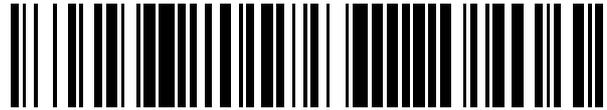


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 527 697**

51 Int. Cl.:

G10L 19/005 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.05.2001 E 08168570 (3)**

97 Fecha y número de publicación de la concesión europea: **29.10.2014 EP 2017829**

54 Título: **Corrección directa de errores en la codificación vocal**

30 Prioridad:

11.05.2000 US 569312

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

28.01.2015

73 Titular/es:

**TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)
(100.0%)
164 83 Stockholm, SE**

72 Inventor/es:

**WESTERLUND, MAGNUS;
NOHLGREN, ANDERS;
UVLIDEN, ANDERS;
SUNDQVIST, JIM y
SVEDBERG, JONAS**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 527 697 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

Corrección directa de errores en la codificación vocal

Antecedentes

5 La presente invención se refiere a un sistema y a un método para realizar una corrección directa de errores en la transmisión de información de audio y, más particularmente, a un sistema y un método para realizar una corrección directa de errores en la transmisión basada en paquetes de información vocalmente codificada.

1. Codificación vocal

Las deficiencias de los métodos de corrección directa de errores (FEC) del estado de la técnica pueden apreciarse del mejor modo por una discusión introductoria de algunos conceptos de codificación vocal convencionales.

10 1.1 Codificación Predictiva Lineal Excitada por Código (CELP)

15 La figura 1 muestra un codificador 100 de análisis por síntesis predictivo lineal excitado por código (CELP) convencional. El codificador 100 incluye unidades funcionales designadas como módulo de tramado 104, módulo de análisis 106 de codificación de predicción lineal (LPC), módulo de cálculo de diferencia 118, módulo de ponderación de error 114, módulo de minimización de error 116 y módulo descodificador 102. El módulo descodificador 102 incluye a su vez un libro de código fijo 112, un filtro 110 predictor a largo plazo (LTP) y un filtro 108 de codificación predictora lineal (LPC) conectados uno a otro en relación de cascada para producir una señal sintetizada (n). El filtro LPC 108 modela la correlación a corto plazo en la voz atribuida a los tractos vocales, correspondiente a la envolvente espectral de la señal vocal. Debe representarse por:

$$1/A(z) = 1 / (1 - \sum_{i=1}^p a_i z^{-i}) \tag{Ec. 1},$$

20 en donde p denota el orden del filtro y a_i denota los coeficientes del filtro. El filtro LTP 110, por otro lado, modela la correlación a largo plazo de la voz atribuida a las cuerdas vocales, correspondiente a la estructura espectral fina similar a periódica de la señal vocal. Por ejemplo, puede tener la forma dada por:

$$1/P(z) = 1 / (1 - \sum_{i=-1}^1 b_i z^{-(D+i)}) \tag{Ec. 2},$$

25 en donde D corresponde generalmente al período de tono de la correlación a largo plazo y b_i pertenece a los coeficientes de ganancia a largo plazo del filtro. El libro de código fijo 112 almacena una serie de secuencias de entrada de excitación. Las secuencias proporcionan señales de excitación al filtro LTP 110 y al filtro LPC 108 y son útiles en la modelación de las características de la señal vocal que no pueden predecirse hasta cierto grado con métodos determinísticos utilizando el filtro LTP 110 y el filtro LPC 108, tal como componentes de audio dentro de música.

30 En funcionamiento, el módulo de tramado 104 recibe una señal vocal de entrada y la divide en tramas sucesivas (por ejemplo, de 20 ms de duración). A continuación, el módulo de análisis LPC 106 recibe y analiza una trama para generar un conjunto de coeficientes LPC. Estos coeficientes son utilizados por el filtro LPC 108 para modelar las características a corto plazo de la señal vocal correspondientes a su envolvente espectral. Puede formarse a continuación un residual LPC alimentando la señal vocal de entrada a través de un filtro inverso, incluyendo los coeficientes LPC calculados. Este residual, mostrado en la figura 2, representa un componente de la señal vocal original que permanece después de la retirada de la redundancia a corto plazo por análisis predictivo lineal. La distancia entre dos pulsos de tono se denota con "L" y se denomina retraso. El codificador 100 puede utilizar a continuación el residual para predecir los coeficientes a largo plazo. Estos coeficientes a largo plazo son utilizados por el filtro LTP 110 para modelar la estructura espectral fina de la señal vocal (tal como retardo de tono y ganancia de tono). Tomados conjuntamente, el filtro LTP 110 y el filtro LPC 108 forman un filtro en cascada que modela las características a largo plazo y a corto plazo de la señal vocal. Cuando se acciona por una secuencia de excitación procedente del libro de código fijo 112, el filtro en cascada genera la señal vocal sintética (n) que representa una versión reconstruida de la señal vocal original s(n).

45 El codificador 100 selecciona una secuencia de excitación óptima generando sucesivamente una serie de señales vocales sintéticas (n), comparando sucesivamente las señales vocales sintéticas (n) con las señales vocales originales s(n) y ajustando sucesivamente los parámetros operativos del módulo descodificador 102 para minimizar la diferencia entre (n) y s(n). Más específicamente, el módulo de cálculo de diferencia 118 forma la diferencia (es decir, la señal de error e(n)) entre la señal vocal original s(n) y la señal vocal sintética (n). Un módulo de ponderación de error 114 recibe la señal de error e(n) y genera una señal de error ponderada e_w(n) sobre la base de factores de ponderación perceptuales. El módulo de minimización de error 116 utiliza un procedimiento de búsqueda para

ajustar los parámetros operativos del descodificador de voz 102 de tal manera que produzca una señal sintetizada (n) que está lo más próxima posible a la señal original s(n).

Tras llegar a una señal sintetizada óptima (n), se transfieren parámetros de codificador relevantes sobre un medio de transmisión (no mostrado) a un sitio de descodificador (no mostrado). Un descodificador en el sitio de descodificador incluye una construcción idéntica al módulo descodificador 102 del codificador 100. El descodificador utiliza los parámetros transferidos para reproducir la señal sintetizada optimizada (n) calculada en el codificador 100. Por ejemplo, el codificador 100 puede transferir índices de libro de código que representan la localización de la señal de excitación óptima en el libro de código fijo 112, junto con parámetros o coeficientes de filtro relevantes (por ejemplo, los parámetros LPC y LTP). La transferencia de los parámetros en vez de una representación más directa de la señal vocal de entrada proporciona una reducción notable en el ancho de banda requerido para transmitir información vocal.

La figura 3 muestra una modificación del codificador 100 de análisis por síntesis mostrado en la figura 1. El codificador 300 mostrado en la figura 3 incluye un módulo de tramado 304, un módulo de análisis LPC 306, un filtro LPC 308, un módulo de cálculo de diferencia 318, un módulo de ponderación de error 314, un módulo de minimización de error 316 y un libro de código fijo 312. Cada una de estas unidades corresponde generalmente a las partes con el mismo nombre mostradas en la figura 1. Sin embargo, en la figura 3, el filtro LTP 110 es sustituido por el libro de código 320 adaptativo. Además, un módulo sumador 322 añade las señales de excitación emitidas desde el libro de código adaptativo 320 y el libro de código fijo 312.

El codificador 300 funciona básicamente de la misma manera que el codificador 100 de la figura 1. Sin embargo, en el codificador 300, el libro de código adaptativo 320 modela las características a largo plazo de la señal vocal. Además, la señal de excitación aplicada al filtro LPC 308 representa una suma total de una entrada del libro de código adaptativo 320 y una entrada del libro de código fijo 312.

1.2 Codificación de Tasa Completa Reforzada GSM (GSM-EFR)

La técnica anterior proporciona numerosas implementaciones específicas del diseño CELP antes descrito. Una de tales implementaciones es el estándar de transcodificación vocal de Tasa Completa Reforzada GSM (GSM-EFR) descrito en el documento "Sistema Global para Comunicaciones Móviles: Sistemas de Telecomunicaciones Celulares Digitales: Transcodificación Vocal de Tasa Completa Reforzada (EFR) (GSM 06.60)", noviembre de 1996, del Instituto Europeo de Normas de Telecomunicaciones (ETSI).

El estándar GSM-EFR modela las propiedades a corto plazo de la señal vocal utilizando:

$$H(z) = 1/\hat{A}(z) = 1 / (1 + \sum_{i=1}^m \hat{a}_i z^{-i}) \quad (\text{Ec. 3}),$$

donde \hat{a}_i representa los parámetros de predicción lineal cuantificada. El estándar modela las características a largo plazo de la señal vocal con:

$$1/B(z) = 1 / (1 - g_p z^{-T}) \quad (\text{Ec. 4}),$$

donde T pertenece al retardo de tono y g_p pertenece a la ganancia de tono. Un libro de código adaptativo implementa la síntesis de tono. Además, el estándar GSM-EFR utiliza un filtro de ponderación perceptual definido por:

$$W(z) = (A(z/1)) / (A(z/2)) \quad (\text{Ec. 5}),$$

donde A(z) define el filtro LPC no cuantificado y 1 y 2 representan factores de ponderación perceptuales. Finalmente, el estándar GSM-EFR utiliza libros de código adaptativos y fijos (innovadores) para proporcionar una señal de excitación. En particular, el libro de código forma un libro de código algebraico estructurado sobre la base de un diseño de permutación de impulso único intercalado (ISPP). Los vectores de excitación constan de un número fijo de impulsos matemáticamente calculados diferentes de cero. Se especifica una excitación por posiciones y signos de impulso seleccionados dentro del libro de código.

En funcionamiento, el codificador GSM-EFR divide la señal vocal de entrada en tramas de 20 ms que, a su vez, son divididas en cuatro subtramas de 5 ms. El codificador realiza entonces el análisis LPC dos veces por trama. Más específicamente, el codificador GSM-EFR utiliza un enfoque de autocorrelación con ventanas asimétricas de 30 ms a fin de calcular los parámetros a corto plazo. No se emplea ninguna anticipación en el análisis LPC. La anticipación se refiere al uso de muestras derivadas de una trama futura en la realización del análisis.

Cada coeficiente LP se convierte entonces en una representación de Par Espectral Lineal (LSP) para la cuantificación y la interpolación utilizando un predictor LSP. El análisis LSP mapea los coeficientes de filtro sobre un círculo unidad en el rango de - a para producir valores de Frecuencia Espectral de Línea (LSF). El uso de valores LSF proporciona mejor robustez y estabilidad contra errores de bits en comparación con el uso de valores LPC. Además, el uso de valores LSF permite una cuantificación más eficiente de la información en comparación con el uso de valores LPC. El GSM-EFR utiliza específicamente la siguiente ecuación predictora para calcular un residual que se cuantifica a continuación:

$$LSF_{res} = LSF - LSF_{media} - predFactor \cdot LSF_{prev, res} \quad (\text{Ec. 6}).$$

El término LSF_{res} se refiere a un vector residual LSF para una trama n. La cantidad $(LSF - LSF_{media})$ define un vector LSF de media eliminada en la trama n. El término $(predFactor \cdot LSF_{prev, res})$ se refiere a un vector LSF predicho en la trama n, en donde predFactor se refiere a una constante de factor de predicción y $LSF_{prev, res}$ se refiere a un segundo vector residual de la trama pasada (es decir, la trama n-1). El descodificador utiliza el proceso inverso, de acuerdo con la Ec. 7 siguiente:

$$LSF = LSF_{res} + LSF_{media} + predFactor \cdot LSF_{prev, res} \quad (\text{Ec. 7}).$$

Para conseguir el resultado predicho, la $LSF_{prev, res}$ residual previa en el descodificador debe tener el valor correcto. Después de la reconstrucción, los coeficientes se convierten en forma de filtro directo y se utilizan cuando se sintetiza la voz.

El codificador ejecuta entonces el denominado análisis de tono de bucle abierto para estimar el retraso de tono en cada mitad de la trama (cada 10 ms) sobre la base de la señal vocal perceptualmente ponderada. A continuación, el codificador realiza una pluralidad de operaciones en cada subtrama. Más específicamente, el codificador calcula una señal diana $x(n)$ restando la respuesta de entrada cero del filtro de síntesis ponderada $W(z)H(z)$ de la señal vocal ponderada. A continuación, el codificador calcula una respuesta de impulso $h(n)$ del filtro de síntesis ponderada. El codificador utiliza la respuesta de impulso $h(n)$ para realizar el denominado análisis de bucle cerrado a fin de encontrar el retraso y la ganancia de tono. El análisis de búsqueda de bucle cerrado implica minimizar el error ponderado cuadrático medio entre la voz original y la sintetizada. La búsqueda de bucle cerrado utiliza el cálculo de retraso de bucle abierto como una estimación inicial. A continuación, el codificador actualiza la señal diana $x(n)$ eliminando la contribución del libro de código adaptativo, y el codificador utiliza la diana resultante para encontrar un vector de innovación óptimo dentro del libro de código algebraico. Los parámetros relevantes de los libros de código se cuantifican a continuación de forma escalar utilizando un predictor de libro de código y las memorias de filtro se actualizan utilizando la señal de excitación determinada para encontrar la señal diana en la siguiente subtrama.

