtro

[0036] En esta etapa, para cada clase, se calcula un número X de vectores de código, referido a continuación como vectores de código centroides CF, lo que minimiza la distancia global entre ellos y los vectores de código FS en la clase bajo consideración. La definición de la distancia puede variar según la clase a la que los vectores de código FS pertenecen. Una posible distancia aplicable es la distancia euclidiana definida por espacios de vector de N dimensiones. Para obtener los vectores de código centroides, es posible aplicar, por ejemplo, un algoritmo conocido como algoritmo *kmeans* (véase "An Efficient k-Means Clustering Algorithm: Analysis and Implementation, IEEE transactions on pattern analysis and machine intelligence", vol. 24, n.º. 7 de Julio de 2002, pág. 881-892). Los vectores de código centroides extraídos CF forman el denominado libro de código de filtro para la clase correspondiente, y el número X de vectores de código centroides CF para cada clase se basa en el detalle de la representación de la muestra codificada. Cuanto mayor sea el número X de vectores de código para cada clase, mayor será el detalle de la representación de la muestra codificada y la ocupación de memoria o ancho de banda de transmisión requerido.

5. Selección de la segunda ventana de análisis de datos

[0037] En esta etapa, según un criterio predefinido, una ventana de análisis WS para la siguiente etapa se determina como un sub-múltiplo de la anchura de la ventana WF determinada en la etapa previa 2. El criterio para determinar de forma óptima la anchura de la ventana de análisis depende del detalle de representación de datos deseado. Cuanto menor sea la ventana de análisis, mayor será el detalle de la representación de los datos codificados y mayor la ocupación de memoria de los datos codificados, y viceversa. La ventana de análisis se aplica a cada tramo, generando de este modo n sub-tramos para cada tramo. El número n de sub-tramos depende del ratio entre las anchuras de las ventanas WS y WF. Una buena elección para la ventana WS puede ser a partir de un cuarto a un quinto de la anchura de la ventana WF. La figura 6 muestra una señal de voz junto con las ventanas de análisis de muestras WF y WS.

6. Extracción de los parámetros de fuente utilizando libros de códigos de filtro

[0038] En esta etapa, la transformación de cada sub-tramo en un vector de parámetro de fuente correspondiente Si se realiza a través de la aplicación de una transformación por filtrado T2 que es, en la práctica, una función de filtrado inversa basada en el libro de código de filtro calculado previamente. En el caso de una señal de voz, el filtrado inverso se aplica a cada sub-tramo para extraer de la señal de voz contenida en el sub-tramo, basándose en el libro de códigos de filtro CF, un conjunto de parámetros de fuente que modela la señal de excitación. Los vectores de parámetro de fuente así calculados se agrupan después en clases, de forma similar a la descrita previamente en referencia a los tramos. Para cada clase Ci, se genera un conjunto correspondiente de vectores de parámetros de fuente S. La figura 7 muestra un diagrama de bloque que representa la transformación T2 de los sub-tramos SBF en parámetros de fuente CF.

7. Generación de libros de códigos de fuente

[0039] En esta etapa, para cada clase C, se calcula un número Y de vectores de parámetro, referido a continuación como parámetros de fuente centroides CSi, lo que minimiza la distancia global entre ellos y los vectores de parámetros de fuente en la clase bajo consideración. La definición de la distancia puede variar según la clase a la que pertenecen los vectores de parámetro de fuente. Una posible distancia aplicable es la distancia euclidiana definida por espacios de vector de N dimensiones. Para obtener los parámetros de fuente centroides, es posible aplicar, por ejemplo, el algoritmo *kmeans* mencionado anteriormente. El parámetro de fuente cetroide extraído forma un libro de código de fuente para la clase correspondiente, y el número Y de parámetros de fuente centroides para cada clase se basa en el detalle de la representación de las muestras codificadas. Cuanto mayor sea el número Y de parámetros de fuente centroides para cada clase, mayor será el detalle de la representación y la ocupación de memoria o ancho de banda de transmisión. Al final de esta etapa, se generan de este modo un libro de código de filtros y un libro de código de fuente para cada clase, en el que los libros de códigos representan los datos obtenidos a partir del análisis mediante la ventana WF y la transformación asociada, y los libros de códigos de fuente representan los datos obtenidos a partir del análisis mediante la ventana WS y la transformación asociada (dependiente de los libros de códigos de filtro).

