

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 767 363**

51 Int. Cl.:

G10L 19/26 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.11.2014 PCT/US2014/067056**

87 Fecha y número de publicación internacional: **18.06.2015 WO15088752**

96 Fecha de presentación y número de la solicitud europea: **24.11.2014 E 14809574 (8)**

97 Fecha y número de publicación de la concesión europea: **13.11.2019 EP 3080805**

54 Título: **Método y aparato para mejorar el índice de modulación de sonidos del habla pasados a través de un codificador de voz digital**

30 Prioridad:

12.12.2013 US 201314104777

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

17.06.2020

73 Titular/es:

**MOTOROLA SOLUTIONS, INC. (100.0%)
500 W. Monroe Street, 43rd Floor
Chicago, IL 60661 , US**

72 Inventor/es:

**KUSHNER, WILLIAM M. y
NOVORITA, ROBERT J.**

74 Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

ES 2 767 363 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

Método y aparato para mejorar el índice de modulación de sonidos del habla pasados a través de un codificador de voz digital

Campo de la divulgación

5 La presente divulgación se refiere en general a comunicaciones de radio y más particularmente al procesamiento de señales del habla en dispositivos de comunicación por radio.

Antecedentes

10 Las radios móviles terrestres que proporciona comunicación de radio bidireccional se utilizan en muchos campos, tales como en la aplicación de la ley, seguridad pública, rescate, seguridad, flotas de camiones y flotas de taxi por nombrar unos pocos. Las radios móviles terrestres incluyen tanto unidades basadas en vehículo como portátiles. Las radios móviles terrestres digitales tienen procesamiento adicional dentro de la radio para convertir la voz analógica original en formato digital antes de transmitir la señal en forma digital en el transcurso de la comunicación. La radio de recepción recibe la señal digital y la convierte de vuelta en una señal analógica por lo que el usuario puede escuchar la voz. Ejemplos de radio digital son radios que cumplen con la norma APCO-25 o la norma TETRA. Sin embargo, en ocasiones se ha percibido que las radios digitales distorsionan ciertos sonidos del habla. En particular, los sonidos del habla que tienen trinos alveolares, tal como la 'r' mordida usada en idiomas español e italiano, pueden percibirse como que suenan distorsionados, planos o arrastrados.

20 En la operación de radio, el habla de audio de entrada en un micrófono se convierte por un convertidor de analógico a digital (A/D) que da como resultado una señal de habla digitalizada que se introduce a un codificador de voz. Los codificadores de voz de banda estrecha se usan en productos de radio digitales. La Figura 1 es un ejemplo 100 gráfico que compara sonidos de trino pre codificador de voz a sonidos de trino pos codificador de voz de acuerdo con la técnica anterior. Los gráficos 102 y 104 muestran el tiempo frente a amplitud para dos muestras del habla. Los trinos 106 y 110 alveolares no codificados (pre codificador de voz) se muestran en el gráfico 102. Los trinos 108 y 112 alveolares codificados/decodificados pos codificador de voz correspondientes se muestran en el gráfico 104. Como se muestra en el gráfico 104, los trinos 108 y 112 alveolares están manchados y por lo tanto no se codifican correctamente por el codificador de voz de banda estrecha que provoca problemas de inteligibilidad, especialmente en italiano y español. Puesto que los codificadores de voz típicamente están regulados por la norma dentro de la que operan, no pueden modificarse fácilmente.

30 El documento EP0764940 A2 (AT&T Corp) proporciona un método de codificación del habla. El habla se digitaliza en tramas definidas temporalmente, incluyendo cada trama una pluralidad de subtramas. El habla digitalizada se particiona en componentes periódicos y una señal residual. Cada subtrama de la señal residual puede a continuación desplazarse en tiempo. El desplazamiento de tiempo depende de la aplicación de interpolación lineal para conocer retardos de desplazamiento que tienen lugar en o cerca de límites de trama a trama de tramas anteriores.

35 Por consiguiente, es necesario un medio para mejorar la fidelidad de sonidos del habla de tasa de modulación superior codificados por voz sin modificar el codificador de voz.

Sumario

40 De acuerdo con un primer aspecto de la invención, se proporciona un aparato que comprende las características de la reivindicación adjunta 1. De acuerdo con un segundo aspecto de la invención, se proporciona un aparato que comprende las características de la reivindicación adjunta 3. De acuerdo con un tercer aspecto de la invención, se proporciona un aparato que comprende las características de la reivindicación adjunta 4. Las reivindicaciones dependientes proporcionan detalles adicionales de las realizaciones de la invención.

Breve descripción de las figuras

45 Las figuras adjuntas, donde números de referencia similares hacen referencia a elementos de funcionalidad idéntica o similar a través de todas las vistas separadas, junto con la descripción detallada a continuación, se incorporan en y forman parte de la memoria descriptiva, y sirven para ilustrar adicionalmente realizaciones de conceptos que incluyen la invención reivindicada, y explican diversos principios y ventajas de estas realizaciones.

La Figura 1 es un ejemplo gráfico que compara sonidos de trino pre codificador de voz a sonidos de trino pos codificador de voz de acuerdo con la técnica anterior;

50 La Figura 2 ilustra un diagrama de bloques de una pluralidad de enfoques de mejora del habla de acuerdo con diversas realizaciones;

La Figura 3 proporciona etapas detalladas para un enfoque de desplazamiento de trama de la Figura 2 de acuerdo con una realización;

La Figura 4 muestra una máquina de estado de alineación de nulos de envolvente de modulación que corresponde con la Figura 3 de acuerdo con una realización;

5 La Figura 5 muestra ejemplos gráficos de señales de trino muestreadas en la salida del codificador de voz con y sin desplazamiento de trama de acuerdo con la realización de desplazamiento de trama.

La Figura 6 muestra un diagrama de bloques más detallado del método de modificación de parámetro de ganancia de codificador de voz de nulos de energía de modulación de acuerdo con una realización;

10 La Figura 7 es un ejemplo ilustrativo de un enfoque de compresión y expansión de tiempo de acuerdo con una realización;

La Figura 8 muestra ejemplos de espectrogramas de muestra que comparan trinos alveolares de acuerdo con las realizaciones de tiempo expandido;

La Figura 9 muestra ejemplos de espectrogramas que comparan trinos alveolares de acuerdo con las realizaciones de filtro de mejora de modulación;

15 La Figura 10 muestra imágenes que comparan trinos alveolares de acuerdo con las realizaciones de filtro de mejora de modulación.

Los expertos apreciarán que en las figuras se ilustran elementos por simplicidad y claridad y no necesariamente se han dibujado a escala. Por ejemplo, las dimensiones de algunos de los elementos en las figuras pueden exagerarse con relación a otros elementos para ayudar a mejorar el entendimiento de las realizaciones de la presente invención.

20 Los componentes del aparato y método se han representado donde sea apropiado mediante símbolos convencionales en los dibujos, mostrando únicamente aquellos detalles específicos que son pertinentes para el entendimiento de las realizaciones de la presente invención para no obstaculizar la divulgación con detalles que serán fácilmente evidentes para los expertos en la materia que tienen el beneficio de la descripción en el presente documento.

25 Descripción detallada

En resumen, se describen en el presente documento métodos y aparatos para mejorar el índice de modulación de sonidos del habla pasados a través de un codificador de voz digital. Se proporcionan métodos para mejorar la codificación de sonido de tasa de alta modulación, particularmente para inteligibilidad de sonido alveolar. Los métodos y aparato tratan errores de codificación de modulación de envolvente del habla provocados por la tasa de análisis de energía de trama lenta inherente en codificadores de voz paramétricos de baja tasa de bits, tal como la clase de codificadores de voz de Excitación Multi-Banda Mejorada (IMBE™) y Excitación de Multi-Banda de Avanzada (AMBE©) de codificadores de voz producidos por DSVI Inc. Se resuelven los errores de codificación de modulación de envolvente del habla y artefactos de solapamiento provocados por la tasa de trama sub-Nyquist usada en codificadores de voz de banda ancha.

35 Los codificadores de voz de banda estrecha se usan en productos de radio digitales. Dependiendo del tipo de técnicas de codificación por voz, el codificador de voz también "comprime" la muestra resultante de modo que puede adaptarse en un ancho de banda más estrecho. El contenido de información del habla humana se codifica por el codificador de voz usando frecuencia acústica y modulación de amplitud. El flujo de información fonémica se descompone en sílabas codificadas como modulación de envolvente de energía. La tasa de modulación silábica del habla es típicamente menor que 16 Hz con la gran mayoría de energía de modulación de amplitud que tiene lugar en el intervalo de 0,5-5 Hz. Sin embargo, como se ha mencionado anteriormente en algunos idiomas, tal como italiano y español, ciertos sonidos, más en particular el trino alveolar (por ejemplo, la "r" con trino), lleva información fonémica importante codificada en modulación de amplitud a una tasa superior de 20-40 Hz. En codificadores de voz paramétricos de baja tasa de bits, el parámetro de energía de señal que codifica la modulación de amplitud de forma de onda se calcula como una tasa de trama baja, típicamente 50 trama/s o menor. Además, se emplea el solapamiento de trama y otras formas de suavizado de parámetro para reducir artefactos de codificación. Para idiomas tales como inglés con tasas de modulación silábica bajas esto no es un problema. Sin embargo, para sonidos que se definen por una tasa de modulación de amplitud superior tal como el trino alveolar, la codificación de voz puede provocar que el componente de modulación de energía se defina de manera pobre debido a suavizado y solapamiento de trama, que reduce la perceptibilidad e inteligibilidad del sonido. Aunque una solución fácil aumentaría la tasa de análisis de trama, esto no puede hacerse sin aumentar la tasa de bits de codificador de voz o modificar la tasa de parámetro de codificador de voz de alguna otra manera. Puesto que los codificadores de voz típicamente están regulados por la norma dentro de la que operan, no pueden modificarse fácilmente.

