

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 743 197**

51 Int. Cl.:

**G10L 19/005** (2013.01)

G10L 19/20 (2013.01)

G10L 25/93 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.04.2015 PCT/FR2015/051127**

87 Fecha y número de publicación internacional: **05.11.2015 WO15166175**

96 Fecha de presentación y número de la solicitud europea: **24.04.2015 E 15725801 (3)**

97 Fecha y número de publicación de la concesión europea: **05.06.2019 EP 3138095**

54 Título: **Corrección de pérdida de trama perfeccionada con información de sonoridad**

30 Prioridad:

**30.04.2014 FR 1453912**

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

**18.02.2020**

73 Titular/es:

**ORANGE (100.0%)  
78, rue Olivier de Serres  
75015 Paris, FR**

72 Inventor/es:

**FAURE, JULIEN y  
RAGOT, STÉPHANE**

74 Agente/Representante:

**ISERN JARA, Jorge**

ES 2 743 197 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 de pérdida de trama perfeccionada con información de sonoridad

5 La presente invención se refiere al campo de la codificación/decodificación de telecomunicaciones y, más particularmente, al de la corrección de la pérdida de trama en la decodificación.

10 Se entiende por "trama" un segmento de audio compuesto por al menos una muestra (aunque la invención se aplica tanto a la pérdida de una o varias muestras en la codificación según la norma G.711 como a una pérdida de uno o varios paquetes de muestras en codificación según las normas G.723, G.729, etc.).

15 Las pérdidas de tramas de audio se dan cuando una comunicación en tiempo real que utiliza un codificador y un decodificador es perturbada por las condiciones de una red de telecomunicación (problemas de radiofrecuencia, congestión de la red de acceso, etc.). En este caso, el decodificador utiliza los mecanismos de corrección de pérdida de tramas para intentar sustituir la señal faltante por una señal reconstruida utilizando informaciones disponibles en el decodificador (por ejemplo la señal de audio ya decodificada para una o varias tramas pasadas). Esta técnica puede mantener una calidad de servicio a pesar de rendimientos de la red degradados.

20 Las técnicas de corrección de pérdida de tramas son muy frecuentemente dependientes del tipo de codificación utilizado.

25 En el caso de una codificación CELP, es normal repetir ciertos parámetros decodificados en la trama precedente (envolvente espectral, tono, ganancias de diccionarios), con unos ajustes como una modificación de la envolvente espectral para converger hacia una envolvente media o la utilización de un diccionario fijo aleatorio.

30 La técnica más empleada para corregir la pérdida de trama en el caso de una codificación por transformada, consiste en repetir la última trama recibida si se pierde una trama y poner la trama repetida a cero a partir de que se pierda más de una trama. Esta técnica se encuentra en varias codificaciones normalizadas (G.719, G.722.1, G.722.1C). Se puede citar también el caso de la codificación normalizada G.711, para la que un ejemplo de corrección de pérdida de trama descrito en el apéndice I de G.711 consiste en identificar un periodo fundamental (llamado "tono") en la señal decodificada y en repetirla teniendo cuidado de hacer una adición con recubrimiento (llamada "overlap-add") entre la señal ya decodificada y la señal repetida. Esta adición con recubrimiento permite "borrar" las artificiosidades de audio pero necesita, para implementarse, un retardo suplementario en el decodificador (correspondiente a la duración del recubrimiento).

35 Por otra parte, en el caso de la codificación normalizada G.722.1, una transformada modulada con superposición (o MLT por "Modulated Lapped Transform"), con una adición con recubrimiento del 50 % y unas ventanas sinusoidales que permiten asegurar una transición entre la última trama perdida y la trama repetida que sea suficientemente lenta para borrar las artificiosidades vinculadas a la simple repetición de la trama en el caso de una única trama perdida. Contrariamente a la corrección de la pérdida de trama descrita en la norma G.711 (Apéndice I), esta realización no necesita retardo suplementario puesto que aprovecha el retardo existente y el repliegue temporal de la transformada MLT para hacer una adición con recubrimiento con la señal reconstituida.

45 Esta técnica es muy poco costosa pero tiene como principal defecto una incoherencia entre la señal decodificada antes de la pérdida de trama y la señal repetida. De donde resulta una discontinuidad de fase que puede producir artificiosidades de audio importantes si la duración del recubrimiento entre las dos tramas es reducida, tal como es el caso cuando las ventanas utilizadas para la transformada MLT son "de reducido retardo" como se describe en el documento FR 1350845 con referencia a las figuras 1A y 1B de ese documento. En este caso, incluso una solución que combinara una búsqueda de tono, en el caso del codificador según la norma G.711 (Apéndice I) y una adición con recubrimiento según la ventana de la transformada MLT no es suficiente para suprimir las artificiosidades de audio.

50 El documento FR 1350845 propone un método híbrido que combina las ventajas de los dos métodos permitiendo mantener la continuidad de fase en el dominio transformado. La presente invención se inscribe en este marco. Una descripción detallada de la solución objeto de este documento FR 1350845 se describe más adelante con referencia a la figura 1.

55 Esta solución, incluso aunque es particularmente prometedora, queda por perfeccionar porque, cuando la señal codificada no incluye más que un periodo fundamental ("mono tono") como por ejemplo un segmento sonoro de una señal de voz, la calidad de audio después de la corrección de trama perdida puede estar degradada y peor que con una corrección de pérdida de trama mediante un modelo de voz de tipo CELP por ejemplo (por "Code-Excited Linear Prediction"). Otra solución se describe en PARIKH V. N. ET AL., "Frame erasure concealment using sinusoidal analysis synthesis and its application to MDCT-based codecs", ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 2000. ICASSP '00. PROCEEDINGS. 2000 IEEE INTERNATIONAL CONFERENCE EL 5-9 DE JUNIO DE 2000, PISCATAWAY, Nueva Jersey, Estados Unidos, IEEE, vol. 2, 5 de junio de 2000, páginas 905-908.

65 La invención se dirige a la mejora de la situación.

Propone con este fin un procedimiento de tratamiento de la señal de audio digital tal como en la reivindicación 1 incluyendo una sucesión de muestras repartidas en tramas sucesivas, implementándose el procedimiento durante una decodificación de dicha señal para sustituir al menos una trama de señal perdida en la decodificación.

5 De este modo, la invención se dirige a mejorar el estado de la técnica en el sentido del documento FR 1350845 modificando diferentes etapas del tratamiento presentado en ese documento (búsqueda de tono, selección de los componentes, inyección de ruido) pero no obstante función en particular de las características de la señal original.

10 Estas características de la señal original pueden codificarse como una información particular en el flujo de datos hacia el decodificador (o "bitstream") en función de la clasificación de la voz y/o de la música y si es necesario de la clase de voz en particular.