8. Codificación

[0040] La codificación se lleva a cabo aplicándole procedimiento CELP arriba mencionado, con la diferencia que cada tramo está asociado a un vector de índices que especifica los vectores de parámetro de filtro centroides y los vectores de parámetro de fuente centroides que representan las muestras contenidas en el tramo y el sub-tramos respectivos para ser codificados. La selección se realiza aplicando una distancia métrica previamente identificada y escogiendo los vectores de parámetro de filtro centroides y los vectores de parámetro de fuente centroides que minimizan la distancia entre las señales de voz originales y las señales de voz reconstruidas o la distancia entre la señal de voz original ponderada con una función que modela la curva de percepción auditiva y la señal de voz reconstruida ponderada con la misma curva de percepción auditiva. El filtro de los libros de códigos de fuente CF y CS se almacenan para que puedan utilizarse en la fase de decodificación. La figura 8 muestra un diagrama de bloque de la fase de decodificación, en el que 10 designa el tramo que se va a codificar, que pertenece a la clase i-th, 11 designa el libro de código de filtro i-th CFi, es decir, el libro de código de filtro asociado con la clase i-th a la que pertenece el tramo, 12 designa el codificador, 13 designa el libro de código de fuente i-th CSi, es decir, el libro de código de fuente asociado a la clase i-t a la que pertenece el tramo, 14 designa el índice del mejor vector de código de filtro del libro de códigos de filtro CFi, y 15 designa los índices de los mejores vectores de código de fuente del libro de código de fuente i-th CSi.

9. Decodificación

[0041] En esta etapa, la reconstrucción de los tramos se lleva a cabo aplicando la transformación inversa aplicada durante la fase de codificación. Para cada tramo y para cada sub-tramo correspondiente, los índices del vector de código de filtro y de los vectores de código de fuente pertenecientes a los libros de códigos de fuente CF y CS que codifican los tramos y los sub-tramos se lee y se reconstruye una versión aproximada de los tramos, aplicando la transformación inversa. La figura 9 muestra un diagrama de bloque de la fase de decodificación, en el que 20 designa el tramo decodificado, que pertenece a la clase i-th, 21 designa el libro de código de filtro i-th CFi, es decir, el libro de código de filtro asociado con la clase i-th a la que pertenece el tramo, 22 designa el decodificador, 23 designa el libro de código de fuente i-th CSi, es decir, el libro de código de fuente asociado a la clase i-th a la que pertenece el tramo, 24 designa el índice del mejor vector de código de filtro del libro de códigos de filtro CFi, y 25 designa los índices de los mejores vectores de código de fuente del libro de código de fuente i-th CSi.

[0042] Las ventajas de la presente invención resultan evidentes en la descripción anterior, En particular, la elección de los vectores de código, la cardinalidad de único libro de códigos y el número de libros de códigos basados en la señal de fuente, así como la elección de técnicas de codificación dependientes del conocimiento del contenido informativo de la señal de fuente permite una mejor calidad que se logra para la señal reconstruida para la misma ocupación de memoria/ancho de banda de transmisión mediante la señal codificada, o una calidad de señal reconstruida lograda que es equivalente a la de los procedimientos de codificación que requieren una mayor ocupación de memoria/ancho de banda de transmisión.

[0043] Finalmente, resulta evidente que varias modificaciones y variantes se pueden hacer en la presente invención, todas dentro del ámbito de la invención, como se define en las reivindicaciones adjuntas.