55 De acuerdo con las diversas realizaciones, se proporcionan enfoques de pre procesamiento y pos procesamiento para mejorar ciertos tipos de sonidos del habla. Se proporciona una pluralidad de módulos de procesador de pre-codificador de voz y módulos de procesador de pos-codificador de voz para mejorar el índice de modulación de sonidos del habla con trino, particularmente el trino alveolar, para hacerles más perceptibles después de pasar a un codificador de voz de banda estrecha. Los codificadores de voz de banda estrecha típicamente emplean una tasa de análisis de trama que es demasiado baja para reproducir de manera precisa modulaciones de amplitud del habla de

frecuencia superior. Puesto que la tasa de trama del codificador de voz no puede aumentarse, los pre y pos procesadores proporcionados en el presente documento se utilizan para mejorar la modulación a través del desplazamiento de tiempo, expansión de tiempo y filtración de dominio de modulación. Se proponen varias técnicas. Alguna de estas técnicas depende de detectar la presencia de un sonido del habla de alta tasa de modulación y determinar la localización de tiempo y de los nulos de modulación. Esta información se usa por métodos posteriores.

La Figura 2 ilustra un diagrama de bloques de diversos enfoques de mejora del habla de acuerdo con algunas realizaciones. El diagrama 200 de bloques mejora la inteligibilidad del sonido para señales procesadas a través de un codificador de voz digital. El codificador de voz digital se muestra en la Figura 2 como el codificador 214 de codificador de voz y el decodificador 220 de codificador de voz para diferenciar entre señales que se transmiten fuera y que se reciben en el codificador de voz. El diagrama 200 de bloques muestra una señal del habla de entrada digitalizada 202 que se procesa por una o más etapas de procesamiento de pre-codificador de voz antes de que se codifique por el codificador 214 de codificador de voz para su transmisión en 216. Para una señal de entrada recibida en 218, el decodificador 220 de codificador de voz decodifica y procesa la señal a través de una o más etapas de pos codificador de voz para generar la señal 234 del habla de salida. Las diversas realizaciones mostrarán que la mejora del habla puede conseguirse con cualquiera del procesamiento pre codificador de voz en solitario, procesamiento pos codificador de voz en solitario y/o una combinación de tanto procesamiento de pre codificador de voz y pos codificador de voz.

El diagrama 200 de bloques se usará para describir cuatro métodos diferentes para potenciar el habla a través del codificador de voz digital. La Tabla a continuación resume estos enfoques:

	Pre-codificador de voz	Pos-codificador de voz
Desplazamiento (210) de trama	x	
Modificación (212) de parámetro de energía	x	
Expansión (210) de tiempo/compresión (222) de tiempo	x	x
Filtro (224) de mejora de modulación		x

Tanto el método 210 de desplazamiento de trama como el método 212 de modificación de parámetro de energía hacen uso de una detección 204 de evento de modulación que comprende el cálculo 206 de la energía de envolvente y el detector 208 de nulos de envolvente de modulación. Estos se describirán adicionalmente en diagramas ampliados de la Figura 3 para desplazamiento de trama y la Figura 6 para modificación de parámetro de energía.

En un primer método, una trama de análisis predeterminado se desplaza en tiempo ligeramente para capturar de manera máxima los nulos de energía de la modulación de trino. Esto es esencialmente un re-muestro de la envolvente de energía con un desplazamiento de fase. En la operación, la señal 202 del habla digitalizada de entrada se recibe y marcha a través de una etapa 210 de procesamiento de pre-codificación de voz, la etapa 210 de pre procesamiento proporciona el método de desplazamiento de trama.

El enfoque de desplazamiento de trama se describe en las Figuras 3 y 4 con etapas detalladas adicionales. Haciendo referencia a la Figura 3, se recibe una señal del habla digitalizada de entrada en 202 a través de una primera tasa de muestreo de ventanas predeterminada. El bloque 204 de procesamiento proporciona cálculos de energía de envolvente y detección de nulos. Las diferencias de envolvente (diferencias de frecuencia de modulación y energía entre la señal de entrada original y aquellas calculadas en la tasa de trama del codificador de voz) se calculan en 304. Este cálculo puede hacerse por un calculador de energía diferencial para determinar diferencias inter-trama. En 306, las diferencias de envolvente $f()$ se muestrean y clasifican para puntos y estados (picos y valles) por un clasificador de diferencia de energía para definir una máquina de estado. La máquina de estado opera en 308 para determinar la localización de nulos de modulación de la envolvente del habla. La máquina de estado identifica nulos de envolvente de energía y los localiza en tiempo y frecuencia. Una memoria intermedia de datos elástica en 310 permite que de desplace hacia delante o atrás una trama de datos en un tiempo relativo al tiempo de muestreo de trama de codificador de voz (se alinea con el desplazamiento 210 de trama de la Figura 2). La trama de análisis por lo tanto puede desplazarse hacia delante o atrás en tiempo para coincidir con los nulos de amplitud de modulación detectados.

La Figura 4 muestra un diagrama 400 del detector de nulos de envolvente de modulación que tiene la máquina de estado de alineación de nulos de envolvente de modulación que corresponde con la Figura 3. De nuevo, se recibe la señal digitalizada en 202 y marcha a través del bloque 204 de procesamiento y una memoria intermedia 410 elástica (desplazamiento 210 de trama de la Figura 2) que puede desplazarse hacia atrás y hacia delante para alienarse con nulos detectados. El desplazamiento hacia delante y hacia atrás se controla por la creación de envolventes de energía en ventanas en 402, energía calculada dentro de la envolvente en ventana en 404, el cálculo de los puntos de diferencias de envolvente en 406, y la clasificación de muestras a estados en 408. La clasificación de estados puede incluir puntos pico, puntos descendentes, puntos ascendentes y puntos nulos como se observa en la máquina

420 de estado finito del detector de modulación de amplitud. Los índices de los nulos se pasan a continuación a través de la memoria intermedia 410 elástica, la memoria intermedia elástica termina en los índices nulos antes de la codificación de la señal de trino mejorada al codificador 214 del codificador de voz.

5 La señal 412 desplazada en trama se codifica a continuación a través del codificador en 214 y se transmite en 216. La Figura 5 muestra ejemplos 500 gráficos de señales de trino muestreadas en la salida del codificador de voz con y sin desplazamiento de trama de acuerdo con la realización de desplazamiento de trama. Las respuestas de envolvente espectral de trino alveolar a diferentes tasas de muestra de trama se muestran en el gráfico 502 (con desplazamiento de trama cero). El tiempo se indica a lo largo del eje 506 horizontal y los niveles de decibelios (dB) en el eje 508 vertical. Las ventanas de tasa de trama (tal como las ventanas creadas en 402 en la Figura 4) se crean en 5 ms (510), 10 ms (512), y 20 ms (514). En el gráfico 504, las respuestas de envolvente espectral de trino alveolar a diferentes tasas de muestra de trama se muestran con un desplazamiento de tiempo de 10 ms. Este desplazamiento de trama se genera en la memoria intermedia 310 elástica de la Figura 3 y 410 de la Figura 4. De nuevo, las ventanas de tasa de trama se crearon a 5 ms (520), 10 ms (522), y 20 ms (524). Sin embargo, el desplazamiento de trama de 10 ms hace una mejora significativa a la señal de retardo de 20 ms, en aproximadamente de 3 a 5 dB. Por lo tanto, el trino que proviene fuera del codificador de voz es ventajosamente bastante más pronunciado con el desplazamiento de trama que sin.

De acuerdo con las diversas realizaciones, el enfoque de desplazamiento de trama puede usarse por sí mismo o en conjunto con el método de filtro de mejora de modulación que se va a describir más adelante.

20 Un segundo enfoque opcional para proporcionar mejora del habla proporciona una variación del re-muestreo modificando el parámetro de energía de trama de codificador de voz directamente para alinearse mejor con los nulos de modulación detectados de manera separada. Este enfoque adicional utiliza la modificación 212 de parámetro de energía mostrado en la Figura 2 que se detalla adicionalmente en la Figura 6 como el método 600 de modificación de parámetro de ganancia de codificador de voz de nulos de energía de modulación de acuerdo con una realización.

25 El habla 602 digitalizada se muestrea como anteriormente, pero a una tasa de trama más rápida (por ejemplo 100 trama/s). Los valores de ganancia se extraen de la trama de voz en 604 mientras que el cálculo de la envolvente de energía se calcula en 606 (se alinea con 206 de la Figura 2). Los nulos de envolvente, en el cálculo de la envolvente, se detectan en el detector 608 de nulos de envolvente de modulación (se alinean con 208 de la Figura 2), basándose en esta tasa muestreada superior. Si la máquina de estado en 608 no detecta un nulo de envolvente, a continuación la ganancia de trama de voz extraída asociada con esa muestra (de 604) se considera satisfactoria. Si se detecta un nulo en 610, la ganancia de trama de voz en 604 se pasa a través de 614 para una comparación de ganancia de trama de voz a cálculo de energía de envolvente. El cálculo de energía en 606 está sincronizado al codificador por el retardo en 618.

35 En 614, la ganancia de trama de voz se compara a la energía en ventana retardada. Si la trama de ganancia de voz se determina que es demasiado grande en 614, entonces la ganancia se reduce en 620 y los parámetros para el codificador de voz se re-empaquetan con la nueva ganancia reducida en 622. La señal a continuación continúa a través del codificador 214 de codificador de voz para su transmisión en 216.

40 Por lo tanto, el enfoque 600 alternativo proporciona procesamiento (212) de pre codificador de voz que recibe la información de detector de nulos de evento de modulación, la compara con la información de parámetro de energía de trama derivada del codificador de voz, y modifica el parámetro de energía de trama de codificador de voz para hacer coincidir con la información de energía de nulos del detector.