El codificador transmite dos conjuntos de coeficientes LSP (que comprenden 38 bits), los parámetros de retardo de tono (que comprenden 30 bits), los parámetros de ganancia de tono (que comprenden 16 bits), los parámetros de código algebraico (que comprenden 140 bits) y los parámetros de ganancia de libro de código (que comprenden 20 bits). El descodificador recibe estos parámetros y reconstruye la voz sintetizada duplicando las condiciones del codificador representadas por los parámetros transmitidos.

1.3 Ocultación de error (EC) en la codificación GSM-EFR

El Instituto Europeo de Normas de Telecomunicaciones (ETSI) propone una ocultación de error para uso en GSM-EFR en el documento "Sistema de Telecomunicaciones Celulares Digitales: Sustitución y Silenciamiento de Tramas Perdidas para Canales de Tráfico Vocal de Tasa Completa Reforzada (EFR) (GSM 06.61)", versión 5.1.2, abril de 1997. La norma referenciada propone un ejemplo de máquina de estado que tiene siete estados, 0 a 6. Una bandera de Indicación de Trama Mala (BFI) indica si la trama vocal actual contiene un error (estado = 0 para ningún error y estado = 1 para errores). Una bandera de Indicación de Trama Mala Previa (PrevBFI) indica si la trama vocal previa contiene errores (estado = 0 para ningún error y estado = 1 para errores). El estado 0 corresponde a un estado en el que las tramas actual y pasada no contienen errores (es decir, BFI = 0, PrevBFI = 0). La máquina avanza al estado 1 cuando se detecta un error en la trama actual. (El error puede detectarse utilizando una verificación de redundancia cíclica de 8 bits en la trama). La máquina de estado avanza sucesivamente a estados más altos (hasta el estado máximo de 6) tras la detección de errores adicionales en tramas posteriores. Cuando se detecta una trama buena (es decir, libre de errores), la máquina de estado vuelve al estado 0, a menos que la máquina de estado esté actualmente en el estado 6, en cuyo caso vuelve al estado 5.

El descodificador realiza diferentes operaciones de ocultación de error que dependen del estado y los valores de las banderas BFI y PrevBFI. La condición BFI = 0 y PrevBFI = 0 (dentro del estado 0) pertenece a la recepción de dos tramas libres de error consecutivas. En esta condición, el descodificador procesa los parámetros vocales de la manera típica expuesta en el estándar GSM-EFR 6.60. El descodificador guarda entonces la trama actual de los parámetros vocales.

La condición BFI = 0 y PrevBFI = 1 (dentro de estados 0 o 5) pertenece a la recepción de una trama libre de errores después de recibir una trama "mala". En esta condición, el descodificador limita la ganancia LTP y la ganancia del

libro de código fijo a los valores utilizados para la última subtrama buena recibida. En otras palabras, si el valor de la ganancia LTP actual (g^p) es igual o menor que la última ganancia LTP buena recibida, se utiliza entonces la ganancia LTP actual. Sin embargo, si el valor de la ganancia LTP actual es mayor que la última ganancia LTP buena recibida, entonces el valor de la última ganancia LTP se utiliza en lugar de la ganancia LTP actual. El valor para la ganancia del libro de código fijo se ajusta de una manera similar.

La condición $BFI = 1$ (dentro de cualquier estado 1 a 6, y $PrevBFI = 0$ o 1) indica que se ha detectado un error en la trama actual. En esta condición, la ganancia LTP actual es sustituida por la siguiente ganancia:

$$g^p = \text{estado}(n) \cdot g^p(-1) \quad \text{si } g^p(-1) \leq \text{mediana}, \text{ si no} \quad (\text{Ec. 8}),$$

$$g^p = \text{estado}(n) \cdot \text{mediana} \quad \text{si } g^p(-1) > \text{mediana}$$

donde g^p designa la ganancia del filtro LTP, $\text{estado}(n)$ designa un coeficiente de atenuación que tiene un efecto de atenuación sucesivamente mayor con incremento en el estado n (por ejemplo, $\text{estado}(1) = 0,98$, mientras que $\text{estado}(6) = 0,20$), "mediana" designa la mediana de los valores g^p para las cinco últimas subtramas, y $g^p(-1)$ designa la subtrama previa. El valor para la ganancia del libro de código se ajusta de una manera similar.

En el estado anteriormente descrito (es decir, cuando $BFI = 1$), el descodificador actualiza también la ganancia del libro de código en la memoria utilizando el valor medio de los últimos cuatro valores en la memoria. Además, el descodificador desplaza las LSFs pasadas hacia su media, es decir:

$$LSF_q1(i) = LSF_q2(i) = \text{pasada_LSF_q}(i) + (1 - \alpha) \cdot \text{media_LSF}(i) \quad (\text{Ec. 9}),$$

donde $LSF_q1(i)$ y $LSF_q2(i)$ son dos vectores de la trama actual, α es una constante (por ejemplo, 0,95), $\text{pasada_LSF_q}(i)$ es el valor de LSF_q2 de la trama previa, y $\text{media_LSF}(i)$ es el valor LSF promedio. Además, el descodificador sustituye los valores de retraso LTP por el valor de retraso pasado a partir de la cuarta subtrama. Y, finalmente, los impulsos de excitación del libro de código fijo recibidos por el descodificador se utilizan como tales a partir de la trama errónea.

1.4 Vocodificadores

La figura 4 muestra otro tipo de descodificador vocal, el vocodificador basado en LPC 400. En este descodificador, se crea el residual LPC a partir del vector de ruido 404 (para sonidos no vocalizados) o una forma de impulso estático 406 (para un mensaje vocalizado). Un módulo de ganancia 406 escala el residual hasta un nivel deseado. La salida del módulo de ganancia se suministra a un bloque de filtro LPC que incluye un filtro LPC 408 que tiene un ejemplo de función definido por:

$$A(z) = \sum_{i=1}^n a_i z^{-i} \quad (\text{Ec. 10}),$$

donde a_i designa los coeficientes del filtro que pueden ser computados minimizando la media cuadrática del error de predicción. Un vocodificador conocido se ha designado como "LPC-10". Este descodificador se desarrolló para el ejército norteamericano a fin de proporcionar comunicación de tasa de bits baja. El vocodificador LPC-10 utiliza tramas de 22,5 ms, correspondientes a 54 bits/trama igual y 2,4 kbits/s.

En funcionamiento, el codificador LPC-10 (no mostrado) toma una decisión de vocalización para utilizar el tren de impulsos o la señal de ruido. En el codificador LPC-10, esto puede realizarse formando una versión filtrada en paso bajo de la señal de entrada muestreada. La decisión se basa en la energía de la señal, la relación de máximo a mínimo de la señal y la pluralidad de cruces por cero de la señal. Se toman decisiones de vocalización para cada mitad de la trama actual, y la decisión de vocalización final se basa en estas dos decisiones de semitrama y las decisiones derivadas de las dos tramas siguientes.

El tono se determina a partir de una señal filtrada en paso bajo y a la inversa. La ganancia de tono se determina a partir del valor medio cuadrático (RMS) de la señal. Se cuantifican los parámetros relevantes que caracterizan la codificación, se les envía al descodificador y se les utiliza para producir una señal sintetizada en el descodificador. Más particularmente, esta técnica de una codificación proporciona codificación con diez coeficientes.

El vocodificador 400 utiliza un modelo de síntesis más simple que la técnica GSM-EFR y, en consecuencia, utiliza menos bits que la técnica GSM-EFR para representar la voz, lo que, sin embargo, da como resultado una calidad inferior. La tasa de bits baja hace adecuados a los vocodificadores como codificadores redundantes para la voz (lo que se describe a continuación). Los vocodificadores trabajan bien modelando mensajes vocalizados y no vocalizados, pero no manejan con precisión las consonantes oclusivas (que representan el cierre completo y la subsiguiente liberación de una obstrucción del tracto vocal) y la información no vocal (por ejemplo, música).

Otros detalles en la codificación vocal convencional pueden deducirse del libro “Voz Digital: Codificación para Sistemas de Comunicación de Tasa de Bits Baja”, A.M. Kondoz, 1994, John Wiley & Sons, que se incorpora aquí por referencia en su totalidad.

2. Corrección Directa de Errores (FEC)

5 Una vez codificado, un sistema de comunicación puede transferir la voz en una variedad de formatos. Las redes basadas en paquetes transfieren los datos de audio en una serie de paquetes discretos.

El tráfico basado en paquetes puede someterse a altas relaciones de pérdida de paquetes, inestabilidad y reordenación. La corrección directa de errores (FEC) es una técnica para abordar el problema de los paquetes perdidos. En general, la FEC implica transmitir información redundante junto con la voz codificada. El descodificador intenta utilizar la información redundante para reconstruir paquetes perdidos. Las técnicas de FEC independientes del medio añaden información redundante sobre la base de los bits dentro de la corriente de audio (independiente del conocimiento de nivel más alto de las características de la corriente vocal). Por otro lado, las técnicas de FEC dependientes del medio añaden información redundante sobre la base de las características de la corriente vocal.

15 La patente U.S. No. 5.870.412 de Schuster et al. describe una técnica independiente del medio. Este método agrega un único código de corrección directa de errores a cada uno de una serie de paquetes de carga útil. El código de corrección de errores se define tomando una suma XOR de un número especificado precedente de paquetes de carga útil. Un receptor puede reconstruir una carga útil perdida a partir de los códigos de corrección de errores redundantes llevados por paquetes sucesivos y puede corregir también la pérdida de múltiples paquetes en una fila. Esta técnica tiene la desventaja de utilizar un retardo variable. Además, el resultado XOR debe ser del mismo tamaño que la carga útil más grande utilizada en el cálculo.

20 La figura 5 muestra una visión de conjunto de una técnica de FEC basada en el medio. El módulo codificador 502 incluye un codificador primario 508 y un codificador redundante 510. Un paquetizador 516 recibe la salida del codificador primario 508 y el codificador redundante 510 y, a su vez, envía su salida sobre el medio de transmisión 506. Un módulo descodificador 504 incluye un descodificador primario 512 y un descodificador redundante 514. La salida del descodificador primario 512 y del descodificador redundante 514 es controlada por una lógica de control 518.

En funcionamiento, el codificador primario 508 genera datos codificados primarios utilizando un modelo de síntesis primario. El codificador redundante 510 genera datos codificados redundantes utilizando un modelo de síntesis redundante. El modelo de síntesis redundante proporciona típicamente una versión más fuertemente comprimida de la voz que la del modelo de síntesis primario (por ejemplo, teniendo un consiguiente ancho de banda inferior y una consiguiente calidad más baja). Por ejemplo, un enfoque conocido utiliza datos codificados PCM como voz codificada primaria, y datos codificados LPC como voz codificada redundante (véase, por ejemplo, V. Hardman et al., “Audio Fiable para Uso en Internet”, Proc. INET’95, páginas 1-8, 1995). Los datos codificados LPC tienen una tasa de bits mucho más baja que los datos codificados PCM.

35 La figura 6 muestra la manera en que los datos redundantes (representados por bloques sombreados) pueden añadirse a los datos primarios (representados por bloques no sombreados). Por ejemplo, con referencia a la fila más alta de paquetes, el primer paquete contiene datos primarios para la trama n. Los datos redundantes para la trama previa, es decir, la trama n-1, se añaden a estos datos primarios. De esta manera, los datos redundantes dentro de un paquete se refieren siempre a datos primarios previamente transmitidos. La técnica proporciona un único nivel de redundancia, pero pueden proporcionarse niveles adicionales (transmitiendo copias adicionales de los datos redundantes).

Se han propuesto formatos específicos para añadir los datos redundantes a la carga útil de datos primarios. Por ejemplo, Perkins et al. proponen un formato específico para añadir datos redundantes codificados LPC a datos de carga útil primarios dentro del Protocolo de Transporte en Tiempo Real (RTP) (por ejemplo, véase C. Perkins et al., “Carga Útil RTP para Datos de Audio Redundantes”, RFC 2198, septiembre de 1997). La cabecera de paquete incluye información que pertenece a los datos primarios e información que pertenece a los datos redundantes. Por ejemplo, la cabecera incluye un campo para proporcionar el sello de tiempo de la codificación primaria que indica el tiempo de codificación primaria de los datos. La cabecera incluye también un sello de tiempo decalado que indica la diferencia de tiempo entre la codificación primaria y la codificación redundante representadas en el paquete.