[0044] En particular, se puede apreciar que la presente invención también se puede aplicar a la codificación de señales distintas a las utilizadas para la generación de los libros de códigos de filtro y fuente CF y CS. En ese aspecto, es necesario modificar la etapa 8 porque la clase a la que pertenece el tramo bajo construcción no se conoce a priori. La modificación proporciona por tanto la ejecución de un ciclo de mediciones para el mejor vector de código utilizando todos los libros de códigos precalculados, determinando de este modo la clase a la que pertenece el tramo destinado a codificarse: la clase a la que pertenece es la que contiene el vector de código con la distancia más corta. En esta aplicación, también se puede explotar un sistema de Reconocimiento de Voz automático (ASR, del inglés, *Automatic Speech Recognition*) para soportar la elección del libro de código, en el sentido que el ASR se utiliza para proporcionar el fonema, y después solo las clases asociadas con el fonema específico se tienen en cuenta.

[0045] Adicionalmente, la velocidad de codificación no tiene que ser necesariamente el mismo para toda la señal de voz que se va a codificar, sino que en general diferentes distorsiones de la señal de voz se pueden codificar con diferentes velocidades. Por ejemplo, las distorsiones de la señal de voz utilizadas más frecuentemente en aplicaciones de texto a voz se podrían codificar con una velocidad mayor, es decir utilizando libros de códigos de filtro y/o fuente con mayor cardinalidad, mientras que las distorsiones de la señal de voz utilizadas menos frecuentemente se podrían codificar con una velocidad menor, es decir, utilizando libros de códigos de filtro y/o fuente con menor cardinalidad, para obtener una mejor calidad en la reconstrucción de la voz en las distorsiones de la señal de voz utilizadas más frecuentemente, aumentando así la calidad total percibida.

[0046] Adicionalmente, la presente invención también se puede utilizar en escenarios concretos como aplicaciones de texto a voz remotas y/o distribuidas, y aplicaciones de voz sobre IP.

[0047] En particular, la voz se sintetiza en un servidor, comprimido utilizando el procedimiento descrito, transmitido remotamente mediante un canal de protocolo de Internet (IP), por ejemplo, GPRS, a un dispositivo móvil como un teléfono agenda personal (PDA), donde la voz sintetizada primero se descomprime y después se reproduce. En particular, una base de datos de voz, en una porción considerable de señal de voz, se procesa previamente no a tiempo real para crear libros de códigos, la cadena fonética del texto que se va a sintetizar se genera a tiempo real durante el proceso de síntesis, p. ej., mediante un proceso automático de reconocimiento de voz, la señal que se va a sintetizar se genera a tiempo real a partir de la base de datos no comprimida, después se codifica a tiempo real en

el servidor, basándose en los libros de códigos creados, transmitidos al dispositivo móvil en forma codificada mediante el canal IP, y finalmente la señal codificada se decodifica a tiempo real en el dispositivo a tiempo real en el dispositivo móvil y la señal de voz finalmente se reconstruye.

REIVINDICACIONES

1. Procedimiento para codificar datos de audio, comprendiendo:

- 5
- agrupar dichos datos de audio en tramos, conteniendo cada tramo un número de muestras igual a la anchura de una ventana de análisis correspondiente;
 - clasificar los tramos en clases; • para cada clase, transformar los tramos pertenecientes a la clase en vectores de parámetro de filtro;
 - para cada clase, calcular un libro de códigos de filtro (CF) basándose en los vectores de parámetro de filtro pertenecientes a la clase;
- 10
- segmentar cada tramo en sub-tramos, definiendo una segunda ventana de análisis de muestra como un sub-múltiplo de la anchura de la primera ventana de análisis de muestra; y segmentar cada tramo en un número de sub-tramos correlacionados con el ratio entre las anchuras de la primera y la segunda ventana de análisis de muestra;
- para cada clase, transformar los sub-tramos pertenecientes a la clase en vectores de parámetros de fuente, que se extraen de los sub-tramos aplicando una transformación por filtrado (T2) basada en el libro de códigos de filtros (CF) calculado para la clase correspondiente;
 - para cada clase, calcular un libro de códigos de fuente (CS) basándose en los vectores de parámetros de fuente pertenecientes a la clase; y
 - codificar los datos basándose en los libros de códigos de filtros (CF) y de fuente (CS) calculados.