15 Esta información en el flujo de la decodificación permite optimizar el compromiso entre complejidad y calidad y, conjuntamente:

- modificar la ganancia del ruido a inyectar en la suma de los componentes espectrales seleccionados para construir la señal de síntesis que reemplaza la trama perdida,
- modificar el número de componentes seleccionados para la síntesis,
- 20 - modificar la duración del segmento de búsqueda del tono.

Una realización de ese tipo puede implementarse en un codificador para la determinación de la información de sonoridad, y más particularmente en un decodificador, particularmente en el caso de pérdida de trama. Puede implementarse en la forma de software en una realización de una codificación/decodificación para los servicios de voz enriquecidos (o "EVS" por "Enhanced Voice Services") especificado por el grupo 3GPP (SA4).

25 Con este fin la presente invención se dirige también a un programa informático tal como en la reivindicación 13. Un ejemplo de ordinograma de un programa de ese tipo se presenta en la descripción detallada a continuación con referencia a la figura 4 para la decodificación en el sentido de la invención y con referencia a la figura 3 para la codificación útil para la invención.

30 La presente invención se dirige también a un dispositivo de decodificación de una señal de audio digital tal como en la reivindicación 14 que incluye una sucesión de muestras repartidas en tramas sucesivas.

35 Aparecerán otras características y ventajas de la invención con el examen de la descripción detallada que sigue y de los dibujos adjuntos en los que:

- la figura 1 recuerda las principales etapas del procedimiento de corrección de pérdida de trama en el sentido del documento FR 1350845;
- 40 - la figura 2 ilustra esquemáticamente las principales etapas de un procedimiento útil para la invención;
- la figura 3 ilustra un ejemplo de etapas implementadas en la codificación, en una forma de realización útil para la invención;
- la figura 4 ilustra un ejemplo de etapas implementadas en la decodificación, en una forma de realización en el sentido de la invención;
- 45 - la figura 5 ilustra un ejemplo de etapas implementadas en la decodificación, para la búsqueda de tono en un segmento de señal válida  $N_c$ ;
- la figura 6 ilustra esquemáticamente un ejemplo de dispositivo codificador útil para la invención y decodificador en el sentido de la invención.

50 Se hace referencia a la figura 1 en la que se han ilustrado las principales etapas descritas en el documento FR 1350845. En la primera etapa S1, se memoriza en una memoria tampón del decodificador (o "buffer") una sucesión de N muestras de audio, indicada por  $b(n)$  en lo que sigue. Estas muestras corresponden a unas muestras ya decodificadas y están por tanto accesibles para la corrección de pérdida de trama en el decodificador. Si la primera muestra a sintetizar es la muestra N, la memoria tampón de audio corresponde a las muestras 0 a N-1 precedentes.

55 En el caso de una codificación por transformada, la memoria tampón de audio corresponde a las muestras de la trama precedente y no son modificables porque este tipo de codificación/decodificación no prevé retardo en la restitución de la señal, de manera que no está previsto realizar un fundido encadenado de duración suficiente para cubrir una pérdida de trama.

60 A continuación, se procede a una etapa de filtrado en frecuencia S2, en el curso de la que la memoria tampón de audio  $b(n)$  se separa en dos bandas, una banda baja BB y una banda alta BA, con una frecuencia de separación indicada por  $F_c$  (por ejemplo  $F_c=4$  kHz). Este filtrado es de manera preferente un filtrado sin retardo. El tamaño de la memoria tampón de audio se reduce ahora a  $N' = N \cdot F_c / f_s$  siguiendo el diezmado de  $f_s$  en  $F_c$ . En variantes de la invención, esta etapa de filtrado puede ser opcional, realizándose las etapas siguientes en banda completa.

65 La etapa siguiente S3 consiste en buscar en la banda baja un punto de bucle y un segmento  $p(n)$  correspondiente al

periodo fundamental (o "tono" en lo que sigue) en el seno de la memoria tampón  $b(n)$  remuestreada a la frecuencia  $f_c$ . Esta realización permite tener en cuenta la continuidad del tono en la o las trama(s) perdida(s) a reconstruir.

5 La etapa S4 consiste en descomponer el segmento  $p(n)$  en una suma de componentes sinusoidales. Por ejemplo, se puede calcular la transformada de Fourier discreta (DFT) de la señal  $p(n)$  en una duración correspondiente a la longitud de la señal. Se obtiene así la frecuencia, la fase y la amplitud de cada una de las componentes sinusoidales (o "pics") que componen la señal. Son posibles otras transformadas distintas a la DFT. Por ejemplo, pueden implementarse unas transformadas de tipo DCT, MDCT o MCLT.

10 La etapa S5 es una etapa de selección de  $K$  componentes sinusoidales de manera que mantenga únicamente los componentes mayores. En un modo de realización particular, la selección de los componentes corresponde primariamente a seleccionar las amplitudes  $A(n)$  para las que  $A(n) > A(n-1)$  y  $A(n) > A(n+1)$  siendo  $n \in \left[0; \frac{P'}{2}\right] - 1$ , lo que asegura que las amplitudes corresponden a unos pics espectrales.

15 Para hacer esto, las muestras del segmento  $p(n)$  se interpolan de manera que se obtenga un segmento  $p'(n)$  compuesto de  $P'$  muestras siendo  $P' = 2^{\lceil \log_2(P) \rceil} > P$ , en la que  $\lceil x \rceil$  es el entero superior o igual a  $x$ . El análisis por transformada de Fourier FFT se realiza por tanto de manera más eficaz sobre una longitud que sea una potencia de 2, sin modificar el periodo de tono efectivo (debido a la interpolación). Se calcula la transformada FFT de  $p'(n)$ :  $\Pi(k) = \text{FFT}(p'(n))$ ; y, a partir de la transformada FFT, se obtienen directamente las fases  $\varphi(k)$  y amplitudes  $A(k)$  de las componentes sinusoidales, viniendo dadas en este caso las frecuencias normalizadas entre 0 y 1 por:

$$f(k) = \frac{2kP'}{P^2} \quad k \in \left[0; \frac{P'}{2} - 1\right]$$

25 A continuación, entre las amplitudes de esta primera selección, se seleccionan las componentes por orden decreciente de amplitud, de manera que la amplitud acumulada de los pics seleccionados sea de al menos el  $x\%$  (por ejemplo  $x=70\%$ ) de la amplitud acumulada sobre típicamente la mitad del espectro en la trama actual.

Es también posible además, limitar el número de componentes (por ejemplo 20) de manera que se haga la síntesis menos compleja.