50 Con referencia a las figuras 5 y 6, el módulo descodificador 504 recibe los paquetes que contienen datos primarios y redundantes. El módulo descodificador 504 incluye una lógica (no mostrada) para separar los datos primarios de los datos redundantes. El descodificador primario 512 descodifica los datos primarios, mientras que el descodificador redundante 514 descodifica los datos redundantes. Más específicamente, el módulo descodificador 504 descodifica una trama n de datos primarios cuando llega el siguiente paquete que contiene los datos redundantes para la trama n. Este retardo se añade a la reproducción y se representa gráficamente en la figura 6 por la leyenda “Retardo extra”.

En el método de la técnica anterior, la lógica de control 518 instruye al módulo descodificador 504 para que utilice la voz sintetizada generada por el descodificador primario 512 cuando se recibe un paquete que contiene datos

codificados primarios. Por otro lado, la lógica de control 518 instruye al módulo descodificador 504 para que utilice la voz sintetizada generada por el descodificador redundante 514 cuando se “pierde” el paquete que contiene datos primarios. En tal caso, la lógica de control 518 sirve simplemente para rellenar intersticios en la corriente recibida de tramas codificadas primarias con tramas codificadas redundantes. Por ejemplo, en la técnica antes referenciada descrita en Hardman et al., el descodificador descodificará los datos codificados LPC en lugar de los datos codificados PCM tras la detección de la pérdida de paquetes en la corriente codificada PCM.

El uso de FEC convencional para mejorar la calidad de la transmisión de audio basada en paquetes no es completamente satisfactorio. Por ejemplo, los modelos de síntesis de voz utilizan los parámetros de estados operativos pasados para generar una síntesis de voz precisa en estados operativos presentes. En este sentido, los modelos “dependen del historial”. Por ejemplo, un modelo de voz de predicción lineal excitada por código algebraico (ACELP) utiliza síntesis previamente producidas para actualizar su libro de código adaptativo. El filtro LPC, los historiales de ocultación de errores y diversos predictores de cuantificación utilizan también estados previos para generar voz con precisión en estados actuales. Así, aun cuando un descodificador pueda reconstruir tramas ausentes utilizando datos redundantes, la “memoria” del modelo de síntesis primaria es deficiente debido a la pérdida de datos primarios. Esto puede crear problemas “persistentes” en la calidad de síntesis de voz. Por ejemplo, un libro de código pobremente actualizado puede causar formas de onda distorsionadas para más de diez tramas. Las técnicas FEC convencionales no hacen nada para abordar estos tipos de problemas persistentes.

Además, las técnicas de codificación de voz basadas en FEC pueden adolecer de una multitud de otros problemas no abordados hasta ahora por las técnicas FEC. Por ejemplo, en técnicas de análisis por síntesis utilizando predictores lineales, las discontinuidades de fase pueden ser muy audibles. En técnicas que utilizan un libro de código adaptativo, un error de fase colocado en el bucle de realimentación puede permanecer para numerosas tramas. Además, en codificaciones vocales que utilizan coeficientes LP que son predichos cuando se codifican, una pérdida del parámetro LPC disminuye la precisión de un predictor. Esto introducirá errores en el parámetro más importante en una técnica de codificación vocal LPC.

Sumario

En consecuencia, es un objetivo general de la presente invención mejorar la calidad de la voz producida utilizando la técnica FEC.

Este y otros objetivos se consiguen por la presente invención a través de una técnica FEC mejorada para codificar datos vocales.

De acuerdo con un primer aspecto, según la reivindicación 1, se proporciona un módulo descodificador para descodificar datos de audio formateados en paquetes que contienen datos codificados primarios y datos codificados redundantes. El módulo descodificador comprende un descodificador primario para descodificar los paquetes utilizando un modelo de síntesis primaria; un descodificador redundante para descodificar los paquetes utilizando un modelo de síntesis redundante; y medios para localizar una posición de impulso de tono en una trama actual localizando la última posición de impulso conocida en una trama previa, y avanzando a continuación desde la última posición de impulso conocida en uno o más valores de retardo de tono a fin de localizar la posición de impulso en la trama actual, en donde la posición de impulso de tono localizada en la trama actual se utiliza para reducir discontinuidades de fase.

De acuerdo con un segundo aspecto de la presente invención, según la reivindicación 3, se proporciona un método para descodificar datos de audio formateados en paquetes que contienen datos codificados primarios y datos codificados redundantes. El método comprende las etapas de descodificar primariamente los paquetes utilizando un modelo de síntesis primaria; descodificar redundantemente los paquetes utilizando un modelo de síntesis redundante; en donde la descodificación primaria o la descodificación redundante comprende la etapa de localizar una posición de impulso de tono en una trama actual localizando la última posición de impulso conocida en una trama previa, y avanzando a continuación desde la última posición de impulso conocida en uno o más valores de retraso de tono a fin de localizar la posición de impulso en la trama actual, en donde la posición de impulso de tono localizada se utiliza para reducir discontinuidades de fase.

Breve descripción de los dibujos

Los anteriores y otros objetos, características y ventajas de la presente invención se entenderán más fácilmente tras la lectura de la siguiente descripción detallada en conjunción con los dibujos, en los que:

La figura 1 muestra un codificador convencional de predicción lineal excitada por código (CELP);

La figura 2 ilustra un residual generado por el codificador CELP de la figura 1;

La figura 3 muestra otro tipo de codificador CELP que utiliza un libro de código adaptativo;

La figura 4 muestra un vocodificador convencional;

La figura 5 muestra un sistema convencional para realizar una corrección directa de errores en una red paquetizada;

La figura 6 muestra un ejemplo de la combinación de información primaria y redundante en el sistema de la figura 5;

La figura 7 muestra un sistema para realizar una corrección directa de errores en una red paquetizada según un ejemplo de la presente invención;

5 La figura 8 muestra un ejemplo de un módulo codificador para uso en la presente invención;

La figura 9 muestra la división de subtramas para un codificador redundante en un ejemplo de la presente invención; y

La figura 10 muestra un ejemplo de una máquina de estado para uso en la lógica de control del módulo descodificador mostrado en la figura 7.

10 Descripción detallada

En la siguiente descripción, para fines de explicación y no de limitación, se exponen detalles específicos a fin de proporcionar una comprensión completa de la invención. Sin embargo, será evidente para un experto en la materia que la presente invención puede ponerse en práctica en otras realizaciones que se aparten de estos detalles específicos. En otros casos, se omiten descripciones detalladas de métodos, dispositivos y circuitos bien conocidos para no oscurecer la descripción de la presente invención con detalle innecesario. En los dibujos, los números iguales representan características idénticas.

La invención se aplica en general al uso de técnicas de corrección directa de errores para procesar datos de audio. Sin embargo, a fin de facilitar la discusión, la siguiente explicación se enmarca en el contexto específico de la codificación de señales vocales.

20 1. Visión de conjunto

La figura 7 muestra una visión de conjunto de un ejemplo de un sistema 700 para implementar la presente invención, que incluye un módulo codificador 702 y un módulo descodificador 704. El módulo codificador 702 incluye un codificador primario 708 para producir datos codificados primarios y un codificador redundante 710 para producir datos codificados redundantes. La lógica de control 720 en el módulo codificador 702 controla aspectos del funcionamiento del codificador primario 708 y del codificador redundante 710. Un paquetizador 716 recibe la salida del codificador primario 708 y el codificador redundante 710 y, a su vez, transmite los datos codificados primarios y los datos codificados redundantes sobre el medio de transmisión 706. El módulo descodificador 704 incluye un descodificador primario 712 y un descodificador redundante 714, ambos controlados por una lógica de control 718. Además, el módulo descodificador 704 incluye una memoria intermedia de recepción (no mostrada) para almacenar temporalmente un paquete recibido al menos hasta que los datos redundantes del paquete recibido lleguen en un paquete posterior.

En funcionamiento, el codificador primario 708 codifica la voz de entrada utilizando una técnica de codificación primaria (sobre la base de un módulo de síntesis primaria), y el codificador redundante 710 codifica la voz de entrada utilizando una técnica de codificación redundante (sobre la base de un modelo de síntesis redundante).

Aunque no es necesario, la técnica de codificación redundante proporciona típicamente un ancho de banda más pequeño que la técnica de codificación primaria. El paquetizador 716 combina los datos codificados primarios y los datos codificados redundantes en una serie de paquetes, en donde cada paquete incluye datos primarios y redundantes. Más específicamente, el paquetizador 716 puede utilizar la técnica FEC ilustrada en la figura 6. En esta técnica, un paquete que contiene datos primarios para una trama actual, es decir, la trama n, se combina con datos redundantes que pertenecen a una trama previa, es decir, la trama n-1. La técnica proporciona un único nivel de redundancia. El paquetizador 716 puede utilizar cualquier formato de paquete conocido para combinar los datos primarios y redundantes, tal como el formato propuesto por Perkins et al. discutido en la sección de Antecedentes (por ejemplo, cuando la cabecera de paquete incluye información que pertenece a cargas útiles primarias y redundantes, incluyendo información de sello de tiempo que pertenece a ambas cargas útiles).

Después del ensamble, el paquetizador 716 reenvía los paquetes a través del medio de transmisión 706. El medio de transmisión 706 puede representar cualquier sistema de transmisión basado en paquetes, tal como una red de Protocolo de Internet (IP). Alternativamente, en lugar de la transmisión, el sistema 700 puede almacenar simplemente los paquetes en un medio de almacenamiento para su recuperación posterior.

El módulo descodificador 704 recibe los paquetes y reconstruye la información vocal utilizando el descodificador primario 712 y el descodificador redundante 714. El módulo descodificador 704 utiliza generalmente el descodificador primario 712 para descodificar los datos primarios y el descodificador redundante 714 para descodificar los datos redundantes cuando no están disponibles los datos primarios. Más específicamente, la lógica de control 718 puede emplear una máquina de estado para gobernar el funcionamiento del descodificador primario 712 y el descodificador redundante 714. Cada estado en la máquina de estado refleja una condición de error diferente experimentada por el módulo descodificador 704. Cada estado define también instrucciones para

descodificador una trama actual de datos. Esto es, las instrucciones especifican diferentes estrategias de descodificación para descodificar la trama actual apropiadas para diferentes condiciones de error. Más específicamente, las estrategias incluyen el uso del modelo de síntesis primaria, el uso del modelo de síntesis redundante y/o el uso de un algoritmo de ocultación de error. Las condiciones de error dependen de la estrategia de codificación utilizada en la trama previa, la disponibilidad de los datos primarios y redundantes en la trama actual, y la recepción o no recepción del siguiente paquete. La recepción o no recepción de paquetes dispara las transiciones entre estados.

A diferencia de los sistemas convencionales, el sistema 700 ofrece varios mecanismos para proporcionar interacción entre los modelos de síntesis primaria y redundante. Más específicamente, la lógica de control 720 del módulo descodificador incluye mecanismos de control para proporcionar interacción entre los modelos de síntesis primaria y redundante utilizados por los codificadores primario y redundante (es decir, los codificadores 708 y 710), respectivamente. Asimismo, la lógica de control 718 del módulo descodificador incluye mecanismos de control para proporcionar interacción entre los modelos de síntesis primaria y redundante utilizados por los descodificadores primario y redundante (es decir, los descodificadores 712 y 714), respectivamente. La figura 7 muestra gráficamente la interacción entre el codificador primario 708 y el codificador redundante 710 utilizando las flechas 750, y la interacción entre el descodificador primario 712 y el descodificador redundante 714 utilizando las flechas 752.

Las siguientes secciones presentan una visión de conjunto de las características utilizadas en el sistema 700 que proporcionan la interacción antes descrita entre los modelos de síntesis primaria y redundante, así como otras nuevas características de codificación vocal FEC.

1.1 Actualización de Estados en el Módulo Descodificador

Como se discute en la sección de Antecedentes, las técnicas FEC convencionales funcionan sustituyendo rudimentariamente datos descodificados redundantes por datos descodificados primarios ausentes, pero no hacen nada para actualizar la "memoria" del modelo de síntesis primaria a fin de reflejar la pérdida de los datos primarios. Para abordar este problema, la presente invención utiliza información deducida del modelo de síntesis redundante a fin de actualizar el estado o los estados del modelo de síntesis primaria. Análogamente, el módulo descodificador 704 puede remediar deficiencias de "memoria" en el modelo de síntesis redundante utilizando información paramétrica obtenida del modelo de síntesis primaria. Así, hablando en general, los dos modelos "se ayudan uno a otro" para suministrar información ausente. Por el contrario, en FEC convencional, los modelos no comparten información.

La estrategia específica utilizada para actualizar los moldeos depende, por supuesto, de los requisitos de los modelos. Algunos modelos pueden tener dependencias de estados pasados más exigentes que las de otros. Esto depende también de las condiciones de error predominantes presentes en el módulo descodificador 704. Para repetir, las condiciones de error se caracterizan por la estrategia utilizada en la trama previa a fin de descodificar la voz (por ejemplo, primaria, redundante, ocultación de error), la disponibilidad de los datos en la trama actual (por ejemplo, primaria o redundante) y la recepción o no recepción de la siguiente trama. En consecuencia, las instrucciones de descodificación asociadas con cada estado de la máquina de estado, que son específicas de las condiciones de error, definen también preferiblemente el método para actualizar los modelos de síntesis. De esta manera, el módulo descodificador 704 adapta la estrategia de actualización a las condiciones de error predominantes.