45 En un tercer método para mejora del habla, la duración del habla de entrada se expande en tiempo para reducir de manera eficaz la frecuencia de modulación de trino para mejorar la codificación a la tasa de trama de codificador de voz fijado. La Figura 2 muestra la expansión de tiempo en el bloque 210 de procesamiento de pre codificador de voz de acuerdo con la tercera realización. En la salida del decodificador 220 de codificador de voz, el habla que puede a continuación expandirse de vuelta a su duración original a través de la compresión de tiempo mostrada en el bloque 222 de pos procesador. El enfoque 700 de expansión y compresión de tiempo se ilustra en la Figura 7. La expansión 702 de tiempo de señal se muestra usando la señal 704 original y la señal 708 extendida. El tiempo expande la señal de trino antes de que la codificación del codificador de voz reduzca la frecuencia de modulación eficaz como se observa en 708. La señal 704 muestra una señal de modulación de envolvente de sonido de un trino con la frecuencia de modulación por encima de una frecuencia de solapamiento de tasa de Nyquist junto con la trama 706 de análisis de codificador de voz, a una tasa de trama fija. Una envolvente de sonido de tiempo expandida del trino mostrada en 708, muestra una frecuencia de modulación por debajo de la de la tasa de Nyquist sin solapamiento. La trama de análisis de codificador de voz sigue siendo la misma en 710. Una señal 712 de modulación de envolvente de sonido de tiempo comprimido tiene la longitud original y ningún solapamiento. Por lo tanto, comprimir en tiempo la señal después de la decodificación del codificador de voz permite que la señal vuelva a su duración de tiempo original. También, la etapa de compresión de tiempo no es necesaria si la expansión de tiempo es menor que el veinte (20) por ciento, puesto que la expansión de tiempo de una señal del habla de menos del (20) por ciento no se percibe fácilmente por un oyente.

Por consiguiente, si la expansión de tiempo es menor que el veinte por ciento (20 %), entonces la etapa de compresión de tiempo no es necesaria sino que puede aplicarse si se desea. Si la expansión de tiempo es mayor de aproximadamente el veinte por ciento (20 %) entonces debería aplicarse la etapa de compresión de tiempo.

5 Hay un número de métodos conocidos para expandir y comprimir de manera inversa una señal del habla en tiempo que puede producir el cambio deseado en frecuencia de modulación necesario para mejorar la modulación de sonido de trino. Un método de este tipo, por ejemplo, es el método PSOLA (Solapamiento y Adición de Tono Síncrono). Pueden usarse también otros métodos de modificación de tiempo similares.

10 La Figura 8 muestra ejemplos de imágenes de espectrograma de muestra que comparan trinos alveolares de acuerdo con las realizaciones de tiempo expandido. La imagen 802 muestra el trino alveolar en un estado no codificado. La imagen 804 muestra el trino alveolar procesado por el codificador de voz sin ninguna expansión de tiempo. La imagen 804 muestra cómo de manchado se vuelve el trino que conduce a problemas con la inteligibilidad. La imagen 806 muestra una expansión de tiempo del diez (10) porciento que se aplica antes del codificador de voz sin ninguna etapa de compresión de tiempo. La imagen 808 muestra una expansión de tiempo del veinte (20) porciento que se aplica antes del codificador de voz. La aplicación de expansión de tiempo antes del codificador de voz mejora por lo tanto enormemente la inteligibilidad del sonido de trino.

15 En un cuarto método, el índice de modulación del sonido de trino puede mejorarse extrayendo la envolvente de modulación de energía del habla, pasándola a través de un filtro de frecuencia selectiva con ganancia positiva aplicado en la frecuencia de modulación de trino. Este cuarto enfoque puede usarse también con un filtro de paso banda o paso bajo de atenuación para ayudar a eliminar componentes de modulación de frecuencia superior que provocan solapamiento. La envolvente de modulación mejorada se imprime a continuación en el flujo de señal del habla decodificada. Este cuarto enfoque se ilustra en la Figura 2 por el filtro 224 de mejora de modulación que comprende un elemento 226 de retardo de tiempo, un elemento 228 de cálculo de envolvente de energía, un filtro 230 de mejora de dominio de modulación, y multiplicador 232 de ganancia de envolvente de energía acoplado en la salida del codificador de voz 220.

20 En la operación, la señal digitalizada proviene del decodificador 220 y el filtro 224 mejora el sonido de trino amplificando las frecuencias de modulación de envolvente en el intervalo 20-40 Hz. El filtro 224 amplifica energía en el intervalo de frecuencia especificado para proporcionar énfasis a la modulación de trino. El componente de retardo de tiempo es necesario para retardar la señal de salida del codificador de voz para contabilizar el retardo de señal provocado por el filtro 230 de mejora de dominio de modulación. Esto asegura que la envolvente de modulación modificada estará alineada en tiempo con la señal de salida del codificador de voz. El calculador 228 de envolvente de energía calcula la envolvente de energía de salida del codificador de voz elevando al cuadrado las muestras de señal. La energía de señal de salida del codificador de voz es una señal únicamente positiva que pasa a través del filtro 230 de dominio de modulación, que puede ser un filtro de paso bajo o de paso banda. Por ejemplo, puede usarse un filtro de paso bajo de dos polos Chebyshev tipo 1, para producir un aumento de ganancia positiva en la banda de modulación de trino mientras se pasan frecuencias de modulación inferior y se suprimen frecuencias de modulación superior de acuerdo con los efectos deseados. El pico de ganancia de filtro tiene lugar a aproximadamente el centro de la banda de modulación de sonido de trino (para este ejemplo 28 Hz, como se mostrará en la Figura 9).

25 Se muestran ejemplos para el método de filtro de mejora de modulación (MEF) en la Figura 9. La respuesta 902 de filtro de mejora de modulación (MEF) muestra la respuesta de magnitud (db) para un filtro Chebyshev tipo 1 de dos polos con un pico 922 de ganancia en la frecuencia de modulación de trino. Este pico de ganancia de filtro tiene lugar a aproximadamente el centro de la banda de modulación de sonido de trino (para este ejemplo 28 Hz). El gráfico 904 muestra el tiempo de respuesta de impulso para el filtro. Este gráfico es representativo del filtro 230 de dominio de modulación.

30 Las formas de onda 906, 908, 910, 911, y 912 se muestran con el tiempo en un eje horizontal y la amplitud (o magnitud para 910, 911) a lo largo de un eje vertical. La forma de onda 906 muestra la señal (202) de habla de entrada original. La forma de onda 908 muestra la señal después de la codificación de voz (220) sin mejora alguna. La forma de onda 910 muestra la envolvente de energía de señal codificada por voz. La forma de onda 911 muestra la envolvente de energía de señal codificada por voz que se está filtrando por el filtro 230 de dominio de modulación. El filtro de mejora de dominio de modulación proporciona una ganancia positiva para las frecuencias de modulación predeterminadas de la envolvente de energía calculada.

35 La forma de onda 912 muestra la señal después de filtrarse por el filtro 230 de dominio de modulación y la aplicación del multiplicador 232 de ganancia de envolvente de energía. Por lo tanto, el multiplicador 232 de ganancia de envolvente de energía impone la envolvente de energía de modulación filtrada en el flujo 226 del habla digitalizada retardado. Como puede observarse por la forma de onda 912, la señal del habla de salida que tiene el filtro 224 de mejora de modulación aplicado a la misma mejora significativamente el índice de modulación y mejora la inteligibilidad del sonido de trino.

La Figura 10 muestra imágenes de espectrograma que comparan trinos alveolares de acuerdo con las realizaciones de filtro de mejora de modulación. El espectrograma 1002 muestra el sonido del trino alveolar en una condición no codificada, que corresponde a la forma de onda 906 de la Figura 9. El espectrograma 1004 muestra el sonido del trino alveolar después de codificarse por voz, que corresponde a la forma de onda 908 de la Figura 9. El espectrograma 1006 muestra el sonido del trino alveolar después de codificarse por voz y el filtro 224 de mejora de modulación que se aplica, que corresponde a la forma de onda 910 de la Figura 9.

El espectrograma 1008 muestra el sonido de trino alveolar después de desplazarse en trama usando el método de desplazamiento de trama, codificarse por voz y que se aplique el filtro 224 de mejora de modulación. Obsérvese que la combinación de los dos métodos de mejora de trino diferentes da como resultado incluso mejor mejora. El método de filtro de mejora de modulación puede usarse con cualquiera de los otros métodos de mejora para efecto aumentado.

Por consiguiente, cuatro métodos/enfoques se han proporcionado para mejorar la mejora del habla en un producto de radio digital. En el primer método, una trama de análisis predeterminada (por ejemplo 20 ms) se desplaza en tiempo ligeramente para capturar de manera máxima los nulos de energía de la modulación de trino. Este desplazamiento de trama proporciona un re-muestreo de la envolvente de energía con un desplazamiento de fase. El segundo método proporciona una variación del re-muestreo para modificar el parámetro de energía de trama de codificador de voz directamente para alinearse mejor con los nulos de modulación detectados de manera separada. En el tercer método, la duración del habla de entrada se expande para reducir de manera eficaz la frecuencia de modulación de trino para mejorar la codificación en la tasa de trama del codificador de voz fijado. En la salida del decodificador el habla puede expandirse de vuelta a su duración original. En un cuarto método, el índice de modulación del sonido de trino puede mejorarse extrayendo la envolvente de modulación de energía del habla, pasándola a través de un filtro de frecuencia selectiva con ganancia positiva aplicado en la frecuencia de modulación de trino. Este cuarto método puede usarse también con un filtro de paso bajo o paso banda de atenuación para eliminar componentes de modulación superpuestos. La envolvente de modulación mejorada se imprime a continuación en el flujo de señal del habla decodificada. Estos métodos pueden usarse de manera única o en combinación para rendimiento mejorado.