2. Procedimiento según la reivindicación 1, en el que los datos son muestras de una señal de voz, y en el que las clases son clases fonéticas.

3. Procedimiento según la reivindicación 1, en el que dicha transformación por filtrado (T2) es una función de filtrado inversa basada en el libro de código de filtro calculado previamente.

4. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que la clasificación de los tramos en clases incluye clasificar cada tramo en solo una clase y, si un tramo coincide con diversas clases, se clasifica el tramo en la clase más cercana según una distancia métrica determinada.

5. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que la etapa de calcular un libro de códigos de filtro para cada clase basándose en los vectores de parámetro de filtro pertenecientes a la clase incluye:

- calcular vectores de parámetros de filtro específicos que minimizan la distancia global entre ellos y los vectores de parámetros de filtro en la clase, y basándose en una distancia métrica dada; y
- calcular el libro de códigos de filtro basándose en los vectores de parámetro de filtro específicos.

6. Procedimiento según la reivindicación 5, en el que la distancia métrica depende de la clase a la que cada vector de parámetro de filtro pertenece.

7. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que el cálculo de un libro de códigos de fuente para cada clase basándose en los vectores de parámetro de filtro pertenecientes a la clase incluye:

- calcular vectores de parámetros de fuente específicos que minimizan la distancia global entre ellos y los vectores de parámetros de fuente en la clase, y basándose en una distancia métrica dada; y
- calcular el libro de códigos de fuente basándose en los vectores de parámetro de fuente específicos.

8. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que la codificación de los datos basándose en los libros de códigos de filtros y de fuente calculados incluye:

- asociar a cada tramo los índices que identifican un vector de parámetro de filtro en el libro de códigos de filtro y los vectores de parámetros de fuente en el libro de códigos de fuente que representan las muestras en el tramo y respectivamente en los sub-tramos respectivos.

9. Un codificador configurado para implementar el procedimiento de codificación según cualquiera de las reivindicaciones anteriores.

10. Codificador según la reivindicación 9, en el que las distorsiones de la señal de voz utilizadas más frecuentemente son codificadas utilizando libros de códigos de filtro y/o fuente con mayor cardinalidad, mientras que las distorsiones de la señal de voz utilizadas con menos frecuencia son codificadas utilizando los libros de códigos de filtro y/o fuente con menor cardinalidad.

11. Codificador según la reivindicación 9, en el que una primera porción de la señal de voz se preprocesa para crear dichos libros de códigos de filtro y de fuente, siendo los mismos libros de códigos de filtro y de fuente utilizados en la codificación a tiempo real de la señal de voz y teniendo parámetros acústicos y fonéticos homogéneos con dicha primera porción.

12. Codificador según la reivindicación 11, en el que dicha señal de voz que se va a codificar está sujeta al reconocimiento de voz automático a tiempo real con el fin de obtener una cadena fonética correspondiente necesaria para la codificación.
- 5 13. Producto de software que se puede cargar en la memoria de un sistema de procesamiento de un codificador y que comprende porciones de código de software para implementar, cuando el programa se ejecuta en el sistema de procesamiento del codificador, el procedimiento de codificación según cualquiera de las reivindicaciones 1 a 8.
- 10 14. Procedimiento para decodificar datos codificados según el procedimiento de codificación de cualquiera de las reivindicaciones anteriores, incluyendo:
- identificar la clase de un tramo que se va a reconstruir basándose en los índices que identifican el vector de parámetro de filtro en el libro de códigos de filtro (CF) y los vectores de parámetro de fuente en el libro de códigos de fuente (CS) que representan las muestras en el tramo y respectivamente en los sub-tramos respectivos;
 - 15 • identificar los libros de códigos de filtro y de fuente, calculados previamente y almacenados durante el proceso de codificación, asociados con la clase identificada;
 - identificar el vector de parámetro de filtro en el libro de códigos de filtro y los vectores de parámetro de fuente en el libros de códigos de fuente identificados por los índices; y
 - 20 • reconstruir el tramo basándose en el vector de parámetro de filtro identificado en el libro de códigos de filtro y los vectores de parámetro de fuente en el libros de códigos de fuente.
15. Descodificador configurado para implementar el procedimiento de codificación según la reivindicación 14.
- 25 16. Producto de software que se puede cargar en la memoria de un sistema de procesamiento de un descodificador y que comprende porciones de código de software para implementar, cuando el programa se ejecuta en el sistema de procesamiento del descodificador, el procedimiento de descodificación según la reivindicación 14.