Unos pocos ejemplos servirán para ilustrar la característica de actualización de la presente invención. Considérese, por ejemplo, el estado en el que el módulo descodificador 704 no ha recibido los datos primarios de la trama actual (es decir, se pierden los datos primarios), pero ha recibido el paquete de la trama posterior que lleva datos redundantes para la trama actual. En este estado, el módulo descodificador 704 descodifica la voz sobre la base de los datos redundantes para la trama actual. Los valores descodificados se utilizan entonces para actualizar el modelo de síntesis primaria. Un modelo basado en CELP, por ejemplo, puede requerir actualizaciones de su libro de código adaptativo, su filtro LPC, sus historiales de ocultamiento de errores y sus diversos predictores de cuantificación. Los parámetros redundantes pueden necesitar alguna forma de conversión para adaptarse al formato de parámetro utilizado en el descodificador primario.

Considérese el caso específico en el que el módulo descodificador 704 utiliza un modelo de síntesis primaria basado en la codificación GSM-EFR. Como se discute en la sección de Antecedentes, el modelo GSM-EFR utiliza un predictor de cuantificación para reducir la dinámica de los parámetros LPC antes de la cuantificación. En este caso, el módulo descodificador 704 utiliza también un modelo de síntesis redundante que no emplee un predictor de cuantificación y, por tanto, proporcione LPCs codificadas "absolutas". En el presente enfoque, el modelo de síntesis primaria proporciona información que pertenece a residuales LSF (es decir, LSF_{res}), mientras que el modelo redundante proporciona información que pertenece a valores LSF absolutos para estos coeficientes (es decir, LSF_{red}). El módulo descodificador 704 utiliza los valores residuales y absolutos para calcular el estado del predictor utilizando la Ec. 11 siguiente, a fin de proporcionar, por tanto, una actualización rápida del predictor:

$$LSF_{prev, res} = (LSF_{red} - LSF_{media} - LSF_{res}) / predFactor$$

(Ec. 11),

donde el término LSF_{media} define un valor LSF medio, el término $predFactor$ se refiere a un constante del factor de predicción y $LSF_{prev,res}$ se refiere a una LSF residual de la trama pasada (es decir, la trama n-1). El módulo descodificador 704 utiliza el estado actualizado del predictor para descodificar los residuales LSF a los coeficientes LPC (por ejemplo, utilizando la Ec. 7 anterior).

- 5 El uso de la Ec. 11 es particularmente ventajoso cuando el estado del predictor ha llegado a ser inseguro debido a una pérdida o pérdidas de paquetes.

1.2 Anticipación del módulo descodificador

Como se ilustra en la figura 6, el módulo descodificador 704 debe retardar la descodificación de los datos primarios contenidos en un paquete hasta que recibe el siguiente paquete. El retardo entre la recepción y la descodificación de los datos primarios permite que el módulo descodificador 704 utilice los datos primarios para cualquier tipo de procesamiento de predescodificación a fin de mejorar la calidad de la síntesis de voz. Esto se denomina aquí "anticipación del descodificador". Por ejemplo, considérese el caso en el que el módulo descodificador 704 deja de recibir el paquete que contiene la trama n codificada primaria, pero recibe seguidamente el paquete que contiene los datos codificados primarios para la trama n+1, que incluye los datos codificados redundantes para la trama n. El módulo descodificador 704 descodificará en consecuencia los datos para la trama n utilizando datos redundantes. Mientras tanto, el módulo descodificador 704 puede utilizar los datos primarios para la trama n+1 (todavía por descodificar) para el procesamiento de la anticipación. Por ejemplo, los datos primarios para la trama n+1 pueden utilizarse para mejorar la interpolación de los niveles de energía a fin de proporcionar una transición más suave entre la trama n y la trama n+1. La anticipación puede utilizarse también en la interpolación LPC a fin de proporcionar resultados de interpolación más precisos cerca del final de la trama.

1.3 Anticipación del módulo codificador

Como se ha explicado previamente, el paquetizador 716 del módulo codificador 702 combina datos primarios que pertenecen a una trama actual con datos redundantes que pertenecen a una trama previa; por ejemplo, el paquetizador combina datos primarios que pertenecen a la trama n con datos redundantes que pertenecen a la trama n-1. En consecuencia, el módulo codificador 702 debe retardar en una trama la transmisión de datos redundantemente codificados. Debido a este retardo de una trama, el codificador redundante 710 puede retardar también su codificación de los datos redundantes de tal manera que todos los datos (primarios y redundantes) combinados en un paquete se descodifican al mismo tiempo. Por ejemplo, el módulo codificador 702 podría codificar los datos redundantes para la trama n-1 al mismo tiempo que codifica los datos primarios para la trama n. En consecuencia, los datos redundantes están disponibles durante un breve tiempo antes de la descodificación. La disponibilidad de avance de los datos redundantes (por ejemplo, la trama redundante n-1) proporciona oportunidades para el procesamiento de anticipación. Los resultados del procesamiento de anticipación pueden utilizarse para mejorar el procesamiento redundante posterior de la trama. Por ejemplo, la decisión de vocalización en un modelo de síntesis de vocodificador (que sirve como el modelo de síntesis redundante) puede mejorarse a través del uso de los datos de anticipación en su cálculo. Esto dará como resultado menos decisiones erróneas con respecto al momento en que comienza realmente el segmento vocalizado.

La anticipación en el módulo codificador 702 puede implementarse de diversas maneras, tal como por el uso de una lógica de control 720 para coordinar la interacción entre el codificador primario 708 y el codificador redundante 710.

1.4 Mantenimiento de la Fase de Impulso de Tono

La fase de tono (es decir, la posición del impulso de tono) proporciona información útil para realizar la técnica FEC. En un primer caso, el módulo descodificador 704 identifica la localización del último impulso en el libro de código adaptativo que pertenece a la trama previa. Más específicamente, el módulo 704 puede localizar la posición del impulso de tono calculando la correlación entre el libro de código adaptativo y un impulso de tono predeterminado. La fase de impulso de tono puede determinarse a continuación localizando el pico o picos de correlación. Sobre la base del conocimiento de la localización del último impulso y el retraso de tono, el módulo descodificador 704 identifica a continuación la localización en la que el impulso subsiguiente deberá colocarse en la trama actual. Hace esto moviéndose hacia delante uno o más periodos de tono dentro de la nueva trama desde la localización del último impulso. Una aplicación específica de esta técnica es aquella en la que GSM-EFR sirve como descodificador primario y un modelo basado en vocodificador sirve como descodificador redundante. El módulo descodificador 704 utilizará los datos redundantes tras dejar de recibir los datos primarios. En esta circunstancia, el módulo descodificador 704 utiliza la técnica para colocar el impulso de tono del vocodificador sobre la base de la información de fase extraída del libro de código adaptativo. Esto ayuda a asegurar que un impulso de tono del vocodificador no se coloque en un periodo completamente incorrecto.

En un segundo caso, el módulo codificador 702 determina y transmite información que pertenece a la fase de tono de la señal vocal original (tal como posición de impulso de tono y signo de impulso de tono) en la codificación redundante. De nuevo, esta información puede obtenerse calculando la correlación entre el libro de código adaptativo y un impulso de tono predeterminado. Tras la recepción, el módulo descodificador 704 puede comparar la información de fase de tono recibida con la información de fase de tono detectada utilizando el libro de código

adaptativo (calculado de la manera descrita anteriormente). Una diferencia entre la información de fase de tono codificada redundante y la información de fase de tono del libro de código adaptativo constituye una discontinuidad de fase. Para abordar este asunto, la técnica puede ajustar periodos de tono sobre el curso de la trama actual con el objetivo de proporcionar la fase correcta al final de la trama. Como consecuencia, el libro de código adaptativo recibirá la información de fase correcta cuando se le actualiza. Una aplicación específica de esta técnica es aquella en la que la técnica GSM-EFR sirve como descodificador primario y un modelo basado en vocodificador sirve como descodificador redundante. De nuevo, el módulo descodificador 704 utilizará los datos redundantes tras dejar de recibir los datos primarios. En esta circunstancia, el vocodificador recibe información con respecto a la posición y signo del impulso desde el codificador redundante. Computa a continuación, en el libro de código adaptativo de la manera descrita anteriormente, la localización en la que deberá tener lugar el impulso. Cualquier diferencia de fase entre la localización recibida y la localización calculada se suaviza en la trama de modo que la fase será correcta al final de la trama. Esto asegurará que el módulo descodificador 704 tendrá información de fase correcta almacenada en el libro de código adaptativo tras volver al uso de la descodificación primaria (por ejemplo, descodificación GSM-EFR) en la siguiente trama.

Como alternativa al segundo caso, el descodificador redundante no recibe ninguna información con respecto a la posición del impulso proveniente del sitio del codificador. En lugar de esto, computa la posición de impulso a partir de los datos primarios descodificados en la siguiente trama. Esto se hace extrayendo información de fase de impulso de la siguiente trama primaria y a continuación retrocediendo hasta la trama actual para determinar la ubicación correcta de los impulsos en la trama actual. Esta información se compara entonces con otra indicación de colocación de impulso calculada a partir de la trama previa según el método descrito anteriormente. Cualquier discrepancia en la posición puede ser corregida según el método descrito anteriormente (por ejemplo, suavizando el error de fase en el curso de la trama actual, de modo que la siguiente trama tendrá la fase correcta, como se refleja en el libro de código adaptativo).

1.5 Selección alternativa de los parámetros redundantes.

La figura 8 muestra un módulo codificador alternativo 800 para uso en la técnica FEC. El codificador 800 incluye un codificador primario 802 conectado a un paquetizador 808. Un extractor 804 extrae información paramétrica del codificador primario 802. Un módulo de retardo 806 retarda los parámetros extraídos en, por ejemplo, una trama. El módulo de retardo 806 reenvía los parámetros redundantes retardados al paquetizador 808.

En funcionamiento, el extractor 804 selecciona un subconjunto de parámetros de entre los parámetros codificados primarios. El subconjunto deberá seleccionarse para permitir la creación de voz sintetizada a partir de los parámetros redundantes, y para permitir la actualización de estados en el modelo de síntesis primaria cuando se requiera. Por ejemplo, los valores de LPC, de retraso LTP y de ganancia serían adecuados para la duplicación en una técnica de codificación de análisis por síntesis. En un caso, el extractor extrae todos los parámetros generados por el codificador primario. Estos parámetros pueden convertirse en un formato diferente para representar los parámetros con ancho de banda reducido (por ejemplo, cuantificando los parámetros por el uso de un método que requiere menos bits que el modelo de síntesis primaria utilizado por el codificador primario 802). El módulo de retardo 806 retarda los parámetros redundantes en una trama, y el paquetizador combina los parámetros redundantes retardados con los parámetros codificados primarios utilizando, por ejemplo, el protocolo FEC ilustrado en la figura 6.

2. Ejemplo

2.1 Codificadores Primario y Redundante para Uso con FEC

El estándar de codificación vocal GSM-EFR, discutida en la sección de Antecedentes, puede utilizarse para codificar la corriente primaria de datos vocales. El estándar GSM-EFR se describe adicionalmente en el documento "Sistema Global para Comunicaciones Móviles: Sistemas de Telecomunicaciones Celulares Digitales: Transcodificación Vocal de Tasa Completa Reforzada (EFR) (GSM 06.60)", noviembre de 1996. Como se describe anteriormente, el estándar de codificación vocal GSM-EFR utiliza un codificador de predicción lineal excitada por código algebraico (ACELP). La ACELP de la GSM-EFR codifica una trama de 20 ms que contiene 160 muestras, correspondientes a 244 bits/trama y una corriente de bits codificada de 12,2 kbits/s. Además, el codificador primario utiliza la técnica de ocultación de errores descrita en el documento "Sistema de Telecomunicaciones Celulares Digitales: Sustitución y Silenciamiento de Tramas Perdidas para Canales de Tráfico Vocal de Tasa Completa Reforzada (EFR) (GSM 06.61)", versión 5.1.2, abril de 1997 (también resumido anteriormente).

Puede utilizarse un vocodificador para codificar la corriente redundante de datos vocales. El vocodificador utilizado en este ejemplo incorpora algunas características del vocodificador LPC-10 discutido en la sección de Antecedentes, y otras características del sistema GSM-EFR. Las características basadas en GSM-EFR hacen que la salida del vocodificador sea más fácilmente compatible con los datos primarios generados por el codificador primario GSM-EFR. Por ejemplo, el vocodificador LPC-10 utiliza tramas de 22,5 ms, mientras que el codificador GSM-EFR utiliza tramas de 20 ms. En consecuencia, el diseño híbrido incorpora el uso de tramas de 20 ms. El vocodificador híbrido diseñado para esta aplicación FEC se denomina vocodificador "GSM-VOC".