Los elementos pre y pos procesamiento proporcionados por las diversas realizaciones aumentan el índice de modulación de sonidos de tasa de alta modulación sin modificar el codificador de voz. Aumentar el índice de modulación de la modulación de trino mejora la perceptibilidad y calidad de los componentes de sonido de frecuencia de modulación altos.

El uso de los pre-/pos-procesadores, de acuerdo con las diversas realizaciones, mejorará el rendimiento de productos de radio que usan codificadores de voz de banda estrecha, particularmente los codificadores de voz de tipo MBE usados en sistemas P25. Adicionalmente, los pre-/pos-procesadores de las diversas realizaciones pueden usarse también para mejorar codificación de tasa de modulación alta por cualquier codificador de voz donde la tasa de trama es insuficiente para codificar de manera precisa tasas de modulación altas. El uso de los pre/pos procesadores que operan de acuerdo con las diversas realizaciones ayudarán a reproducir sonidos alveolares (es decir con trino) 'r' y otros fomentando la aceptación y venta de sistemas de radio digital de banda estrecha.

El codificador de voz IMBE/AMBE es una norma requerida para compatibilidad en radios de sistema P25 (DMR). La inteligibilidad mejorada para ciertos sonidos del habla mejorará la capacidad de comercialización de productos que incorporan los enfoques de mejora del habla proporcionados por las diversas realizaciones. La tecnología de pre y pos procesamiento mejora la calidad e inteligibilidad del habla codificada por voz que proporciona un rendimiento mejorado y ventaja de comercialización. Otros codificadores de voz de tasa de trama baja, tales como el codificador de voz ACELP usados en sistemas TETRA pueden también aprovecharse de la inteligibilidad mejorada.

Las realizaciones proporcionadas en el presente documento pertenecen a una mejora de sonido de trino de filtración de envolvente de modulación. Las realizaciones tratan los nulos de amplitud de dominio de tiempo del habla para afectar la envolvente de modulación del habla. La acción del filtro de envolvente de modulación (es decir filtro de mejora de trino) es para operar en la envolvente de energía del habla a diferencia del contenido espectral de tramas de análisis individuales en el dominio de la frecuencia. La envolvente de amplitud de forma de onda del habla se analiza ventajosamente como un grupo de múltiples tramas. Las realizaciones utilizan el análisis de energía para identificar nulos de envolvente de energía del habla en el dominio del tiempo para el fin de ajustar la trama de entrada al codificador de voz desplazándola en el tiempo a diferencia de sistemas que manipulan parámetros de dominio de frecuencia.

En la memoria descriptiva anterior, se han descrito realizaciones específicas. Sin embargo, un experto en la materia aprecia que pueden realizarse diversas modificaciones y cambios sin alejarse del alcance de la invención como se exponen en las reivindicaciones a continuación. Por consiguiente, la memoria descriptiva y figuras han de considerarse en un sentido ilustrativo en lugar de restrictivo, y todas tales modificaciones se pretende que estén incluidas dentro del alcance de las presentes enseñanzas.

Los beneficios, ventajas, soluciones a los problemas y cualquier elemento o elementos que provocan que tenga

lugar o sea más pronunciado algún beneficio, ventaja o solución no han de interpretarse como unas características o elementos críticos, requeridos o esenciales de cualquiera o todas las reivindicaciones. La invención se define solamente por las reivindicaciones adjuntas que incluyen cualesquiera modificaciones realizadas durante la tramitación de esta solicitud y todos los equivalentes de aquellas reivindicaciones emitidas.

5 Además en este documento, los términos relacionales tales como primero y segundo, superior e inferior, y similares pueden usarse solamente para distinguir una entidad o acción de otra entidad o acción sin requerir o implicar necesariamente ninguna relación u orden tal actual entre tales entidades o acciones. Los términos "comprende", "que comprende", "tiene", "que tiene", "incluye", "que incluye", "contiene", "que contiene" o alguna otra variación de los mismos, se pretende que cubran una inclusión no exclusiva, de manera que un proceso, método, artículo, o aparato que comprende, tiene, incluye, contiene una lista de elementos no incluye únicamente estos elementos sino que puede incluir otros elementos no expresamente enumerados o inherentes a tal proceso, método, artículo o aparato. Un elemento seguido por "comprende...un", "tiene...un", "incluye...un", "contiene...un" no excluye, sin más restricciones, la existencia de elementos idénticos adicionales en el proceso, método, artículo, o aparato que comprende, tiene, incluye, contiene el elemento. Los términos "un/una" se definen como uno o más a menos que se indique explícitamente de otra manera en el presente documento. Los términos "sustancialmente", "esencialmente", "aproximadamente", "alrededor de" o cualquier otra versión de los mismos, se definen como que están cerca como se entiende por un experto en la materia, y en una realización no limitante el término se define para que esté dentro del 10 %, en otra realización dentro del 5 %, en otra realización dentro del 1 % y en otra realización dentro del 0,5 %. El término "acoplado" como se usa en el presente documento se define como conectado, aunque no necesariamente de manera directa y no necesariamente de manera mecánica. Un dispositivo o estructura que está "configurado" de una cierta manera está configurado en al menos esa manera, pero puede también estar configurado en maneras que no están enumeradas.

Se apreciará que algunas realizaciones pueden estar comprendidas de uno o más procesadores genéricos o especializados (o "dispositivos de procesamiento") tales como microprocesadores, procesadores de señales digitales, procesadores personalizados y campos de matrices de puertas programables (FPGA) e instrucciones de programa almacenadas únicas (que incluyen tanto software como firmware) que controlan el uno o más procesadores para implementar, en conjunto con ciertos circuitos no de procesador, alguna, la mayoría o todas las funciones del método y/o aparato descritos en el presente documento. Como alternativa, algunas o todas las funciones podrían implementarse por una máquina de estado que no tiene almacenadas instrucciones de programa, o uno o más circuitos integrados específicos de la aplicación (ASIC), en el que cada función o algunas combinaciones de ciertas de las funciones se implementan como lógica personalizada. Por supuesto, podría usarse una combinación de los dos enfoques.

Además, una realización puede implementarse como un medio de almacenamiento legible por ordenador que tiene código legible por ordenador almacenado en el mismo para programar un ordenador (por ejemplo, que comprende un procesador) para realizar un método como se describe y reivindica en el presente documento. Ejemplos de tales medios de almacenamiento legibles por ordenador incluyen, pero sin limitación, un disco duro, un CD-ROM, un dispositivo de almacenamiento óptico, un dispositivo de almacenamiento magnético, una ROM (Memoria de Solo Lectura), una PROM (Memoria de Solo Lectura Programable), una EPROM (Memoria de Solo Lectura Programable Borrable), una EEPROM, Memoria de Solo Lectura Programable Electrónicamente Borrable) y una memoria flash. Además, se espera que un experto en la materia, a pesar de esfuerzo posiblemente significativo y muchas elecciones de diseño motivadas por, por ejemplo, tiempo disponible, tecnología actual y consideraciones económicas, cuando se guíe por los conceptos y principios desvelados en el presente documento, sea fácilmente capaz de generar tales instrucciones de software y programas y CI con experimentación mínima.

El resumen de la divulgación se proporciona para permitir al lector determinar rápidamente la naturaleza de la divulgación técnica. Se presenta con el entendimiento de que no se usará para interpretar o limitar el alcance o el significado de las reivindicaciones. Además, en la descripción detallada anterior, puede observarse que se agrupan juntas diversas características en diversas realizaciones para el fin de simplificar la divulgación. Este método de divulgación no ha de interpretarse como que refleja una intención de que las realizaciones reivindicadas requieran más características que las expresamente indicadas en cada reivindicación. En su lugar, como reflejan las siguientes reivindicaciones, la materia objeto inventiva radica en menos de todas las características de una única realización desvelada. Por lo tanto las siguientes reivindicaciones se incorporan en la presente en la descripción detallada, indicando cada reivindicación por sí misma una materia objeto reivindicada por separado.

REIVINDICACIONES

1. Una radio, que comprende:

5 un codificador de voz digital que tiene una tasa de muestreo de trama de datos predeterminada;
al menos un procesador para mejorar un índice de modulación de un evento de sonido de tasa de modulación
alta predeterminada, detectando el al menos un procesador nulos de energía del evento de sonido de tasa de
modulación alta predeterminada en un flujo del habla digitalizada, en el que el al menos un procesador
comprende:

10 un procesador de pre-codificador de voz que comprende un desplazador (210) de trama para desplazar una
trama de datos del flujo del habla digitalizada hacia delante o hacia atrás en el tiempo con relación al tiempo
de muestreo de trama de codificador de voz para coincidir con los nulos de energía detectados; y
en el que el desplazador de trama comprende adicionalmente:

15 un calculador (302) de energía de trama de voz para calcular energía de trama de voz a una tasa de
muestreo de trama de datos más alta que el codificador de voz;
un calculador (304) de energía diferencial para determinar diferencias inter-trama;
un clasificador (306) de diferencia de energía que define una máquina (308) de estado;
identificando y localizando la máquina (308) de estado los nulos; y
una memoria intermedia (310) para desplazar la trama de datos del flujo del habla digitalizada hacia atrás
o hacia delante basándose en los nulos de energía identificados y detectados.

20 2. La radio de la reivindicación 1, en la que el evento de sonido de tasa de modulación alta predeterminada
comprende un sonido de trino.