30 La etapa S6 de síntesis sinusoidal consiste en generar un segmento  $s(n)$  de longitud al menos igual al tamaño de la trama perdida ( $T$ ). La señal de síntesis  $s(n)$  se calcula como una suma de las componentes sinusoidales seleccionadas:

$$s(n) = \sum_{k=0}^{K-1} A(k) \sin(\pi f(k)n + \varphi(k)) \quad n \in \left[0; 2T + \frac{LF}{2}\right]$$

35 en la que  $k$  es el índice de los  $K$  pics seleccionados de la etapa S5.

40 La etapa S7 consiste en "inyectar ruido" (llenar las zonas espectrales correspondientes a las rayas no seleccionadas) de manera que se compense la pérdida de energía vinculada a la omisión de ciertos pics de frecuencia en la banda baja. Un modo de realización particular consiste en calcular el residuo  $r(n)$  entre el segmento correspondiente al tono  $p(n)$  y la señal sintetizada

$$s(n), \text{ siendo } n \in [0; P-1], \text{ tal que:} \\ r(n) = p(n) - s(n) \quad n \in [0; P-1]$$

45 Este residuo de tamaño  $P$  se transforma, por ejemplo en ventana y se repite haciendo unos recubrimientos entre unas ventanas de tamaños variables, como se describe en el documento FR 1353551:

$$r'(k) = f(r(n)) \quad n \in [0; P-1] \quad y \quad k \in \left[0; 2T + \frac{LF}{2}\right]$$

la señal  $s(n)$  se combina a continuación en la señal  $r'(n)$ :

$$s(n) = s(n) + r'(n) \quad n \in \left[0; 2T + \frac{LF}{2}\right]$$

50 La etapa S8 aplicada sobre la banda alta puede consistir simplemente en repetir la señal pasada.

55 En una etapa S9, la señal se sintetiza remuestreando la banda baja a su frecuencia  $f_c$  de origen, después de haberse mezclado en la etapa S8 a la banda alta filtrada (simplemente repetida en la etapa S11).

La etapa S10 es una adición con recubrimiento que permite asegurar la continuidad entre la señal anterior a la pérdida de trama y la señal sintetizada.

60 Se describen ahora los elementos añadidos al procedimiento de la figura 1, en una realización útil para la invención.

Según un planteamiento general presentado en la figura 2, una información de sonoridad de la señal antes de la pérdida de trama, transmitida a al menos una velocidad del codificador, se utiliza en la decodificación (etapa DI-1) para

determinar cuantitativamente una proporción de ruido a añadir a la señal de síntesis que sustituye una o varias tramas perdidas. De este modo, el decodificador utiliza la información de sonoridad, para disminuir, en función de la sonoridad, la cantidad general de ruido mezclado a la señal de síntesis (asignando una ganancia  $G(\text{res})$  más reducida a la señal de ruido  $r'(k)$  procedente de un residuo en la etapa DI-3 y/o seleccionando por adelantado componentes de amplitudes  $A(k)$  a utilizar para la construcción de la señal de síntesis en la etapa DI-4).

El decodificador puede ajustar además sus parámetros, particularmente la búsqueda de tono, para optimizar el compromiso calidad/complejidad del tratamiento, en función de la información de sonoridad. Por ejemplo, para la búsqueda de tono, si la señal es sonora, la ventana de búsqueda de tono  $N_c$  puede ser mayor (en la etapa DI-5), como se verá más adelante con referencia a la figura 5.

Para la determinación de la sonoridad, puede suministrarse por el codificador una información, de dos maneras, a al menos una velocidad del codificador:

- en la forma de un bit de valor 1 o 0 según un grado de sonoridad identificado en el codificador (recibido del codificador en la etapa DI-1 y leído en la etapa DI-2 en caso de pérdida de trama para el tratamiento posterior), o
- en la forma de un valor de amplitud media de los pics que componen la señal en la codificación, comparada con un ruido de fondo.

Este dato de "planitud"  $PI$  del espectro puede recibirse sobre varios bits en el decodificador en la etapa opcional DI-10 de la figura 2, y posteriormente compararse con un umbral en la etapa DI-11, lo que se convierte en determinar en las etapas DI-1 y DI-2 si la sonoridad es superior o inferior a un umbral y deducir los tratamientos adecuados, particularmente para la selección de pics y para la elección de duración del segmento de búsqueda de tono.

Esta información (tanto si está bajo la forma de un bit único o de un valor sobre varios bits) se recibe del codificador (al menos a una velocidad de la codificación), en el ejemplo descrito en el presente documento.

En efecto, con referencia a la figura 3, en el codificador, la señal de entrada presentada en la forma de tramas  $C1$  se analiza en la etapa  $C2$ . La etapa de análisis consiste en determinar si la señal de audio de la trama actual presenta características que necesitarían un tratamiento particular en caso de pérdida de tramas en el decodificador, como es el caso por ejemplo de las señales de voz sonoras.

En un modo de realización particular, se utiliza ventajosamente una clasificación (voz/música u otras) ya efectuada en el codificador de manera que no se aumente la complejidad global del tratamiento. En efecto, en el caso de codificadores de conmutación de modos de codificación entre la voz o la música, una clasificación en el codificador permite ya adaptar la técnica empleada para la codificación en función de la naturaleza de la señal (voz o música). Asimismo, en el caso de la voz, unos codificadores de tipo predictivo como por ejemplo el codificador según la norma G.718 utiliza también una clasificación de manera que se adapten los parámetros del codificador a la naturaleza de la señal (sonidos sonoros / no sonoros, transitorios, genéricos, inactivos).

En un primer modo particular de realización, no se reserva más que un único bit de "caracterización para la pérdida de trama". Se añade al flujo codificado (o "bitstream") en la etapa  $C3$  para indicar si la señal es una señal de voz (sonora o genérica). Este bit se pone por ejemplo a 1 o a 0 según el caso de la tabla a continuación:

- de la decisión del clasificador voz/música,
- y además de la decisión del clasificador sobre el modo de codificación de la voz.

Decisión del clasificador del codificador	Voz	Música
Valor del bit de caracterización para la pérdida de trama	Decisión del clasificador	
	Modo de codificación:	
	Sonora	1
	No sonora	0
	Transitoria	0
	Genérica	1
Inactiva	0	

Se entiende en este caso por "genérica" una señal de la voz habitual (que no es un transitorio vinculado a la pronunciación de una plosiva, que no es inactivo y que no es necesariamente puramente sonora como la pronunciación de una vocal sin consonante).