El decodificador GSM-VOC incluye la configuración conceptual básica mostrada en la figura 4. A saber, el GSM-VOC incluye una funcionalidad para aplicar una señal de excitación que comprende un vector de ruido (para sonidos no vocalizados) o una forma de impulso estático (para mensajes vocalizados). La excitación se procesa a continuación por un bloque de filtro LPC para producir una señal sintetizada.

- 5 En funcionamiento, el codificador GSM-VOC divide la voz de entrada en tramas de 20 ms, y filtra la voz en paso alto utilizando un filtro con una frecuencia de corte de 80 Hz. Se calcula a continuación el valor medio cuadrático (RMS) de la energía de la voz. El GSM-VOC calcula y cuantifica entonces un único conjunto de coeficientes LP utilizando el método expuesto en el estándar GSM-EFR. (Por el contrario, sin embargo, el estándar GSM-EFR descrito anteriormente computa dos conjuntos de coeficientes). El codificador GSM-VOC deriva el único conjunto de
10 coeficientes sobre la base de la ventana que tiene más peso en las últimas muestras, como en el estándar GSM-EFR 06.60. Después de que el codificador encuentre los coeficientes LP, éste calcula el residual.

- El codificador realiza entonces una búsqueda de tono de bucle abierto en cada mitad de la trama. Más específicamente, el codificador realiza esta búsqueda calcula la autocorrelación sobre 80 muestras para retrasos en el rango de 18 a 143 muestras. El codificador pondera entonces las correlaciones calculadas a favor de pequeños
15 retrasos. Esta ponderación se hace dividiendo el vano de las muestras de 18 a 143 en tres sectores, a saber, un primer vano de 18-35, un segundo vano de 36-71 y un tercer vano de 72-143 muestras. El decodificador determina y pondera entonces el valor máximo de cada sector (para favorecer pequeños retrasos) y selecciona el más grande. A continuación, el codificador compara los valores máximos asociados con las dos mitades de trama, y selecciona el retraso LTP de la mitad de la trama con la correlación más grande. La ponderación favorable de pequeños retrasos
20 es útil para seleccionar un valor de retraso primario (básico) cuando están presentes múltiplos del valor de retraso en la correlación.

- El codificador calcula la vocalización sobre la base de la correlación máxima no ponderada derivada de la búsqueda de bucle abierto. Más específicamente, como se muestra en la figura 9, el codificador basa la decisión de
25 vocalización en el rango de muestras que abarca las dos semitramas previas, la semitrama actual y las dos semitramas siguientes (para un total de cinco correlaciones). Para calcular las correlaciones para la siguiente trama, el codificador requiere una anticipación de 20 ms. La técnica FEC proporciona la anticipación sin añadir un retardo extra al codificador. A saber, el módulo codificador combina datos primarios que pertenecen a una trama n con datos redundantes que pertenecen a una trama anterior, es decir, la trama $n-1$. Codificando la trama redundante $n-1$ al mismo tiempo que la trama primaria n , el codificador redundante tiene acceso a la trama de anticipación. En otras
30 palabras, el codificador redundante tiene una oportunidad de "investigar" la trama redundante $n-1$ antes de su codificación redundante.

- Para determinar si el mensaje está vocalizado o no, el codificador compara las cinco correlaciones mostradas con tres umbrales diferentes. En primer lugar, el codificador calcula una mediana a partir de la presente trama y las dos
35 semitramas siguientes, y compara la mediana con un primer umbral. El codificador utiliza el primer umbral para reaccionar rápidamente al inicio de un segmento vocalizado. En segundo lugar, el codificador calcula otra mediana formada a partir de las cinco correlaciones, y compara a continuación esta mediana con un segundo umbral. El segundo umbral es inferior al primer umbral y se utiliza para detectar la vocalización durante un segmento vocalizado. En tercer lugar, el codificador determina si la semitrama previa estaba vocalizada. Si es así, el
40 codificador compara también la mediana formada a partir de las cinco correlaciones con un tercer umbral. El valor del tercer umbral es el más bajo de los tres umbrales. El codificador utiliza el tercer umbral para extender los segmentos vocalizados hasta el punto verdadero de transición o más allá de éste (por ejemplo, para crear una "persistencia"). El tercer umbral asegurará que el codificador marcará la semitrama en la que la transición del mensaje vocalizado al no vocalizado tiene lugar como mensaje vocalizado. La información enviada al decodificador incluye la vocalización anteriormente computada para ambas subtramas.

- 45 El codificador utiliza una técnica codificadora vocal GSM-EFR 06.60 modificada (o una técnica IS-641 modificada) para cuantificar los coeficientes LP. Como se detalla, GSM-EFR 06.60 describe un predictor que utiliza un factor de predicción basado en las frecuencias espectrales de línea LSFs de la trama previa. Por el contrario, el predictor de la presente técnica utiliza valores LSF medios (en donde los valores medios se computan según el estándar GSM-EFR 06.60). Esto elimina dependencias de la trama previa en la cuantificación de las LPCs. La técnica agrupa tres
50 vectores basados en residuales (por ejemplo, 10 residuales) obtenidos a partir de la predicción. La técnica compara entonces los vectores con una tabla estadísticamente producida para determinar la mejor coincidencia. Se devuelve un índice de la tabla que presenta la mejor coincidencia. Los tres índices correspondientes a los tres vectores utilizan 26 bits.

- Además, el codificador conviene el valor RMS en dB y, a continuación, lo cuantifica linealmente utilizando siete bits,
55 aunque pueden utilizarse menos bits (por ejemplo, cinco o seis bits). El estado de vocalización utiliza dos bits para representar la vocalización en cada semitrama. El tono tiene un rango de (18 a 143) muestras. Se resta un valor de 18 de modo que los números válidos encajen en siete bits (es decir, para proporcionar un rango de 0 a 125 muestras).

La Tabla 1 siguiente resume la asignación de bit anteriormente discutida en el GSM-VOC.

Tabla 1

Parámetro	Número de Bits
LPC	26
Retraso de Tono	7
Valor RMS	7
Estado de Vocalización	2
Posición de Impulso de Tono	8
Signo de Impulso de Tono	1
Total (ancho de banda)	51 (2550 b/s)

La posición de impulso de tono y su señal proporcionan información útil para realizar la técnica FEC. Estos parámetros indican, con una resolución de una muestra, la posición de partida del impulso de tono en una trama. El uso de esta información permite que la técnica mantenga la excitación y su síntesis en la fase con la voz original. Estos parámetros se encuentran correlacionando primero el residual y una forma de impulso fija. La posición y el signo se localizan a continuación en la curva de correlación con ayuda de la decisión de vocalización, que se utiliza para identificar la mitad de trama correcta (por ejemplo, la decisión de vocalización podría utilizarse para descartar un impulso "falso" detectado en una mitad de trama no vocalizada). Por el contrario, un codificador autónomo (es decir, un codificador no acoplado con otro codificador para realizar FEC) no especifica ninguna información que pertenezca a la posición de impulso (es decir, la fase de impulso). Esto se debe a que la fase de tono es irrelevante en un codificador autónomo en tanto que una época de tono tenga la distancia de retraso de tono dada.

Volviendo ahora al descodificador, el descodificador GSM-VOC crea un vector de excitación a partir de la decisión de vocalización y el paso. La vocalización tiene seis estados diferentes, incluyendo dos estados permanentes y cuatro estados de transición. Los estados permanentes incluyen un estado vocalizado y un estado no vocalizado. Los estados de transición incluyen un estado que pertenece a la transición desde un estado no vocalizado hasta un estado vocalizado, y un estado que pertenece a la transición desde un estado vocalizado hasta un estado no vocalizado. Estos estados de transición tienen lugar en cada mitad de la trama, definiendo así los cuatro estados diferentes. Para las partes vocalizadas de la trama, el descodificador utiliza el tono dado a fin de determinar las épocas que se calcula (en donde el término "épocas" se refiere a los vanos correspondientes de las muestras, por ejemplo a un periodo de tono). Por otro lado, el descodificador divide las tramas no vocalizadas en cuatro épocas de 40 muestras, cada una para fines de interpolación.

Para cada época de paso, el descodificador interpola los valores viejos y nuevos de RMS y tono (es decir, de la trama previa y las tramas actuales, respectivamente) para proporcionar transiciones más suaves. Además, para el mensaje vocalizado, la técnica de descodificación crea una excitación derivada de un impulso de 25 muestras de longitud y un ruido de baja intensidad. Para el mensaje no vocalizado, la señal de excitación incluye solamente ruido. Más específicamente, en una época de tono vocalizado, el bajo paso del descodificador filtra el impulso y el paso alto filtra el ruido. Un filtro definido por $1 + 0,7 A(z)$ filtra entonces la excitación creada, en donde $A(z)$ es la ganancia de $A(z)$. Esto reduce la naturaleza con picos de la voz sintética, como se discute en el documento Tremain, T., "El Algoritmo de Codificación Predictiva Lineal del Estándar del Gobierno: LPC-10", Tecnología de Voz, abril de 1982, págs. 40-48. El descodificador añade una oclusiva para tramas no vocalizadas, en donde el valor RMS se incrementa en más de ocho veces el valor de la trama previa. La posición de la oclusiva es aleatoria en la primera época de tono no vocalizado y consta de un doble impulso formado por un impulso positivo (añadido) y un impulso negativo (restado) consecutivos. El doble impulso proporciona la respuesta máxima del filtro. A continuación, la técnica ajusta el valor RMS de la época para que case con el valor interpolado (por ejemplo, un valor RMS interpolado formado a partir de los valores RMS de la trama pasada, actual y, si estuviera disponible, siguiente). Esto se hace calculando el valor RMS presente de una excitación filtrada por síntesis.

El descodificador interpola entonces las LPCs en el dominio LSF para cada subtrama de 40 muestras y aplica a continuación el resultado a la excitación. El impulso utilizado para la excitación vocalizada incluye polarización. Un filtro de paso alto elimina esta polarización utilizando una frecuencia de corte de 80 Hz.

Habiendo expuesto las características del codificador y el descodificador redundantes GSM-VOC, se describirá ahora el funcionamiento del conjunto de la técnica FEC utilizando GSM-EFR (para codificación y descodificación primarias) y GSM-VOC (para codificación y descodificación redundantes).

2.2 Utilización de los Codificadores Primario y Redundante en FEC

La figura 10 muestra un diagrama de estado de la máquina de estado prevista en la lógica de control 718 (de la figura 7). La llegada o no llegada de cada paquete da lugar a que la máquina de estado transite entre estados (o permanezca en el mismo estado). Más específicamente, la llegada del siguiente paquete define una transición

etiquetada con "0" en la figura. La no llegada del siguiente paquete (es decir, la pérdida de un paquete) define una transición etiquetada con "1" en la figura. Las características de los estados mostrados en la figura 10 se identifican a continuación.

Estado: Norma EFR

- 5 El estado "Norma EFR" indica que el módulo descodificador ha recibido tanto el paquete actual como el siguiente paquete.

El módulo descodificador descodifica la voz utilizando el descodificador primario según el protocolo estándar expuesto en, por ejemplo, GSM-EFR 06.60.

Estado: EFR Sig E

- 10 El estado "EFR Sig E" indica que el módulo descodificador ha recibido el paquete actual, pero no el siguiente paquete (nótese que el diagrama de estado en la figura 10 etiqueta la transición del estado "Norma EFR" a "EFR Sig E" como "1", indicando que se ha perdido un paquete).

- 15 En este estado, el módulo descodificador descodifica la voz como en el estado "Norma EFR". No obstante, debido a que los datos redundantes para esta trama están ausentes, no se proporciona ningún valor del parámetro RMS. Por tanto, el módulo descodificador calcula el valor RMS y lo introduce en el historial. Análogamente, debido a que el parámetro de estado de vocalización no está disponible, el módulo descodificador calcula la vocalización de la trama (por ejemplo, a partir de la voz sintetizada generada) tomando el máximo de la autocorrelación y suministrándolo al módulo de decisión de vocalización utilizado en el codificador. Dado que no se utiliza ninguna anticipación, puede resultar una decisión menos precisa.

- 20 Estado: Error Único Red

El estado "Error Único Red" indica que el módulo descodificador no ha recibido los datos primarios de la trama actual (es decir, se pierden los datos primarios), pero ha recibido el paquete de la trama siguiente que lleva datos redundantes para la trama actual.

- 25 En este estado, el módulo descodificador descodifica la voz utilizando los datos redundantes para la trama actual y los datos primarios para la siguiente trama. Más específicamente, el módulo descodificador descodifica las LPCs para la subtrama cuatro de la trama actual a partir de la trama redundante. Los valores descodificados se utilizan entonces para actualizar el predictor del descodificador LPC primario (es decir, el predictor para la cuantificación de los valores LPC). El módulo descodificador hace este cálculo de actualización sobre la base del residual LSF de la trama previa (como se discutirá con detalle adicional más adelante con respecto al estado "EFR R+C"). El uso de datos redundantes (en vez de datos primarios) puede introducir un error de cuantificación. El módulo descodificador computa los otros valores LPC de la subtrama por interpolación en el dominio LSF entre valores descodificados en la trama actual y las LPCs de la trama previa.