3. Una radio, que comprende:

25 un codificador de voz digital que tiene una tasa de muestreo de trama de datos predeterminada;
al menos un procesador para mejorar un índice de modulación de un evento de sonido de tasa de modulación
alta predeterminada, detectando el al menos un procesador nulos de energía del evento de sonido de tasa de
modulación alta predeterminada en un flujo del habla digitalizada, en el que el al menos un procesador
comprende:

30 un procesador (210) de pre-codificador de voz para recibir el flujo de entrada del habla digitalizada y expandir
en tiempo el evento de sonido de tasa de modulación alta predeterminada del flujo de entrada del habla
digitalizada antes del codificador de voz, reduciendo la expansión de tiempo frecuencias de modulación de
envolvente del evento de sonido de tasa de modulación alta predeterminada más allá de la tasa de muestreo
predeterminada del codificador de voz; y
un procesador de pos-codificador de voz para comprimir en tiempo un flujo de salida de habla digitalizada del
codificador de voz, invirtiendo de esta manera la expansión de tiempo.

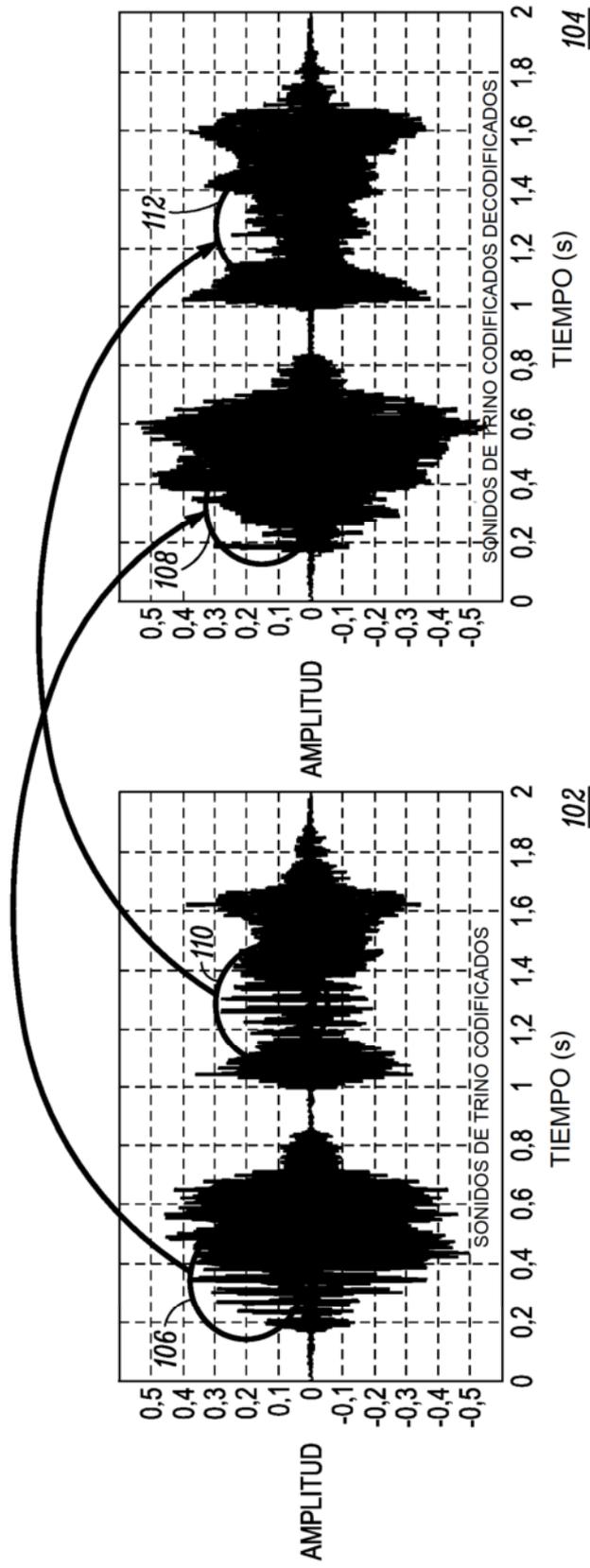
4. Una radio, que comprende:

35 un codificador de voz digital que tiene una tasa de muestreo de trama de datos predeterminada; y
al menos un procesador para mejorar un índice de modulación de un evento de sonido de tasa de modulación
alta predeterminada, detectando el al menos un procesador nulos de energía del evento de sonido de tasa de
modulación alta predeterminada en un flujo del habla digitalizada, en el que el al menos un procesador
comprende:

40 un procesador pos-codificador de voz que proporciona un filtro (224) de mejora de modulación que filtra una
envolvente de energía de un flujo del habla digitalizada emitido del codificador de voz (220) para potenciar el
índice de modulación del evento de sonido de tasa de modulación alta predeterminada, en el que el filtro (224) de
mejora de modulación comprende:

45 un elemento (226) de retardo de tiempo para retardar el flujo del habla digitalizado emitido del codificador de
voz (220);
un elemento (228) de cálculo de envolvente de energía para calcular la envolvente de energía de modulación
del flujo del habla digitalizada del codificador de voz (220);
un filtro (230) de mejora de dominio de modulación que proporciona una ganancia positiva para frecuencias
de modulación predeterminadas de la envolvente de energía calculada; y
50 un multiplicador (232) de ganancia de envolvente de energía para imponer la envolvente de energía de
modulación filtrada en el flujo del habla digitalizada retardado emitido del elemento (226) de retardo de
tiempo.

5. La radio de la reivindicación 3, en la que el evento de sonido de tasa de modulación alta predeterminada comprende un sonido de trino.



100

(TÉCNICA ANTERIOR)

FIG. 1

102

TIEMPO (s)

TIEMPO (s)