- 5 En un segundo modo de realización, alternativo, La información transmitida al decodificador en el flujo codificado no es binaria sino que corresponde a una cuantificación de la relación entre los niveles de picos y los niveles de valles en el espectro. Esta relación puede expresarse por una medida de "plenitud" del espectro, indicada por PI:

$$Pl = \log_2 \left( \frac{\exp\left(\frac{1}{N} \sum_{k=0}^{N-1} \ln(x(k))\right)}{\frac{1}{N} \sum_{k=0}^{N-1} x(k)} \right)$$

- 10 En esta expresión, x(k) es el espectro de amplitud de tamaño N resultante del análisis de la trama actual en el dominio de la frecuencia (después de la FFT).

- 15 En una alternativa, está disponible un análisis sinusoidal que descompone la señal en el codificador en componentes sinusoidales y ruido y la medida de la plenitud se obtiene por relación entre los componentes sinusoidales y la energía global en la trama.

- 20 A continuación en la etapa C3 (que incluye la información de sonoridad en un único bit o por la medida de plenitud sobre varios bits), la memoria tampón de audio del codificador se codifica clásicamente en una etapa C4 antes de la transmisión posterior eventual al decodificador.

Se hace referencia ahora a la figura 4 para describir las etapas implementadas en el decodificador, en un ejemplo de realización de la invención.

- 25 En el caso en el que no hay pérdidas de trama en la etapa D1 (flecha KO en la salida de la prueba D1 de la figura 4), el decodificador lee las informaciones contenidas en el flujo codificado, comprendidas las informaciones de "caracterización para la pérdida de trama" en la etapa D2 (al menos a una velocidad de la codificación). Estas últimas se almacenan en la memoria de manera que se reutilicen en el caso de que faltara una trama siguiente. El decodificador continúa entonces las etapas clásicas de la decodificación D3, etc. de manera que obtenga la trama de salida sintetizada FR SYNTH.

- 30 En el caso en el que interviene una pérdida de trama(s) (flecha OK en la salida de la prueba D1), se aplican las etapas D4, D5, D6, D7, D8 y D12, correspondientes respectivamente a las etapas S2, S3, S4, S5, S6 y S11 de la figura 1. No obstante, se realizan algunas modificaciones con relación a las etapas S3 y S5, respectivamente a las etapas D5 (búsqueda de un punto de bucle para la determinación del tono) y D7 (elección de las componentes sinusoidales). Por otra parte, la inyección de ruido en la etapa S7 de la figura 1 se realiza con una determinación de ganancia según dos etapas D9 y D10 en la figura 4 del decodificador en el sentido de la invención.

- 40 En efecto, en el caso en el que la información de "caracterización para la pérdida de trama" es conocida (cuando se ha recibido la trama precedente), la invención consiste en modificar el tratamiento de las etapas D5, D7 y D9-D10, como sigue.

En un primer ejemplo de realización, la información de "caracterización para la pérdida de trama" es binaria y de valor:

- 45 - igual a 0 en el caso de la señal no sonora, de tipo música, de tipo transitoria,  
- igual a 1 si no (tabla anterior).

- 50 La etapa D5 consiste en buscar un punto de bucle y un segmento p(n) correspondiente al tono en el seno de la memoria tampón de audio remuestreada a la frecuencia Fc. Esta técnica, descrita en el documento FR 1350845, se ilustra en la figura 5, en la que:

- 55 - la memoria tampón de audio en el decodificador es de tamaño de muestras N',  
- se determina el tamaño de la memoria tampón objetivo BC de Ns muestras,  
- la búsqueda de correlación se efectúa sobre Nc muestras,  
- la curva de correlación "Correl" presenta un máximo en mc,  
- el punto de bucle se designa Pt Bucl y se sitúa a Ns muestras del máximo de correlación,  
- el tono se determina entonces sobre las p(n) muestras restantes a N'-1.

- 60 Se calcula en particular una correlación normalizada corr(n) entre el segmento de memoria tampón objetivo de tamaño Ns comprendido entre N'-Ns y N'-1 (de una duración por ejemplo de 6 ms) y el segmento deslizante de tamaño Ns que comienza entre la muestra 0 y Nc (siendo Nc>N'-Ns):

$$Corr(n) = \frac{\sum_{k=0}^{k=N_s} b(n+k)b(N' - N_s + k)}{\sqrt{\sum_{k=0}^{k=N_s} b(n+k)^2 \sum_{k=0}^{k=N_s} b(N' - N_s + k)^2}} \quad n \in [0; N_c]$$

Para señales de música, debido a la naturaleza de la señal, el valor  $N_c$  no tienen necesidad de ser demasiado grande (por ejemplo  $N_c=28$  ms). Esta limitación permite economizar la complejidad del cálculo durante la búsqueda de tono.

5 Por el contrario, la información de sonoridad de la última trama válidamente recibida anteriormente permite determinar si la señal que se busca reconstruir es una señal de voz sonora (mono tono). Es posible por tanto, en este caso y gracias a esta información, aumentar el tamaño del segmento  $N_c$  (por ejemplo  $N_c=33$  ms) de manera que se optimice la búsqueda de tono (y potencialmente encontrar un valor de correlación más elevado).

10 Por otra parte, en la etapa D7 de la figura 4, se seleccionan unas componentes sinusoidales de manera que se mantengan únicamente las componentes mayores. En un modo de realización particular presentado también en el documento FR 1350845, la primera selección de componentes se convierte en seleccionar las amplitudes  $A(n)$  para las que  $A(n) > A(n-1)$  y  $A(n) > A(n+1)$  siendo

15  $n \in [0; \frac{P'}{2} - 1]$

En el caso de la invención, se sabe ventajosamente si la señal que se busca reconstruir es una señal de voz (sonora o genérica) por tanto con unos pics marcados y un reducido nivel de ruido. En estas condiciones, es preferible seleccionar no solamente los pics  $A(n)$  para los que  $A(n) > A(n-1)$  y  $A(n) > A(n+1)$  como se ha presentado anteriormente, sino también ampliar la selección  $A(n-1)$  y  $A(n+1)$  de manera que los pics seleccionados representen una gran parte de la energía total del espectro. Esta modificación permite particularmente reducir el nivel de ruido (y particularmente el nivel de ruido inyectado en las etapas D9 y D10 presentadas a continuación) con relación al nivel de la señal sintetizada por síntesis sinusoidal en la etapa D8, mientras se conserva un nivel global de energía suficiente para no provocar artificiosidades audibles vinculadas a fluctuaciones de energía.

25 A continuación, en el caso en que la señal está exenta de ruido (al menos en las bajas frecuencias), tal como es el caso en una señal de voz sonora o genérica, se observa que la adición de ruido correspondiente al residuo transformado  $r'(n)$  en el sentido del documento FR 1350845, degrada de hecho la calidad.