- 30 La técnica de codificación extrae el retraso LTP, el valor RMS, la posición de impulso de tono y el signo de impulso de tono, y descodifica los valores extraídos en valores paramétricos descodificados. La técnica extrae también las decisiones de vocalización de la trama para uso en la creación de un estado de vocalización. El estado de vocalización depende de la decisión de vocalización hecha en la semitrama previa, así como la decisión en las dos semitramas actuales. El estado de vocalización controla las acciones tomadas al construir la excitación.

- 35 La descodificación en este estado hace uso también de la posibilidad de buscar previamente datos primarios. Más específicamente, el módulo descodificador aplica una corrección de errores (EC) a la ganancia LTP y a la ganancia del libro de código algebraico (Alg CB) para la trama actual (lo que comprende promediar y atenuar las ganancias según el estándar GSM 06.61 GSM anteriormente discutido). El módulo descodificador descodifica entonces los parámetros de la siguiente trama cuando el predictor y los historiales han reaccionado a la trama actual. Estos valores se utilizan para predecir el RMS de la siguiente trama. Más específicamente, la técnica realiza la predicción utilizando la ganancia LTP media (es decir, $LTP_{ganancia,media}$), el valor RMS previo (prevRMS) y la energía del vector Alg CB con la ganancia aplicada (es decir, $RMS(AlgCB \text{ Algganancia})$) según la siguiente ecuación:

$$RMS = [LTP_{ganancia,media} \text{ prevRMS}^2 + (RMS(AlgCB \text{ Algganancia}))^2]^{1/2} \quad (\text{Ec. 12}).$$

- 40 En tramas con un estado de vocalización que representa un mensaje vocalizado en estado permanente, el módulo descodificador crea la excitación de una manera diferente a la de los otros estados. A saber, el módulo descodificador crea la excitación de la manera expuesta en el estándar GSM-EFR. El módulo crea el vector LTP interpolando los retrasos LTP entre los valores de los datos redundantes y la trama previa, y copiando el resultado en el historial de excitación. Esto se realiza solamente si la diferencia entre los valores de los datos redundantes y la trama previa está por debajo de un umbral prescrito, por ejemplo menor que el ocho. En caso contrario, el módulo de descodificación utiliza el nuevo retraso en todas las subtramas (a partir de los datos redundantes). El módulo realiza la verificación de umbral para evitar que se interpole un intersticio resultante de que el codificador elija un retardo

LTP de dos periodos de longitud. La técnica aleatoriza el Alg CB para evitar resonancias, y calcula la ganancia de modo que el vector Alg CB tenga una décima parte del valor de ganancia del vector LTP.

El módulo descodificador forma la excitación sumando el vector LTP y el vector Alg CB. El módulo descodificador ajusta entonces la amplitud del vector de excitación con un valor RMS para cada subtrama. Tal ajuste sobre una base de subtrama puede no representar la mejor opción debido a que la distribución de energía de impulso de tono no es uniforme. Por ejemplo, dos partes de alta energía de impulsos de tono en una subtrama recibirán una amplitud más pequeña en comparación con una parte de alta energía en una subtrama. Para evitar este resultado no óptimo, el módulo descodificador puede realizar en lugar de ello un ajuste sobre una base de impulso de tono. La técnica interpola el valor RMS en las tres primeras subtramas entre el valor RMS en la última subtrama en la trama previa y el valor RMS de la trama actual. En la última subtrama de la trama actual, la técnica interpola el valor RMS entre el valor de la trama actual y el valor predicho de la siguiente trama. Esto da como resultado una transición más suave a la siguiente trama.

En tramas con estados de vocalización distintos del estado vocalizado en régimen permanente, el módulo descodificador crea la excitación de una manera específica de GSM-VOC. A saber, en un estado no vocalizado en régimen permanente, la excitación constituye ruido. El módulo descodificador ajusta la amplitud del ruido de modo que las subtramas reciban el RMS correcto. En transiciones a un estado no vocalizado, la técnica de codificación localiza la posición del último impulso de tono correlacionando la síntesis de la trama previa con una forma de impulso. Esto es, la técnica localiza sucesivamente el siguiente máximo de impulso local a partir del máximo de correlación utilizando escalones del tamaño del retraso LTP hasta que encuentre el último máximo posible. La técnica actualiza a continuación el módulo de excitación del vocodificador para comenzar al final del último impulso, en algún lugar en la trama actual. Además, la técnica de codificación copia las muestras ausentes de las posiciones poco antes del inicio del último impulso. Si esta posición no está más allá de la posición en la que comienza el segmento no vocalizado, el módulo descodificador añade uno o más impulsos de vocodificador e interpola los valores RMS hacia el valor de la trama. A partir del final del último impulso vocalizado, el módulo descodificador genera ruido en el límite de la trama. El módulo descodificador interpola también el RMS del ruido de modo que la técnica proporcione una transición suave a una condición no vocalizada.

Si el estado de vocalización representa una transición a un estado vocalizado, la técnica de codificación depende crucialmente de la posición y signo del impulso. La excitación consta de ruido hasta la posición de impulso de tono dada. El módulo descodificador interpola este RMS de ruido hacia el valor recibido (a partir de los datos redundantes). La técnica coloca el impulso del vocodificador en la posición de impulso de tono, con un valor RMS interpolado. Todos los impulsos utilizan el retraso recibido. La técnica forma la interpolación de RMS entre el valor de la última subtrama de la trama previa y el valor recibido en la primera mitad de la trama y entre el valor recibido y el valor predicho en la segunda mitad.

Cuando se calcula el valor RMS para la excitación, el módulo descodificador filtra con síntesis la excitación con los estados de filtro correctos para tener en cuenta la ganancia del filtro. Después del ajuste de la energía, la técnica filtra en paso alto la excitación para eliminar la parte polarizada del impulso del vocodificador. Además, el módulo descodificador introduce la excitación creada en el historial de excitación para proporcionar al LTP algo para trabajar en la siguiente trama. El módulo descodificador aplica entonces el modelo de síntesis una vez final para crear la síntesis. La síntesis a partir de un estado vocalizado en régimen permanente es también filtrada posteriormente.

Estado: EFR Después de Red

En el estado "EFR Después de Red" el módulo descodificador ha recibido los paquetes de las tramas actual y siguiente, aunque el módulo descodificador utiliza solamente datos redundantes para descodificar la trama previa.

En este estado, la técnica utiliza la descodificación GSM-EFR convencional. Sin embargo, el módulo descodificador utiliza parámetros de ganancia que ya se han descodificado. La síntesis creada tiene su amplitud ajustada de modo que el valor RMS de toda la trama corresponda al valor recibido de los datos redundantes. Para evitar discontinuidades en la síntesis que puedan producir ruido de alta frecuencia, el módulo descodificador realiza el ajuste sobre la excitación. El módulo alimenta a continuación la excitación al historial de excitación para su consistencia con la siguiente trama. Además, el módulo restablece el filtro de síntesis al estado que tenía inicialmente en la trama actual, y a continuación utiliza de nuevo el filtro sobre la señal de excitación.

Estado: EFR Red Sig E

En el estado "EFR Red Sig E", el módulo descodificador ha recibido los datos primarios de la trama actual, pero no ha recibido el paquete de la trama siguiente (es decir, se ha perdido la siguiente trama). Además, el módulo descodificador descodifica la trama previa utilizando datos redundantes.

Este estado carece de datos redundantes para uso en la corrección del nivel de energía de la síntesis. En lugar de esto, el módulo descodificador realiza una predicción utilizando la ecuación 12.

Estado: EFR EC

En el estado EFR EC, el módulo descodificador ha dejado de recibir múltiples paquetes en secuencia. En consecuencia, no existen datos primarios ni redundantes para uso en la descodificación de la voz en la trama actual.

5 Este estado intenta remediar la carencia de datos utilizando técnicas de ocultación de error GSM-EFR (por ejemplo, descritas en la sección de Antecedentes). Esto incluye tomar la media de los historiales de ganancia (LTP y Alg CB), atenuar los valores medios y suministrar de nuevo los valores medios al historial. Debido a que los datos se pierden en lugar de distorsionarse por errores de bit, el módulo descodificador no puede utilizar el vector del libro de código algebraico recibido. En consecuencia, el módulo descodificador aleatoriza un nuevo vector de libro de código. Este método se utiliza en el GSM-EFR adaptado para redes basadas en paquetes. Si, por el contrario, el módulo descodificador copió el vector de la última trama, podrían producirse resonancias en la voz. La técnica de codificación calcula el valor RMS y el estado de vocalización de la voz sintetizada como en el estado "EFR sig E.". El uso del tono de la última trama buena puede dar como resultado una deriva de fase grande de las posiciones del impulso en el historial de excitación.

Estado: Red después de EC

15 En el estado "Red después de EC", el módulo descodificador ha recibido el paquete de la trama siguiente que contiene los datos redundantes de la trama actual. El módulo descodificador aplicó una corrección de error a una o más tramas anteriores (y este estado puede distinguirse del estado "Error Único Red" sobre esta base).

20 En este estado, el historial de excitación es muy incierto y no deberá utilizarse. El módulo descodificador crea la excitación en un estado vocalizado en régimen permanente a partir del impulso de tono del vocodificador y el módulo descodificador interpola la energía RMS a partir de: el valor de la trama previa, el valor actual y la predicción para la siguiente trama. El módulo descodificador toma la posición y el signo de los impulsos de los datos recibidos (redundantes) para hacer a la fase del historial de excitación tan precisa como sea posible. El módulo descodificador copia los puntos antes de la posición dada del historial de excitación de una manera que relacionada con el procesamiento del estado vocalizado en régimen permanente del estado "Error Único Red". (Si los datos redundantes carecieran de la información de fase del impulso de tono, la colocación del impulso de tono podría determinarse utilizando la técnica mencionada en primer lugar discutida en la Sección No. 1.4 anterior).

Estado: ERF R+EC Sig E

En el estado "EFR R+EC Sig E", el módulo descodificador deja de recibir el paquete de la trama siguiente. Además, el módulo descodificador descodificó la trama previa con solamente datos redundantes, y la trama antes de ésta con EC.

30 El módulo descodificador descodifica la trama actual con datos primarios. Pero este estado representa el peor estado entre la clase de estados que descodifican datos primarios. Por ejemplo, el predictor LSF actúe probablemente de forma pobre en esta circunstancia (por ejemplo, el predictor está "fuera de línea") y no puede corregirse con los datos disponibles. Por tanto, el módulo descodificador descodifica las GSM-EFR LPCs de la manera estándar y, a continuación, el ancho de banda expande ligeramente las LPCs. Más específicamente, esto se realiza de la manera estándar de la corrección de errores GSM-EFR, pero en una extensión menor para evitar que se cree otro tipo de inestabilidad (por ejemplo, los filtros llegarán a ser inestables utilizando demasiado la media). El módulo descodificador realiza el ajuste de energía de la excitación y la síntesis contra un valor predicho, por ejemplo con referencia a la Ec. 12. Después de esto, el módulo descodificador calcula el RMS y la vocalización para la trama actual a partir de la síntesis.

40 Estado: ERF R+EC

En el estado "ERF R+EC", el módulo descodificador ha recibido el paquete de la trama siguiente, pero descodificó la trama previa con solamente datos redundantes, y la trama antes de ésta con EC.

45 En este estado, el módulo descodificador descodifica en general la trama actual utilizando datos primarios y redundantes. Más específicamente, después de que la EC ha sido aplicada a los coeficientes LP, el predictor pierde su capacidad de proporcionar predicciones precisas. En este estado, el módulo descodificador puede corregirse con los datos redundantes. A saber, el módulo descodificador descodifica los coeficientes LPC redundantes. Estos coeficientes representan el mismo valor que la segunda serie de coeficientes LPC proporcionada por el estándar GSM-EFR. La técnica de codificación utiliza ambos para calcular una estimación del valor del predictor para la trama actual, por ejemplo utilizando las siguientes ecuaciones. (La Ec. 13 es la misma que la Ec. 11 reproducida aquí por motivos de conveniencia).

$$LSF_{prev, res} = (LSF_{red} - LSF_{media} - LSF_{res}) / predFactor. \quad (Ec. 13)$$

$$LSF = LSF_{res} + LSF_{media} + predFactor \quad LSF_{prev, res} \quad (Ec. 14)$$

En el presente enfoque, el modelo de síntesis primaria proporciona información que pertenece a residuales LSF (es decir, LSF_{res}), mientras que el modelo redundante proporciona información que pertenece a valores LSF

redundantes para estos coeficientes (es decir, LSF_{red}). El módulo descodificador utiliza estos valores para calcular el estado del predictor utilizando la Ec. 13 a fin de proporcionar una actualización rápida del predictor. En la Ec. 13, el término LSF_{media} define un valor LSF medio, el término $predFactor$ se refiere a un constante del factor de predicción y $LSF_{prev,res}$ se refiere a una LSF residual de la trama pasada. El módulo descodificador utiliza entonces el estado actualizado del predictor para descodificar los residuales LSF a coeficientes LPC utilizando la Ec. 14 anterior. Esta estimación asegura ventajosamente que los coeficientes LP para la trama actual tengan un error igual al error de cuantificación LPC redundante. En caso contrario, el predictor habría sido correcto en la siguiente trama si se le hubiera actualizado con los residuales LSF de la trama actual.