104

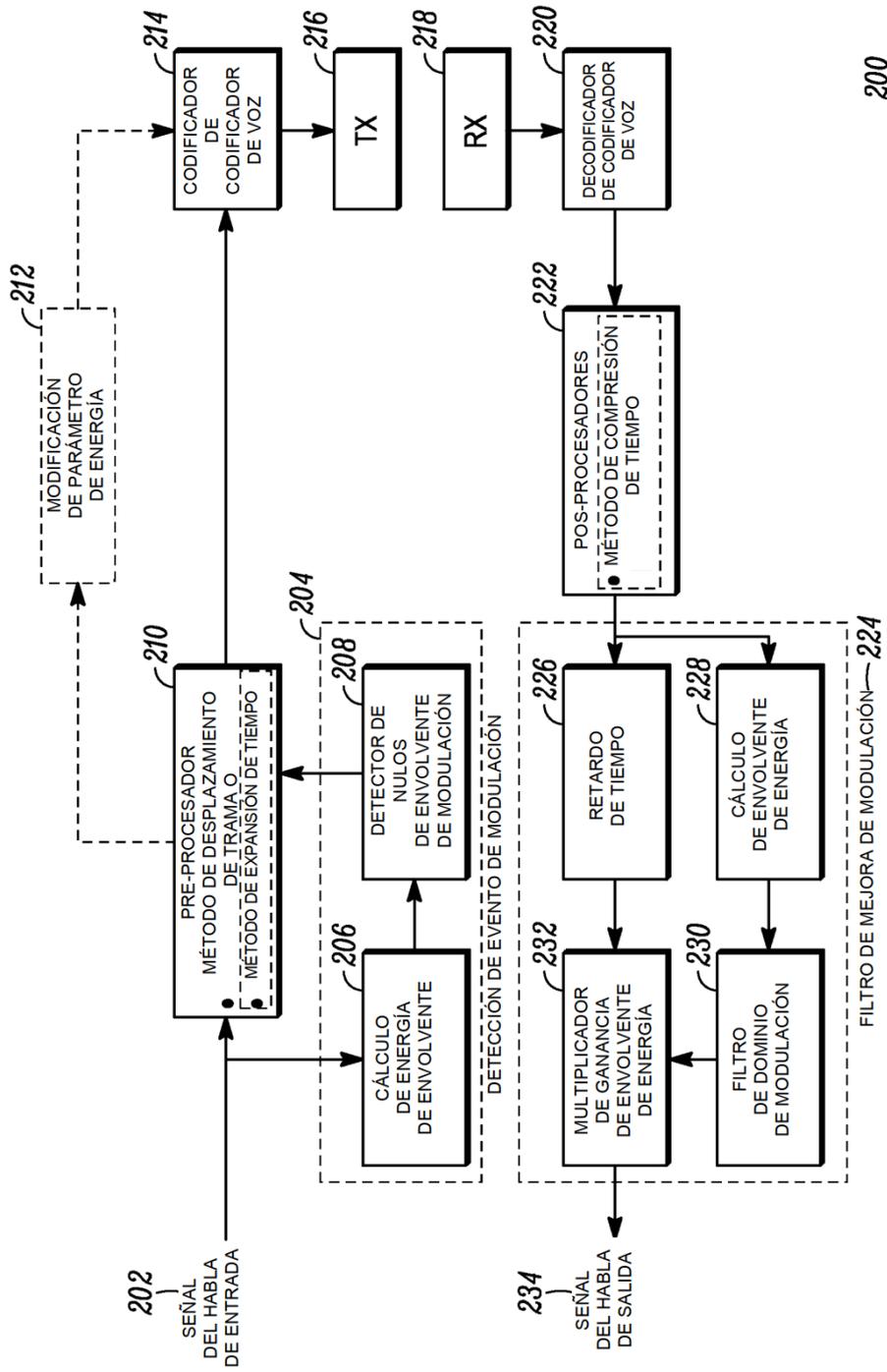


FIG. 2

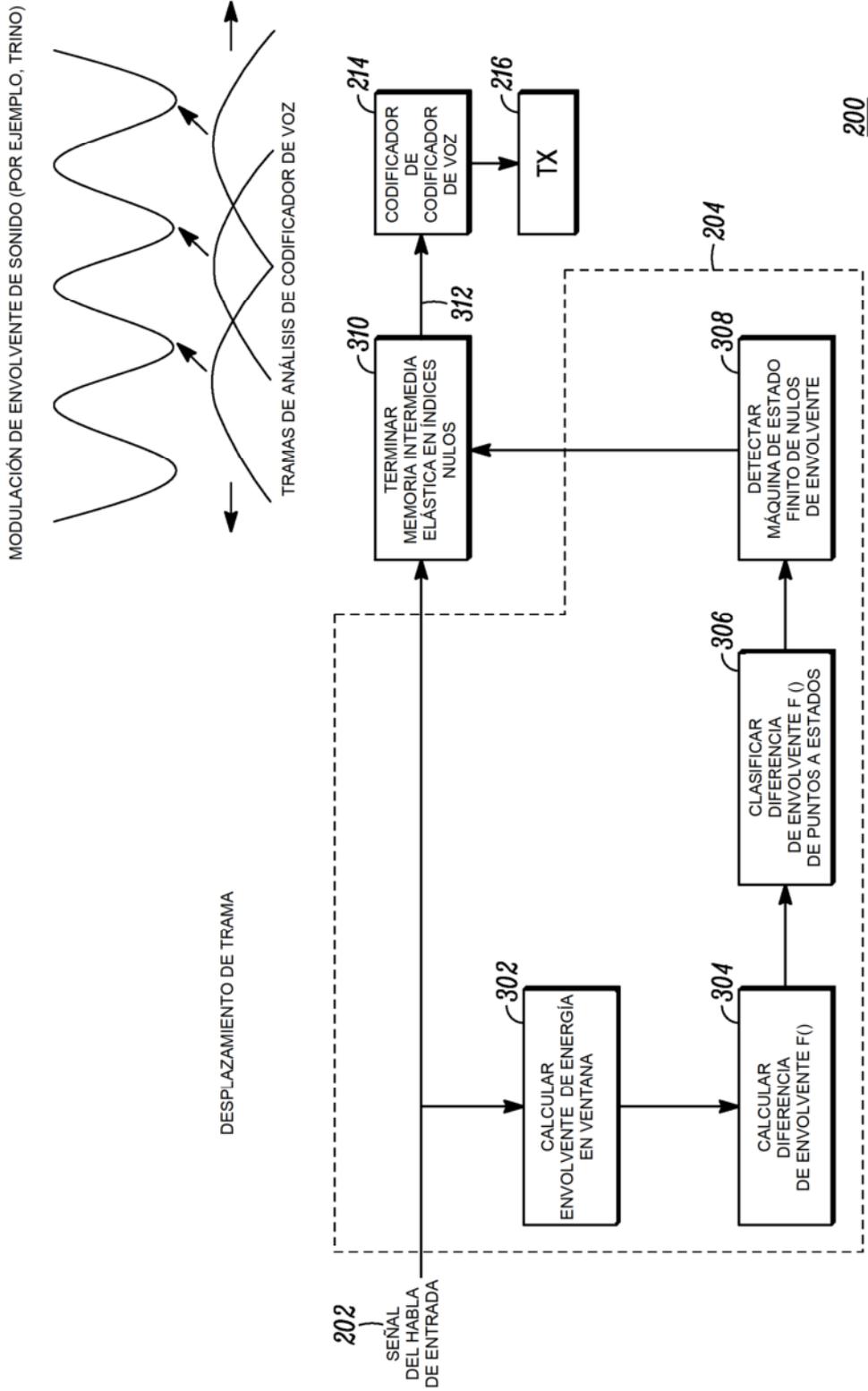


FIG. 3

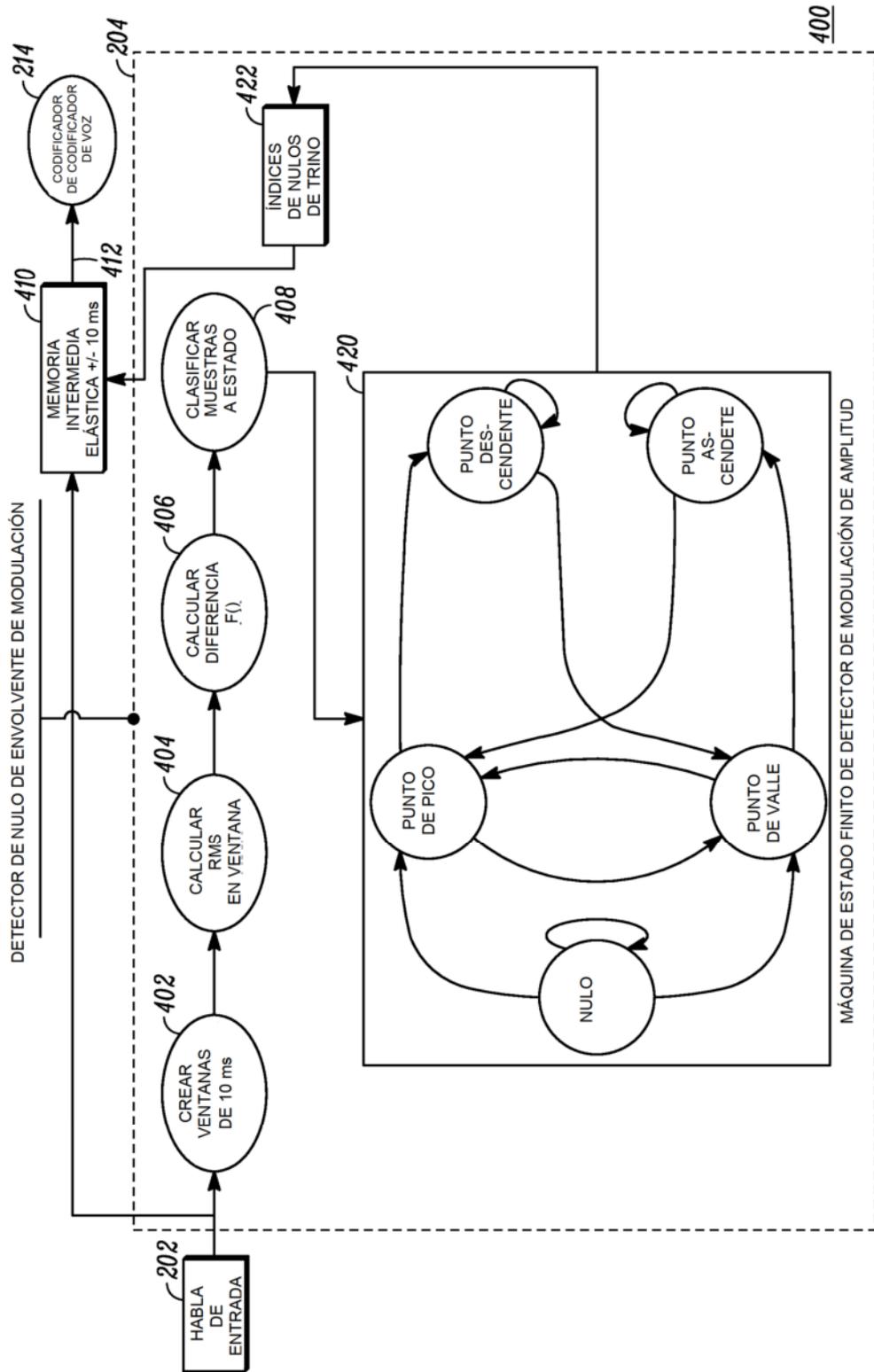


FIG. 4