30 De este modo, se utiliza en este caso ventajosamente la información de sonoridad para atenuar el ruido aplicándole una ganancia  $G$  en la etapa D10. La señal  $s(n)$  procedente de la etapa D8 se mezcla a la señal de ruido  $r'(n)$  procedente de la etapa D9 aplicando sin embargo en este caso una ganancia  $G$  que depende de la información de "caracterización para la pérdida de trama" procedente del flujo codificado de la trama precedente, es decir:

35  $s(n) = s(n) + G x r'(n) \quad n \in \left[0; 2T + \frac{LF}{2}\right]$

En este modo de realización particular,  $G$  puede ser una constante igual a 1 o 0,25 en función de la naturaleza sonora o no sonora de la señal de la trama precedente, según la tabla dada a continuación a título de ejemplo:

Valor del bit de "caracterización para la pérdida de trama"	0	1
Ganancia $G$	1	0,25

40 En el modo de realización alternativo en el que la información de "caracterización para la pérdida de trama" posee varios niveles discretos que caracterizan la planitud  $PI$  del espectro. La ganancia  $G$  puede expresarse directamente en función del valor  $PI$ . Es lo mismo para el límite del segmento  $N_c$  para la búsqueda de tono y/o el número de pics  $A_n$  a tener en cuenta para la síntesis de la señal.

45 Se puede definir a título de ejemplo un tratamiento como sigue.

Se define ya la ganancia  $G$  directamente en función del valor  $PI$ :  $G(PI) - 2^{PI}$

50 Además, se compara el valor  $PI$  con un valor medio 3 dB, entendiéndose que el valor 0 corresponde a un espectro plano y -5 dB corresponde a un espectro de pics pronunciados.

Si el valor  $PI$  es inferior al valor medio de umbral -3 dB (correspondiente por tanto a un espectro de pics pronunciados, típicamente de una señal sonora), entonces se puede fijar la duración del segmento de búsqueda de tono  $N_c$  a 33 ms y seleccionar los pics  $A(n)$  tales que  $A(n) > A(n-1)$  y  $A(n) > A(n+1)$ , así como los pics primeros vecinos  $A(n-1)$  y  $A(n+1)$ .

55 Si no (si el valor  $PI$  es superior al umbral, lo que corresponde a unos pics menos marcados, más el ruido de fondo como por ejemplo una señal de música), la duración  $N_c$  puede elegirse más corta, por ejemplo de 25 ms y solo se

seleccionan los pics  $A(n)$  tales que  $A(n) > A(n-1)$  y  $A(n) > A(n+1)$ .

5 La decodificación puede proseguirse a continuación mediante la mezcla de ruido cuya ganancia se obtiene así con los componentes de ese modo seleccionados para obtener la señal de síntesis en las bajas frecuencias en la etapa D13, que se añade a la señal de síntesis en las altas frecuencias obtenida en la etapa D14, para obtener en la etapa D15 la señal global sintetizada.

10 Con referencia a la figura 6, se ha ilustrado una implementación posible en la que, un decodificador DECOD (que incluye por ejemplo software y hardware tal como una memoria MEM razonablemente programada y un procesador PROC que coopera con esta memoria o como variante un componente tal como un ASIC u otro, así como una interfaz de comunicación COM) implantado por ejemplo en un dispositivo de telecomunicación tal como un teléfono TEL, utiliza, para la implementación del procedimiento de la figura 4, una información de sonoridad que recibe de un codificador COD. Este codificador incluye por ejemplo un software y hardware tal como una memoria MEM razonablemente programada para determinar la información de sonoridad y un procesador PROC' que coopera con esta memoria o, como variante, un componente tal como un ASIC u otro, así como una interfaz de comunicación COM'. El codificador COD se implanta en un dispositivo de telecomunicación tal como un teléfono TEL'.

15 Por supuesto, la presente invención no se limita a las formas de realización descritas anteriormente a título de ejemplo; se extiende a otras variantes.

20 De este modo, por ejemplo, se comprenderá que la información sobre la sonoridad puede tomar diferentes formas susceptibles de variantes. En el ejemplo anteriormente descrito, se puede tratar de un valor binario sobre un único bit (sonoridad o no) o también un valor sobre varios bits que puede ser relativo a un parámetro tal como la planitud del espectro de la señal o cualquier otro parámetro que permita caracterizar (cuantitativamente o cualitativamente) una sonoridad. Aún más, este parámetro puede determinarse en la decodificación, por ejemplo en función del grado de correlación que puede medirse durante la identificación del periodo de tono.

25 Por otra parte, se ha presentado anteriormente a título de ejemplo una realización que incluye una separación en una banda de frecuencias altas y una banda de frecuencias bajas, de la señal procedente de tramas válidas precedentes, en particular una selección de los componentes espectrales en la banda de frecuencias bajas. No obstante, esta realización es opcional aunque ventajosa en el sentido de que permite reducir la complejidad del tratamiento. El procedimiento de sustitución de trama asistido por la información de sonoridad en el sentido de la invención puede sin embargo realizarse considerando todo el espectro de la señal válida, como variante.

30 Por otra parte, se ha descrito anteriormente un ejemplo de realización en el que la invención se implementaba en el marco de una codificación por transformada con adición y recubrimiento. No obstante, ese tipo de procedimiento puede adaptarse a cualquier otro tipo de codificación (particularmente CELP).

35 Se ha de observar que en el marco de una codificación por transformada con adición y recubrimiento (en el que típicamente la señal de síntesis se construye sobre al menos dos duraciones de tramas debido al recubrimiento), la señal de ruido antes citada puede obtenerse por el residuo (entre la señal válida y la suma de los pics) ponderando este residuo temporalmente. Puede ponderarse por ejemplo mediante unas ventanas de recubrimiento, como en el marco habitual de una codificación/decodificación por transformada con recubrimiento.

40 Se entenderá entonces que la aplicación de la ganancia en función de la información de sonoridad tiende a añadirse además a otra ponderación, esta vez en función de la sonoridad.