El estándar GSM-EFR proporciona otro predictor para una ganancia del libro de código algebraica. Los valores de la ganancia GSM-EFR representan una información bastante estocástica. Ningún parámetro redundante disponible casa con tal información, impidiendo la estimación de la ganancia Alg CB. El predictor toma aproximadamente una trama antes de que llegue a ser estable después de una pérdida de trama. El predictor podría actualizarse sobre la base de cambios de energía presentes entre tramas. El módulo codificador podría medir la distribución (por ejemplo, relación) entre la ganancia LTP y la ganancia algebraica y enviarla con muy pocos bits, por ejemplo dos o tres. La técnica para actualizar el predictor deberá considerar también el estado de vocalización. En la transición al estado vocalizado, la ganancia algebraica es frecuentemente demasiado grande para desarrollar un historial para que el LTP lo utilice en tramas posteriores. En el estado permanente, la ganancia es más moderada y, para el estado vocalizado, produce la mayoría de la aleatoriedad encontrada en el estado no vocalizado.

2.4 Variaciones

Se contemplan una pluralidad de variaciones del ejemplo anteriormente descrito. Por ejemplo, la medición de RMS en la última subtrama podría cambiarse para medir la última época de tono completa de modo que solamente se mida un impulso de tono. Con la medición actual sobre la última subtrama, pueden estar presentes cero, una o dos partes de alta energía dependiendo de la posición del impulso y el retraso de tono. Es posible una modificación similar para la distribución de la energía en el estado "Error Único Red" y el estado vocalizado en régimen permanente. En estos casos, puede ajustarse la interpolación de la energía sobre la base de la cantidad de impulsos de tono.

Puede modificarse la búsqueda de posición de impulso en el módulo codificador de modo que éste utilice la decisión de vocalización sobre la base de la anticipación.

Cuando se encuentra en el estado de error "Red después de EC", la técnica puede ajustar la colocación del primer impulso de tono. Este ajuste deberá considerar tanto la posición de impulso recibido como la información de fase en la síntesis de la trama previa. Para minimizar las discontinuidades de fase, la técnica deberá utilizar toda la trama para corregir el error de fase. Esto supone que la síntesis de la trama previa consiste en un mensaje vocalizado.

La interpolación utilizando técnicas polinomiales puede sustituir la interpolación lineal. La técnica deberá casar los valores polinomiales con los siguientes valores: RMS total de la trama previa, RMS para el último impulso de la trama previa, RMS de la trama actual y RMS predicho de la trama siguiente.

La técnica puede emplear una predicción más avanzada de la energía. Por ejemplo, existen suficientes datos para determinar la envolvente de energía para la siguiente trama. La técnica puede modificarse para predecir la energía y su derivada al inicio de la siguiente trama a partir de la envolvente. La técnica puede utilizar esta información para mejorar la interpolación de la energía a fin de proporcionar un límite de trama aún más suave. En el caso de que la técnica proporcione una predicción ligeramente imprecisa, la técnica puede ajustar el nivel de energía en la siguiente trama. Para evitar discontinuidades, la técnica puede utilizar alguna clase de ajuste irregular. Por ejemplo, la técnica puede establecer el ajuste de ganancia en casi cero al comienzo de una trama y aumentar el ajuste al valor requerido hacia el centro de la trama.

Para reducir la cantidad de datos redundantes (sobrecarga) transmitidos por la red, la técnica de codificación puede descartar algunos parámetros. Más específicamente, la técnica puede descartar diferentes parámetros dependiendo del estado de vocalización.

Por ejemplo, la Tabla 2 identifica parámetros apropiados para un mensaje no vocalizado. La técnica requiere las LPCs para conformar las propiedades espectrales del ruido. La técnica necesita el valor RMS para transportar la energía del ruido. La tabla consigna el estado de vocalización, pero este parámetro puede descartarse. En su lugar, la técnica puede utilizar el tamaño de datos como indicador de un mensaje no vocalizado. Esto es, sin el estado de vocalización, el parámetro establecido en la Tabla 2 proporciona un tamaño de trama de 33 bits y una tasa de bits de 1650 b/s. Este tamaño de datos (33 bits) puede utilizarse como indicador de un mensaje no vocalizado (en el caso de que la técnica de paquetización especifique esta información de tamaño, por ejemplo en la cabecera de los paquetes). Adicionalmente, la técnica de codificación puede no requerir valores precisos para uso en la conformación espectral del ruido (en comparación con segmentos vocalizados). A la vista de ello, la técnica puede utilizar un tipo menos preciso de cuantificación para reducir adicionalmente el ancho de banda. Sin embargo, tal modificación puede afectar a la efectividad de la operación de actualización del predictor para el descodificador LPC primario.

Tabla 2

Parámetro	Número de Bits
LPC	26
Valor RMS	7
Estado de Vocalización	2
Total (Ancho de banda)	35 (1750 b/s)

5 En las transiciones de un mensaje vocalizado a otro no vocalizado, la técnica requiere todos los parámetros de la Tabla 1 (anterior). Esto es debido a que los parámetros LPC cambian típicamente de manera drástica en esta circunstancia. El mensaje vocalizado incluye un tono y existe un nuevo nivel de energía en la trama. La técnica utiliza así el impulso y el signo de tono para generar una fase correcta para la excitación.

10 En el estado vocalizado en régimen permanente y en las transiciones al estado no vocalizado, la técnica puede eliminar la posición y el signo del impulso de tono, reduciendo así la cantidad total de bits a 42 bits (es decir, 2100 b/s). En consecuencia, el módulo descodificador no recibe ninguna información de fase en estas tramas, lo que puede tener un impacto negativo en la calidad de su salida. Esto forzará al descodificador a buscar la fase en la trama previa, lo que, a su vez, puede dar como resultado errores de fase mayores, puesto que el algoritmo no puede detectar la fase debido a la pérdida de una ráfaga de paquetes. Esto hace también imposible corregir cualquier deriva de fase que haya ocurrido durante un periodo de ocultación de error.

15 En lugar del GSM-VOC anteriormente descrito, el descodificador redundante descrito anteriormente puede utilizar codificación multiimpulsos. En la descodificación multiimpulsos, la técnica de codificación codifica los impulsos más importantes a partir del residual. Esta solución reaccionará mejor a cambios en las transiciones de los estados no vocalizados a los vocalizados. Además, no surgirá ninguna complicación de fase cuando se combine esta técnica de codificación con GSM-EFR. Por otro lado, esta técnica utiliza un ancho de banda mayor que el GSM-VOC anteriormente descrito.

20 El ejemplo descrito anteriormente proporciona un único nivel de redundancia. Sin embargo, la técnica puede utilizar múltiples niveles de redundancia. Además, el ejemplo antes descrito combina preferiblemente los datos primarios y redundantes en el mismo paquete. Sin embargo, la técnica puede transferir los datos primarios y redundantes en paquetes independientes u otros formatos alternativos.

Otras variaciones de los principios anteriormente descritos serán evidentes a los expertos en la materia. El alcance de la presente invención se define por las siguientes reivindicaciones.

25

REIVINDICACIONES

1. Un módulo descodificador (704) para descodificar datos de audio formateados en paquetes que contienen datos codificados primarios para una trama actual y datos codificados redundantes para una trama previa, que comprende:
- 5 un descodificador primario (712) para descodificar los datos codificados primarios de los paquetes utilizando un modelo de síntesis primaria;
- un descodificador redundante (714) para descodificar los datos codificados redundantes de los paquetes utilizando un modelo de síntesis redundante, cuando los datos codificados primarios no están disponibles, y
- 10 medios para localizar una posición de impulso de tono en la trama actual localizando la última posición de impulso conocida en la trama previa, y avanzando entonces desde la última posición de impulso conocida en uno o más valores de retraso de tono para localizar la posición del impulso en la trama actual, utilizando los datos codificados redundantes al dejar de recibir los datos codificados primarios, en donde la posición del impulso de tono localizada en la trama actual se utiliza para reducir las discontinuidades de fase.
2. Un módulo descodificador para descodificar datos de audio según la reivindicación 1, en el que los medios para localizar la posición del impulso de tono están configurados además para recibir un valor de posición de impulso de
- 15 tono desde un sitio de codificación, comparar el valor recibido con la posición localizada del impulso de tono, y suavizar a continuación cualquier discrepancia de fase detectada en el curso de la trama actual.
3. Un método para descodificar datos de audio formateados en paquetes que contienen datos codificados primarios para una trama actual y datos codificados redundantes para una trama previa, que comprende:
- 20 descodificar primariamente los datos codificados primarios de los paquetes utilizando un modelo de síntesis primaria; descodificar redundantemente los datos codificados redundantes de los paquetes utilizando un modelo de síntesis redundante, cuando los datos codificados primarios no están disponibles, en donde la descodificación primaria o la descodificación redundante comprende el paso de localizar una posición de impulso de tono en la trama actual localizando la última posición de impulso conocida en la trama previa, y, a continuación, avanzar desde
- 25 la última posición de impulso conocida en uno o más valores de retraso de tono para localizar la posición del impulso en la trama actual, utilizando los datos codificados redundantes al dejar de recibir los datos codificados primarios, y en donde la posición localizada del impulso de tono se utiliza para reducir las discontinuidades de fase.
4. Un método para descodificar datos de audio según la reivindicación 3, en el que el paso de localizar la posición del impulso de tono comprende además recibir un valor de posición del impulso de tono desde un sitio de
- 30 codificación, comparar el valor recibido con la posición localizada del impulso de tono, y suavizar entonces cualquier discrepancia de fase detectada en el curso de la trama actual.

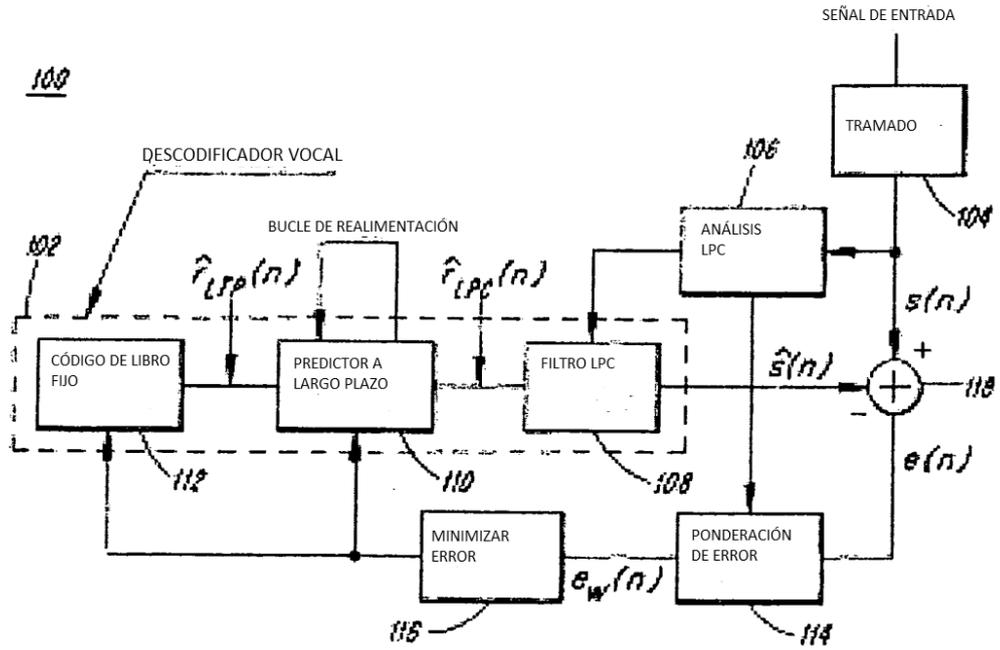


Fig. 1
(TÉCNICA ANTERIOR)

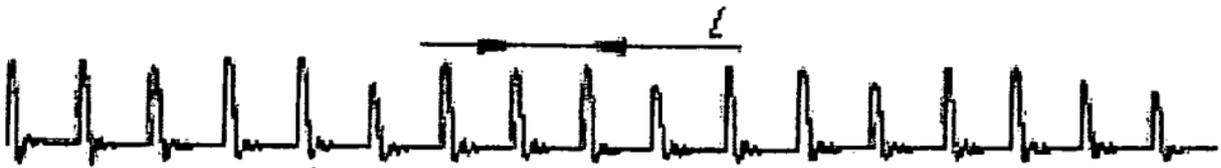


Fig. 2
(TÉCNICA ANTERIOR)