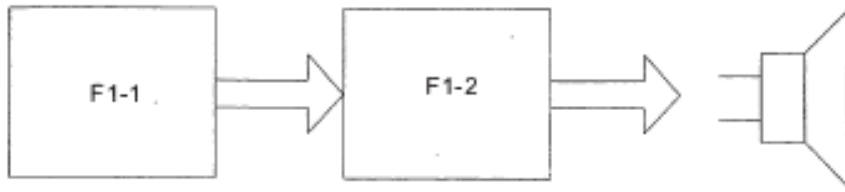


Fig. 1

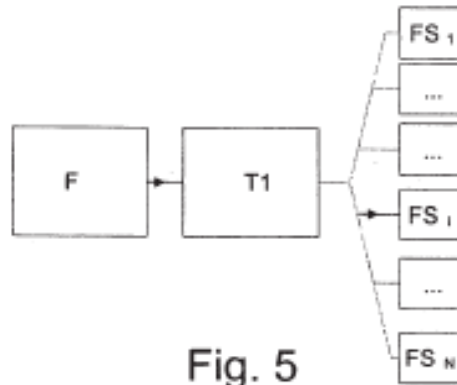


Fig. 5

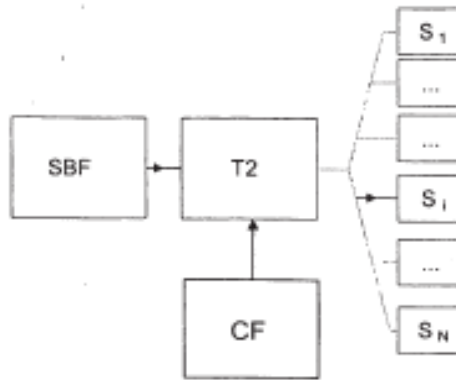


Fig. 7

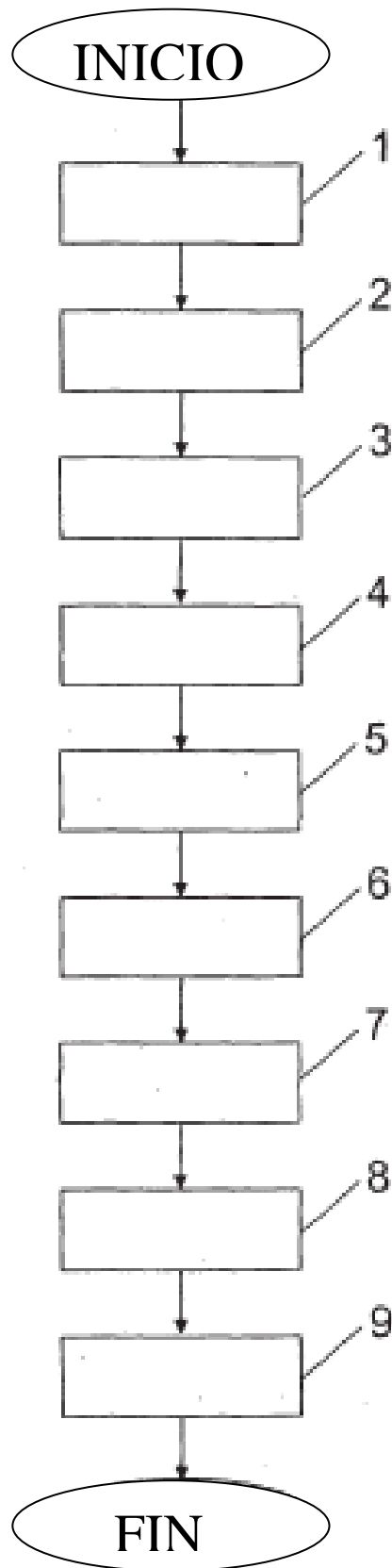


Fig. 2

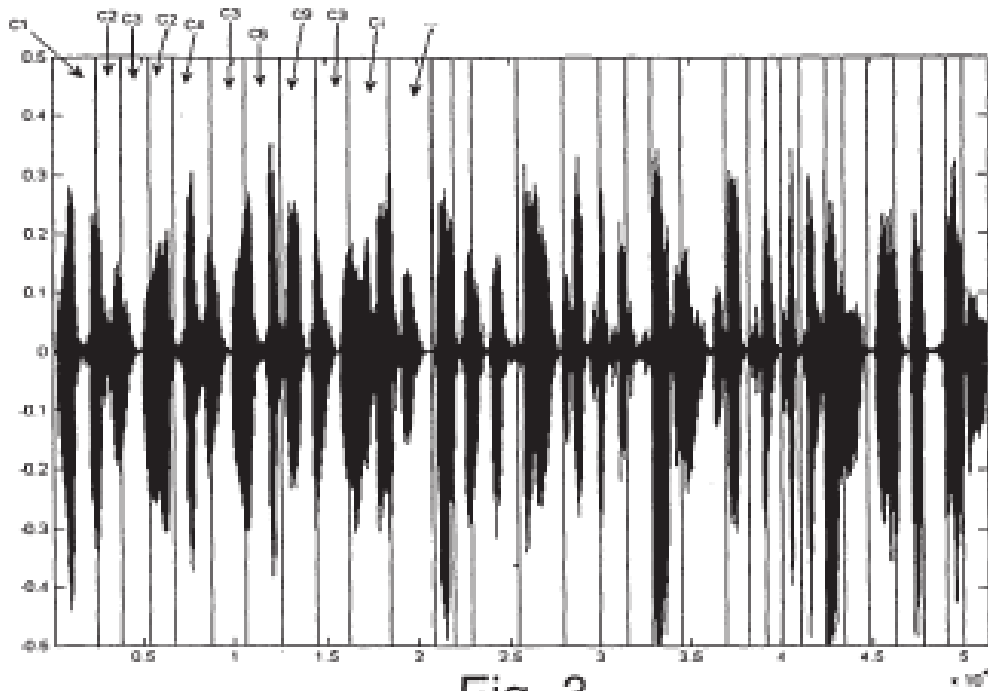