**REIVINDICACIONES**

1. Procedimiento de tratamiento de una señal de audio digital que incluye una sucesión de muestras repartidas en tramas sucesivas, implementándose el procedimiento durante una decodificación de dicha señal para sustituir al menos una trama de señal perdida en la decodificación, incluyendo el procedimiento las etapas:
- a) búsqueda, en un segmento de señal válida disponible en la decodificación (Nc), de al menos un periodo de la señal, determinado en función de dicha señal válida,
  - b) análisis de la señal en dicho periodo, para una determinación de componentes espectrales de la señal en dicho periodo,
  - c) síntesis de al menos una trama de sustitución de la trama perdida, por construcción de una señal de síntesis a partir:
    - de una adición de componentes seleccionados entre dichos componentes espectrales determinados y
    - de un ruido añadido a la adición de componentes,
- en el que la cantidad de ruido añadida a la adición de componentes se pondera en función de una información de sonoridad de la señal válida, determinándose dicha información de sonoridad por un codificador y suministrándose posteriormente en un flujo codificado, correspondiente a dicha señal, entregado por dicho codificador y recibido en la decodificación, de manera que, en caso de pérdida de trama en la decodificación, se utilice la información de sonoridad contenida en una trama de señal válida precedente a la trama perdida.
2. Procedimiento según la reivindicación 1, caracterizado por que una señal de ruido añadida a la adición de componentes se pondera mediante una ganancia más pequeña en caso de sonoridad de la señal válida.
3. Procedimiento según la reivindicación 2, caracterizado por que la señal de ruido se obtiene por un residuo entre la señal válida y la adición de los componentes seleccionados.
4. Procedimiento según una de las reivindicaciones anteriores, caracterizado por que el número de componentes seleccionados para la adición es mayor en caso de sonoridad de la señal válida.
5. Procedimiento según una de las reivindicaciones anteriores, caracterizado por que, en la etapa a), el periodo se busca en un segmento de señal válida (Nc) de duración mayor en caso de sonoridad de la señal válida.
6. Procedimiento según una de las reivindicaciones anteriores, caracterizado por que la información de sonoridad se codifica sobre un único bit en el flujo codificado.
7. Procedimiento según la reivindicación 6, tomada en combinación con la reivindicación 2, caracterizado por que, si la señal es sonora, el valor de la ganancia es de 0,25 y es de 1 si no.
8. Procedimiento según una de las reivindicaciones anteriores, caracterizado por que la información de sonoridad procede de un codificador que determina un valor de planitud del espectro (PI), obtenido por comparación con un ruido de fondo de las amplitudes de los componentes espectrales de la señal, entregando el codificador dicho valor en la forma binaria en el flujo codificado.
9. Procedimiento según una de las reivindicaciones anteriores, tomada en combinación con la reivindicación 2, caracterizado por que el valor de la ganancia es función de dicho valor de planitud.
10. Procedimiento según una de las reivindicaciones 8 y 9, caracterizado por que dicho valor de planitud se compara con un umbral para determinar:
  - que la señal es sonora si el valor de planitud es inferior al umbral y
  - que la señal no es sonora si no.
11. Procedimiento según una de las reivindicaciones 6 y 10, tomadas en combinación con la reivindicación 4, caracterizado por que:
  - si la señal es sonora, se seleccionan las componentes espectrales cuyas amplitudes son superiores a las de las primeras componentes espectrales vecinas, así como las primeras componentes espectrales vecinas y
  - no se seleccionan más que los componentes espectrales cuyas amplitudes son superiores a las de las primeras componentes espectrales vecinas, si no.
12. Procedimiento según una de las reivindicaciones 6 y 10, tomada en combinación con la reivindicación 5, caracterizado por que:

- si la señal es sonora, el periodo se busca en un segmento de señal válida de duración superior a 30 milisegundos,
- y, si no, el periodo se busca en un segmento de señal válida de duración inferior a 30 milisegundos.

5 13. Programa informático caracterizado por que incluye unas instrucciones para la implementación del procedimiento según una de las reivindicaciones 1 a 12, cuando este programa es ejecutado por un procesador.

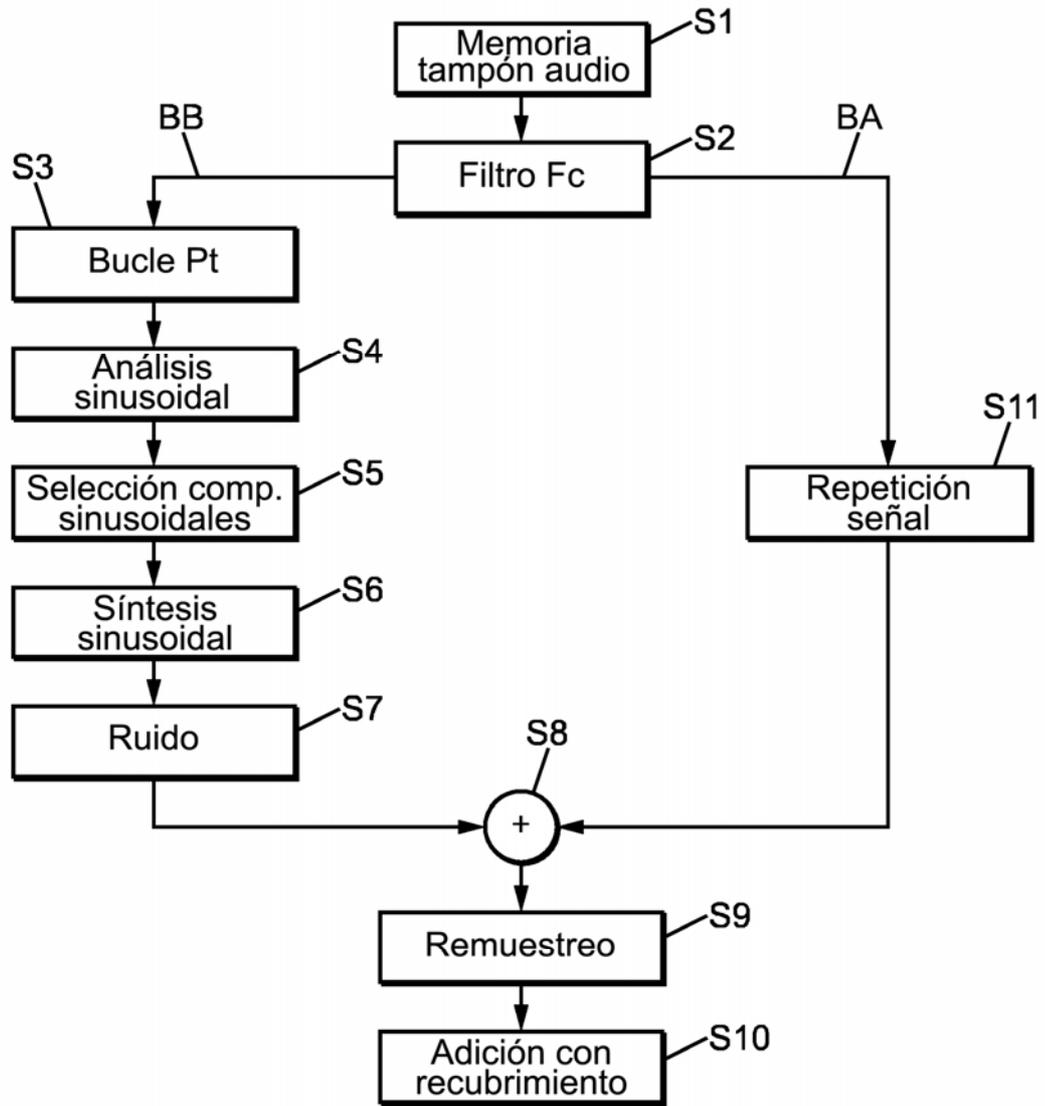
14. Dispositivo de decodificación de una señal de audio digital que incluye una sucesión de muestras repartidas en tramas sucesivas, incluyendo el dispositivo unos medios (MEM, PROC) para sustituir al menos una trama de señal perdida, por:

- 10
- a) búsqueda, en un segmento de señal válida disponible en la decodificación ( $N_c$ ), de al menos un periodo de la señal, determinado en función de dicha señal válida,
  - b) análisis de la señal en dicho periodo, para una determinación de componentes espectrales de la señal en dicho periodo,
  - 15 c) síntesis de al menos una trama de sustitución de la trama perdida, por construcción de una señal de síntesis a partir:

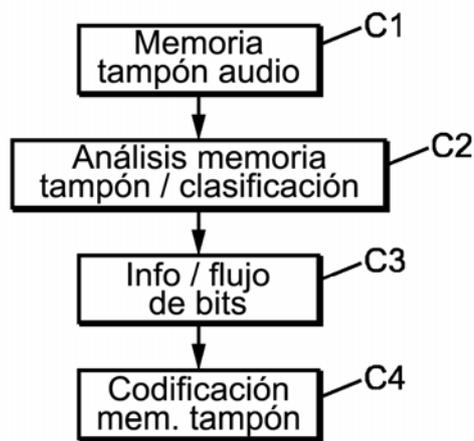
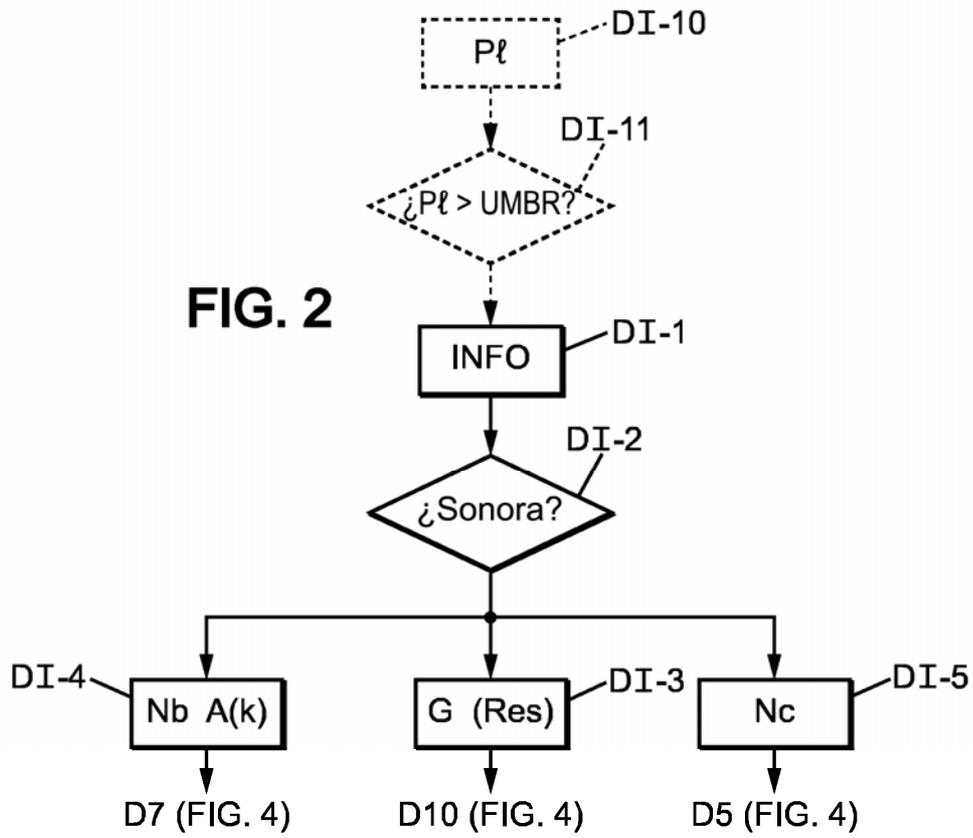
- de una adición de componentes seleccionados entre dichos componentes espectrales determinados y
  - de un ruido añadido a la adición de componentes,
- 20

ponderándose la cantidad de ruido añadida a la adición de componentes en función de una información de sonoridad de la señal válida, determinándose dicha información de sonoridad por un codificador y suministrándose posteriormente en un flujo codificado, correspondiente a dicha señal, entregada por dicho codificador y recibida en la decodificación de manera que, en caso de pérdida de trama en la decodificación, se utilice la información de sonoridad contenida en una trama de señal válida precedente a la trama perdida.