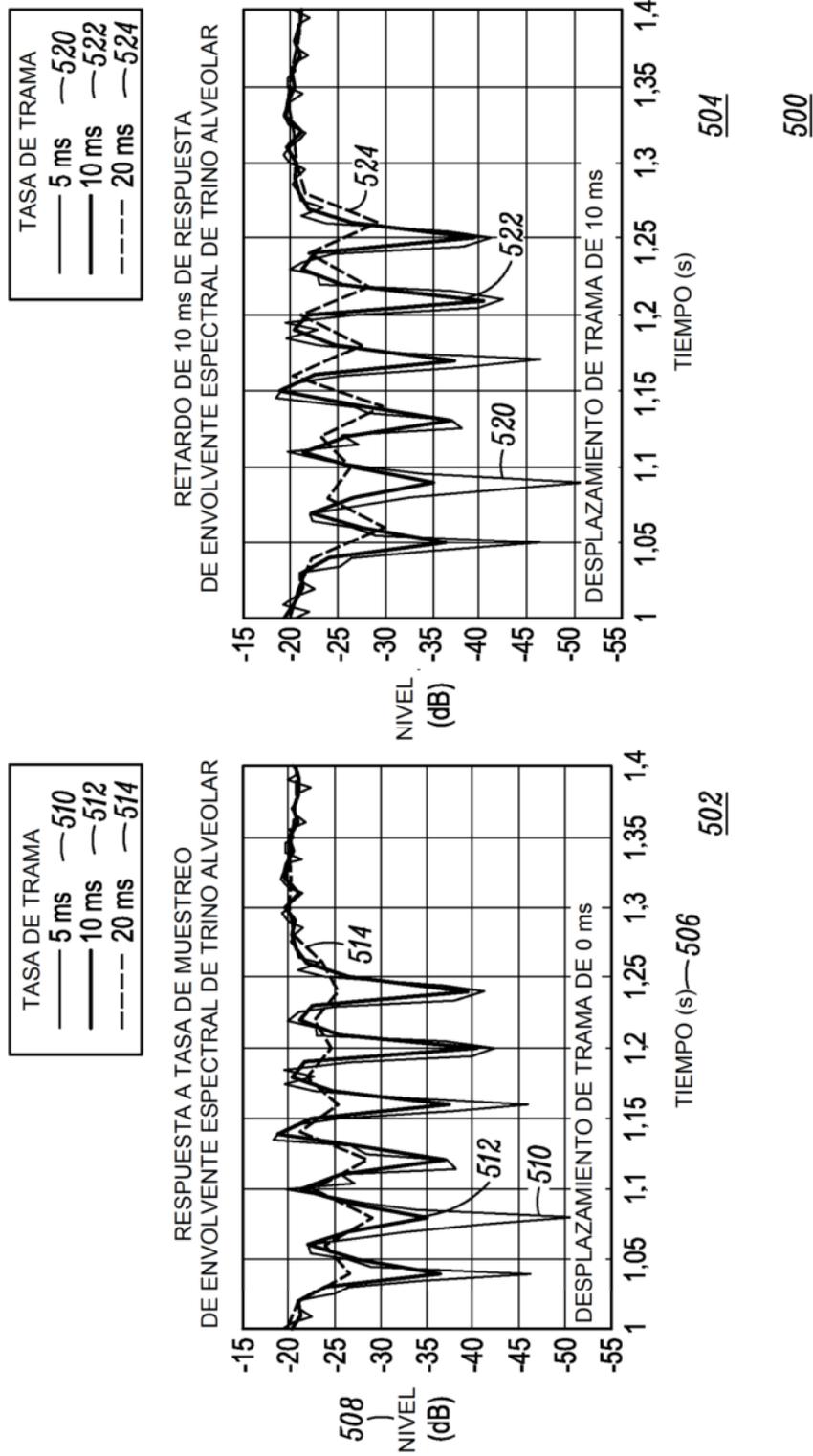


FIG. 5

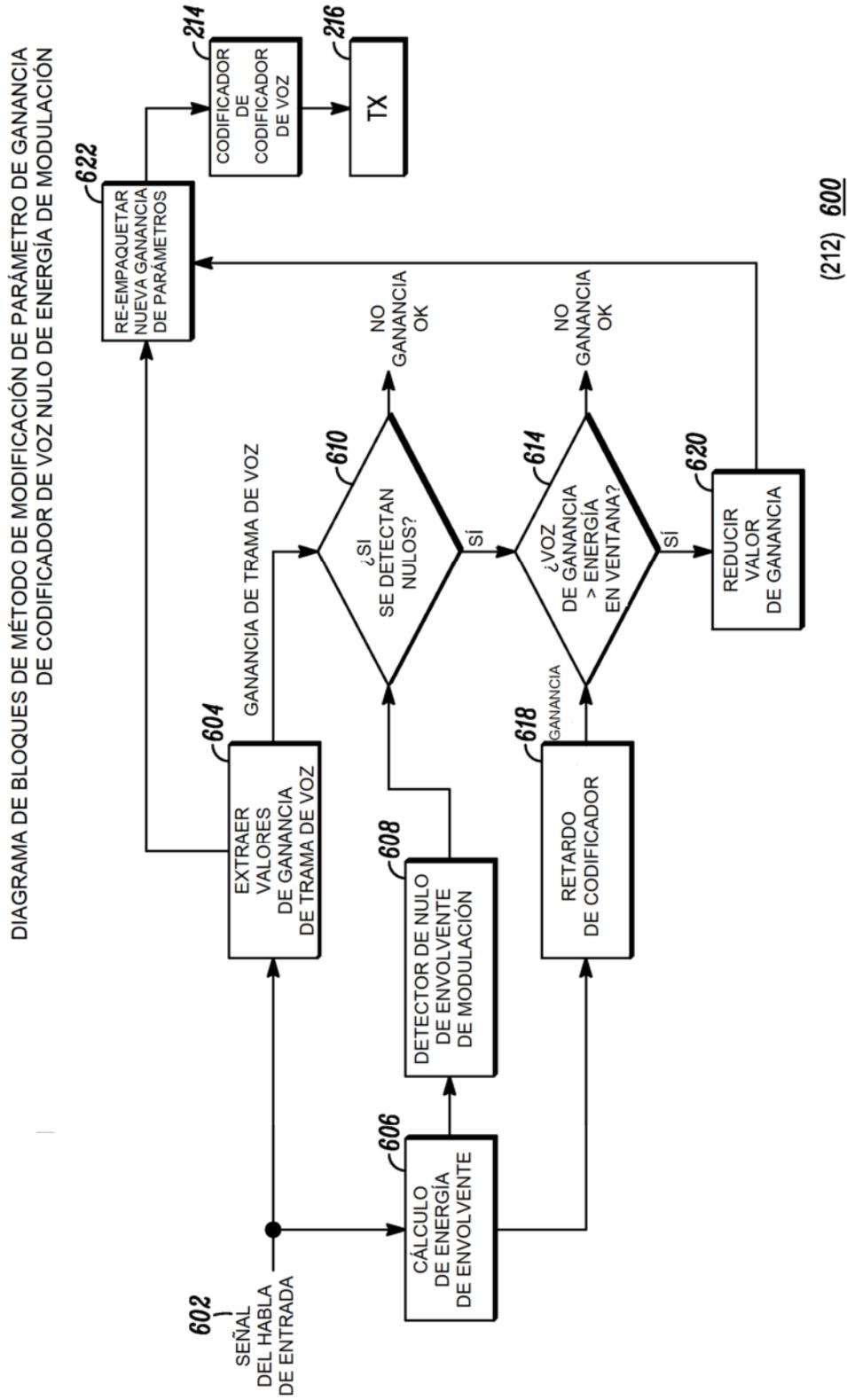


FIG. 6

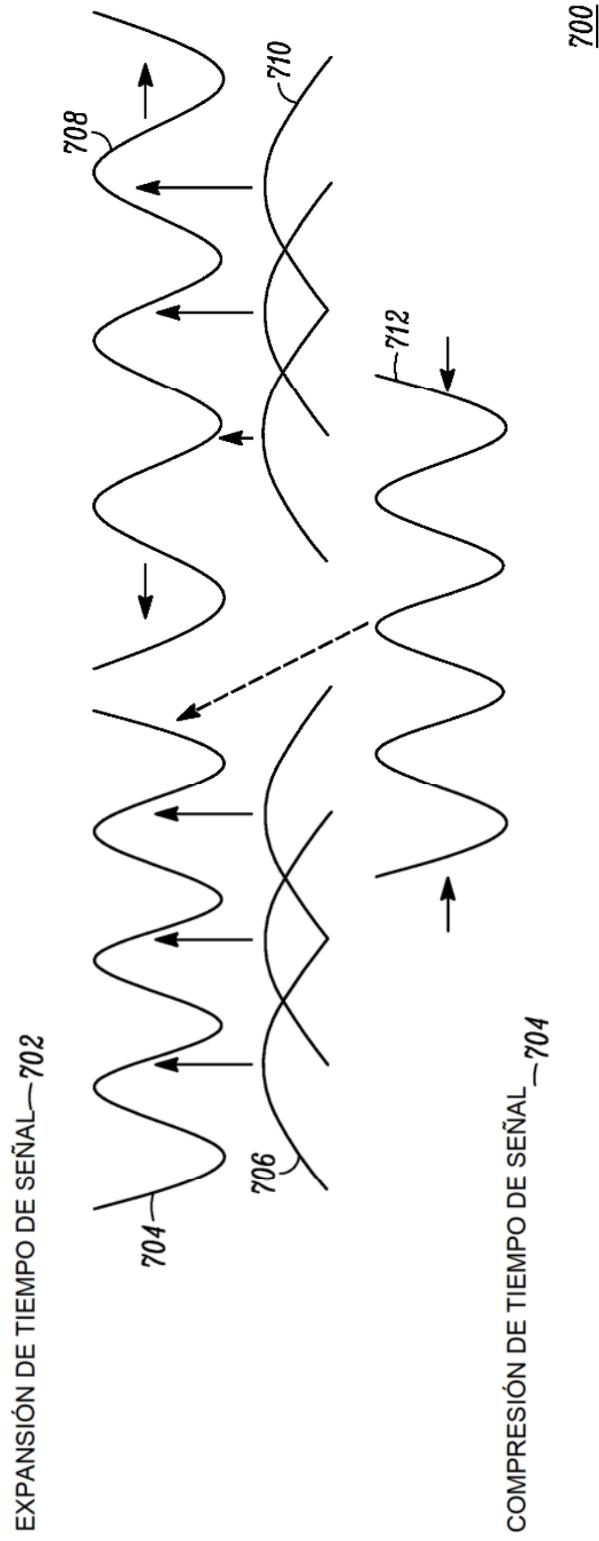


FIG. 7