Fig. 3

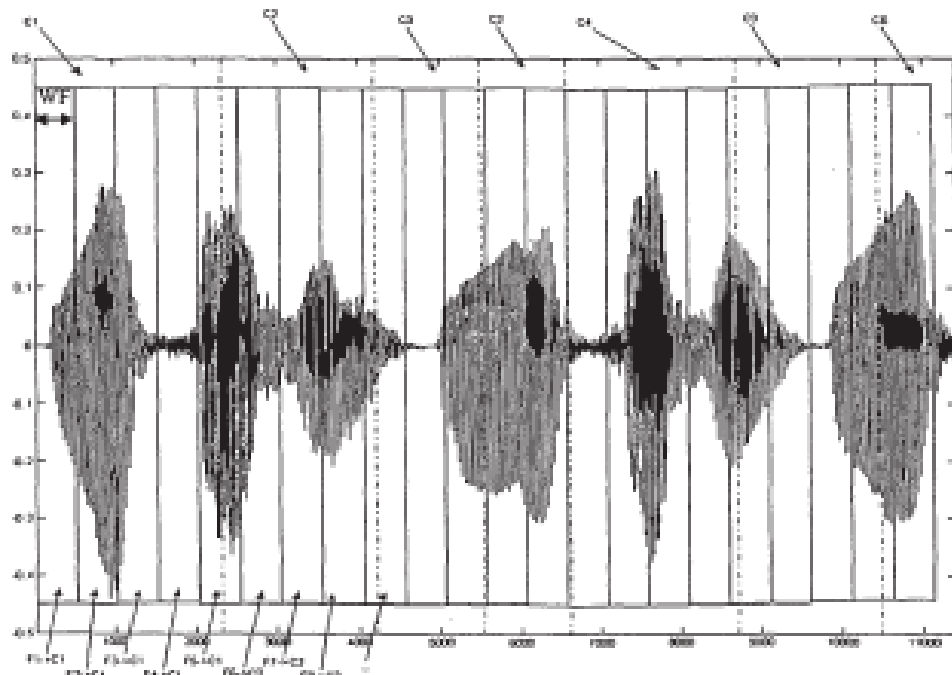


Fig. 4

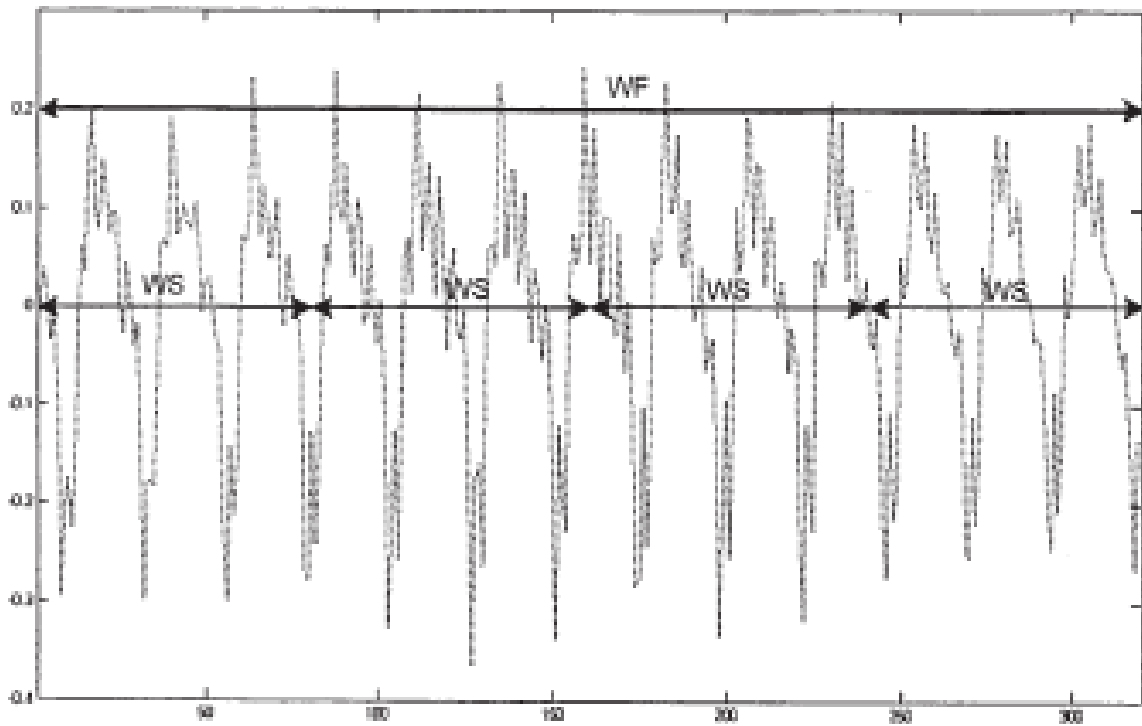