25



**FIG. 1**



**FIG. 3**

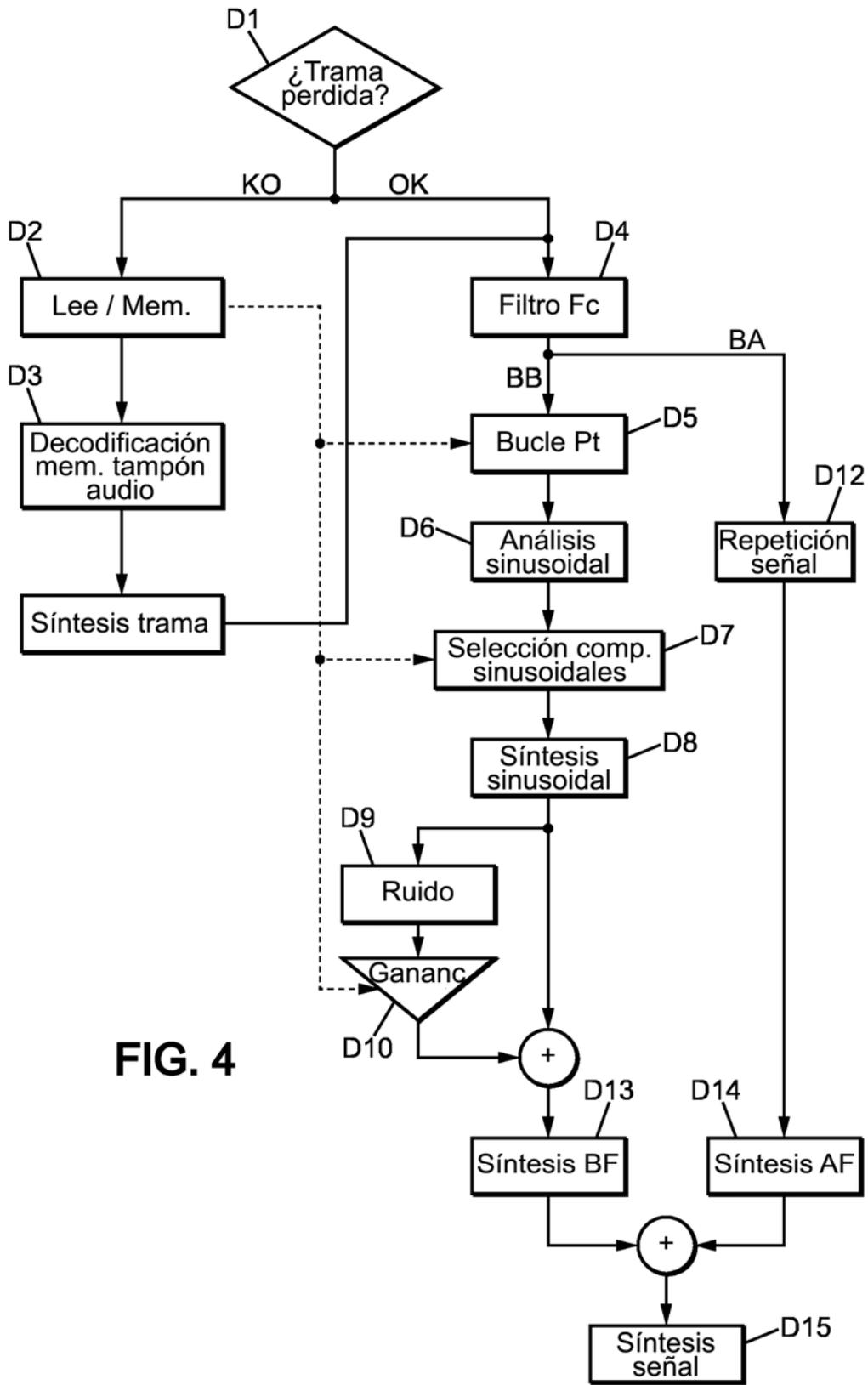
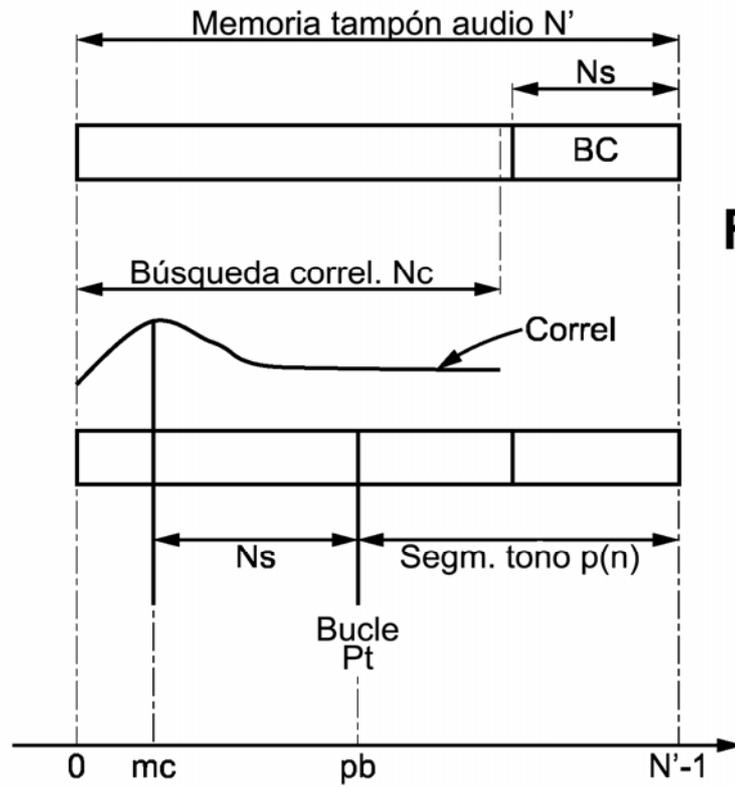
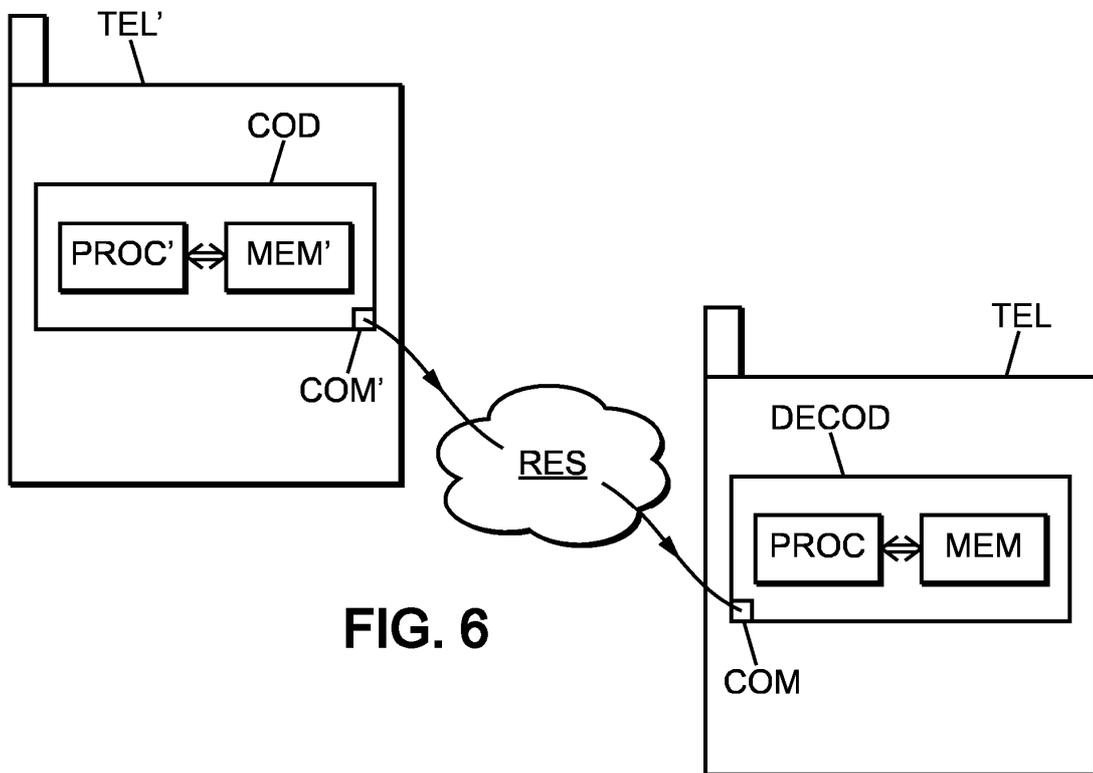


FIG. 4



**FIG. 5**



**FIG. 6**