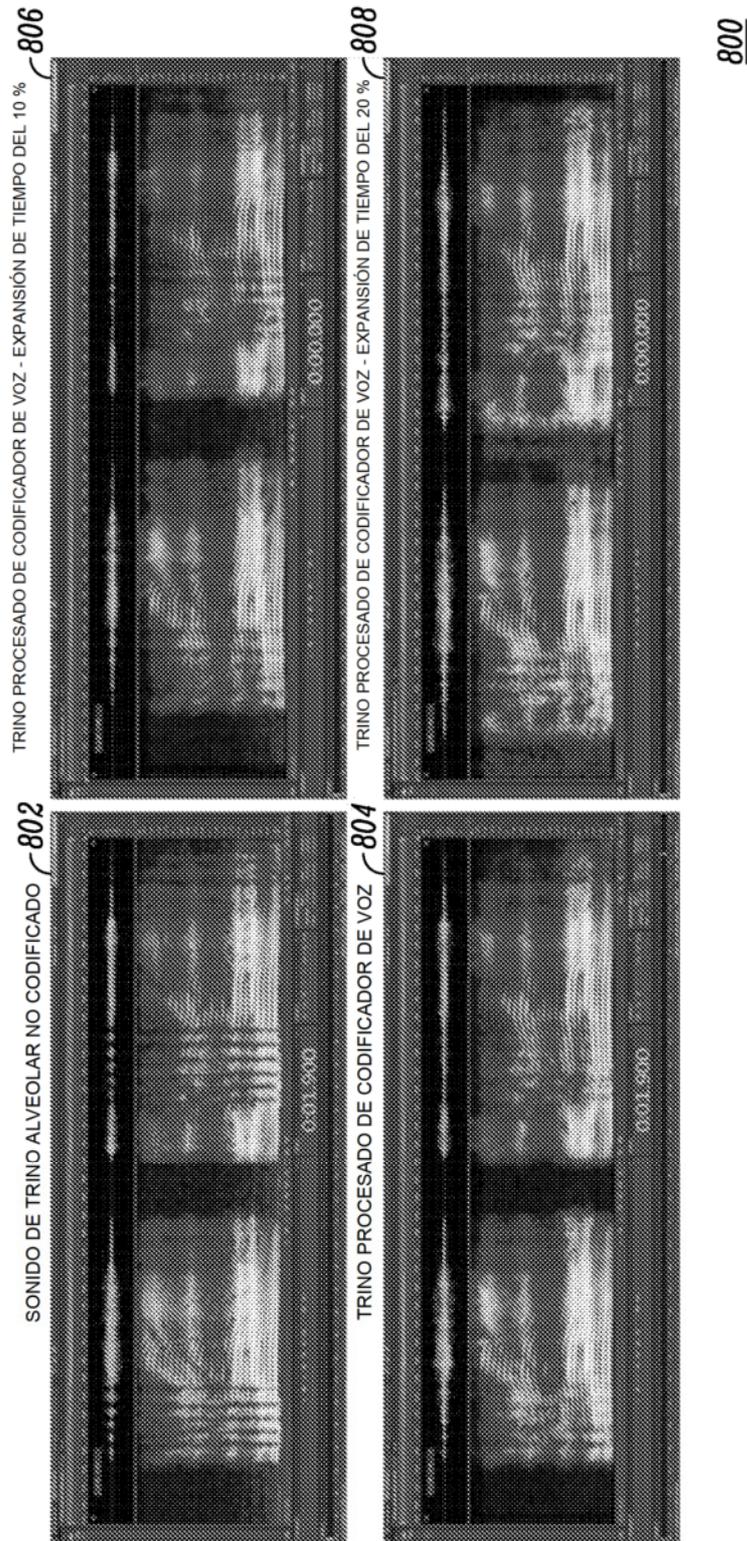


FIG. 8

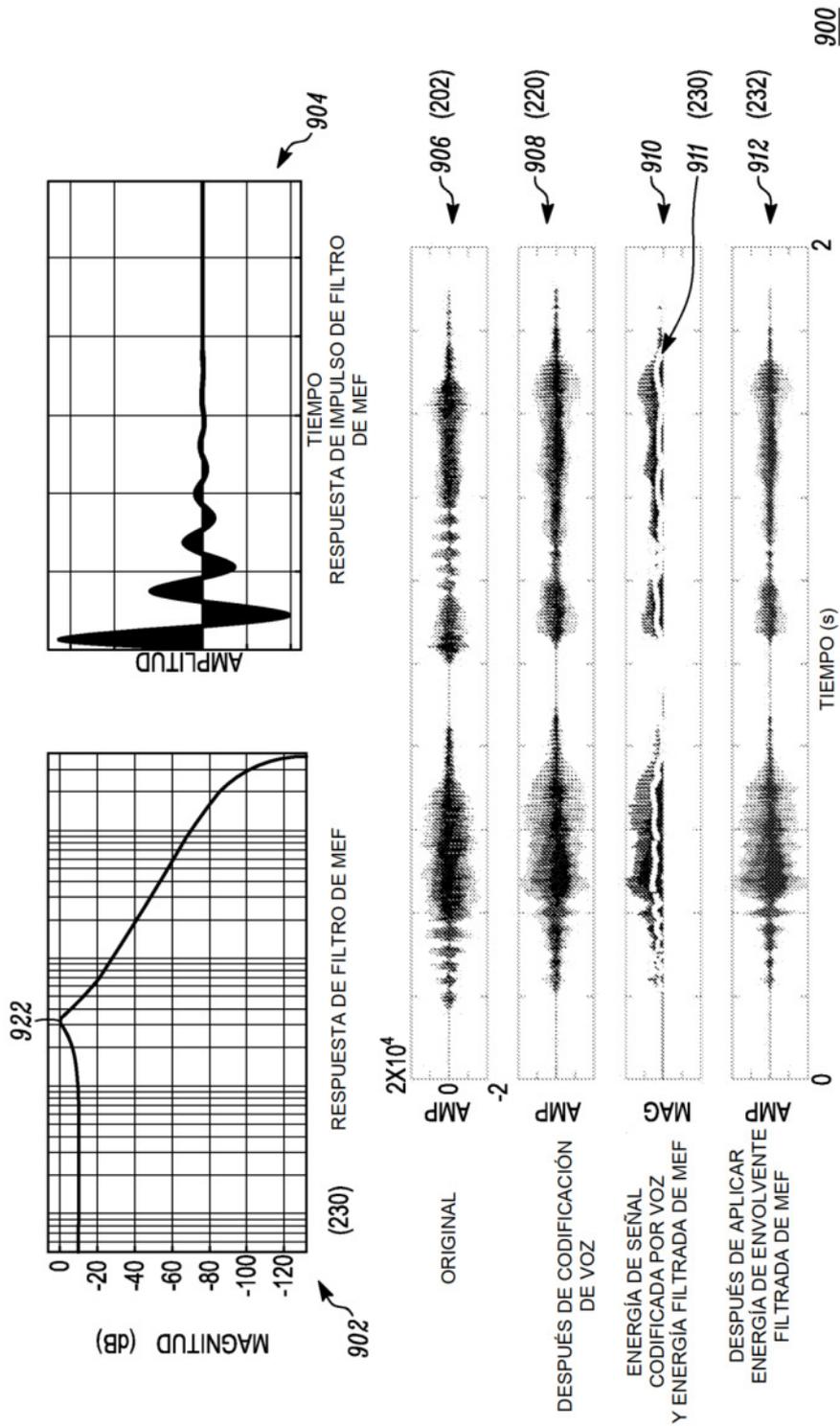


FIG. 9

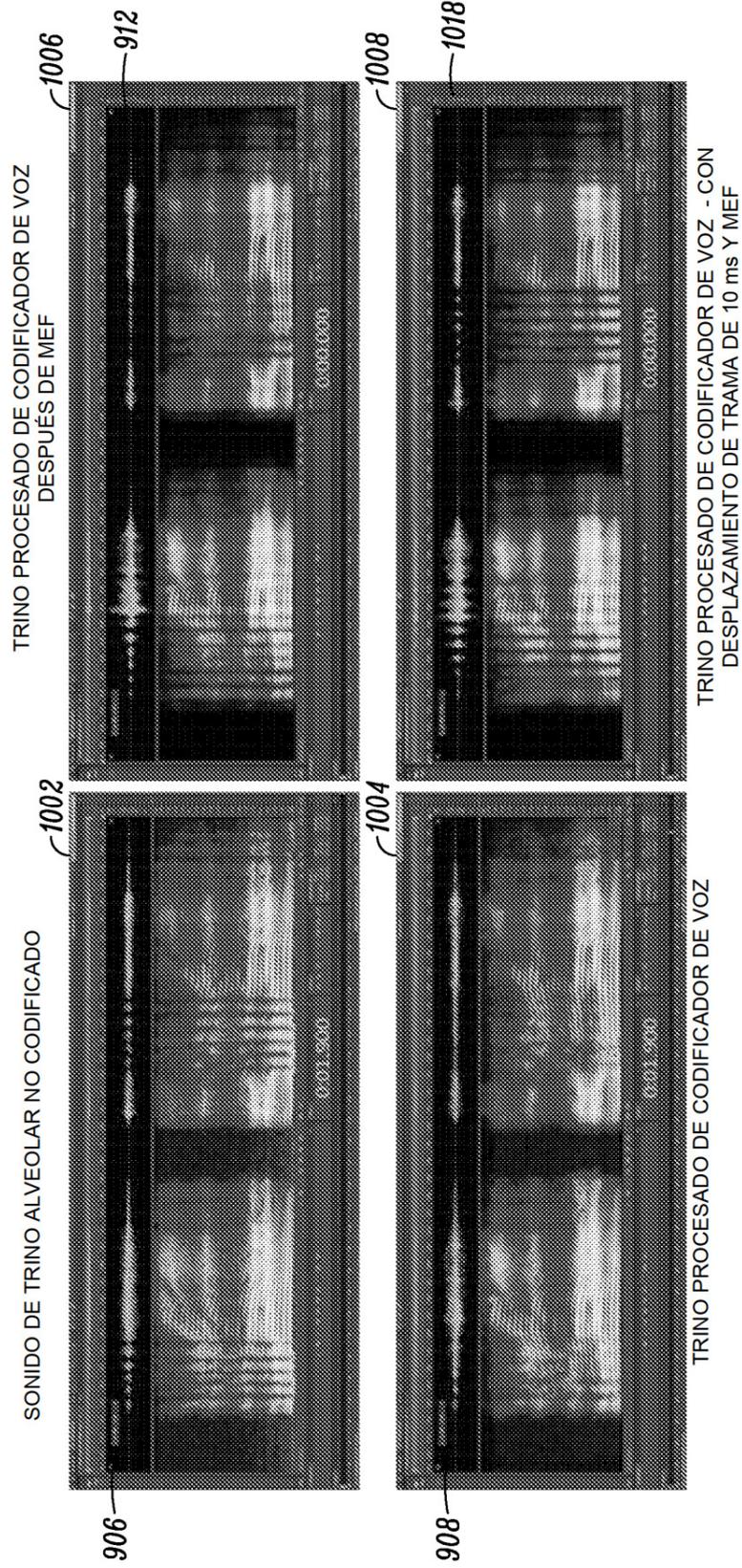


FIG. 10

1000