Fig. 6

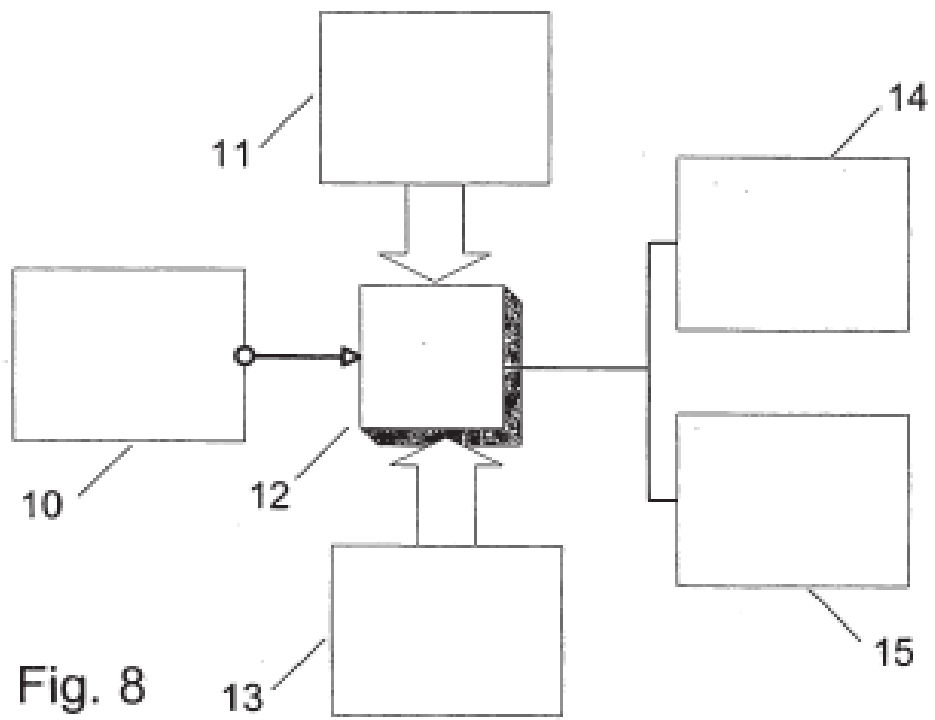


Fig. 8

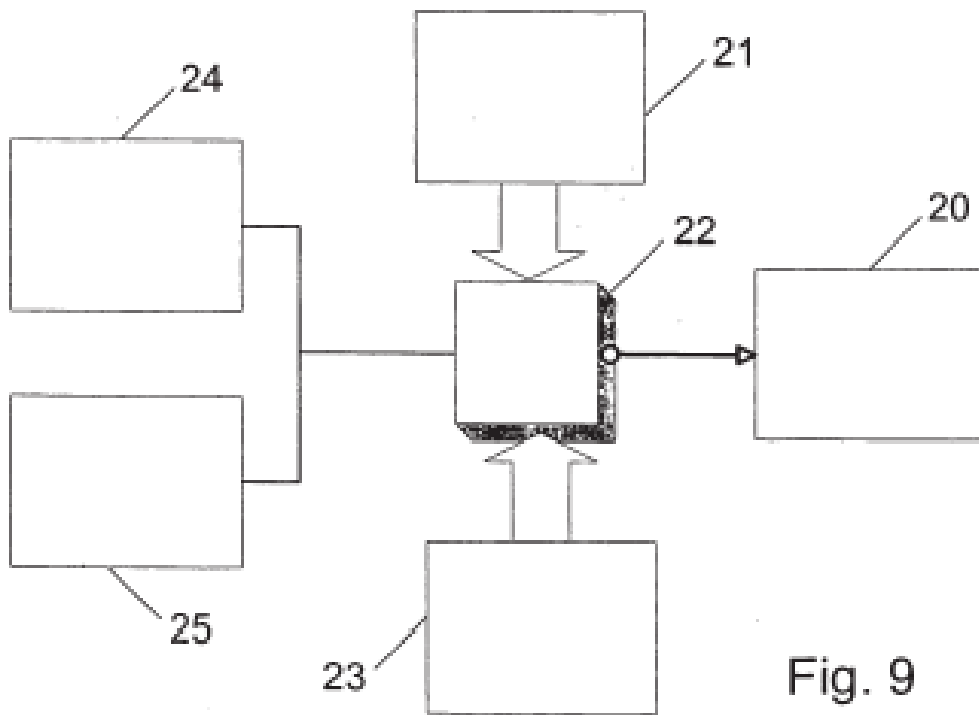


Fig. 9