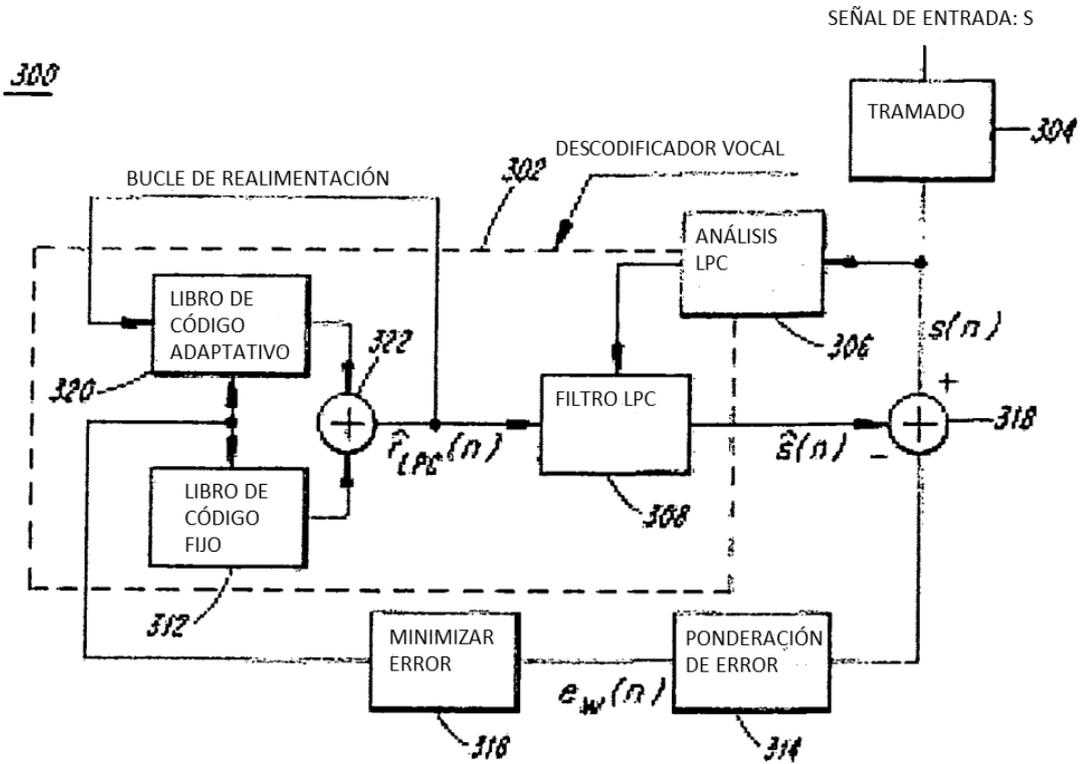


Fig. 3
(TÉCNICA ANTERIOR)

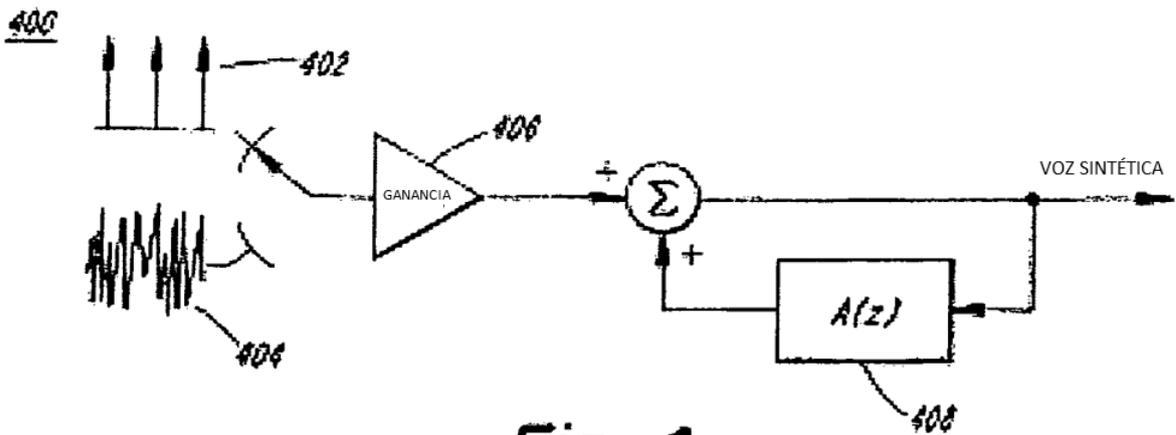
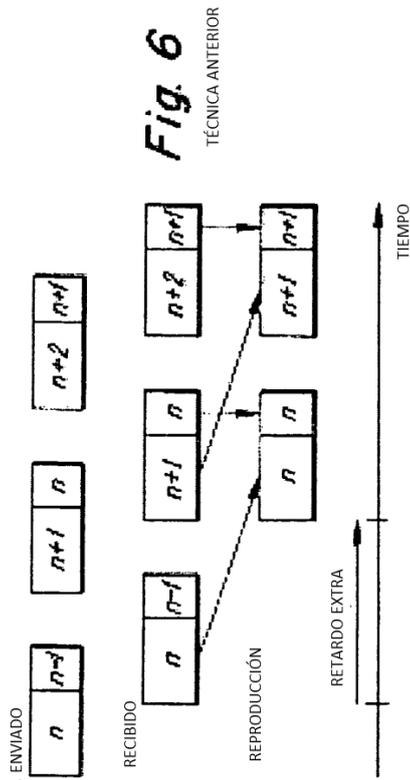
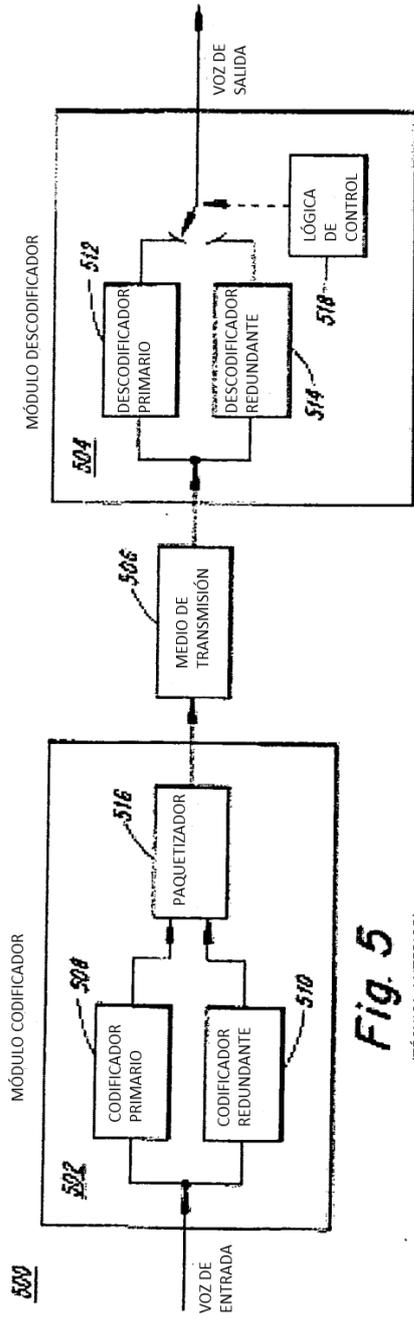


Fig. 4
(TÉCNICA ANTERIOR)



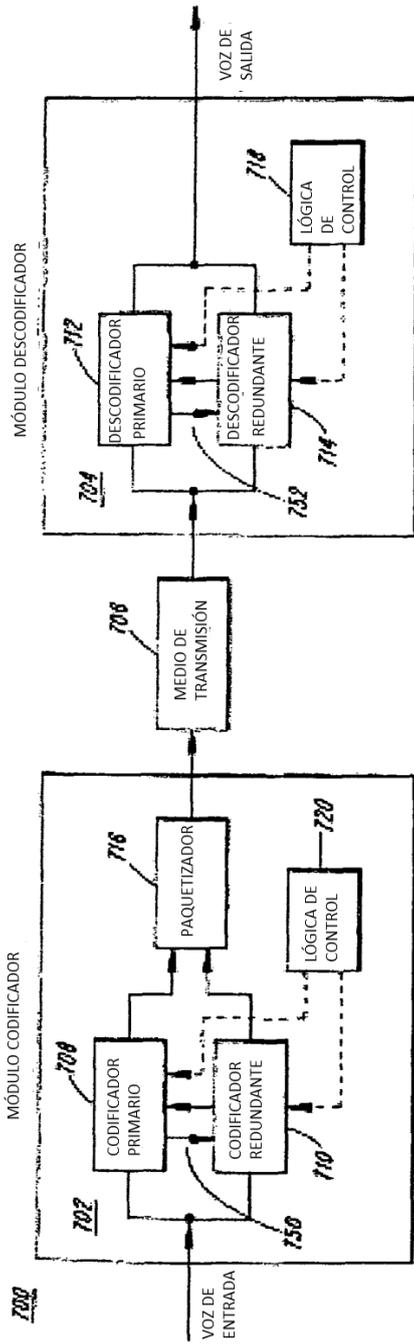


Fig. 7

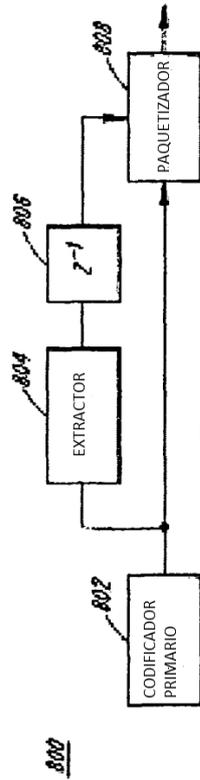


Fig. 8

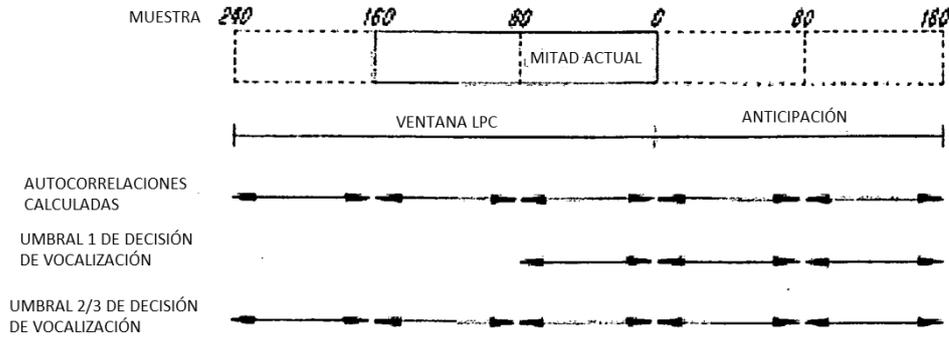


Fig. 9

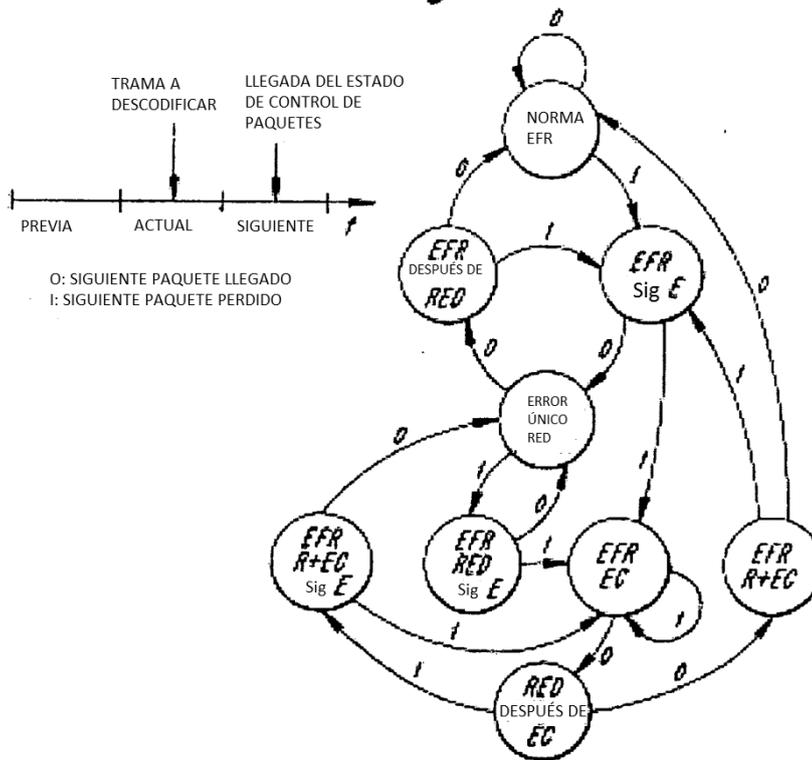


Fig. 10